# PREFACE

Since the first edition of this series, we have witnessed the birth of "molecular medicine," using biochemistry, cell biology, and genetics to diagnose and treat disease. Consequently, the basic sciences are becoming more important to the practice of medicine. This puts a new pressure on the student—to understand the basis of molecular medicine and the molecular sciences. I still think that it's easier to remember things that you understand, things that make sense. That's the idea behind the Basic Concepts series and that's why I have been so pleased with the expansion of the Basic Concepts series beyond Biochemistry.

The revisions in the second edition include two new chapters, "Membranes and Membrane Proteins" and "Signal Transduction Pathways." These topics are related to the explosion of new information about cell signaling and signal transduction pathways. In addition, I've added some tables of information that I think will be helpful in seeing the big picture (and remembering some of the more important details). As before, the major topics and things to remember are set off in boxes so that if you already know everything in the box, you can skip the rest of the section.

# · C O N T E N T S ·


Preface xiii

Prologue xv

**CHAPTER 1 WHERE TO START 1**

Instructions 1

What Do I Need to Know? 2

Instructions for Use 2

Studying and Exams 2

Trivia Sorter 4

**CHAPTER 2 PROTEIN STRUCTURE 6**

Amino Acid Structure 6

Interactions 8

Water 9

Hydrophobic Interaction 9

van der Waals Interactions and London Dispersion Forces 11

Hydrogen Bonds 11

Secondary Structure 12

Protein Stability 15

Favorable (Good) Interactions 17

Unfavorable (Bad) Interactions 17

Temperature-Sensitive Mutations 19

Ligand-Binding Specificity 20

Global Conclusion 21

**CHAPTER 3 MEMBRANES AND MEMBRANE PROTEINS 22**

General Membrane Function 22

Membrane Composition 23

Phospholipid Bilayer 24

Membrane Structure 25

Posttranslational Modification 26

Membrane Fluidity 27

Diffusion in Membranes 28

Movement of Ions and Molecules Across Membranes 28

Transport Across Membranes 29
The Nernst Equation 31

**CHAPTER 4 DNA-RNA STRUCTURE 35**
DNA Structure 35
DNA Stability 37
RNA Secondary Structure 38

**CHAPTER 5 EXPRESSION OF GENETIC INFORMATION 40**
Information Metabolism 40
Directions and Conventions 41
DNA Replication 42
Types of DNA Polymerase 45
Recombination 47
Regulation of Information Metabolism 49
Transcription 53
Regulation of Transcription 55
Translation 57
Use of High-Energy Phosphate Bonds During Translation 60

**CHAPTER 6 RECOMBINANT-DNA METHODOLOGY 61**
Restriction Analysis 61
Gels and Electrophoresis 65
Blotting 67
Restriction Fragment-Length Polymorphism 69
Cloning 70
Sequencing 73
Mutagenesis 75
Polymerase Chain Reaction 76

**CHAPTER 7 ENZYME MECHANISM 80**
Active Site 81
Transition State 81
Catalysis 83
Lock and Key 83
Induced Fit 83
Nonproductive Binding 85
Entropy 87
Strain and Distortion 88

Transition-State Stabilization 88
Transition-State Analogs 91
Chemical Catalysis 93

**CHAPTER 8 ENZYME KINETICS 95**
S, P, and E (Substrate, Product, Enzyme) 96
Amounts and Concentrations 96
Active Site 97
Assay 98
Velocity 98
Initial Velocity 100
Mechanism 101
Little k's 102
Michaelis-Menten Equation 103
$V_{max}$ 106
$k_{cat}$ 106
$K_m$ 107
Special Points 107
$k_{cat}/K_m$ 107
Rate Accelerations 108
Steady-State Approximation 109
Transformations and Graphs 111
Inhibition 112
Allosterism and Cooperativity 117
The Monod-Wyman-Changeaux Model 119

**CHAPTER 9 SIGNAL TRANSDUCTION PATHWAYS 123**
Signal Transduction Pathways 123
Organization 124
Signals 125
Receptors 126
Soluble Receptors 126
Transmembrane Receptors 128
Enzyme Coupled Receptors 128
G-Protein Coupled Receptors 131
Ion-Channel Coupled Receptors 132
Second Messengers 133
Amplifiers 136
Integrators 137
Inhibitors 138

**CHAPTER 10 GLYCOLYSIS AND GLUCONEOGENESIS** **141**
Glycolysis Function 143
Glycolysis Location 143
Glycolysis Connections 143
Glycolysis Regulation 143
Glycolysis ATP Yields 144
Glycolysis Equations 144
Effect of Arsenate 144
Lactate or Pyruvate 145
Gluconeogenesis Function 145
Gluconeogenesis Location 145
Gluconeogenesis Connections 145
Gluconeogenesis Regulation 146
Gluconeogenesis ATP Costs 146
Gluconeogenesis Equations 146

**CHAPTER 11 GLYCOGEN SYNTHESIS AND DEGRADATION** **147**
Function 147
Location 147
Connections 148
Regulation 148
ATP Yield 150
ATP Cost 150
Molecular Features 150

**CHAPTER 12 TCA CYCLE** **153**
TCA Cycle 153

**CHAPTER 13 FAT SYNTHESIS AND DEGRADATION** **155**
Fatty Acid Synthesis Function 156
Fatty Acid Synthesis Location 156
Fatty Acid Synthesis Connections 157
Fatty Acid Synthesis Regulation 157
Fatty Acid Synthesis ATP Costs (for $C_{16}$) 157
Fatty Acid Synthesis Equation 159
Elongation and Desaturation 160
Triglyceride and Phospholipid Synthesis 162

β-Oxidation Function 164
β-Oxidation Location 164
Carnitine Shuttle 164
β-Oxidation Connections 164
β-Oxidation Regulation 164
β-Oxidation ATP Yield 166
β-Oxidation Equation 167
β-Oxidation of Unsaturated Fatty Acids 168
β-Oxidation of Odd-Chain-Length Fatty Acids 172

**CHAPTER 14 ELECTRON TRANSPORT AND OXIDATIVE PHOSPHORYLATION 173**
Oxidation and Reduction 173
The Electron Transport Chain 174
Connections 176
Regulation 178
P/O Ratios 178
Uncouplers 179
Inhibitors 180

**CHAPTER 15 PENTOSE PHOSPHATE PATHWAY 183**
Pentose Phosphate Pathway 183

**CHAPTER 16 AMINO ACID METABOLISM 186**
Nonessential Amino Acid Synthesis 186
Essential Amino Acids 187
Amino Acid Degradation 187
Generalities of Amino Acid Catabolism 187
Products of Amino Acid Degradation 188

**CHAPTER 17 INTEGRATION OF ENERGY METABOLISM 190**
Integrating Metabolic Pathways 191
ATP 192
Glucose 192
Storage Molecules 193
Metabolic States and Signals 194
Insulin 195

Glucagon 196
Epinephrine 197
Secondary Signals 198
Generalities of Metabolism 199
Phosphorylation 202
Glycogen 204
Metabolic Movements of Glycogen 205
Fat 207
Metabolic Movements of Fat 207
Protein 209
Metabolic Movements of Protein 209
Tissue Cooperation 211
Liver 212
Muscle 212
Adipose 212
Brain 213
Connection of Storage Pools 213
Feeding 214
Fasting 214
Starvation 217
Excitement 219
Interorgan Cycles 221
Cori Cycle 221
Alanine Cycle 222
Ketone Bodies 223

**CHAPTER 18 UREA CYCLE 225**
Urea Cycle 225

**CHAPTER 19 PURINE METABOLISM 227**
Purine Synthesis 227
Purine Salvage 228
Deoxynucleotides 228
Purine Degradation 229

**CHAPTER 20 PYRIMIDINE METABOLISM 230**
Pyrimidine Synthesis 230
Pyrimidine Salvage 231
Pyrimidine Degradation 232

**CHAPTER 21 ONE-CARBON METABOLISM** **233**
One-Carbon Metabolism 233
Oxidation States of Carbon 233

**CHAPTER 22 TRACKING CARBONS** **236**
Glucose to Pyruvate 236
TCA Cycle 238

**CHAPTER 23 ph, p$K_A$, pROBLEMS** **241**
Proton: $H^+$ or $H_3O^+$ 242
Acid 242
Base 242
Not All Acids and Bases Are Created Equal 243
$pK_a = -\log(K_a)$ 244
Weak Acids Make Strong Bases (and Vice Versa) 244
Who Gets the Proton? 245
Don't Forget Stoichiometry 245
The Sadistic Little *p* 246
Taking $\log_{10}(x)$ 247
Taking $-\log_{10}(x)$ 247
$pH = -\log_{10}[H^+]$ 248
$pK_a = -\log_{10}(K_a)$ 248
Buffers 248
Henderson-Hasselbalch Equation 249
Titration Curves 250
p*I*—Isoelectric Point 254
The Bicarbonate Buffer 255
Imbalance in Blood pH 257
Acidosis and Alkalosis 258

**CHAPTER 24 THERMODYNAMICS AND KINETICS** **261**
Thermodynamics 261
Free Energy 263
Adding Free-Energy Changes 268
Coupling Free Energies 268
Thermodynamic Cycles 269
$\Delta G = \Delta H - T\Delta S$ 272
Driving Force 273

Kinetics 275
Velocity 275
Transition State Theory 276
Rate Constants 277
Rate Constants and Mechanism 283

Appendix 284

Glossary 287

Index 319

• C H A P T E R • 1 •

# WHERE TO START

•

**Instructions**

**What Do I Need to Know?**

**Instructions for Use**

**Studying and Exams**

**Trivia Sorter**

• • • • • • • • • • • •

## INSTRUCTIONS

Read for understanding. Read only what you don't know. Organize, organize, organize.

The first page of each chapter presents an index. A title-summary box for each section presents a short summary and memory jogger intended to be helpful for review. If you already know what the boxed terms mean and feel comfortable with them, don't bother to read the text section that follows—proceed until you find a heading you don't understand, and then read till you understand. The first rule (it may not really be the first rule, but it is a rule) is not to waste time reading things you already know.

Keep on not reading the text until you find something you don't understand—then read the text till you do. The sections are generally arranged in order of increasing complexity and build on previous sections. So if you screwed up and jumped in over your head, back up a section or two. Another option is just to look at the pictures. Pictures and diagrams, if extensively annotated and carefully designed (by you), can be an enormous review aid.

## WHAT DO I NEED TO KNOW?

You need to know only the things you will need later.

Medicine and biology are becoming increasingly molecular in nature, so one answer to the question is that you need to know things down to the last atom. *Everything* is not the right answer. You can't possibly learn it all. Therefore, you will have to be selective.

Another answer is that you just need to know the things on the exam. *Later* ends at the final. In reality, later may be longer than this. Try to pick out the major concepts of biochemistry as you go along. Concepts are generally easier to remember than factual details—particularly if the concepts make sense.

## INSTRUCTIONS FOR USE

Understand the concepts first. Make notes. Never use a colored highlighter.

General concepts don't need to be memorized. Once you understand them, they provide a framework to hang the rest of the material on. Since they don't need to be memorized, they can be learned (or thought about) almost anywhere. To remember something, write it down. Don't just highlight it with a colored pen or pencil. Highlighting is a great way to forget to read the material.

## STUDYING AND EXAMS

Organize, understand, condense, memorize.

**• 1. ALWAYS REMEMBER THAT IT IS POSSIBLE TO BE A WORTHWHILE HUMAN BEING REGARDLESS OF (OR IN SPITE OF) HOW MUCH BIOCHEMISTRY YOU KNOW.** This won't necessarily help you with biochemistry, but it may help you keep your sanity.

**• 2. MINIMIZE THE AMOUNT OF MATERIAL THAT YOU HAVE TO MEMORIZE.** If you understand a general concept, you can often figure out the specific details rather than memorize them. For example,

does phosphorylation activate or inactivate acetyl-CoA carboxylase? You could just memorize that it inactivates the enzyme. However, this would not help when it came to the phosphorylation of glycogen synthase. Try the following line of reasoning. We store energy after eating and retrieve it between meals. Storage and retrieval of energy do not happen at the same time. Protein phosphorylation generally increases when you're hungry. Since both acetyl-CoA carboxylase and glycogen synthase are involved in energy storage (fat and glucose, respectively), they will both be inactivated by phosphorylation. For just two enzymes, it might be easier to just memorize all the regulatory behaviors—but for several hundred?

**• 3. ARRANGE NOTES AND STUDY TIME IN ORDER OF DECREASING IMPORTANCE.** During the first (or even second and third) pass, you can't possibly learn everything biochemistry has to offer. Be selective. Learn the important (and general) things first. If you have enough gray matter and time, then pack in the details. Organize your notes the same way. For each topic (corresponding to about a chapter in most texts) write down a *short* summary of the really important concepts (no more than one to two pages). Don't write down the things that you already know, just the things you're likely to forget. Be really cryptic to save space, and use lots of diagrams. These don't have to be publication-quality diagrams; they only have to have meaning for you. The idea is to minimize the sheer volume of paper. You can't find yourself at finals time with a yellow-highlighted 1000-page text to review 2 days before the exam. An enormous amount of information can be crammed onto a diagram, and you learn a significant amount by creating diagrams. Use them extensively.

**• 4. SORT OUT THE TRIVIA AND FORGET ABOUT IT.** The most difficult part may be deciding what the important things actually are. After all, if you've never had biochemistry, it all sounds important (or none of it does). Use the following trivia sorter (or one of your own invention) to help with these decisions. To use this sorter, you must first set your trivia level. Your trivia level will depend on whether you just want to pass or want to excel, whether you want to devote a lot of time or a whole lot of time to biochemistry, and your prior experience. Once you set this level, make sure you know *almost everything* above this level and ignore almost everything below it. Setting your trivia level is not irreversible; the setting can be moved at any time. You should consider levels 7 to 10 as the minimal acceptable trivia level (passing). The trivia sorter shown here is generic. You can make your own depending on the exact demands of the course you're taking. Levels 21 and 22 might be too trivial for anybody to spend time learning (again, this is opinion).

## TRIVIA SORTER

1. Purpose of a pathway—what's the overall function?
2. Names of molecules going into and coming out of the pathway
3. How the pathway fits in with other pathways
4. General metabolic conditions under which the pathway is stimulated or inhibited
5. Identity (by name) of control points—which steps of the pathway are regulated?
6. Identity (by name) of general regulatory molecules and the direction in which they push the metabolic pathway
7. Names of reactants and products for each regulated enzyme and each enzyme making or using ATP equivalents
8. Names of molecules in the pathway and how they're connected
9. Structural features that are important for the function of specific molecules in the pathway (this includes DNA and proteins)
10. Techniques in biochemistry, the way they work, and what they tell you
11. Molecular basis for the interactions between molecules
12. Genetic diseases and/or specific drugs that affect the pathway
13. Essential vitamins and cofactors involved in the pathway
14. pH
15. Enzyme kinetics
16. Specific molecules that inhibit or activate specific enzymes
17. Names of individual reactants and products for nonregulated steps
18. Chemical structures (ability to recognize, not draw)
19. Structures of individual reactants and products for all enzymes in pathway
20. Reaction mechanism (chemistry) for a specific enzyme
21. Cleavage specificity for proteases or restriction endonucleases
22. Molecular weights and quaternary structures

**• 5. DON'T WASTE TIME ON ABSOLUTE TRIVIA UNLESS YOU HAVE THE TIME TO WASTE.** It is possible to decide that something is just not worth remembering; for example, cleavage specificities of proteases or restriction endonucleases, and protein molecular weights, are

obvious choices. You can set the "too trivial to bear" level anywhere you want. You could decide that glycolysis is just not worth knowing. However, if you set your limits totally in the wrong place, you will get another chance to figure this out when you repeat the course. The trivia line is an important line to draw, so think about your specific situation and the requirements of the course before you draw it.

• C H A P T E R • 2 •

# PROTEIN STRUCTURE

•

**Amino Acid Structure**

**Interactions**

**Water**

**Hydrophobic Interaction**

**Van der Waals Interactions and London Dispersion Forces**

**Hydrogen Bonds**

**Secondary Structure**

**Protein Stability**

**Favorable (Good) Interactions**

**Unfavorable (Bad) Interactions**

**Temperature-Sensitive Mutations**

**Ligand-Binding Specificity**

**Global Conclusion**

• • • • • • • • • • • •

Proteins start out life as a bunch of amino acids linked together in a head-to-tail fashion—the primary sequence. The one-dimensional information contained in the primary amino acid sequence of cellular proteins is enough to guide a protein into its three-dimensional structure, to determine its specificity for interaction with other molecules, to determine its ability to function as an enzyme, and to set its stability and lifetime.

## AMINO ACID STRUCTURE

Remember a few of the amino acids by functional groups. The rest are hydrophobic.

Remembering something about the structures of the amino acids is just one to those basic language things that must be dealt with since it crops up over and over again—not only in protein structure but later in metabolism. You need to get to the point that when you see Asp you don't think snake but see a negative charge. Don't memorize the amino acids down to the last atom, and don't spend too much time worrying about whether glycine is polar or nonpolar. Methylene groups ($–CH_2–$) may be important, but keeping track of them on an individual basis is just too much to ask. Organize the amino acids based on the functional group of the side chain. Having an idea about functional groups of amino acids will also help when you get to the biosynthesis and catabolism of amino acids. Might as well bite the bullet early.

## HYDROPHILIC (POLAR)

• **CHARGED POLAR** *Acidic* ($–COO^-$) and *basic* ($–NH_3^+$) amino acid side chains have a charge at neutral pH and strongly "prefer" to be on the exterior, exposed to water, rather than in the interior of the protein. The terms *acidic* and *basic* for residues may seem a little strange. Asp and Glu are called acidic amino acids, although at neutral pH in most proteins, Asp and Glu are not present in the acidic form (–COOH) but are present in the basic form ($–COO^-$). So the acidic amino acids, Asp and Glu, are really bases (proton acceptors). The reason that Asp and Glu are called acidic residues is that they are such strong acids (proton donors) they have already lost their protons. Lys, Arg, and His are considered basic amino acids, even though they have a proton at neutral pH. The same argument applies: Lys, Arg, and His are such good bases (proton acceptors) that they have already picked up a proton at neutral pH.

| FUNCTIONAL GROUP | | | AMINO ACID |
|---|---|---|---|
| | | **Hydrophilic, Polar** | |
| **Acidic** | Carboxylates | $—COO^-$ | Asp, Glu |
| **Basic** | Amines | $—NH_3^+$ | Lys, Arg, His |
| **Neutral** | Amides | $—CONH_2$ | Asn, Gln |
| | Alcohols | —OH | Ser, Thr, Tyr |
| | Thiol | —SH | Cys |
| | | **Hydrophobic, Apolar** | |
| **Aliphatic** | | $—CH_2—$ | Ala, Val, Leu, Ile, Met |
| **Aromatic** | | C Rings | Phe, Trp, Tyr |
| **Whatever** | | | Pro, Gly |

Charged groups are usually found on the surface of proteins. It is very difficult to remove a charged residue from the surface of a protein and place it in the hydrophobic interior, where the dielectric constant is low. On the surface of the protein, a charged residue can be solvated by water, and it is easy to separate oppositely charged ions because of the high dielectric constant of water.[1] If a charged group is found in the interior of the protein, it is usually paired with a residue of the opposite charge. This is termed a *salt bridge*.

• **NEUTRAL POLAR** These side chains are uncharged, but they have groups (–OH, –SH, NH, C═O) that can hydrogen-bond to water. In an unfolded protein, these residues are hydrogen-bonded to water. They prefer to be exposed to water, but if they are found in the protein interior they are hydrogen-bonded to other polar groups.

## HYDROPHOBIC (APOLAR)

*Hydrocarbons* (both aromatic and aliphatic) do not have many (or any) groups that can participate in the hydrogen-bonding network of water. They're greasy and prefer to be on the interior of proteins (away from water). Note that a couple of the aromatics, Tyr and Trp, have O and N, and Met has an S, but these amino acids are still pretty hydrophobic. The hydrophobic nature usually dominates; however, the O, N, and S atoms often participate in hydrogen bonds in the interior of the protein.

## INTERACTIONS

A few basic interactions are responsible for holding proteins together. The properties of water are intimately involved in these interactions.

[1] The dielectric constant is a fundamental and obscure property of matter that puts a number on how hard it is to separate charged particles or groups when they're in this material. In water, charge is easy to separate (water has a high dielectric constant). The charge distribution on water is uneven. It has a more positive end (H) and a more negative end (O) that can surround the charged group and align to balance the charge of an ion in water. This dipolar nature of water makes it easy for it to dissolve ionic material. Organic solvents like benzene or octane have a low dielectric constant and a more uniform distribution of electrons. They do not have polar regions to interact with ions. In these types of solvents, just as in the interior of a protein, it is very difficult to separate two oppositely charged residues.

## WATER

Water's important. Polar amino acid chains can participate in hydrogen bonding to water, or hydrophobic side chains can interfere with it.

The properties of water dominate the way we think about the interactions of biological molecules. That's why many texts start with a lengthy, but boring, discussion of water structure, and that's why you probably do need to read it.

Basically, water is a polar molecule. The H—O bond is polarized—the H end is more positive than the O end. This polarity is reinforced by the other H—O bond. Because of the polarity difference, water is both a hydrogen-bond donor and a hydrogen-bond acceptor. The two hydrogens can each enter into hydrogen bonds with an appropriate acceptor, and the two lone pairs of electrons on oxygen can act as hydrogen-bond acceptors. Because of the multiple hydrogen-bond donor and acceptor sites, water interacts with itself. Water does two important things: It squeezes out oily stuff because the oily stuff interferes with the interaction of water with itself, and it interacts favorably with anything that can enter into its hydrogen-bonding network.

## HYDROPHOBIC INTERACTION

Proteins fold in order to put as much of the greasy stuff out of contact with water as possible. This provides much of the "driving force" for protein folding, protein–protein interactions, and protein–ligand interactions (Fig. 2-1).

The *driving force* for a chemical reaction is what makes it happen. It's the interaction that contributes the most to the decrease in free energy. For protein (and DNA) folding, it's the hydrophobic interaction that provides most of the driving force. As water squeezes out the hydrophobic side chains, distant parts of the protein are brought together into a compact structure. The hydrophobic core of most globular proteins is very compact, and the pieces of the hydrophobic core must fit together rather precisely.

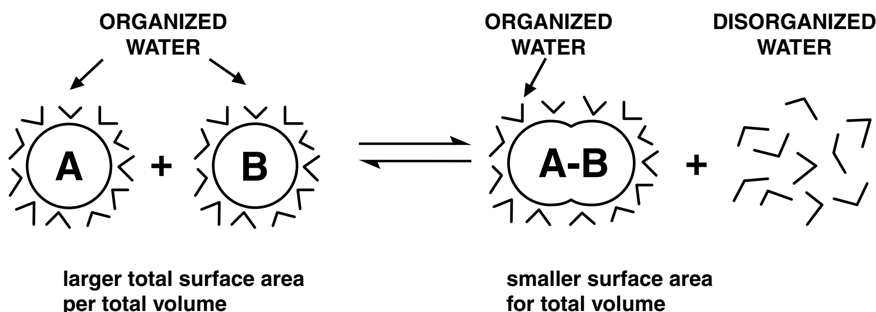



**Figure 2-1 The Hydrophobic Interaction**
As hydrophobic surfaces contact each other, the ordered water molecules that occupied the surfaces are liberated to go about their normal business. The increased entropy (disorder) of the water is favorable and drives (causes) the association of the hydrophobic surfaces.

Putting a hydrophobic group into water is difficult to do (unfavorable). Normally, water forms an extensive hydrogen-bonding network with itself. The water molecules are constantly on the move, breaking and making new hydrogen bonds with neighboring water molecules. Water has two hydrogen bond donors (the two H—O bonds) and two hydrogen bond acceptors (the two lone electron pairs on oxygen), so a given water molecule can make hydrogen bonds with neighboring water molecules in a large number of different ways and in a large number of different directions. When a hydrophobic molecule is dissolved in water, the water molecules next to the hydrophobic molecule can interact with other water molecules only in a direction away from the hydrophobic molecule. The water molecules in contact with the hydrophobic group become more organized. In this case, organization means restricting the number of ways that the water molecules can be arranged in space. The increased organization (restricted freedom) of water that occurs around a hydrophobic molecule represents an unfavorable decrease in the entropy of water.[2] In the absence of other factors, this increased organization (decreased entropy) of water causes hydrophobic molecules to be insoluble.

The surface area of a hydrophobic molecule determines how unfavorable the interaction between the molecule and water will be. The big-

[2] As with most desks and notebooks, disorder is the natural state. Order requires the input of energy. Reactions in which there is an increasing disorder are more favorable. Physical chemists (and sometimes others) use the word *entropy* instead of *disorder*. There's a discussion of entropy at the end of this book.

ger the surface area, the larger the number of ordered water molecules and the more unfavorable the interaction between water and the hydrophobic molecule. Bringing hydrophobic residues together minimizes the surface area directly exposed to water. Surface area depends on the square of the radius of a hydrophobic "droplet," while volume depends on the cube of the radius. By bringing two droplets together and combining their volume into a single droplet of larger radius, the surface area of the combined, larger droplet is less than that of the original two droplets. When the two droplets are joined together, some of the organized water molecules are freed to become "normal." This increased disorder (entropy) of the liberated water molecules tends to force hydrophobic molecules to associate with one another. The hydrophobic interaction provides most of the favorable interactions that hold proteins (and DNA) together. For proteins, the consequence of the hydrophobic interaction is a compact, hydrophobic core where hydrophobic side chains are in contact with each other.

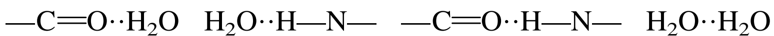
## VAN DER WAALS INTERACTIONS AND LONDON DISPERSION FORCES

These are very short-range interactions between atoms that occur when atoms are packed very closely to each other.

When the hydrophobic effect brings atoms very close together, van der Waals interactions and London dispersion forces, which work only over very short distances, come into play. This brings things even closer together and squeezes out the holes. The bottom line is a very compact, hydrophobic core in a protein with few holes.

## HYDROGEN BONDS

*Hydrogen bonding* means sharing a hydrogen atom between one atom that has a hydrogen atom (donor) and another atom that has a lone pair of electrons (acceptor):


$—C{=}O \cdot\cdot H_2O \quad H_2O \cdot\cdot H—N— \quad —C{=}O \cdot\cdot H—N— \quad H_2O \cdot\cdot H_2O$

The secondary structure observed in proteins is there to keep from losing hydrogen bonds.

A hydrogen bond is an interaction between two groups in which a weakly acidic proton is shared (not totally donated) between a group that has a proton (the donor) and a group that can accept a proton (the acceptor). Water can be both a hydrogen-bond donor and a hydrogen-bond acceptor. In an unfolded protein, the hydrogen-bond donors and acceptors make hydrogen bonds with water. Remember that the polar amino acids have groups that can form hydrogen bonds with each other and with water. The peptide bond [–C(═O)–NH–] that connects all the amino acids of a protein has a hydrogen-bond donor (NH) and a hydrogen-bond acceptor (═O). The peptide bond will form hydrogen bonds with itself (secondary structure) or with water.

Everything is just great until the hydrophobic interaction takes over. Polar peptide bonds that can form hydrogen bonds connect the amino acid side chains. Consequently, when hydrophobic residues aggregate into the interior core, they must drag the peptide bonds with them. This requires losing the hydrogen bonds that these peptide bonds have made with water. If they are not replaced by equivalent hydrogen bonds in the folded structure, this costs the protein stability. The regular structures (helix, sheet, turn) that have become known as *secondary structure* provide a way to preserve hydrogen bonding of the peptide backbone in the hydrophobic environment of the protein core by forming regular, repeating structures.

## SECONDARY STRUCTURE

Secondary structure is not just hydrogen bonds.

**α Helix:** Right-handed helix with 3.6 amino acid residues per turn. Hydrogen bonds are formed parallel to the helix axis.

**β Sheet:** A parallel or antiparallel arrangement of the polypeptide chain. Hydrogen bonds are formed between the two (or more) polypeptide strands.

**β Turn:** A structure in which the polypeptide backbone folds back on itself. Turns are useful for connecting helices and sheets.

Secondary structure exists to provide a way to form hydrogen bonds in the interior of a protein. These structures (helix, sheet, turn) provide ways to form regular hydrogen bonds. These hydrogen bonds are just replacing those originally made with water.

As a protein folds, many hydrogen bonds to water must be broken. If these broken hydrogen bonds are replaced by hydrogen bonds within

the protein, there is no net change in the number of hydrogen bonds (Fig. 2-2). Because the actual number of hydrogen bonds does not change as the secondary structure is formed, it is often argued that hydrogen bonds don't contribute much to the stability of a protein. However, hydrogen bonds that form after the protein is already organized into the correct structure may form more stable hydrogen bonds than the ones to water. Hydrogen bonding does contribute somewhat to the overall stability of a protein; however, the hydrophobic interaction usually dominates the overall stability.

Small peptides generally do not form significant secondary structure in water (there are some that do). For small peptides that do not form stable secondary structure, there are often other favorable interactions within the peptide that stabilize the formation of the helix or sheet structure.

The stability of secondary structure is also influenced by surrounding structures (Fig. 2-3). Secondary structure may be stabilized by interactions between the side chains and by interactions of the side chains with other structures in the protein. For example, it is possible to arrange the amino acid sequence of a protein or peptide into a helix that has one face that is hydrophobic and one that is hydrophilic. The helix wheel shown in Fig. 2-3 illustrates how this is possible. View the helix as a long cylinder. The peptide backbone spirals up and around the cylinder. The

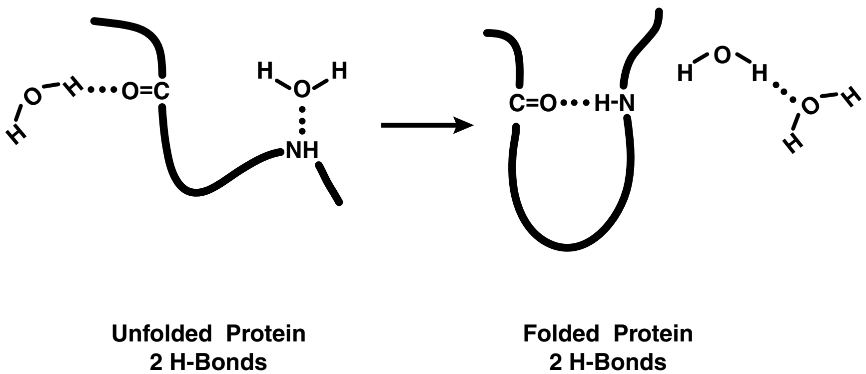
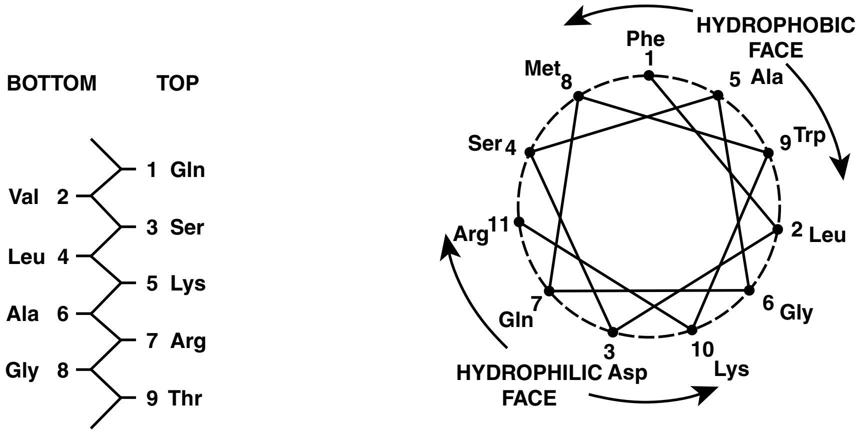



**Figure 2-2 Solvation in Protein Folding**
In an unfolded protein, water makes hydrogen bonds to all the donors and acceptors. As the protein folds and some polar groups find themselves inside, many of the hydrogen bonds with the solvent are replaced by hydrogen bonds between the different donors and acceptors in the protein. Because hydrogen bonds are being replaced rather than gained or lost as the protein folds, there is not a large net stabilization of the protein by the hydrogen bonds.




Looking at the side of a β-sheet. Every other residue is on the same face of the sheet.

Looking down the axis of an α-helix. Residue sequence is numbered. The angle between residues is 360°/3.6 residues or 100°.

**Figure 2-3**
**SECONDARY-STRUCTURE STABILIZATION** is not provided by just the hydrogen bonds. On the left, you're looking at a representation of a β sheet in which the amino acid side chains alternately stick up and down. If every other side chain is hydrophobic, one side of the sheet will be hydrophobic and the other side will be hydrophilic. Interaction of the hydrophobic side with a hydrophobic region on the protein will add stability to the β sheet. On the right an α helix is shown with a hydrophobic and a hydrophilic face. Again, putting the hydrophobic face (or surface) up against another hydrophobic region of the protein will stabilize the helix. In the helix representation, there is a 100° angle (360°/3.6 residues) between residues. Side chains would stick out from the side of the cylinder defined by the helix.

side chains of the amino acid residues point out from the helix. Each amino acid residue moves up the helix and around the helix at an angle of 100° (360°/turn = 3.6 residues/turn = 100°/residue). What you see in Fig. 2-3 is a view looking down the helix axis. The side chains are on the side of the circle (cylinder). One surface of the helix has only hydrophobic side chains, while the other side has hydrophilic side chains. This is termed an *amphipathic* helix (or *amphiphilic*, depending on whether you're a lover or a hater). With these kinds of helices, the hydrophobic face is buried in the interior while the hydrophilic face is exposed to water on the surface. There are two ways to look at this. The formation of the helix allows it to interact in a very specific way with

the rest of the protein. Alternatively, you could suppose that the interaction with the rest of the protein allows the helix to form. These are equivalent ways to view things, and energetically it doesn't make any difference (see linked thermodynamic functions in Chap. 24 if you dare)—the result is that the presence of a hydrophobic and a hydrophilic side of a helix and a complementary hydrophobic region in the interior of the protein makes it more favorable to form a helix. Secondary structure can be stabilized by interactions with other parts of the protein.

β Sheets can also have a hydrophobic face and a hydrophilic face. The backbone of the β sheet is arranged so that every other side chain points to the same side of the sheet. If the primary sequence alternates hydrophobic–hydrophilic, one surface of the sheet will be hydrophobic and the other will be hydrophilic.

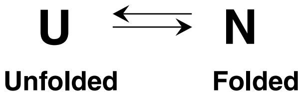
## PROTEIN STABILITY

Protein stability is proportional to the free-energy difference between an unfolded protein and the native structure (Fig. 2-4).

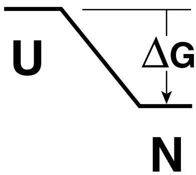
It's a miracle that we're here at all. Most proteins are not very stable even though there are a large number of very favorable interactions that can be seen in the three-dimensional structure. The reason is that the favorable interactions are almost completely balanced by unfavorable interactions that occur when the protein folds. A reasonably small net protein stability results from a small net difference between two large numbers. There are lots of favorable interactions but also lots of unfavorable interactions.

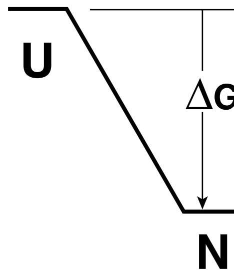
Protein stability is just the difference in free energy between the correctly folded structure of a protein and the unfolded, denatured form. In the denatured form, the protein is unfolded, side chains and the peptide backbone are exposed to water, and the protein is conformationally mobile (moving around between a lot of different, random structures). The more stable the protein, the larger the free energy difference between the unfolded form and the native structure.

You can think about the energy difference in terms of an equilibrium constant if you want. For the folding reaction, the equilibrium constant $K_{eq}$ = [native]/[denatured] is large if the protein is stable. Proteins can be denatured (unfolded) by increasing the temperature, lowering the pH, or adding detergents, urea, or guanidine hydrochloride. Urea and guanidine hydrochloride denature proteins by increasing the solubility of the hydrophobic side chains in water. Presumably these compounds, which




$$K_{eq} = \frac{[N]}{[U]}$$




**More stable protein**
**More favorable equilibrium constant**
**More negative ΔG**

**Figure 2-4**
The **FREE-ENERGY CHANGE** during a reaction such as the folding of a protein is related to how big the equilibrium constant is. For reactions that are downhill and favorable, the free energy of the product is lower than that of the reactant. The change in free energy (products − reactants) is less than zero (negative). Very downhill reactions have very large equilibrium constants.

are polar, alter water structure in some way to make it easier to dissolve hydrophobic molecules.[3]

Protein structure (and also the interactions between proteins and small molecules) is a compromise. It may be necessary to sacrifice a hydrogen bond or two in order to gain two or three hydrophobic interactions. In contrast, it may be necessary to place a hydrophobic residue in contact with water in order to pick up a few more hydrogen bonds in

[3] You may have figured out from this sentence that it's not exactly known how urea and guanidine denature proteins.

secondary structure. So it's all a compromise—a constant game of give and take. The game involves getting as many favorable interactions as you can while doing as few of the unfavorable things as possible.

## FAVORABLE (GOOD) INTERACTIONS

Try to get as many of these as possible:

**1.** Hydrophobic interactions
**2.** van der Waals interactions
**3.** London dispersion forces
**4.** Hydrogen bonds
**5.** Charge–charge interactions

These are the favorable interactions that were discussed above. They work together to provide stabilizing interactions that hold the structure together.

## UNFAVORABLE (BAD) INTERACTIONS

Avoid as many of these as possible:

**1.** Organizing anything into a structure (decreasing entropy)
**2.** Removing a polar group from water without forming a new hydrogen bond to it
**3.** Removing a charged group from water without putting an opposite charge nearby or putting two like charges close together
**4.** Leaving a hydrophobic residue in contact with water
**5.** Putting two atoms in the same place (steric exclusion)

There are numerous bad things (energetically speaking) that can happen when proteins fold into a three-dimensional structure. The worst thing that has to happen is that lots of covalent bonds in the protein must assume relatively fixed angles. They're no longer free to rotate as they were in the unfolded form. Protein folding requires a large loss in the conformational entropy (disorder) of the molecule. Restriction of the conformational freedom is probably the biggest unfavorable factor opposing the folding of proteins.

When a protein folds, most of the hydrophobic side chains pack into the interior. As they move into the interior, they must drag the polar amides of the polypeptide backbone with them. These backbone amides must lose contact with water and break hydrogen bonds to the solvent.[4] If these hydrogen bonds that were formed with the solvent aren't replaced by new hydrogen bonds between the different polar groups that now find themselves in the interior, there will be a net loss in the number of hydrogen bonds upon folding—this is not good. Secondary structure provides a way to allow much of the polypeptide backbone to participate in hydrogen bonds that replace the ones made with water. But then there's the odd residue that just may not be able to find a suitable hydrogen-bonding partner in the folded protein. This costs energy and costs the protein stability. The same thing happens with charged residues (although they're almost always ion-paired). By the same token, it may occasionally be necessary to leave a hydrophobic group exposed to water. It may not be possible to bury all the hydrophobic residues in the interior. If not, this is also unfavorable and destabilizes the protein. All these unfavorable interactions sum up to make the protein less stable.

Don't get the impression that proteins need to be as stable as possible and that the unfavorable interactions are necessarily bad. Proteins shouldn't live forever. A good bit of metabolism is regulated by increasing and decreasing the amount of a specific enzyme or protein that is available to catalyze a specific reaction. If a protein were too stable, it might not be possible to get rid of it when necessary.

The net result of all the favorable and unfavorable interactions is that they're almost balanced. For a 100-residue protein, it is possible to estimate roughly that the sum of all the favorable interactions that stabilize the three-dimensional, native structure is on the order of −500 kcal/mol. This comes from all the favorable hydrophobic, van der Waals, hydrogen-bonding, and electrostatic interactions in the native protein. In contrast, the sum of all the unfavorable interactions that destabilize the structure is probably near +490 kcal/mol. These come from conformational entropy losses (organization of the protein into a structure) and other unfavorable effects such as leaving a hydrophobic group exposed to water or not forming a hydrogen bond in the interior after having lost one that was made to water in the unfolded state. The net result is that the three-dimensional structure of a typical protein is only about −5 to −15 kcal/mol more stable than the denatured, structureless state.

[4] The same argument applies to polar groups on the side chains of the amino acids.

> **TEMPERATURE-SENSITIVE MUTATIONS**
>
> These are mutations that decrease the stability of a protein so that the denaturation temperature is near 40°C.

A single methylene group (–$CH_2$–) involved in a hydrophobic interaction may contribute as much as −1.5 to −2 kcal/mol to the stability of a protein that is only stable by −10 kcal/mol. A single hydrogen bond might contribute as much as −1.5 to −3.5 kcal/mol. If a mutation disrupts interactions that stabilize the protein, the protein may be made just unstable enough to denature near body (or culture) temperature. It might strike you as strange that we were talking earlier about how hydrogen bonds didn't contribute much to the net stability of proteins and now I'm telling you they contribute −1.5 to −3.5 kcal/mol. Both statements are more or less right. In the first case we were considering the folding process in which a hydrogen bond to solvent is replaced by a hydrogen bond in the folded protein—the result is a small contribution of a hydrogen bond to stability. What we're talking about now is messing up a protein by changing one amino acid for another by mutation. Here we're destroying an interaction that's present in the intact, folded protein. For any hydrogen-bonded group in the folded protein, there must be a complementary group. A donor must have an acceptor, and vice versa. Making a mutation that removes the donor of a hydrogen bond leaves the acceptor high and dry, missing a hydrogen bond. In the unfolded protein, the deserted acceptor can be accommodated by water; however, in the folded protein the loss of the donor by mutation hurts. It costs a hydrogen bond when the protein folds. The result: a loss in stability for the protein. Loss in stability means that the protein will denature at a lower temperature than before.

Temperature-sensitive mutations usually arise from a single mutation's effect on the stability of the protein. Temperature-sensitive mutations make the protein just unstable enough to unfold when the normal temperature is raised a few degrees. At normal temperatures (usually 37°C), the protein folds and is stable and active. However, at a slightly higher temperature (usually 40 to 50°C) the protein denatures (melts) and becomes inactive. The reason proteins unfold over such a narrow temperature range is that the folding process is very cooperative—each interaction depends on other interactions that depend on other interactions.

For a number of temperature-sensitive mutations it is possible to find (or make) a seond mutation in the protein that will suppress the effects

of the first mutation. For example, if the first mutation decreased the protein stability by removing a hydrogen-bond donor, a second mutation that changes the acceptor may result in a protein with two mutations that is just as stable as the native protein. The second mutation is called a *suppressor mutation*.

## LIGAND-BINDING SPECIFICITY

This is also a compromise (Fig. 2-5).

The specificity of the interaction between a protein and a small molecule or another protein is also a compromise. We've just said that charge–charge and hydrogen-bond interactions don't contribute a lot to the stability of a protein because their interaction in the folded protein simply replaces their individual interaction with water. The same may be said of the interaction between an enzyme and its substrate or one protein and another. However, there is a huge amount of specificity to be gained in these kinds of interactions. For tight binding, the protein and its ligand must be complementary in every way—size, shape, charge, and hydrogen-bond donor and acceptor sites.

Both the protein and the ligand are solvated by water when they are separated. As the two surfaces interact, water is excluded, hydrogen bonds are broken and formed, hydrophobic interactions occur, and the protein and ligand stick to each other. As in protein folding and for the same reasons, the hydrophobic interaction provides much of the free energy for the association reaction, but polar groups that are removed

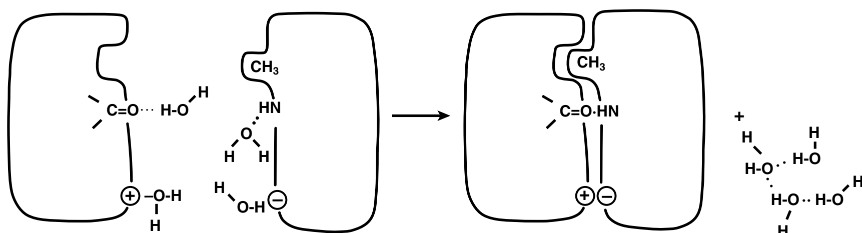



**Figure 2-5**
The **ASSOCIATION** of two molecules uses the same interactions that stabilize a protein's structure: hydrophobic interactions, van der Waals interactions, hydrogen bonds, and ionic interactions. To get the most out of the interaction, the two molecules must be complementary.

from water by the interaction must find suitable partners in the associated state.

Consider what happens when a nonoptimal ligand binds to the protein. The binding of this modified ligand is much weaker not because it's not the right size to fit into the protein-binding site, but because the complementary group on the protein loses a favorable interaction with water that is not replaced by an equally favorable interaction with the ligand (Fig. 2-6).

As with the formation of secondary structure, the multiple, cooperative hydrogen bonds that can be formed between the ligand and the protein may be stronger and more favorable than hydrogen bonds that the ligand might make to water. Hydrogen bonding may, in fact, make some contribution to the favorable free energy of binding of ligands to proteins.

## GLOBAL CONCLUSION

Now that you understand the basis for the interactions between functional groups in water, you also understand the basis for most interactions: DNA–DNA, DNA–RNA, DNA–protein, RNA–protein, protein–protein, protein–ligand, enzyme–substrate (Get the picture?), antibody–antigen, protein–chromatography column—it's all the same stuff.

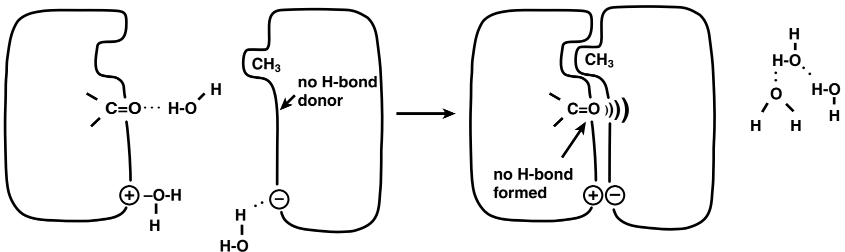



**Figure 2-6**
**SPECIFICITY** in the association of two proteins or a protein and a small molecule results from the requirement that the two interacting molecules must be complementary—complementary in charge, hydrogen bonding, and hydrophobic patches as well as shape. If any of the possible interactions are not satisfied, the strength of the interaction suffers.

· C H A P T E R · 3 ·

# MEMBRANES AND MEMBRANE PROTEINS

•

**General Membrane Function**

**Membrane Composition**

**Phospholipid Bilayer**

**Membrane Structure**

**Posttranslational Modification**

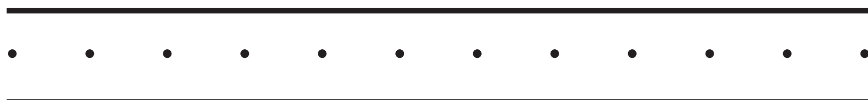
**Membrane Fluidity**

**Diffusion in Membranes**

**Movement of Ions and Molecules Across Membranes**

**Transport Across Membranes**

**The Nernst Equation**



## GENERAL MEMBRANE FUNCTION

1. Separates one area of the cell from another
2. Provides a diffusion barrier
3. Concentrates membrane-associated molecules
4. Enables ion and concentration gradients

Membranes separate one part of the cell from the other. Proteins and other molecules can be localized in the membrane. Membrane localization concentrates the molecules and makes it easier for them to find each other (two-dimensional diffusion) than it is for two molecules in solution (three-dimensional diffusion). Because most molecules can't pass through the membrane by themselves, the cell machinery can create con-

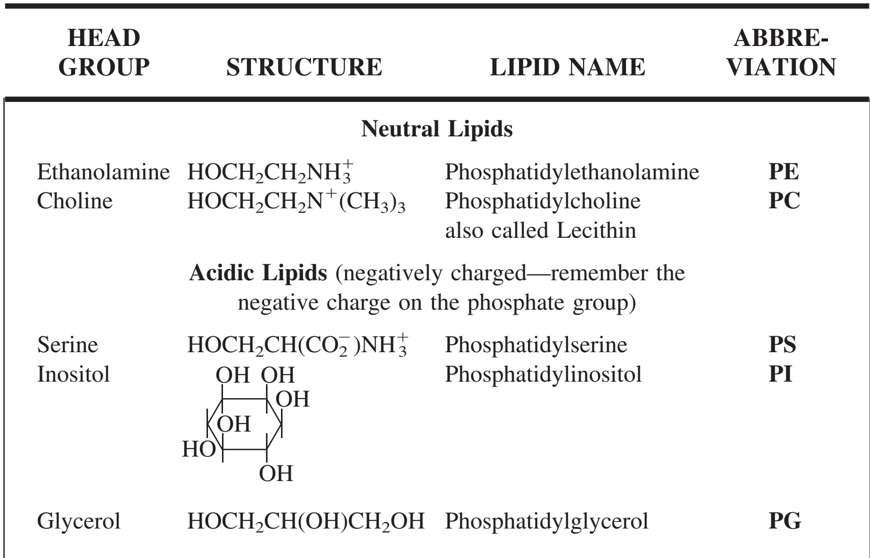
centration gradients across membranes by pumping specific molecules out of the cell and/or by allowing specific molecules into the cell. As we'll see later, these gradients are a source of energy for the cell and can be used for signaling.

## MEMBRANE COMPOSITION

This includes negative phospholipids (PG, PS, PI), neutral phospholipids (PC, PE, sphingolipids), cholesterol, and asymmetric structure.

Lipids are biological molecules that are soluble in certain organic solvents (whether or not something is a lipid is operationally defined by the solubility). Lipids include a variety of molecules such as triglycerides, phospholipids, and cholesterol. The major type of lipid in membranes is the phospholipid. They're called phospholipids because they all contain a phosphate diester.

### COMMON PHOSPHOLIPIDS



| HEAD GROUP | STRUCTURE | LIPID NAME | ABBREVIATION |
|---|---|---|---|
| | **Neutral Lipids** | | |
| Ethanolamine | $HOCH_2CH_2NH_3^+$ | Phosphatidylethanolamine | **PE** |
| Choline | $HOCH_2CH_2N^+(CH_3)_3$ | Phosphatidylcholine also called Lecithin | **PC** |
| | **Acidic Lipids** (negatively charged—remember the negative charge on the phosphate group) | | |
| Serine | $HOCH_2CH(CO_2^-)NH_3^+$ | Phosphatidylserine | **PS** |
| Inositol | (cyclohexane ring with OH, OH, OH, OH, HO, OH) | Phosphatidylinositol | **PI** |
| Glycerol | $HOCH_2CH(OH)CH_2OH$ | Phosphatidylglycerol | **PG** |

The other phospholipids that you may encounter are based on sphingosine. They are derived from serine instead of glycerol but the concept is the same. They have two long, fatty acid chains, a phosphate diester, and a choline-like charged group. This is a neutral lipid.

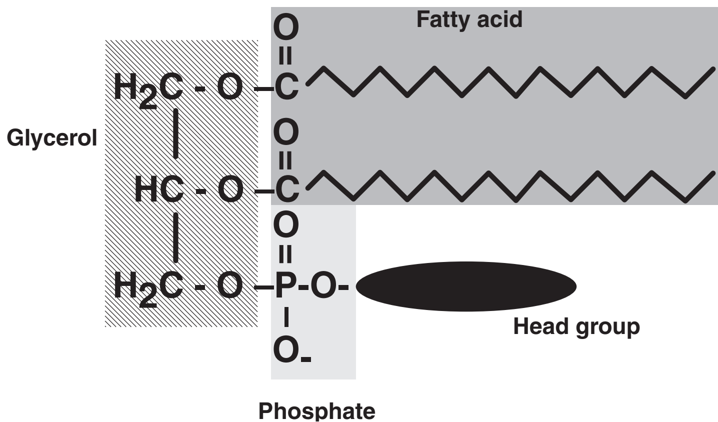
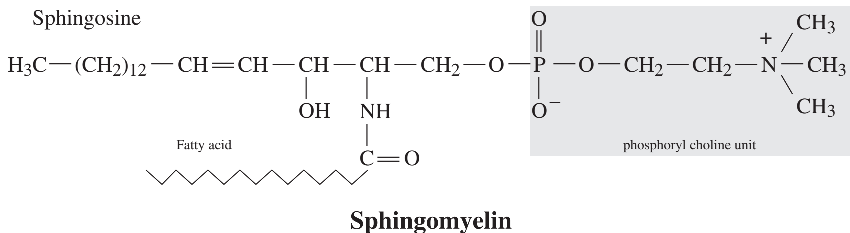



**Figure 3-1 Structure of Phospholipids**
The hydrophobic tail is provided by long-chain fatty acids attached to a glycerol backbone. The head group contains oxygen and may be positively charged or neutral. The name of the phospholipid is dictated by the head group. The head and tail are attached through a phosphate diester.




**Sphingomyelin**

*Glycolipids* are derived from sphingosine, but have a sugar unit, such as glucose or galactose attached instead of the choline unit. The carbohydrate can be extended to form more complex structures, including branches. The sugars point out from the cell surface and are involved in cell-cell recognition.

*Cholesterol* is an essential component of mammalian membranes. It is obtained from the diet or can be synthesized from acetyl-CoA.

## PHOSPHOLIPID BILAYER

This consists of two layers: tails inside, heads outside. The hydrophobic part is 30 Å thick.

Phospholipids are detergents; they have a hydrophobic part (the fatty acid tail) and a hydrophilic part (the head) (Fig. 3-1). The phospholipids

associate with each other through hydrophobic interaction, forming two layers (leaflets) of phospholipid (Fig. 3-2). This buries the hydrophobic fatty acid tails and exposes the polar part (head) to water. Because of the thick layer of hydrocarbon, any molecule that may try to penetrate the bilayer must pass through this hydrophobic region. For polar molecules and ions this is very difficult because they must lose the strong interactions they make with water in order to pass through the bilayer.

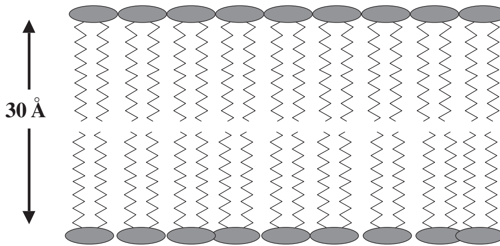



**Figure 3-2**
**PHOSPHOLIPIDS** associate to form a bilayer consisting of a hydrophobic core (phospholipid tails) and a polar surface (phospholipid heads).

## MEMBRANE STRUCTURE

Membranes are asymmetric. Integral membrane proteins can't be washed off. Peripheral membrane proteins can be washed off. Membrane spanning segments and lipid modification (fatty acylation and prenylation), anchor proteins in a fluid bilayer (Singer fluid mosaic model).

The membrane establishes in and out. The membrane is asymmetric because the inner and outer leaflets can have a different lipid composition and contain different proteins (Fig. 3-3). Proteins can be associated with either side of the membrane, or they can pass through the membrane using membrane-spanning segments. The functional part of the protein can be on the cytosolic side, the external side, or even in the membrane itself. A common structure for spanning a membrane is an $\alpha$-helix (but there are examples of sheets spanning a membrane). It takes about 20 amino acid residues arranged in a helix to span to a 30 Å hydrophobic interior of the bilayer.

Proteins that can be removed from membranes by washing them with salt solutions or low pH solutions (disrupts ionic interactions) are called *peripheral membrane proteins*. Proteins that cannot be removed without disrupting the membrane with detergents are called *integral*

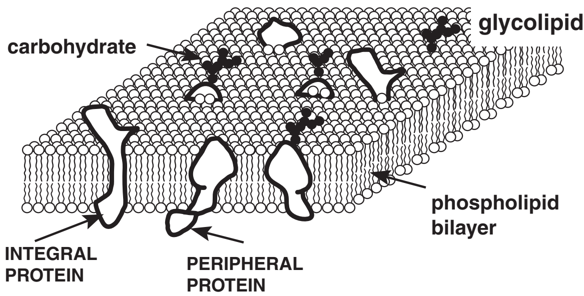



**Figure 3-3**
**FLUID-MOSAIC MODEL** of membrane structure. Proteins and lipids that are embedded in the lipid bilayer diffuse rapidly in the plane of the membrane.

*membrane proteins*. Remember that the distinction between integral and peripheral membrane proteins is operational rather than structural. All proteins that pass through the membrane one or more times will be integral membrane proteins, but not all integral membrane proteins will pass through the membrane—it depends on whether or not the protein can be removed by salt or low pH washes. Peripheral membrane proteins associate with the membrane or, more usually, with integral membrane proteins.

## POSTTRANSLATIONAL MODIFICATION

Posttranslational modification can affect membrane association by prenylation (adding C15 or C20 unsaturated hydrocarbons) or fatty acylation (C14 or C16). Glycoproteins and glycolipids on the exterior face of the membrane have carbohydrates attached.

Some proteins can be posttranslationally modified by the addition of prenyl groups. Prenyl groups are long-chain, unsaturated hydrocarbons that are intermediates in isoprenoid synthesis. The *farnesyl* group has 15 carbons, and the *geranylgeranyl* has 20 carbons. They are attached to a cysteine residue near the end of the protein as a thiol ether (Protein-S-R). Other proteins can have a long-chain fatty acid (C14=myristoyl, C16=palmitoyl) attached to the amino terminus as an amide. These fatty acid modifications can increase the association of proteins with the membrane.

Glycoproteins and glycolipids have complex sugar residues attached. Since they are attached in the ER and Golgi compartments, the sugar

coating will point outward from the cell (will be on the outside surface of the membrane). Membrane proteins as well as phospholipids and glycolipids are embedded in the lipid bilayer and move around in the plane of the bilayer very rapidly.

## MEMBRANE FLUIDITY

Increasing fluidity makes lateral diffusion faster. Fluidity increases with increased temperature, increased content of short-chain fatty acids, and increased content of *cis*-fatty acids. Cholesterol increases the fluidity of membranes that are not very fluid, but decreases the fluidity of membranes that are already fluid.

The membrane is a dynamic assembly and things are diffusing rapidly in the plane of the bilayer. The middle of the bilayer has been likened to olive oil. As with oil, cooling the lipid bilayer will cause the hydrocarbons to become more ordered (structured). The side chains pack closer to each other, and the fluidity of the membrane is lower. Things that disrupt the ability of the side chains to pack in a regular fashion make the membrane more fluid (Fig. 3-4). These include high temperature, lipids with shorter chains (<C16), and lipids with *cis*-double bonds. The shorter lipids and the *cis*-double bonds cause the occurrence of holes (packing defects).

Cholesterol has a funny effect on membrane fluidity. Because of its shape, cholesterol prevents long-chain fatty acids from packing close to each other. When cholesterol is added to a membrane composed largely

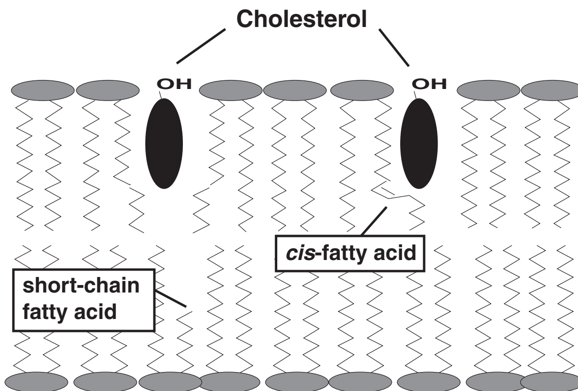



**Figure 3-4**
**MEMBRANE FLUIDITY** is regulated by altering the chain length of fatty acids, the presence of *cis*-unsaturations, and the content of cholesterol.

of saturated, long-chain fatty acids, it will cause the fluidity to increase. However, cholesterol is just the right size to pack into the defects caused by *cis*-fatty acids. In a membrane (like most mammalian membranes) that contains significant *cis*-fatty acids, adding cholesterol will cause the membrane fluidity to decrease.

## DIFFUSION IN MEMBRANES

Lateral diffusion is in the plane of the membrane, and transverse (flip-flop) diffusion is perpendicular to the membrane (through the membrane). Lateral diffusion (in two dimensions) is fast, and transverse diffusion is slow (or nonexistent) except for gases ($CO_2$, $NH_3$) and hydrophobic, uncharged, small molecules (such as cholesterol)

To diffuse rapidly in the plane of the membrane (lateral diffusion), a molecule must simply move around in the lipid environment (including the polar head groups). It need not change how it interacts with phospholipids or with water since it is constantly exposed to pretty much the same environment. Lateral diffusion can be slowed (or prevented) by interactions between membrane proteins and the cellular cytoskeleton. This spatially restricts a plasma membrane protein to a localized environment.

To move through the membrane (change sides or transverse diffusion), a molecule must be able to pass through the hydrophobic portion of the lipid bilayer. For ions and proteins, this means that they must lose their interactions with water (desolvation). Because this is extremely difficult, ions and proteins do not move through membranes by themselves. Small molecules such as $CO_2$, $NH_3$ (but not $NH_4^+$), and water can diffuse through membranes; however, most other small molecules pass through the lipid bilayer very slowly, if at all. This permeability barrier means that cells must develop mechanisms to move molecules from one side of the membrane to the other.

## MOVEMENT OF IONS AND MOLECULES ACROSS MEMBRANES

This requires the participation of a protein transporter. Molecules move spontaneously toward lower concentration (chemical gradient) and opposite charge (electrical gradient). Moving in the opposite direction requires the input of energy.

Because the cell membrane is not permeable to ions and most molecules, the cell can regulate the concentrations of things on either side of the membrane. There are two factors that influence the movement of ions and molecules through a membrane. These are the concentration gradient across the membrane (also called the "chemical potential") and the electrical potential of the membrane.

A *concentration gradient* (chemical potential) exists if the concentration of a given molecule or ion is different on the two sides of the membrane. If you punch a hole in the membrane, the concentration of the molecule will try to equalize itself on the two sides of the membrane (if it is an uncharged molecule).

Normally, cells maintain a slight excess of negative ions inside the cell. This costs energy, but you'll see it's worth it. This uneven distribution of charge across the membrane results in a *membrane potential* (or electrical potential). The membrane potential is negative, indicating that the inside of the cell is negatively charged as opposed to the outside which is positive. It has a normal value of about −0.06 V (−60 mV). As long as the membrane potential is maintained, it will affect how ions move (the movement of molecules with no charge is not sensitive to the membrane potential). Moving an ion toward the opposite charge (moving a positive ion from outside to inside the cell) will be easier than moving the ion toward the same charge.

The membrane potential and the concentration gradient can reinforce each other or they can be in opposition to each other. The total force tending to move a molecule or ion through a membrane is called the *electrochemical potential*. When the concentration gradient and the electrical potential work to oppose each other, the stronger effect wins. If someone forces you to get quantitative (this may be a physiologist rather than a biochemist), see the section on the Nernst equation at the end of the chapter.

## TRANSPORT ACROSS MEMBRANES

*Facilitated diffusion (channel)*: molecule moves down its electrochemical gradient. *Active transport (pump)*: molecule moves up its electrochemical gradient (requires energy input). Pumps use energy (usually ATP hydrolysis). $Na^+$ high outside/$K^+$ high inside.

Because membranes are impermeable to most molecules, you must have a transporter (a protein) in the membrane to help molecules or ions move through it (Fig. 3-5). They are also called *channels* because they behave like selective holes in the membrane. Transporters are selective

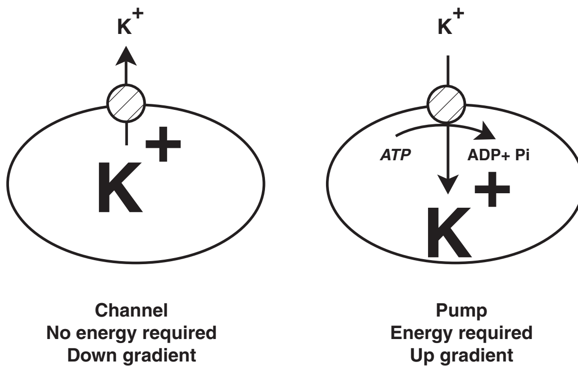



**Figure 3-5**
Moving from high to low concentration **(CHANNEL)** does not require the input of energy. Moving from low to high concentration **(PUMP)** does require some input of energy, usually in the form of ATP hydrolysis of another ion moving down its concentration gradient at the same time.

because they usually allow only one type of ion to pass through the membrane. For example, the calcium channel that releases calcium from the endoplasmic reticulum, releases only calcium but not other divalent cations. What distinguishes a channel from a pump is that the movement of ions through a channel does *not* require any input of energy.

Channels can be gated. This means that something causes the channel to open. For example, a ligand-gated channel opens when a specific ligand binds to a receptor in the membrane. The acetylcholine receptor is a ligand-gated sodium channel that initiates the flow of sodium into the cell (and potassium out) when the receptor binds the neurotransmitter, acetylcholine. Voltage-gated channels open and close in response to changes in the membrane potential. The acetylcholine receptor is also voltage-gated. When the membrane potential becomes positive, it opens the channel further, increasing the rate of membrane depolarization.

Pumps move ions and molecules up their electrochemical gradient. Pumps require energy, usually in the form of ATP hydrolysis. Sodium-potassium ATPase is an example of a pump. Cells maintain a higher concentration of potassium inside the cell than they do outside the cell. Sodium is maintained low inside, high outside. Sodium-potassium ATPase pumps three sodium ions from inside the cell to outside. This is the unfavorable direction—$Na^+$ moves from low concentration to a higher one and against the membrane potential. At the same time, it also

pumps two potassium ions from outside the cell to inside (against the concentration gradient but with the electrical gradient). Both ion movements are unfavorable so that the transport process requires energy.

Pumps work by changing their structure and binding characteristics during the cycle of ATP binding, hydrolysis, and release of ADP and Pi. Exactly when and what triggers these changes will vary from pump to pump, but the essential feature is that there must be a cycle that moves a binding site from one side of the membrane to the other while it also changes the affinity for the ligand. When it faces outside, the sodium-binding site must have a low affinity for sodium. This allows sodium to be released where the concentration is high. The potassium-binding site will have a high affinity for potassium when it faces outside. This allows potassium to be taken up outside where the concentration is low. During the movement to the other side of the membrane (this doesn't happen at the same time for the potassium and sodium sites), the affinity for substrate changes. When it faces inside, the sodium-binding site has a high affinity so that it can take up sodium at the low concentration inside. The potassium site that faces inside has a low affinity so that it can release potassium at the high concentration inside the cell.

Ion gradients can also be used to transport other molecules. During digestion, glucose is concentrated in the intestinal epithelium (moves from low concentration outside to a high concentration inside). The energy for this process comes from coupling the transport of glucose to the cotransport of sodium from outside the cell to inside (down its concentration gradient; Fig. 3-6).

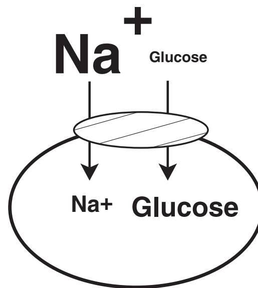



**Figure 3-6**
Intestinal epithelial cells concentrate **GLUCOSE** from the intestinal lumen by using the energy from moving sodium down its electrochemical gradient to provide the energy for moving glucose against (up) its concentration gradient.

## THE NERNST EQUATION

$V = RT/zF \cdot \ln (C_o/C_i)$
$\Delta G_{ion} = zFV = z(23\ \text{kcal/mol/V})V$
$\Delta G_{conc} = -RT\ln (C_o/C_i) = -1.36 \log (C_o/C_i)$
$\Delta G_{total} = \Delta G_{ion} + \Delta G_{conc}$

The Nernst equation tells you quantitatively about the energetics of ions moving through membranes. There are two things you need to consider to decide which direction an ion will move spontaneously or whether an ion movement will require energy or not. The first thing to consider is the concentration difference across the membrane. There is an energy associated with maintaining this. Like other energies we've talked about, it will be a free energy. The free energy difference for having different concentrations across a membrane is the following:

$$\Delta G_{conc} = -RT1n\left(\frac{C_{out}}{C_{in}}\right) = -1.36 \cdot \log\left(\frac{C_{out}}{C_{in}}\right)$$

Here R is the gas constant and T is the temperature (in kelvin). Like the other energetic things we've done (see Chapter 24), using R and T and converting from natural to base 10 logs, the factor in front of the concentration ratio is 1.36 kcal/mol. A concentration ratio of 10 between the outside and inside corresponds to a free energy of −1.36 kcal/mol. The sign here can be confusing—it's a convention. Having higher concentration outside than inside is considered favorable. If something moves from higher to lower concentration, that's in the favorable direction and will contribute a negative ΔG. The ΔG is independent of whether or not the molecule is charged.

If the species has a charge on it, there's another ΔG associated with moving it through a membrane.

$$\Delta G_{ion} = zFV$$

Here z is the charge, F is the faraday (23 kcal/V/mol), and V is the potential difference (voltage) across the membrane. Positive ions ($z = +1$) will move toward places where the charge is negative. Cells (except when they depolarize) are usually negative inside. The membrane potential is maintained by using energy (ATP hydrolysis) to move ions across the membrane. A membrane potential of −60 mV (negative inside) means that the $\Delta G_{ion}$ required to set up this gradient would be (1)*23*0.06 = −1.38 kcal/mol. This would amount to maintaining a factor of 10 difference in the concentration of one positively charged ion

(higher outside than inside). The unequal charge distribution is discussed in terms of volts and is called the *membrane potential*. In reality, multiple ions contribute to the net membrane potential.

The direction that something will move spontaneously is determined by the free energy ($\Delta G$) for that movement, which is going to depend on direction. The $\Delta G$ for moving from out to in will be of the opposite sign of the $\Delta G$ for moving from in to out. When you're trying to decide which way something might move spontaneously (where $\Delta G$ is negative), you have to assume a direction, calculate the $\Delta G$ for that direction, and look at the sign. If it's negative (favorable), things will move in the direction you assumed. If it's positive, things will move in the opposite direction.

For any molecule distributed across a membrane at equilibrium, the overall $\Delta G$ will be the sum $\Delta G_{ion} + \Delta G_{conc}$. Thus, at equilibrium $zFV = -1.36 \log (C_{out}/C_{in})$. If $C_{out}/C_{in}$ is 10, then V will be $-1.36(\text{kcal/mol})(1)/(1 \cdot 23\ \text{kcal/V/mol}) = -1.36/23 = -0.06$ V (or 60 mV). The conclusion is that a ratio of 10:1 for a positive ion outside the cell will create a voltage difference of $-60$mV (negative side). By convention, the potential difference is defined with respect to the inside—a more negative inside will have a negative membrane potential while a more positive inside will have a positive membrane potential.

There are lots of questions that can be asked about your understanding of membrane potential. One common question is to give you a membrane potential, a few concentrations of ions, and ask which way will things move spontaneously. Which way will an ion move? It will move so that the $\Delta G$ is $< 0$ (favorable). Let's do an example. Assume that the membrane potential of the cell is $-60$ mV (inside). This is near the potential that the cell normally maintains and denotes that there are more negative ions inside the cell than outside. Let's also assume that the concentration of $Na^+$ is 20 mM outside and 100 mM inside. Now, if we were to punch a hole in the membrane that would allow $Na^+$ to move (a channel), which way would it move?

The signs of $\Delta G$ are dictated by what you call inside and outside, so let's try to figure out the magnitude of $\Delta G$ and determine the sign of $\Delta G$ by intuition. Because the cell is negative inside, the membrane potential would tend to cause the $Na^+$ to move from outside to inside (+ likes −). If we assume that the $Na^+$ would move from outside to inside, this would correspond to a $\Delta G_{ion} = -(1)23(0.06) = -1.38$ kcal/mol. The sign would be negative because moving inside would be good (remember + likes −). Now you have to consider the concentration gradient. Since the $Na^+$ is higher inside than outside, the concentration gradient would make it harder to move the $Na^+$ from outside to inside (what we assumed originally). The $\Delta G_{conc}$ would be positive and equal to $-1.36{*}\log(20/100) = -1.36{*}\log(20/100) = -1.36(-0.7) = +0.95$ kcal/mol. The easiest thing to do here is to forget whether outside or inside is on top of the

ratio of concentrations. If you make a mistake here, all it will do is change the sign of $\Delta G$, not the magnitude of the number. So be sure that you determine if each $\Delta G$ makes sense. If we had made that mistake here, the $\Delta G_{conc}$ we calculated would be $-0.95$ kcal/mol instead of $+0.95$ kcal/mol. But this must be wrong—we know that with a higher concentration inside, the $Na^+$ would like to move down its concentration gradient. Since we're calculating the $\Delta G$ for moving from outside to inside (we assumed that), then the $\Delta G_{conc}$ must be positive.

We've just found that the $\Delta G_{ion}$ is $-1.38$ kcal/mol for moving $Na^+$ from outside to inside and that $\Delta G_{conc}$ is $+0.95$ kca/mol. Since the total $\Delta G$ is negative ($-1.38 + 0.95$), the $Na^+$ will move in the direction we assumed (outside to inside). Even though the concentration gradient is opposing the movement, the strength of the membrane potential is sufficient to cause the ion to move up its concentration gradient. This requires energy—which in this case must come from maintaining the membrane potential. You've just seen how a pump works.

• C H A P T E R • 4 •

# DNA-RNA STRUCTURE

•

**DNA Structure**

**DNA Stability**

**RNA Secondary Structure**

• • • • • • • • • • • •

## DNA STRUCTURE

Double helix
A = Adenine = purine
T = Thymine = pyrimidine (DNA only)
G = Guanine = purine
C = Cytosine = pyrimidine
U = Uracil = pyrimidine (RNA only)
AT/GC base pairs
Antiparallel strands
Major groove–minor groove
A-, B-, and Z-DNA

The two complementary strands of the DNA double helix run in antiparallel directions (Fig. 4-1). The phosphodiester connection between individual deoxynucleotides is directional. It connects the 5′-hydroxyl group of one nucleotide with the 3′-hydroxyl group of the next nucleotide. Think of it as an arrow. If the top strand sequence is written with the 5′ end on the left (this is the conventional way), the bottom strand will have a complementary sequence, and the phosphate backbone will run in the opposite direction; the 3′ end will be on the left. The antiparallel direc-

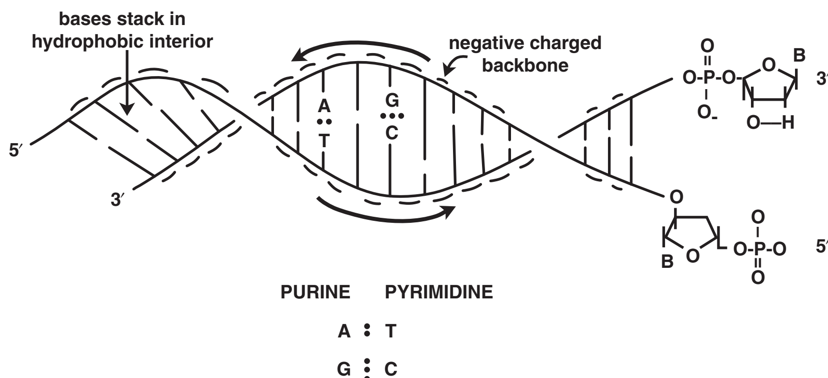



**Figure 4-1 Structural Features of DNA**

tionality of DNA is an important concept (i.e., it always appears on exams). Either of the two strands could be written on top (just rotate the paper by 180°), but if the DNA codes for a protein, the top strand is usually arranged so that it matches the sequence of the RNA that would be made from the DNA (see later). In Fig. 4-2, you're looking at a base pair as it would be seen from above, looking down the helix axis. The DNA double helix has two grooves—the major and the minor. If the helix were flat, the major and minor grooves would correspond to the two different flat surfaces represented by the front and back of the flat sheet. The major and minor grooves are different size because the two strands come together so that the angle between corresponding points on the phosphate backbone is not 180°. Many of the sequence-specific interactions of proteins with DNA occur along the major groove because the bases (which contain the sequence information) are more exposed along this groove.

The structures shown in Fig. 4-1 are for B-form DNA, the usual form of the molecule in solution. Different double-helical DNA structures can be formed by rotating various bonds that connect the structure. These are termed *different conformations*. The A and B conformations are both right-handed helices that differ in pitch (how much the helix rises per turn) and other molecular properties. Z-DNA is a left-handed helical form of DNA in which the phosphate backbones of the two antiparallel DNA strands are still arranged in a helix but with a more irregular appearance. The conformation of DNA (A, B, or Z) depends on the temperature and salt concentration as well as the base composition of the DNA. Z-DNA appears to be favored in certain regions of DNA in which the sequence is rich in G and C base pairs.

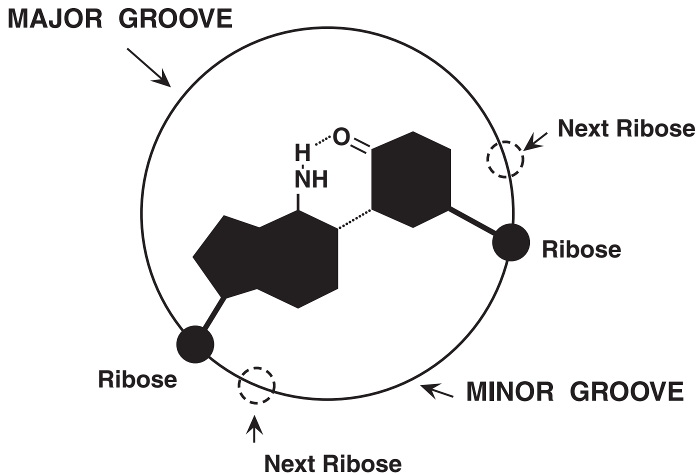



**Figure 4-2**
DNA has a **MAJOR AND MINOR GROOVE** because the bases attach at an angle that is not 180° apart around the axis of the helix. The major groove has more of the bases exposed. Sequence-specific interactions with DNA often occur along the major groove. Since the helix is right-handed, the next ribose shown is above the last one.

## DNA STABILITY

Melting is denaturation.
Annealing is renaturation.
Hydrophobic stacking provides stability.
Intercalating agents stack between bases.

## STABILITY INCREASED BY

Decreased temperature
Increased GC content (three hydrogen bonds)
Increased salt (ionic strength)

The DNA double helix is stabilized by hydrophobic interactions resulting from the individual base pairs' stacking on top of each other in the nonpolar interior of the double helix (Figs. 4-1 and 4-2). The hydrogen bonds, like the hydrogen bonds of proteins, contribute somewhat to the overall stability of the double helix but contribute greatly to the specificity for forming the correct base pairs. An incorrect base pair would not

be able to form as many hydrogen bonds as a correct base pair and would be much less stable. The hydrogen bonds of the double helix ensure that the bases are paired correctly.

The double helix can be denatured by heating (melting). Denatured DNA, like denatured protein, loses its structure, and the two strands separate. Melting of DNA is accompanied by an increase in the absorbance of UV light with a wavelength of 260 nm. This is termed *hyperchromicity* and can by used to observe DNA denaturation. DNA denaturation is reversible. When cooled under appropriate conditions, the two strands find each other, pair correctly, and reform the double helix. This is termed *annealing*.

The stability of the double helix is affected by the GC content. A GC base pair has three hydrogen bonds, while an AT base pair has only two. For this reason, sequences of DNA that are GC-rich form more stable structures than AT-rich regions.

The phosphates of the backbone, having a negative charge, tend to repel each other. This repulsion destabilizes the DNA double helix. High ionic strength (high salt concentration) shields the negatively charged phosphates from each other. This decreases the repulsion and stabilizes the double helix.

Intercalating agents are hydrophobic, planar structures that can fit between the DNA base pairs in the center of the DNA double helix. These compounds (ethidium bromide and actinomycin D are often-used examples) take up space in the helix and cause the helix to unwind a little bit by increasing the pitch. The pitch is a measure of the distance between successive base pairs.

## RNA SECONDARY STRUCTURE

**Stem** A stretch of double-stranded RNA
**Loop:** A loop of RNA
**Hairpin loop:** A very short loop
**Pseudoknot:** Interaction between one secondary structure element and another part of the same RNA molecule

RNA is often depicted as a single-stranded molecule. However, in many RNA's, internal complementarity may result in secondary (and tertiary) structure in which one part of the RNA molecule forms a double-stranded region with another part of the same molecule. There are usually a number of mismatches in these structures. Names have been given to some of these structural features (Fig. 4-3).

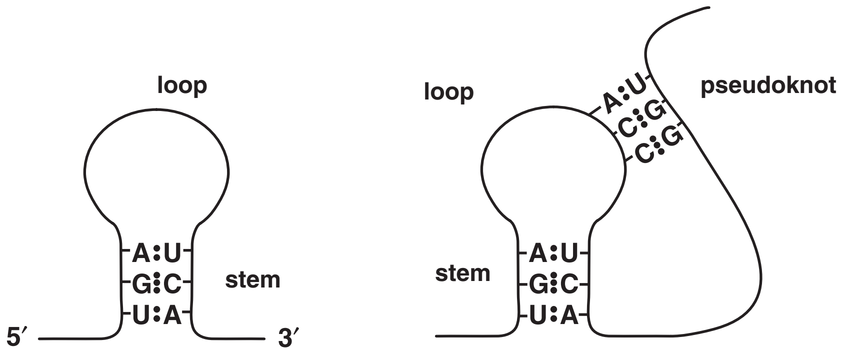



**Figure 4-3 RNA Secondary Structure**
A single molecule of RNA often contains segments of sequence that are complementary to each other. These complementary sequences can base-pair and form helical regions of secondary structure. Interactions between the secondary structures give RNA a significant folded, three-dimensional structure.

• C H A P T E R • 5 •

# EXPRESSION OF GENETIC INFORMATION

•

**Information Metabolism**

**Directions and Conventions**

**DNA Replication**

**Types of DNA Polymerase**

**Recombination**

**Regulation of Information Metabolism**

**Transcription**

**Regulation of Transcription**

**Translation**

**Use of High-Energy Phosphate Bonds During Translation**

• • • • • • • • • • • •

## INFORMATION METABOLISM

DNA → RNA → protein → structure.

Information metabolism provides a way to store and retrieve the information that guides the development of cellular structure, communication, and regulation. Like other metabolic pathways, this process is highly regulated. Information is stored by the process of DNA replication and meiosis, in which we form our germ-line cells. These processes are limited to specific portions of the cell cycle. Information is retrieved by the transcription of DNA into RNA and the ultimate translation of the signals in the mRNA into protein.

Regulation of information metabolism occurs at each stage. The net result is that specific proteins can be made when their activities are needed.

## DIRECTIONS AND CONVENTIONS

The 5′ end of the top (sense) strand is on the left.
Top strand = RNA sequence.
Decoded RNA sequence in 5′ to 3′ direction gives protein sequence in N to C direction (Fig. 5-1).

DNA is a double-stranded molecule in which the structure of the second strand can be deduced from the structure of the first strand. The second strand is complementary (A's in the first strand match T's in the

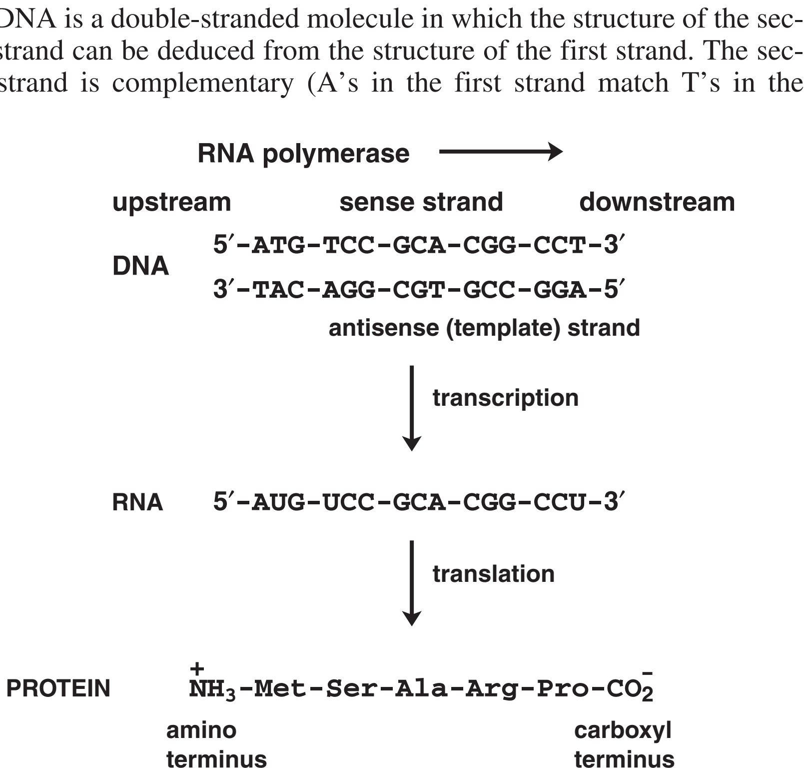



**Figure 5-1**
**DIRECTIONALITIES** in the flow of information from DNA to RNA to protein. All new DNA or RNA chains grow by adding new nucleotides to a free 3′ end so that the chain lengthens in the 5′ to 3′ direction. Protein is made by reading the RNA template starting at the 5′ end and making the protein from the N to the C terminus.

second and G's match C's) and runs in the opposite direction. This means that you don't have to write both strands to specify the structure—one will do.

When you see a sequence written with only one strand shown, the 5′ end is written on the left. Usually this sequence is also identical to that of the RNA that would be made from this piece of DNA when transcribed left to right. The DNA strand that has the same sequence (except U for T) as the RNA that is made from it is called the *sense strand*. The sense strand has the same sequence as the mRNA. The antisense strand serves as the template for RNA polymerase.

The protein synthesis machinery reads the RNA template starting from the 5′ end (the end made first) and makes proteins beginning with the amino terminus. These directionalities are set up so that in prokaryotes, protein synthesis can begin even before the RNA synthesis is complete. Simultaneous transcription-translation can't happen in eukaryotic cells because the nuclear membrane separates the ribosome from the nucleus.

When writing protein sequences, you write the amino terminus on the left. If you have to use the genetic code tables to figure out a protein sequence from the DNA sequence, it is not necessary to write down the complementary RNA sequence first; it's the same as that of the sense strand (the one on top) with the Ts replaced by Us.

## DNA REPLICATION

**Origin** is the beginning.
New chains grow **5′ to 3′**.
**Bidirectional** synthesis.
**Leading strand** = continuous synthesis.
**Lagging strand** = discontinuous synthesis.
**Order of action:**
- Unwinding proteins.
- Single-strand binding proteins.
- Primase makes RNA primer.
- DNA polymerase makes DNA.
- RNAse H removes RNA primer.
- DNA polymerase fills in gaps.
- DNA ligase joins gaps.

Keeping your direction in mind is never a bad idea, but with replication, transcription, and translation it's absolutely essential—these types of questions are just too easy to write, and you'll see them for certain

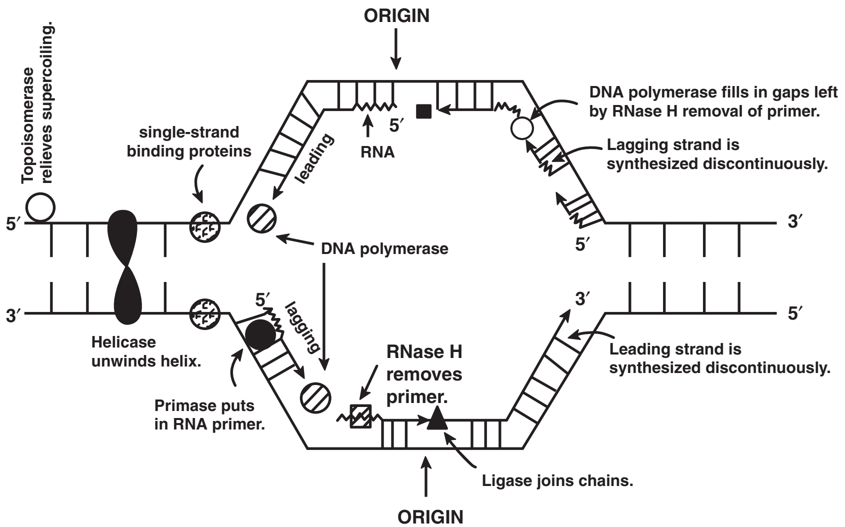



**Figure 5-2**
**DNA REPLICATION** begins at a defined origin, is bidirectional, and is semi-conservative (one new chain, one old chain in daughter DNA), and chain growth occurs in the 5′ to 3′ direction.

(Fig. 5-2).

All DNA polymerases are single-minded—they can do it only one way. Each dNTP (deoxynucleoside triphosphate) is added to the 3′-OH group of the growing chain so that all chains grow from the 5′ end in the dirction 5′ to 3′. Since strands are antiparallel, the template strand is read in the 3′ to 5′ direction. This is true of both DNA and RNA synthesis. Most of what you need to know about DNA replication can be summarized in a single picture.

To remember the order in which things happen, you must understand the structure of chromosomal DNA, and directions. Then it's just a matter of developing a mechanical picture of how things must be done in order to get access to the information and make a copy. Chromosomal DNA is normally packaged around histones. At unique DNA sites called *origins of replication*, unwinding proteins (helicases) unwind the helix in an ATP-dependant manner. Single-strand binding proteins then bind to and stabilize the single-stranded and DNA regions to keep them single-stranded.

In addition to a template (a DNA sequence that specifies the order in which the nucleotides will be jointed), DNA polymerase requires a primer. A primer is a short piece of DNA or RNA that is complemen-

tary to the template and has a free 3′ end onto which the growing strand can be elongated. DNA polymerase can't prime itself—it must have a 3′ end to get started. These primers are actually RNA. A special RNA polymerase (primase) puts them in. Later the RNA primer is removed (by RNAse H in eukaryotes and DNA polymerase I in prokaryotes), and the gaps are filled in by DNA polymerase. This may be a mechanism to enhance the fidelity of DNA replication.

DNA replication proceeds in both directions from the replication origin (bidirectional), which means you need to form two sets of replication complexes. Each replication complex moves away from the origin (in opposite directions), unwinding and replicating both strands at each replication fork.

The strands of DNA are wound around each other like the strands of a rope. As the strands are pulled apart during the movement of a replication fork, this unwinding tends to make the ends of the DNA turn (imagine unwinding the strands of a rope). Since the DNA is very long, twisted, and wrapped around histones, the DNA really can't turn—its ends are rather tied down. The unwinding of one region around the replication forks introduces strain into the regions of DNA that are still double-stranded, tending to make them wind tighter. This is called *supercoiling* of the DNA. Proteins (topoisomerases) are present to relieve the strain associated with helix unwinding by nicking and rejoining the DNA in the double-stranded regions.

At each replication fork there are two DNA polymerase complexes. As the double-stranded DNA is unwound, two template strands are exposed. One of the templates can be replicated in a continuous fashion by DNA polymerase since a continuous synthesis of new strands can occur in the 5′ to 3′ direction as the template strand is exposed. Since all growing chains must be synthesized in the 5′ to 3′ direction, the lagging chain must be continuously reinitiated as new template is exposed. The lagging strand is then synthesized discontinuously, in pieces that must be joined together later.

After synthesis, the RNA primers must be removed, gaps filled in, and the strands joined to give a linear, duplex DNA. New histones are added to the lagging strand (which is now a duplex) while the old histones remain with the leading strand. As the smoke slowly clears, we have a copy of the original DNA.

Since all DNA polymerases require a primer and work only in the 5′ to 3′ direction, there's a problem with replicating the 5′ ends of the DNA. If an RNA primer has to be laid down and later removed, these ends can't get replicated. For bacteria with a circular genome, this isn't a problem. Eukaryotes have specialized structures called *telomeres* at the

ends of the chromosome to solve this problem. The exact details aren't known, but telomeres at the ends of each chromosome consist of a larger number (3000 to 12,000 base pairs) of a tandem (side by side) repeat of a G-rich sequence, $(TTAGGG)_n$, in human DNA. A specialized enzyme, telomerase, that also contains an RNA cofactor is responsible for replication of at least one strand of these telomeric sequences.

## TYPES OF DNA POLYMERASE

Polymerize in 5′-3′ direction
Proofread (check back for mistakes and remove them) 3′ → 5′ exonuclease
Remove RNA primer or remove damaged DNA 5′ → 3′ exonuclease

There's not just one DNA polymerase; there's a whole army. DNA replication actually occurs in large complexes containing many proteins and sometimes many polymerases. In eukaryotic cells we have to replicate both mitochondrial and nuclear DNA, and there are specific DNA polymerases for each. In addition to DNA replication, you have to make new DNA when you repair. Consequently, the function may be specialized for repair or replication. There can also be specialization for making the leading or lagging strand. Some of the activities of DNA polymerases from eukaryotes and prokaryotes are shown in the table on the next page.

DNA polymerases all synthesize new DNA using a template and make the new DNA in a 5′ → 3′ direction (new nucleotides are added to the 3′ end). In addition to making DNA, some of the DNA polymerases can also hydrolyze it. An exonuclease works only on the ends of the DNA (or RNA), and like everything else about DNA, exonuclease activity has a direction too. The 3′ → 5′ exonuclease activity removes nucleotides from the 3′ end (by hydrolyzing the phosophodiester bond). Since the chain grows in the 5′ → 3′ direction, polymerases that have a 3′ → 5′ exonuclease activity can look back over their work and remove what they just put in if it was wrong. This is the only direction that proofreading will work. The 5′ → 3′ exonuclease activity looks forward in the same direction as the new chain is growing. Therefore, it can only remove things that it finds in front of it (such as RNA primers on the lagging strand).

## DNA POLYMERASES FROM PROKARYOTES AND EUKARYOTES

| DNA POLY-MERASE | SOURCE | $3' \rightarrow 5'$ EXONUCLEASE (PROOFREADING) | $5' \rightarrow 3'$ EXONUCLEASE (EXCISION) | ROLE |
|---|---|---|---|---|
| **I** | Prokaryotes | Yes | Yes | Gap repair<br>RNA primer removal |
| **II** | Prokaryotes | Yes | No | Repair |
| **III** | Prokaryotes | Yes | No | Replication |
| α | Eukaryotes | No | No | Nuclear replication<br>Continuous (lagging) strand synthesis |
| β | Eukaryotes | No | No | Nuclear repair |
| γ | Eukaryotes | Yes | No | Mitochondrial replication |
| δ | Eukaryotes | Yes | No | Nuclear replication<br>Continuous (leading) strand replication |
| ε | Eukaryotes | Yes | No | Nuclear repair |

## RECOMBINATION

Recombination rearranges genetic information by breaking and joining DNA.

**Homologous:** Two DNA sequences that are very similar or identical. Homologous recombination occurs between two genes that have very similar or identical sequences.

**Nonhomologous:** Two DNA sequences that are very different. Nonhomologous recombination can occur between two unrelated genes.

**Aligned:** Recombination occurs between the same genes and at the same location within each gene. Gene order is not altered.

**Nonaligned:** Recombination occurs between two different genes. The order of genes is altered by nonaligned recombination.

There are lots of ways of moving genetic information around. All contribute to genetic diversity in the population. The result of recombination can be pictured as breaking two DNA strands into two pieces, swapping the ends, and rejoining. At the level of the individual strands, it's a little more complicated, but for our purposes it's good enough.

Recombination can occur in regions of sequence homology. If these homologous regions correspond to the same position in the same gene, this is an aligned recombination (also called "recombination with equal crossing over"). If all the genes on the two chromosomes are the same, then recombination won't have any affect. But if one of the genes contains a mutation, recombination results in two new chromosomal structures in which different genes are linked to the site of the mutation. Note that in recombination between two chromosomes, no information is actually lost—all the DNA ends up somewhere. However, each offspring receives only one of the two new chromosomes (Fig. 5-3).

If recombination occurs between two regions of homology that are in different genes (unaligned recombination or unequal crossing over), individual genes can be duplicated or lost in the resulting daughter DNA. A good example is the globin gene family. There are several α- and β-globin genes that share some sequence homology. If recombination occurs between two similar (but not identical) genes, the resulting DNA will have been rearranged so that one progeny is a gene or two short while the other offspring has a few too many. Again, no DNA has actually been lost; it's just been redistributed between offspring (Fig. 5-4).

Gene deletion may cause genetic disease if the gene product is essential, and gene duplication, which creates an extra copy of the gene, can

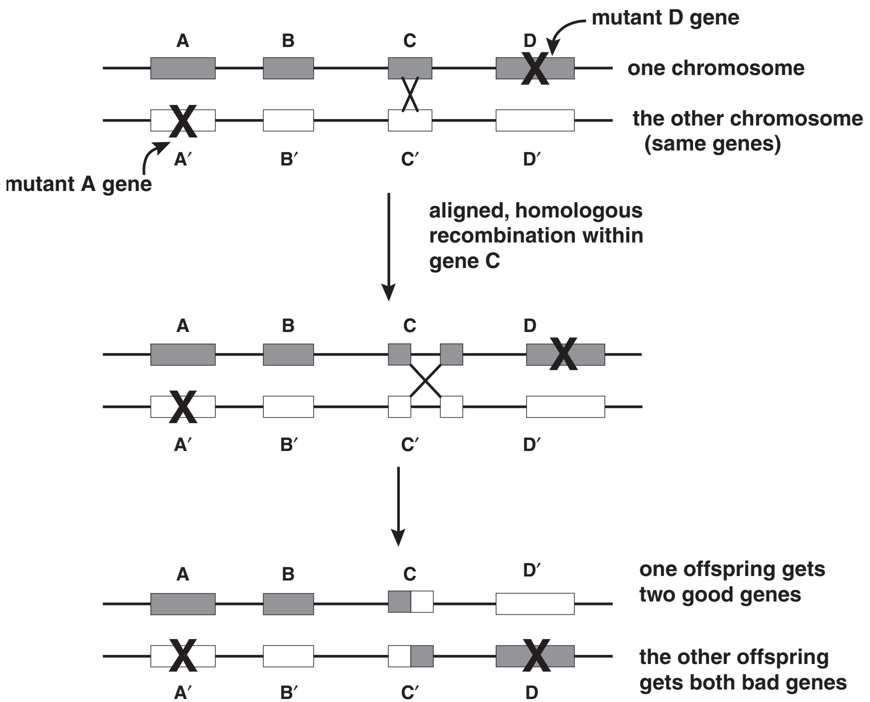



**Figure 5-3**
**ALIGNED, HOMOLOGOUS RECOMBINATION** swaps information between the same genes on two copies of the same chromosome. Genes are not lost or duplicated, nor is their order changed. Different combination of specific alleles (copies of same gene) does occur.

be used to help create new genes by mutation. If you've got two copies of a gene, you can afford to fool around changing one of them, and maybe you'll invent a new and improved gene in the process.

During the generation of genes that direct the synthesis of antibody molecules, recombination within the same chromosome is used to bring distant segments of the gene together and to generate the diversity of recognition sites that allow different antibodies to recognize different antigens. Immunoglobulins consist of two copies of a light chain and two copies of a heavy chain. The heavy and light chains combine to generate the antigen-recognition site. The genes for the different parts of the light chain are arranged in three different clusters: a large number of gene segments for the variable regions of the light chains, a series of joining genes (J), and the constant region. A given variable region is joined to the constant region by a nonaligned recombination that deletes the DNA

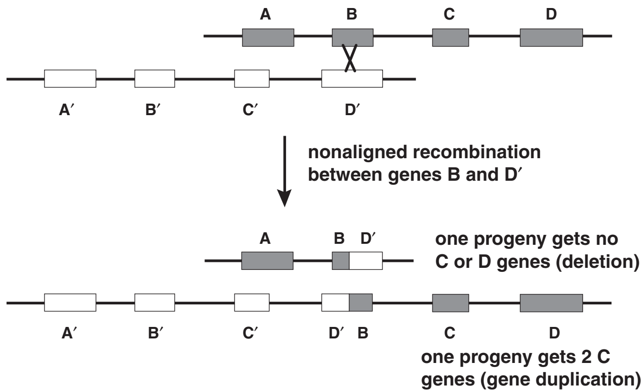



**Figure 5-4 Nonaligned Homologous Recombination**
Genes may be duplicated or deleted when recombination occurs between two different genes on the two copies of the same chromosome. Recombination can occur between two regions of two different genes with some sequence homology.

between the two points of recombination. A similar mechanism is used in making the heavy-chain gene [except there's another type of segment (D) and a few more types of constant regions]. These genetic rearrangements within the same piece of DNA actually cause DNA to be lost. Once the recombination is done it's done, and this cell and its offspring are committed to producing one specific light-chain protein. If the antibody made by a specific cell actually recognizes something foreign, the cell is saved and copied; if not, the cell dies. The large number of different antibody-recognition sites is made possible by the random joining of one of the many variable (V) segments to one of the joining segments by recombination (Fig. 5-5 and 5-6).

## REGULATION OF INFORMATION METABOLISM

**Inducible**: Genes turned on by the presence of a substrate for a catabolic (degradative) pathway.

**Repressible:** Genes turned off by the presence of a product of a biosynthetic pathway.

**Positive regulators (enhancers):** Turn on transcription when a specific effector protein binds to a specific enhancer sequence in the DNA.

**Negative regulators (repressors):** Turn off transcription when a specific effector protein binds to a specific repressor sequence in the DNA.

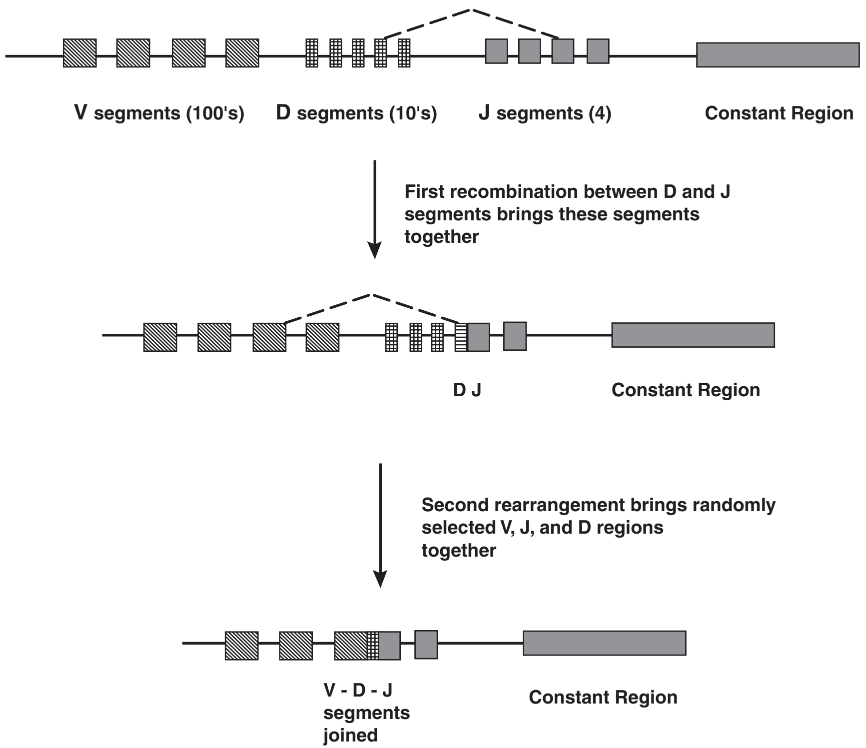



**Figure 5-5 Recombination in Immunoglobulin Genes**
Recombination is used to randomly combine a variable and two joining segments of the immunoglobulin heavy-chain genes. These rearrangements generate a new DNA that codes for an immunoglobulin heavy chain with a single antigenic specificity. A later recombination joins the selected VDJ region to an appropriate constant-gene segment. Similar rearrangements are used to generate the light chain.

One way to control how much of something a cell uses or makes is to control the levels of the enzymes that are required to metabolize it (Fig. 5-7). Whether or not transcription happens is controlled by the binding of specific proteins to the DNA. When they bind to DNA, these proteins can either help or hinder the transcription process. *Positive* and *negative* refer only to the effect a protein has when it binds to the DNA. A positive effect is when the protein binds to the DNA and turns on the transcription of the gene. A negative effect is when the binding of the protein to the DNA turns off transcription.

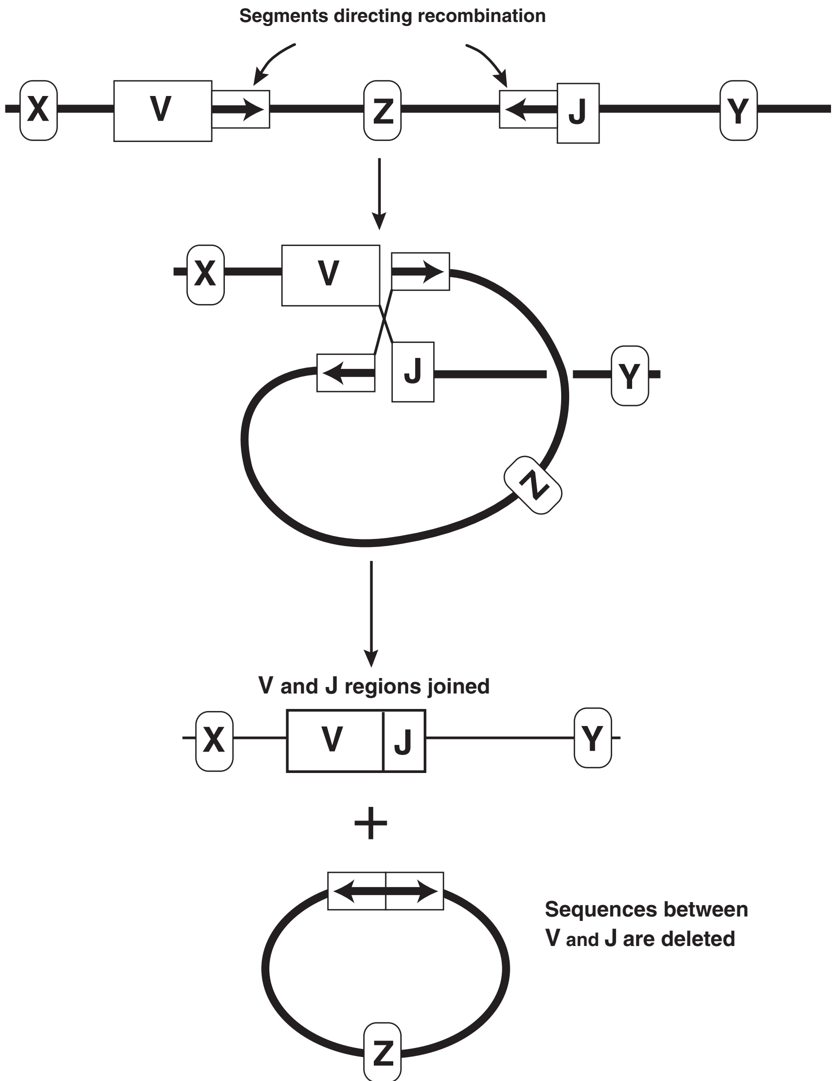



**Figure 5-6**
The **RECOMBINATION THAT JOINS** the V, D, and J gene segments of the immunoglobulin heavy chain occurs between specific regions that precede and follow the V, D, and J regions. Intragene recombination between these regions results in deletion of the intervening DNA and joining of the two segments.

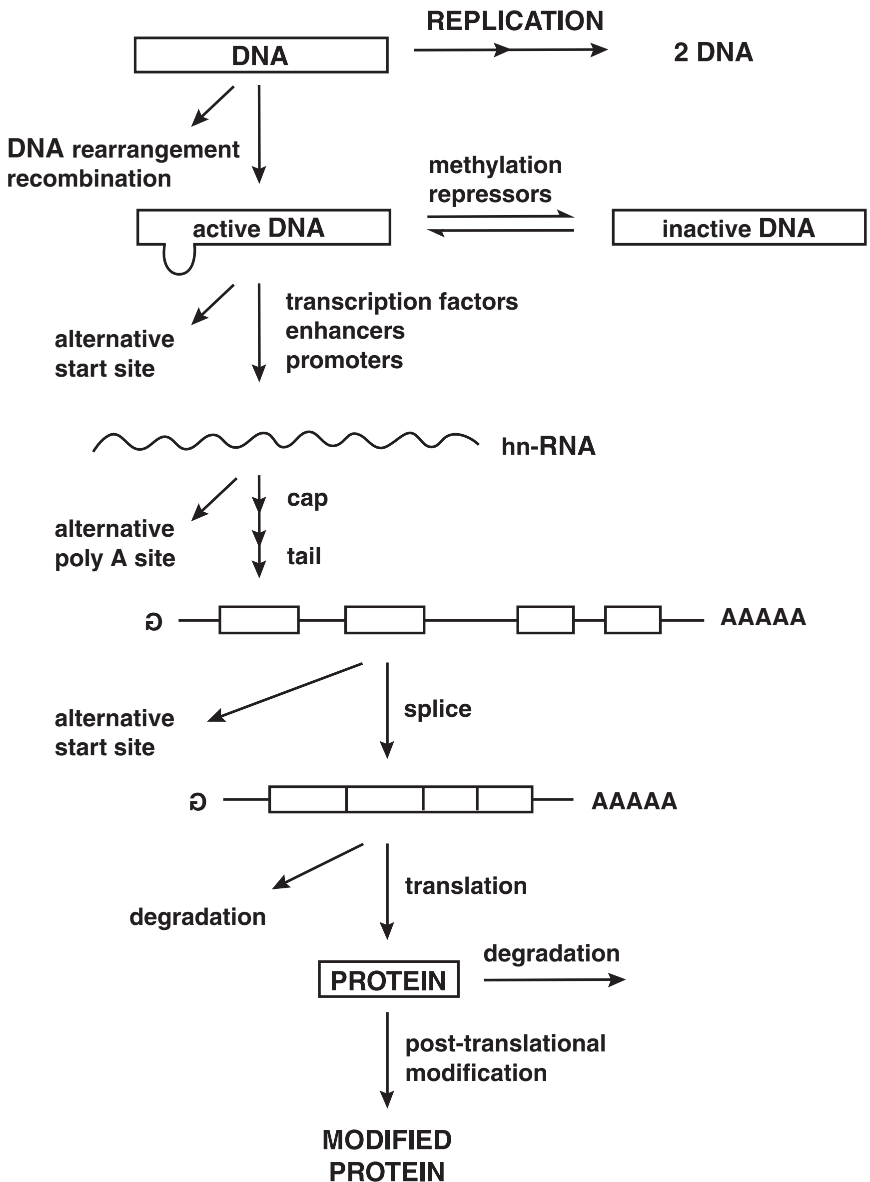



**Figure 5-7**
**REGULATION OF INFORMATION FLOW** from DNA to RNA to protein. Every aspect of the process is controlled, and alternatives are available that affect which information is expressed at what time.

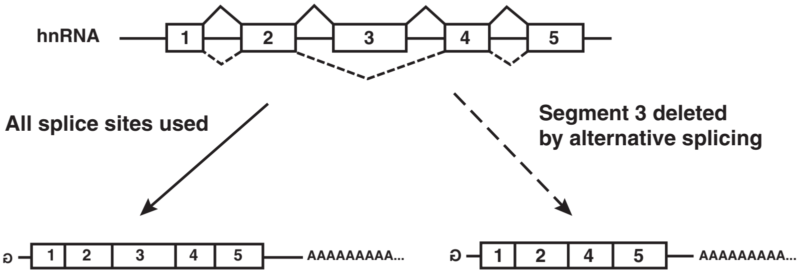
*Inducible* or *repressible* refers to the type of response the system makes to the presence of a metabolite. Inducible genes are turned on when they sense the presence of a metabolite. Usually, this means that the metabolite is a precursor of something the cell needs. If the precusor is present, inducible genes are turned on to metabolize it. Repressible genes are turned off by the presence of a metabolite. These genes are usually involved in the synthesis of the metabolite. If the cell has enough of the metabolite, the pathway is turned off (repressed). If the metabolite is not present, the pathway is turned on.

Operons are clusters of genes located next to each other. The proteins they make are usually required at the same time and for the same overall function. The transcription of genes in an operon is regulated by a common regulatory site(s) on the DNA. Inducible or repressible operons may be created by either positive or negative regulatory elements. The concepts of inducible/repressible and positive/negative control are related but independent. There are then two possibilities for regulation of inducible pathways. If a regulatory protein binds to DNA when it senses the metabolite and then activates transcription, this is a positive way of inducing RNA synthesis. An inducible gene can also function by negative regulation. If a regulatory protein binds to DNA and shuts off transcription when the metabolite is absent, and the protein is released from the DNA when it binds the metabolite, the net effect is the same (increased transcription). Inducible genes can be regulated by either positive or negative effectors. There are also two ways to have repressible genes using positive and negative regulation.

## TRANSCRIPTION

RNA polymerase uses the antisense strand of DNA as a template.
RNA is synthesized in the 5′ to 3′ direction.
The 5′ end is capped with inverted 7-methyl-G.
Poly(A) tail is added.
Introns are spliced out and exons joined.
RNA is exported from the nucleus.
RNA is translated into protein.
RNA is degraded.

RNA polymerase makes a copy of the sense strand of the DNA using the antisense strand as a template (Fig. 5-8). The sequence of the primary transcript is the same as that of the sense strand of the DNA. RNA polymerase needs no primer—only a template. Either of the two DNA strands can serve as the template strand. Which DNA strand is used as the tem-



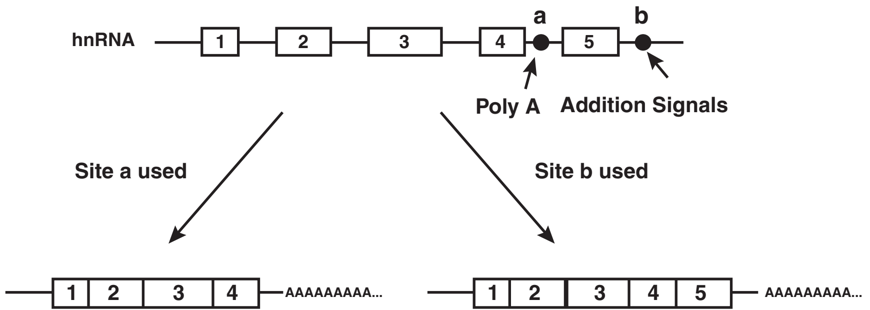




**Figure 5-8**
**ALTERNATIVE SPLICING OR ALTERNATIVE USE OF MULTIPLE POLY(A) SITES** can be used to generate an RNA (and protein) that is missing a portion of the information present in the gene. These mechanisms are useful in generating two proteins from the same gene. A soluble and a membrane-bound form of the same protein can be made from the same RNA by simply splicing out or skipping the membrane-anchor sequences during RNA processing.

plate depends on the direction in which the gene is transcribed. Within the genome, some genes are transcribed left to right while other genes in the same chromosome are transcribed right to left. The direction depends on which strand actually contains the signals that form the binding site for RNA polymerase (the *promoter*). Regardless of the direction of transcription, the new RNA strand is synthesized in the 5′ to 3′ direction and the antisense strand is read in the 3′ to 5′ direction.

After synthesis, the primary transcript (hnRNA—for *heterogeneous nuclear*) is capped on the 5′ end with an inverted G residue. The G is not actually backward or inverted; the *inverted* refers to the fact that the 5′ end is capped by forming a phosphate ester between the 5′ end of the DNA and the 5′-triphosphate of 7-methyl-GTP rather than the normal 5′—3′ bond. This stabilizes the message against degradation from exonucleases and provides a feature that is recognized by the ribosome. Next the message is tailed on the 3′ end with a stretch of A's of variable length (100 to 200 nucleotides). There is not a corresponding set of T's in the DNA template. Poly(A) addition requires a sequence (AAUAAA) in the RNA that helps direct the cleavage of the transcript and the addition of the poly(A) tail by poly(A) polymerase, an RNA polymerase that does not use a template.

To make mRNA, the primary transcript must be spliced to bring the protein-coding sequences (exons) together and to remove the intervening sequences (introns). The splice signals consist of a 5′ and a 3′ set of sequences that are always found at splice junctions. However, this is generally believed to provide too little information to recognize a splice site specifically and correctly. Some sequences in the intron are also important.

After synthesis, the mRNA exits the nucleus through a nuclear pore and proceeds to the ribosome for translation into protein. Competing with export and translation is the process of message degradation by cellular ribonucleases. The competition between degradation and translation provides another mechanism to regulate the levels of individual messages.

## REGULATION OF TRANSCRIPTION

Template availability
Methylation
Exposure of DNA
Attraction for RNA polymerase
Promoter: TATA, CAT
Transcription factors
Alternative Poly(A) tailing
Alternative splicing
Alternative translation start

As goes RNA, so goes protein. Higher levels of mRNA are associated with higher levels of the encoded protein. There is a definite need to regulate the amounts of different proteins during development, differ-

entiation, and metabolism, so there are a lot of controls on the synthesis and degradation of RNA. Control of some sort is exerted at virtually every step of mRNA synthesis.

• **TEMPLATE AVAILABILITY:** The DNA template must be available. This may be controlled by DNA methylation, histone arrangement on the DNA, and interactions of the DNA with the nuclear matrix (a catch phrase for a bunch of protein that's always found in the nucleus).

• **ATTRACTION FOR RNA POLYMERASE:** RNA polymerase binds to DNA at specific sites (called *promoters*) to initiate transcription. A major site is the TATA box (named for the consensus sequence[1] that is often found there) that is located about 25 nucleotides upstream (on the 5′ side) of the translation start site. Not all genes have TATA boxes, and not all promoters have the same efficiency—some are better than others. For many genes, there are other DNA sequences that regulate transcription by binding specific proteins (transcription factors). These transcription factors (enhancers and repressors) may help or hinder transcription. The transcription factor binding sites may be located at varying distances from the transcription start site, and a given promoter region may be affected by more than one of these enhancer or repressor sites. The binding of transcription factors to a specific site on the DNA regulates the transcription by enhancing or inhibiting the formation of the complex structure that is required to initiate transcription. The rules of this regulatory game are not totally sorted out. Transcription factor binding is important in the tissue-specific expression of an mRNA, the regulation of expression during development, and who knows what else.

Eukaryotes have a specific signal for termination of transcription; however, prokaryotes seem to have lost this mechanism. Once started, RNA polymerase keeps going, making a primary transcript [pre-mRNA or hnRNA (for *heterogeneous nuclear*)] until far past the end of the final mRNA message.

• **POLY(A) TAILING:** Most RNAs that code for protein are poly(A)-tailed. Having a poly(A) tail helps direct the RNA to the cytoplasm and may increase the stability of the message. One mechanism of regulation of transcription involves the alternative use of different poly(A) addition sites. Some genes have more than one poly(A) addition signal. Which signal is used can depend on the type of cell or the stage of development,

[1] Consenus sequences are sequences that agree with each other more or less. Often there are a few differences found among the different genes that might have a given consensus sequence. It can be viewed as an "average" sequence.

or it can be used to make two kinds of protein from the same message. Alternative poly(A) addition site usage has the same effect as alternative splicing, except that it deletes terminal exons from the message and creates proteins with different COOH-terminal sequences.

• **ALTERNATIVE SPLICING:** Most primary transcripts must by spliced to connect the proper exons. Some genes contain alternative splice sites that can be used to bring two different exons together and make different gene products depending on need. Alternative splicing changes the sequence of the actual protein that's made. It's useful for making two proteins that share a common sequence. For example, during immunoglobulin synthesis, IgM is made in two forms. One has a membrane-spanning domain so that the IgM with its antigen-recognition site is anchored to the cell plasma membrane. The other form simply lacks the membrane anchor and is secreted in a soluble form. These two forms of the IgM molecule are generated by using alternative splice sites. If the membrane-spanning region is spliced out, the protein loses the ability to bind to the membrane.

• **ALTERNATIVE START SITES:** If all of the above didn't provide enough diversity, some messages contain two AUG initiation codons separated by some intervening information. Protein synthesis can initiate at either site. This is useful for making proteins with or without $NH_2$- terminal signal sequences.

## TRANSLATION

Translation reads the RNA template in the 5′ to 3′ direction.
The amino terminus is synthesized first.
AUG = start = Met in eukaryotes and fMet in prokaryotes.

Protein synthesis (translation) is a two-component system—a system for activating individual amino acids into a chemically reactive form and a system that directs exactly which amino acid is to be used when (Fig. 5-9).

Activation of individual amino acids occurs in the synthesis of aminoacyl tRNA. This process burns two ATP equivalents (forms pyrophosphate and AMP) and connects a specific amino acid to a specific tRNA.

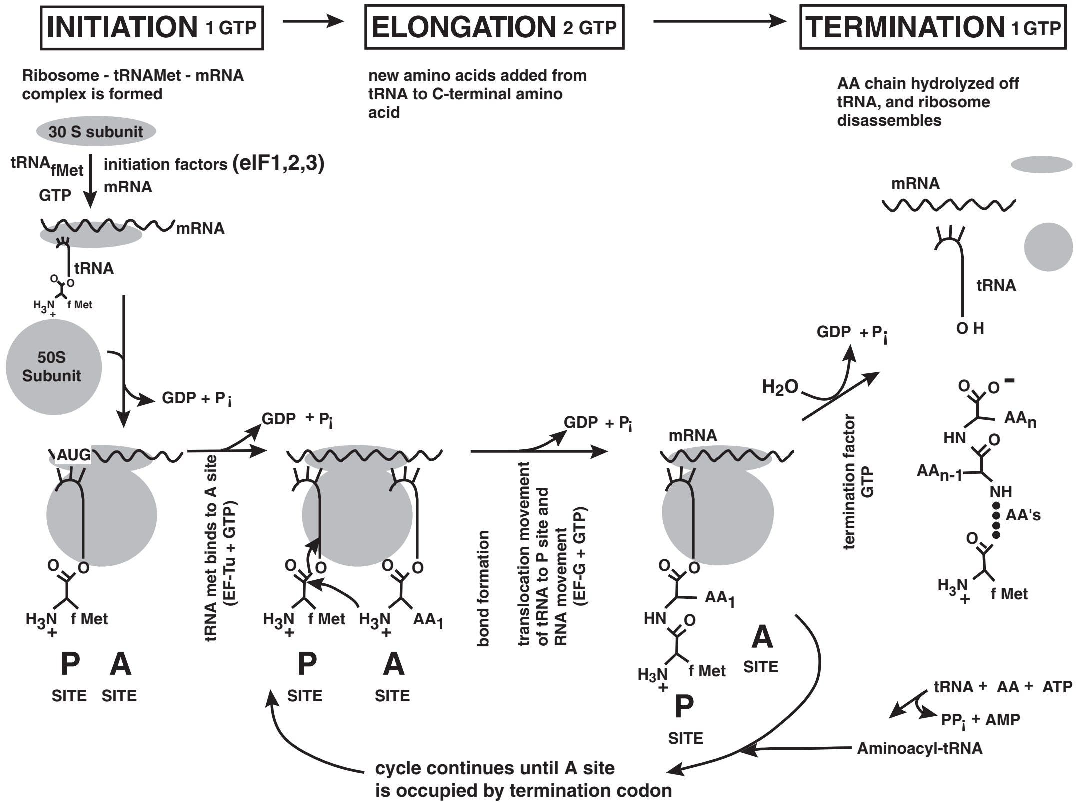
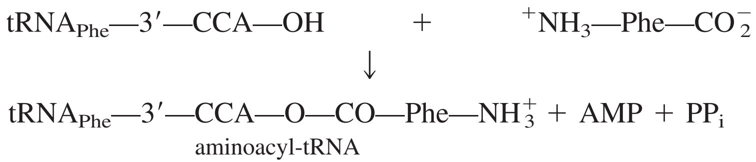



**Figure 5-9 Translation**




The tRNA synthetases may provide a check to make sure that the correct amino acid has been attached to the correct tRNA. If an incorrect amino acid is attached to the tRNA, it will be incorporated into the protein at the position specified by the identity of the tRNA. At least some of the aminoacyl tRNA synthetases have a "proofreading" function that hydrolyzes any incorrect aminoacyl tRNAs (for example, a Val residue attached to an Ile tRNA).

Each tRNA has a different sequence at the anticodon loop that is complementary to the codon sequence in the RNA. The recognition structure that is formed is analogous to double-stranded, antiparallel DNA. If the codon (in the RNA) is GCA (written 5′ to 3′), the anticodon loop in the tRNA would have the sequence UGC (again written to 5′ to 3′).

There are 64 different three-letter codons, but we don't have to have 64 different tRNA molecules. Some of the anticodon loops of some of the tRNAs can recognize (bind to) more than one condon in the mRNA. The anticodon loops of the various tRNAs may also contain modified bases that can read (pair with) multiple normal bases in the RNA. This turns out to be the reason for the "wobble hypothesis," in which the first two letters of a codon are more significant than the last letter. Look in a codon table and you'll see that changing the last base in a codon often doesn't change the identity of the amino acid. A tRNA that could recognize any base in codon position 3 would translate all four codons as the same amino acid. If you've actually bothered to look over a codon table, you realize that it's not quite so simple. Some amino acids have single codons (such as AUG for Met), some amino acids have only two codons, and some have four.

After attachment of amino acids to tRNA, the amino acids are assembled beginning with the amino terminus and proceeding in the direction of the carboxy terminus. The ribosome is the machinery that translates the mRNA into protein. The ribosome is a very complex protein that contains ribosomal RNA as a functional and structural component. The ribosome assembles around the mRNA, and the cap and other signals allow alignment of the mRNA into the correct position. The initial assembly of the mRNA into the ribosome requires association of the small ribosomal subunit with an initiator tRNA (Met or fMet). "Small" is a misstatement, because the small ribosomal subunit is a large, complex assembly of numerous smaller proteins—it's just smaller than the

*large* subunit. This association requires a specific initiation factor and the hydrolysis of GTP. The reactions of translation are driven by the hydrolysis of GTP, not ATP. Throughout the process, elongation factors come and go, GTP gets hydrolyzed, and finally the completed protein is released from the ribosome.

Several key concepts are worth remembering. GTP is used as an energy source for translation, but ATP is used to form the aminoacyl-tRNA. The ribosome effectively has two kinds of tRNA binding sites. Only tRNAMet can bind to the P (for *peptide*) site, and this only occurs during the initial formation of the functional ribosome (initiation). All other aminoacyl-tRNAs enter at the A (for *amino acid*) binding site. After formation of the peptide bond (this doesn't require GTP hydrolysis), the tRNA with the growing peptide attached is moved (translocated) to the other site (this does require GTP hydrolysis).

## USE OF HIGH-ENERGY PHOSPHATE BONDS DURING TRANSLATION

Four high-energy phosphates are used for each amino acid that is incorporated into a protein.

How many high-energy phosphate bonds are required for the synthesis of a protein from amino acids during translation?

**ENERGY REQUIREMENTS FOR THE SYNTHESIS OF A 100-RESIDUE PROTEIN:**

| | | |
|---|---|---|
| 100 aminoacyl-tRNAs | (2 P each) | 200 |
| Initiation complex | (1 P each) | 1 |
| 99 tRNAs binding to A site | (1 P each) | 99 |
| 99 peptide bonds | (0 P each) | 0 |
| 99 translocations (A to P) | (1 P each) | 99 |
| 1 termination (hydrolysis) | (1 P each) | 1 |
| Total per 100 amino acids | | 400 |

Bottom line: 4 high-energy phosphates used per amino acid incorporated into a protein.

• C H A P T E R • 6 •

# RECOMBINANT-DNA METHODOLOGY

•

**Restriction Analysis**

**Gels and Electrophoresis**

**Blotting**

**Restriction Fragment-Length Polymorphism**

**Cloning**

**Sequencing**

**Mutagenesis**

**Polymerase Chain Reaction**

• • • • • • • • • • • •

Much of what we know about the regulation of information flow (gene expression) has been made possible by the ability to manipulate the structures of DNA, RNA, and proteins and see how this affects their function. The ability to manipulate DNA (recombinant-DNA methods) has generated a new language filled with strange-sounding acronyms that are easy to understand if you know what they mean but impossible to understand if you don't. Understand?

## RESTRICTION ANALYSIS

Restriction enzymes are sequence-specific endonucleases that cut double-stranded DNA at specific sites.

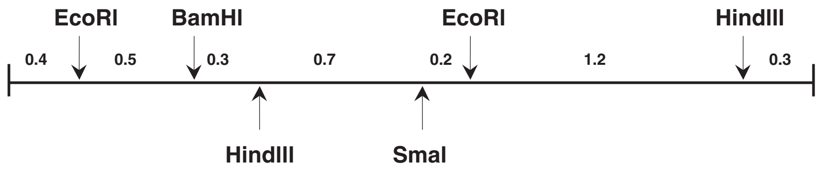
Most useful restriction enzymes cut DNA at specific recognition sites, usually four to six nucleotides in length. There can be multiple restriction sites for a single endonuclease within a given piece of DNA, there can be only one (a unique restriction site), or there can be none. It all depends on the sequence of the specific piece of DNA in question.

Cutting with restriction endonucleases is very useful for moving specific pieces of DNA around from place to place. It's also a useful way to name pieces of DNA. For example, a piece of DNA that is cut from a bigger piece of DNA is often named by size and given a surname that corresponds to the two restriction enzymes that did the cutting—the 0.3-kb EcoRI-BamHI fragment. Restriction enzymes themselves are named for the bacterial strains from which they were initially isolated.

A restriction map shows the location of restriction sites in a given DNA sequence.

When digested with two (or more) restriction enzymes at the same time, most large pieces of DNA give a specific pattern of different-sized DNA fragments depending on the distance separating the different cleavage sites. These different fragments can then be separated by size on an agarose gel. By working backward (biochemists are good at this) from the sizes of the different DNA fragments, it is possible to construct a map that locates the different restriction sites along a given piece of DNA. For example, if we cut the 3.6-kb piece of DNA in Fig. 6-1 with SmaI, we would see two bands on the agarose gel—1.9 and 1.7 kb. This would tell us that the SmaI site is very near the middle of the fragment. We could start constructing our map by putting the 1.7-kb fragment on the left side or the right side—it doesn't matter, and we can't know which is right (or left). In Fig. 6-1, the DNA is arbitrarily put down with the smaller fragment on the right. If we cut with BamHI, we get fragments that are 0.9 and 2.7 kb. Again we wouldn't know whether to put the BamHI site on the right or left of the map, but here it does matter because we already have the SmaI site on the map. The way to decide where to put the BamHI site is to cut with both BamHI and SmaI. Let's say that you get fragments of 0.9, 1.0, and 1.7 kb. Notice that the 1.7-kb fragment is the same size as in the digest with SmaI alone. This tells you that the BamHI site is in the 1.9-kb SmaI fragment, that is, on the left side of our map. By going through this kind of reasoning over and over, it is possible to construct a map of restriction sites along your piece of DNA.

Restriction enzymes that recognize a specific sequence of five nucleotides should cut the DNA, on average, every $4^5$ base pairs (this is the frequency with which a given sequence of five nucleotides would occur by chance), or every 1024 base pairs. As a result, the average size




Restriction Map

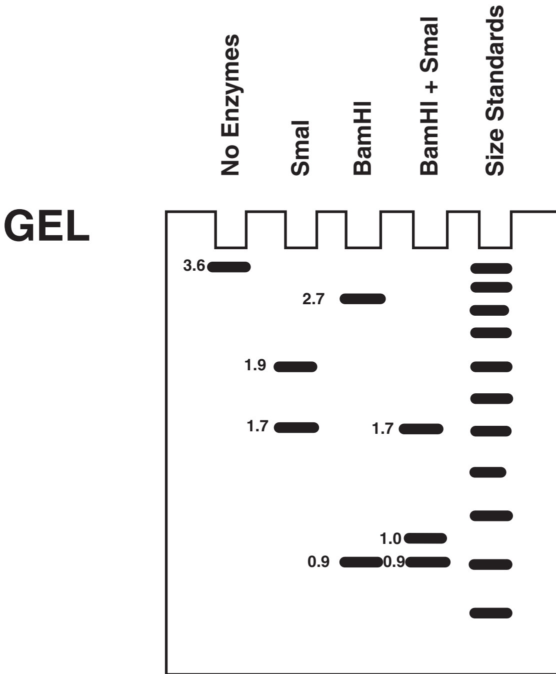
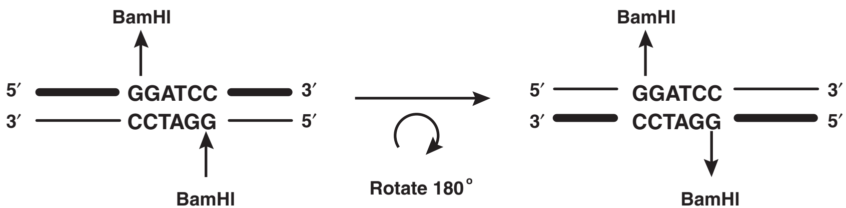



**Figure 6-1**
A **RESTRICTION MAP** is used to identify and locate specific restriction sites on a given piece of DNA. The size of a fragment is determined by running the restriction digest on an agarose gel. Fragments separate by size—the smaller ones move farther toward the bottom of the gel.

of most restriction fragments is near this length. Fortunately, they are not exactly this length, or they wouldn't be very useful.

The sequence of DNA recognized by a specific restriction endonuclease is often palindromic. A *palindrome* is something that reads the same way backward and forward (Fig. 6-2). The sequence of the bottom strand read in the 5′ to 3′ direction is the same as that of the top strand read in the 5′ to 3′ direction. The usual analogy for a verbal palindrome is a sentence that reads the same way backward and forward. "Madam, I'm Adam" is the usual example. It's not exactly the same way for DNA palindromes. The top strand does not read the same from the left as from the right; the top strand read from left to right is the same as the *bottom* strand read from right to left.

BamHI and other restriction endonucleases are dimeric enzymes that bind to a DNA palindrome and cut both strands at equivalent positions. The cut leaves two ends with complementary overhangs that will




The BamHI site is palindromic—rotate it in the plane of the paper by 180° and the recognition sequence doesn't change.

Cutting with a restriction enzyme generates two ends that are complementary to each other

**Figure 6-2**
Useful **RESTRICTION ENZYMES** cleave DNA at symmetrical sites, leaving ends that are complementary.

hybridize to each other. The two ends can be rejoined later, or the fragment can be combined with other pieces of DNA cut with the same restriction endonuclease.

There can be a problem when using a single endonuclease to cut and rejoin different DNA fragments. The DNA fragments that result from cutting with a single restriction enzyme are the same at the two ends.[1] They can (and do) recombine with other pieces of DNA cut with the same restriction enzyme in either of two orientations—forward and backward. Since the DNA between the different restriction sites is not palindromic, the two orientations are not really equivalent, particularly if you're trying to make a protein by translating this region.

This problem can be solved by cutting the two pieces of DNA you want to join with two different restriction enzymes. This way the two ends of the DNA are not equivalent and the two cut pieces can be joined so that the DNA fragments can combine in only one orientation. This approach is very useful for joining different DNA fragments and inserting one specific piece of DNA into another specific piece of DNA. As we'll see a little later, putting inserts (translate as the piece of DNA you're interested in) into vectors (translate as something to carry your DNA around in) is essential to using recombinant-DNA techniques for sequencing, expressing, and mutating your protein (Fig. 6-3).

## GELS AND ELECTROPHORESIS

These separate molecules by size—smaller ones move farther.

Gels are indispensable tools for the molecular biologist. Agarose or polyacrylamide can be formed into hydrophilic polymers that form hydrated gels in water. The gels are usually cast into thin, flat sheets between two plates of glass. The porous network in these gels retards the movement of macromolecules through them so that smaller molecules move faster. The size of the holes in the polymer can be changed by varying the amount of agarose or polyacrylamide in the gel. An electric field

[1] Try rotating the DNA fragment by 180° in the plane of the paper (this means don't pick it up and flip it over—just turn the page upside down). You'll see that the ends look exactly the same as without the rotation. However, the middle, which is not palindromic, will be different.

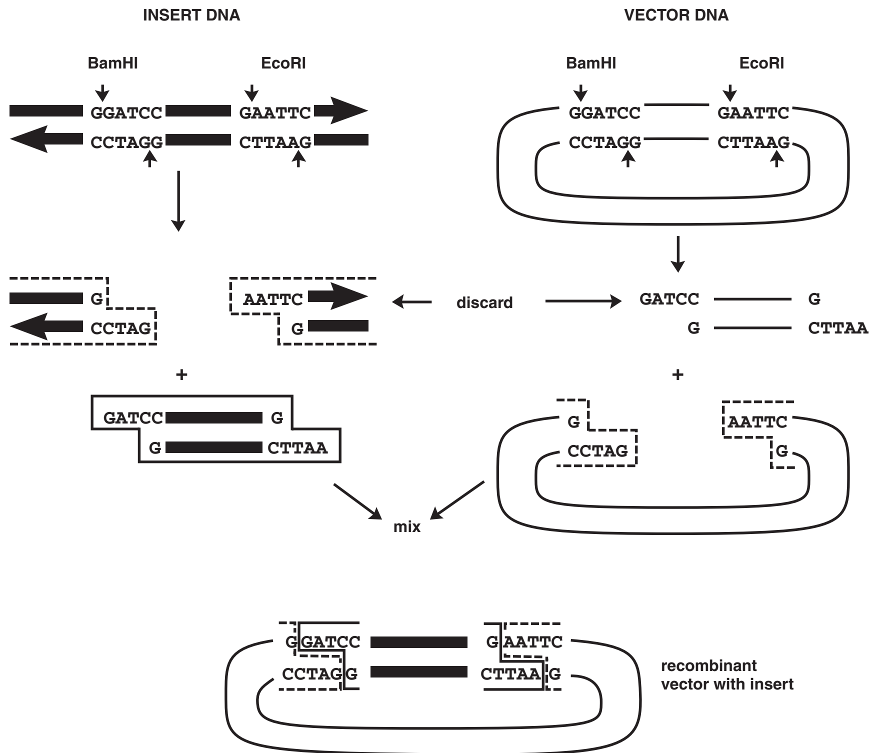



**Figure 6-3**
Generating a **RECOMBINANT-DNA** molecule using restriction enzymes to generate ends that can be joined in a specific fashion.

applied across the ends of the gel causes the macromolecules to move. (DNA is negative and moves to the + electrode, which is at the bottom of the gel.) Molecules of the same size move the same distance, forming a band. Samples of DNA are applied to the top of the gel by putting them in slots (wells) formed during the casting operation. After electrophoresis, the molecules can be visualized by staining. A number of different stains can be used. Commonly, DNA is visualized by staining the gel with ethidium bromide, a dye that becomes intensely fluorescent when it intercalates into DNA. Radioactive nucleic acid fragments can be visualized by placing a piece of x-ray film against the gel. By comparing the distance a given band moves to the mobility of a series of standards of known size, the length of the DNA can be estimated.

## BLOTTING

This means looking at specific molecules on gels even though there are many other molecules present that have the same size.

| MOLECULE ON GEL | LABELED PROBE | NAME OF BLOT |
|---|---|---|
| DNA | DNA | Southern |
| RNA | DNA | Northern |
| Protein | Antibody | Western |

The beauty of blotting techniques is that they let you see only what you're interested in. Take a whole gene's worth of DNA and make fragments with a restriction enzyme. Then separate these fragments by size on an agarose gel. Since there's lots of DNA in a genome, there will be lots of different DNA fragments of almost every size. Usual staining methods would show only a smear over the whole gel. What blotting techniques allow you to do is to detect only the molecules you've interested in.

After separating the molecules based on size, all the DNA fragments are transferred from the agarose gel to a piece of nitrocellulose paper.[2] The paper is actually placed against the gel, and the DNA molecules in the gel migrate from the gel to the paper, where they stick. The paper is then removed and heated to denature the DNA (it still sticks to the paper), and then the blot is cooled in the presence of a large excess of a radiolabeled, single-stranded DNA molecule (the probe) that contains the complement of the specific sequence that you want to detect. DNA fragments on the paper that contain sequences complementary to sequences in the probe will anneal to the radiolabeled probe. The excess probe is washed off, and the blot is placed against a piece of film. Only DNA fragments that have annealed to the probe will be radioactive, and a band will "light up" on the film everywhere there was a DNA molecule that contained sequences complementary to the probe. Conditions of hybridization (salt and temperature) can be changed to make the hybridization more selective (this is called *increased stringency*) so that the extent of sequence complementary between the probe and the DNA that is detected must be quite high.

[2] Special paper that actually reacts chemically with the DNA to cross-link it to the paper can also be used.

As long as the probe can find enough homology, it will stick (anneal) to DNA fragments on the blot that are longer or shorter than the probe itself. In the example shown in Fig. 6-4, the DNA fragment of interest (center lane) shows up as a single band. In this sample, there is only one size of DNA that has a sequence complementary to the probe sequence. In the digest of genomic DNA, two bands light up with this probe. In the genomic DNA, the probe sequence occurs in two different EcoRI fragments of different size. This could mean that there is sequence homology between two different genes (coding for two different proteins) or that an EcoRI restriction site is missing in one of the two copies of the gene present in the genome, reflecting a heterozygous gene pattern (in which the gene is different on each of the two diploid chromosomes).

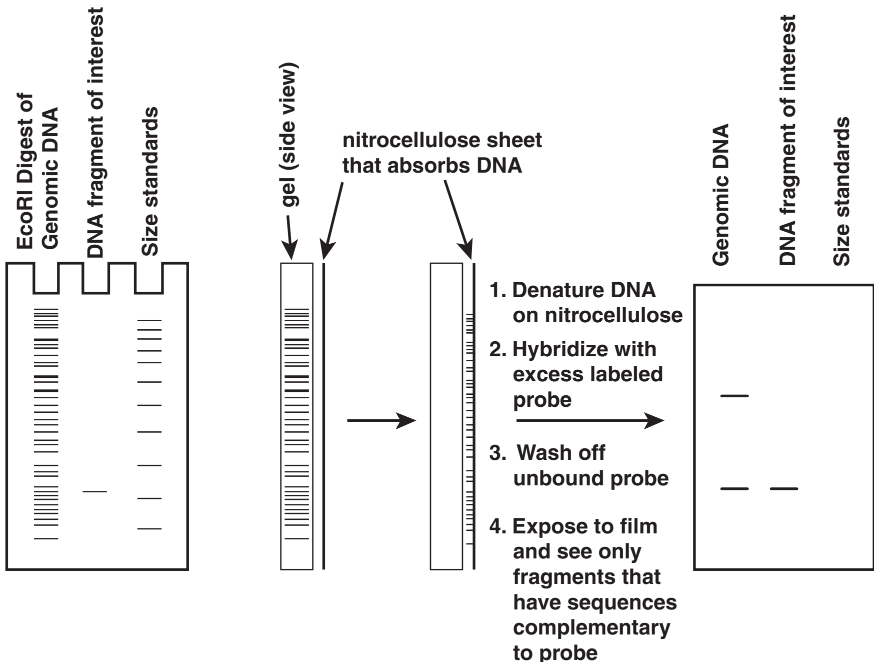



**Figure 6-4**
**BLOTTING** is a method to detect specific DNA (or RNA) fragments that contain sequences that are complementary to sequences in the labeled probe molecule. Only a few of the many DNA fragments on a gel will contain the sequence of interest, and only these will be seen (light up) on the blot. Specific proteins can also be visualized by blotting techniques using a specific antibody to detect a specific protein.

These blotting techniques are known by the names of compass directions (Southern, Northern, Western). Since Southern is a person's name, there's no logic in how the different blots were named. Southern developed a blot in which DNA on the blot is detected by a labeled DNA probe. It was then fairly logical that the next technique developed, detecting RNA on the blot with a DNA probe, should be called a Northern blot. Then things got carried away with the Western, and now the Southwestern, and so on and so on.

If the gel separates DNA and the DNA is detected with a DNA probe, it is called a *Southern blot*. If RNA is separated on the gel and then detected by a DNA probe, it is a Northern. A Western uses specific antibodies to detect specific protein molecules on a blot of a protein gel. In the Western blot, the role of the DNA probe is filled by an antibody that recognizes a specific protein.

## RESTRICTION FRAGMENT-LENGTH POLYMORPHISM

RFLP is a Southern blot used to detect genetic disease.

For the diagnosis of genetic disease, some specific way of detecting a single mutation in DNA from the fetus must be used. The most obvious way to do this would be to use a restriction enzyme that cuts the wild-type sequence but does not cut the mutant sequence (or vice versa). A restriction site right at the site of the mutation would come in handy. If the fetal DNA has the normal sequence, the DNA will be cut and the restriction pattern will be identical to the wild type. If not, not. For many genetic diseases, the mutation does not conveniently occur right at a restriction site. However, in many cases, it just happens that the mutation that's being diagnosed is associated with another, nearby mutation (polymorphism) that does alter some endonuclease cleavage site. This second site is closely linked genetically to the mutation that leads to the genetic disease. If the patient has this secondary restriction site, it's a good bet he or she has the mutation as well. The patterns that are observed when genomic DNA is digested with different endonucleases and the DNA is probed with a specific sequence can then be used to determine if a particular patient is homozygous or heterozygous for the specific mutation—a useful diagnostic tool.

More modern techniques for detecting mutations or differences in DNA sequences in different people can be used. These include PCR (see later) that can distinguish mutations by the length and pattern of

products. DNA chips have specific sequences linked to a solid support in small, solid-state plates (chips). Genomic DNA can be hybridized to the DNA on the chip (if it matches the sequence on the chip). Many (thousands) of sequences can be detected at the same time.

## CLONING

Cloning is manipulating a specific piece of DNA so that it can be used to generate multiple copies of itself or the RNA and protein that it encodes.

## STEPS IN CLONING DNA:

1. Identify the DNA you want.
2. Put the DNA into a vector.
3. Change the sequence of the DNA (this is optional).
4. Put your DNA back into cells.
5. Grow the cells with your DNA/RNA/protein.

There are many different ways to clone a specific piece of DNA, but basically, they all involve (1) identifying and isolating the DNA you are interested in; (2) putting this DNA into something (a vector) to move it around from cell to cell; (3) altering the DNA sequence; (4) introducing this new DNA back into cells; and (5) growing the cells that have your DNA, RNA, or protein. You often do all this randomly to millions and millions of cells and then just select the few cells that got the piece of DNA you're interested in.

• **1. IDENTIFYING YOUR DNA:** There's lots of DNA out there, and finding just the right piece of DNA can be like finding a word in a dictionary that's arranged randomly.[3] The way you go about finding your DNA may depend on the reason for wanting the DNA in the first place.[4] The DNA you want will be contained in the genome of some cell. A frequent strategy is to take all the DNA in a specific cell, cut it into small fragments with restriction endonucleases, and put all these fragments into individual vectors (this is called a *genomic library*). A vector is a piece of DNA that makes it easy to capture other DNA fragments and move them around. Each individual vector will have only one piece of DNA

[3] You're right: "Arranged randomly" contradicts itself.

[4] The Mt. Everest rationale, "Because it's there," is not usually selective enough.

inserted; however, the collection of vectors will contain much of the original cellular DNA. The same can be done with all the cell's mRNA, making cDNA first using reverse transcriptase[5] (this is called a *cDNA library*). The DNA in a genomic library will contain introns, promoters, enhancers, and so forth; however, the DNA in a cDNA library will not contain introns or promoters, but it will contain a strength of A's from the poly(A) tail of the mRNA. After introducing all this DNA into cells under conditions under which each cell will get only one of the DNA fragments in the library, the few cells that have your specific DNA will be identified.

Identification is easiest if your DNA confers some selective advantage to the cell (that is, if it expresses drug resistance or directs a function that is essential for cell survival under the conditions of your culture). Under selective conditions, only the cells with your DNA will survive. Killing cells (or, more mercifully, letting them die) that don't have the desired piece of DNA is called *selection*. A large number of cells (a million or so) can be spread on a culture plate, and only the ones that survive selection will continue to grow. These surviving colonies can be selected individually. If your DNA codes for a protein and you have an antibody to the protein or the protein has an activity that is not present in the host cell, the cells with your DNA can be detected by looking for the cells that make the protein or have the activity. Finding the cells with your DNA by detecting the DNA directly with a Southern blot, or by detecting the protein or RNA product of the gene, is called *screening*.

It's also possible to select your DNA before you put it in the vector. If you know the sequence (or even part of it), DNA pieces (from genomic DNA or cDNA) with this sequence can be purified on a gel and identified by hybridization to an oligonucleotide using a Southern blot. Alternatively, if you know the sequence of the ends of your DNA, you can amplify it specifically by the polymerase chain reaction. There are lots of clever ways to find your DNA.

**• 2. PUTTING YOUR DNA INTO A VECTOR:** Vectors are specialized pieces of DNA used to move other pieces of DNA around. Modern vectors are usually either bacterial plasmids or viral genomes. The act of isolating your DNA in the first place usually involves putting it into a vector and then selecting the vector that has your DNA in it. DNA pieces (called *inserts* when they are placed in a vector) are usually placed in vectors using restriction endonucleases. The vector is cut with two restriction enzymes of different specificity (Fig. 6-3). This removes a

[5] Reverse transcriptase is an enzyme isolated from viruses that contain a genome that is RNA. This viral enzyme makes DNA using RNA as a template.

chunk of the vector DNA and leaves two different ends. You then cut your DNA with the same two enzymes so that it will have the same complementary ends. Restriction enzymes that don't cut in an essential part of the vector or insert must be used. You can also make suitable cloning sites by cutting with just one restriction enzyme; however, because of the palindromic nature of restriction enzyme specificity, the ends of the piece of DNA will be the same. The DNA can then go into the vector in either one of two orientations. Sometimes this matters and sometimes it doesn't. If you want RNA or protein expressed from your DNA, direction will matter if the promoter site is provided by the vector. After mixing your cut DNA with the cut vector under conditions under which the ends will anneal, DNA ligase (and ATP) is added to join the strands with a covalent bond.

Vectors are often designed to contain a drug-resistance marker to aid in the selection of cells that have incorporated your vector (not all cells do). They can also have a variety of other goodies depending on the type of vector. An expression vector is used to express RNA or protein from the DNA, and these vectors usually contain a good promoter region and some way to turn the promoter on and off. Many expression vectors have been engineered to contain a convenient set of unique restriction sites (termed a *polylinker*) near the promoter to make it easy to put your insert in the right place. Sequencing vectors, designed to make it easy to sequence your DNA, usually have a defined site for the sequencing primer to bind that is adjacent to a polylinker region.

**• 3. CHANGING THE SEQUENCE OF YOUR DNA.** The sequence of the DNA can be changed in lots of ways. Large chunks can be deleted or added (deletion or insertion mutagenesis) by mixing and matching endonuclease fragments. Sequences of DNA from one gene can be combined with sequences from another gene (chimeric DNA—named for the Chimera, a mythological beast with the head of a lion, the tail of a serpent, and the body of a goat). If protein product is going to be made from the mutant DNA, care must be taken to preserve the reading frame. Deleting or inserting a number of bases that is not divisible by 3 will cause a shift in the reading of the triplet codons and a jumbling of the protein sequence. Individual nucleotides can be changed at any specific site by the use of site-directed mutagenesis. The reason for changing the DNA sequence is to change the function of the DNA itself or its RNA or protein product.

**• 4. PUTTING YOUR DNA BACK INTO CELLS.** Vectors can be isolated and then added back to cells. DNA can be introduced into cells in a variety of ways: by infection with a virus containing your DNA, by

poking holes in the cells with specific salt solutions, by precipitating the DNA with calcium phosphate and having cells take up the precipitate, by blowing holes in the cells with an electric discharge and allowing pieces of DNA to enter the cells through the holes *(electroporation)*, or by directly microinjecting the DNA with a very small glass capillary.

Not all cells that are exposed to your vector will take it up. That's where selection is helpful. You just kill all the cells you're not interested in.

## SEQUENCING

Sequencing is determining the sequential order of DNA bases in a given piece of DNA.

Sequencing DNA is relatively easy these days, at least for small pieces (a few thousand nucleotides). In the Sanger dideoxynucleotide method, a specific primer is used that is complementary to one of the two DNA strands you want to sequence. The primer can be a vector sequence so that you can sequence any piece of DNA cloned into the vector. The primer is a synthetic oligonucleotide that is radiolabeled (or fluorescently labeled) so that you can see all new DNA molecules that have the primer attached to the 5′ end. Alternatively one of the deoxynucleotides used in the DNA synthesis can be labeled. After denaturing the double-stranded DNA that you want to sequence and annealing the primer, the DNA is elongated from the primer (in the 5′ to 3′ direction) using DNA polymerase. The reaction is run for a short time with all four deoxynucleotides. There will be pieces of DNA that are at all stages of the replication process—the newly synthesized DNA will be of all different lengths. The reaction is then stopped by adding it to four separate tubes, each of which contains a different 2′,3′-dideoxynucleotide. When a dideoxynucleotide is incorporated by the polymerase, the elongation stops (there's no 3′-hydroxyl group on the dideoxynucleotide). Alternatively, you can include a small quantity of a dideoxynucleotide during the polymerase reaction so that some DNA stops when a dideoxynucleotide is added and the rest goes on to stop later on. The trick is that only one of the four dideoxynucleotides will stop the reaction at any given point in the random mixture of newly synthesized DNA. The synthesized DNA is then run on a high-resolution acrylamide gel that can separate DNA molecules that differ in length by one nucleotide. Four lanes are run, one for each type of dideoxynucleotide used to stop the reaction. A ladder of bands will be seen. The shorter bands, at the bottom of the gel, will correspond to termination nearest the primer (near

the 5′ end). The sequence is then read from the bottom (5′ end) to the top (3′ end) of the gel by noting which dideoxynucleotide stopped the reaction at that length (that is, simply which one of the four lanes has a band in it at that length) (Fig. 6-5).

Automated methods for doing this are available that can sequence >500 bp in one run and automatically read out the sequence. Each type of dideoxynucleotide product is marked with a different color so that all four sequencing reactions can be run in one lane of the gel. These automated methods are being used to complete the sequence of the whole human genome.

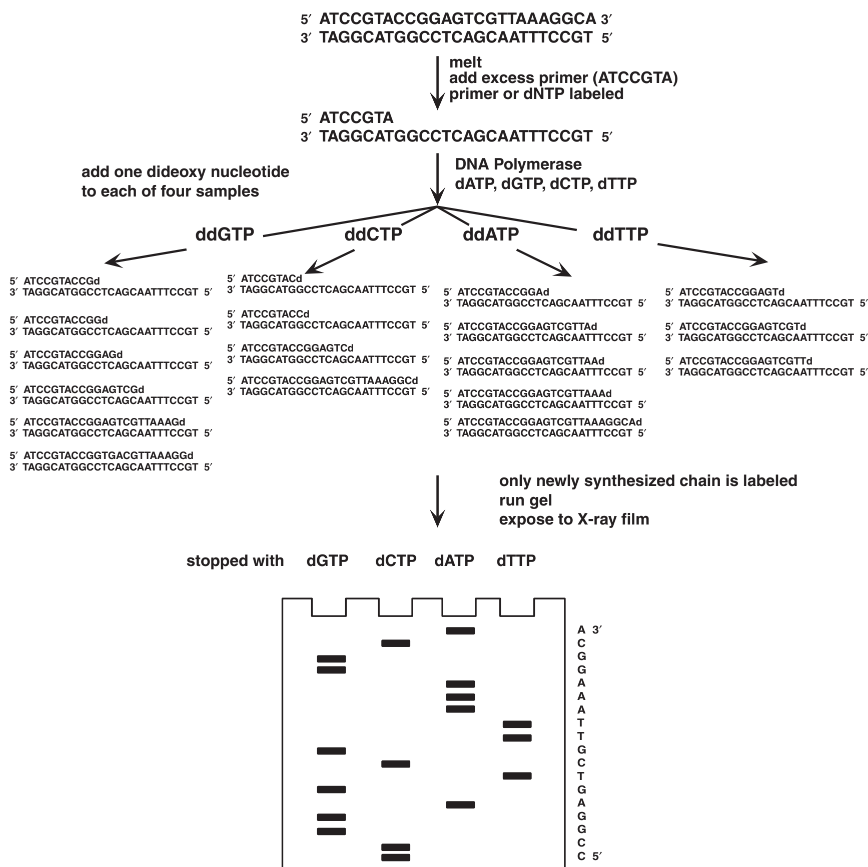



**Figure 6-5 DNA Sequencing**

## MUTAGENESIS

Making a mutant DNA
Deletion: Deletes a hunk of DNA
Insertion: Inserts a hunk of DNA
Site-directed: Modifies a specific nucleotide
Random: Introduces random changes in the DNA

Mutagenesis is used to alter the DNA structure (sequence) in a known way by either deleting nucleotides, inserting nucleotides, or changing a single nucleotide at a defined location. Random mutagenesis of DNA may be performed over the whole piece of DNA by exposing the DNA to chemicals (mutagens) that react with the DNA and change the specificity for base pairing or by using oligonucleotides that contain random, deliberate mistakes in the sequence. Mutants are then selected or screened for changes in function of the protein or RNA product. This technique allows you to define specific amino acids that are essential to the function of the protein and to determine which amino acids can be replaced by which other amino acids and still conserve function.

In deletion or insertion mutagenesis, restriction enzymes are used to generate DNA with a specific fragment missing or with another piece of DNA inserted. This change in the DNA sequence can then be used to produce RNA or protein containing a deletion or insertion of amino acids in the protein. This is useful in determining the gross features of the gene structure that are necessary to preserve a functional gene or to express a functional protein product. For example, a signal sequence that directs the synthesized protein to the mitochondrial matrix can be placed in the sequence of a protein that is normally cytosolic. This mutant protein will be expressed as a mitochondrial matrix protein.

With site-directed mutagenesis, a change in the DNA sequence can be introduced at any specific site. An oligonucleotide is annealed to a single-strand copy of the DNA that you want to mutate. This oligonucleotide contains the correct (complementary) base at every position except the one you want to change. At the mutated position, there is a mismatch. After the oligonucleotide is annealed at the proper position, the DNA is fully replicated using DNA polymerase and then sealed with ligase. When the vector is introduced into the host cell and replicated during cell division, some of the progeny cells will get DNA that has used the mutant strand as the template for DNA replication. There are clever ways to increase your chances of getting only cells containing the mutant DNA. These involve selectively destroying the wild-type (nonmutated)

strand. Site-directed mutagenesis is used to change single amino acids in proteins or single bases in RNA or DNA. The technique has been very useful in determining the function of a specific amino acid residue in enzyme catalysis, binding of a ligand, or stabilizing a protein. It has also been possible to selectively change the activity and specificity of some enzymes using this technique (Fig. 6-6).

## POLYMERASE CHAIN REACTION

PCR amplifies DNA sequences that lie between specific 5′ and 3′ sequences.

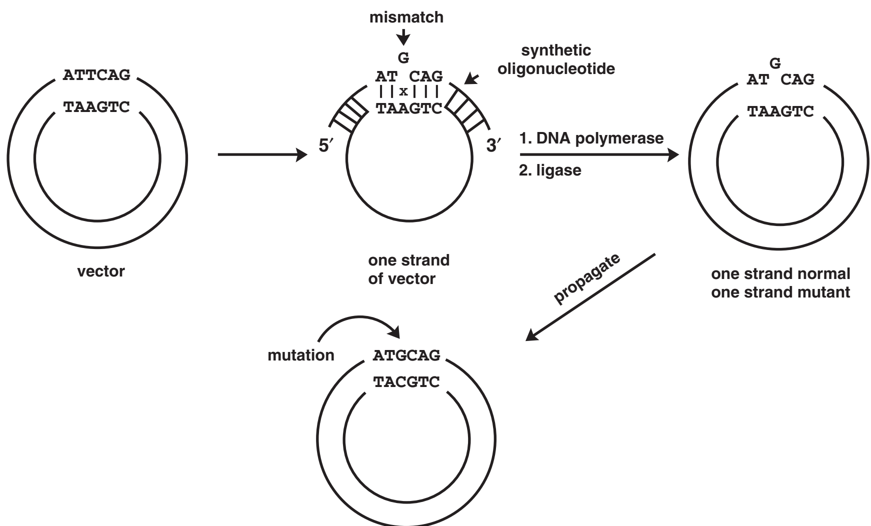
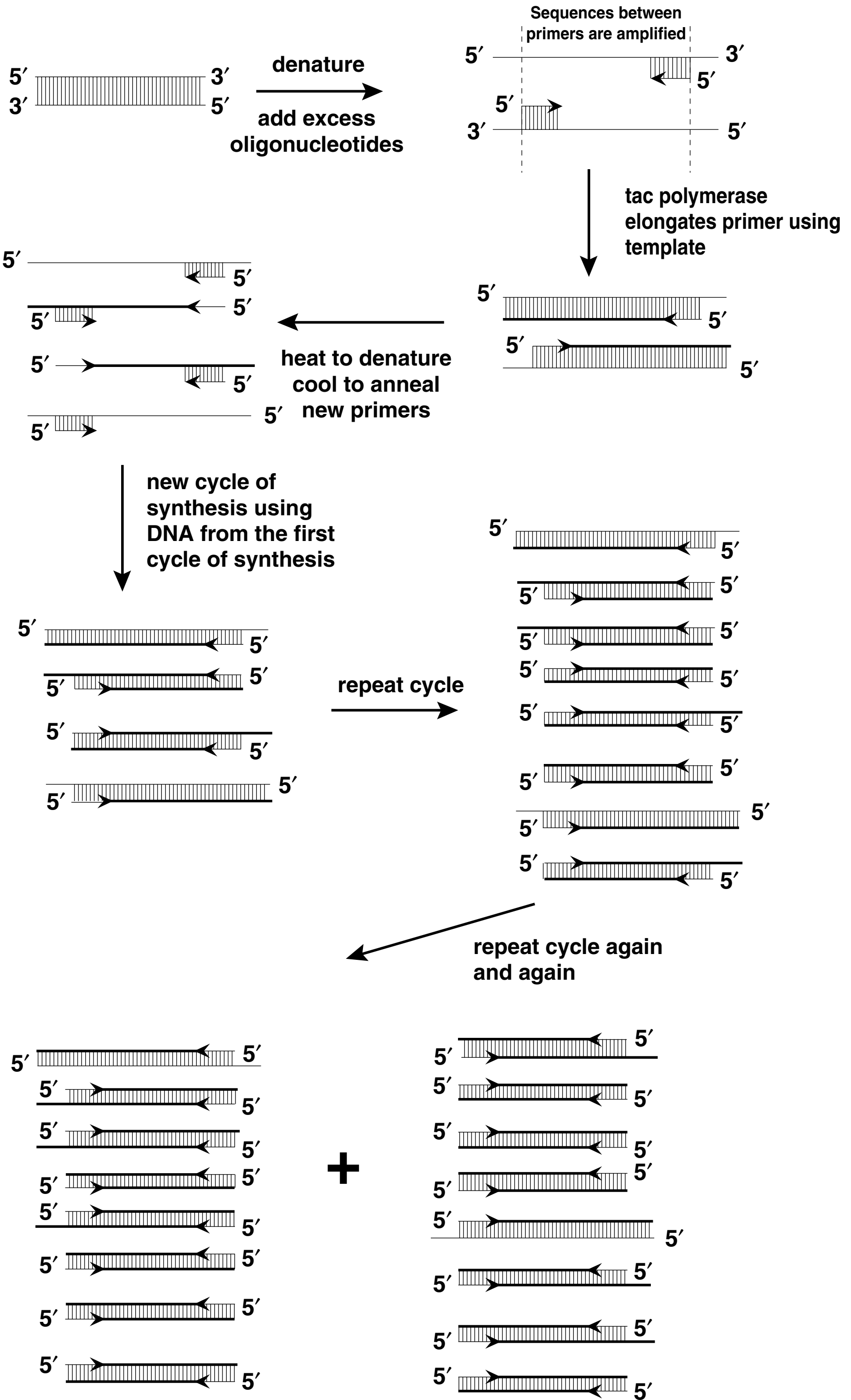



**Figure 6-6**
**SITE-DIRECTED MUTAGENESIS** can be used to change one or more base pairs in the DNA resulting in a change in the amino acid that appears in the protein produced from this DNA.

This is a technique for amplifying a specific segment of DNA. Oligonucleotide primers are synthesized that are complementary to one strand at the 5′ end of your DNA and complementary to the opposite strand at the 3′ end. After the DNA is denatured and the oligonucleotide primers (in excess) are annealed, the DNA is elongated using DNA polymerase and deoxynucleotides. A new double-stranded DNA molecule will be generated starting from each primer. DNA sequences behind (to the 5′ side of) the primer will not be replicated. The DNA is then heated to denature it and reannealed to the primer again. Another round of replication is performed. This cycle is repeated over and over, with a twofold increase in the amount of DNA on each cycle. Because two primers are used, only the sequence *between* the two primers will be amplified. Since the cycle is carried out multiple times with a twofold increase in the amount of DNA each time, a geometric amplification results (10 cycles would result in a $2^{10}$ increase in the DNA concentration). The limitation on the amount of DNA that is produced is the amount of primer and deoxynucleotides added. The cleverness of this technique is extended by using a heat-stable DNA polymerase that is not inactivated by the temperatures needed to denature the DNA. Multiple cycles can be performed simply by heating (denaturing) and cooling (renaturing and polymerizing) a tube containing the DNA, the primers, deoxynucleotide triphosphates, and the DNA polymerase. Because of the extreme amount of amplification, PCR can be used to amplify sequences from very small amounts of DNA. New restriction sites can be easily generated by including them in the 5′ end of the oligonucleotide primer even though they are not present in the original DNA. As long as the primer is still long enough to hybridize to the DNA through complementary sequences, the dangling 5′ ends containing the restriction site sequence will be amplified in the next round. PCR can also be used to remove inserts from vectors and to introduce site-specific mutants (Fig. 6-7).

By first making a DNA copy of an RNA molecule, one can also amplify RNA sequences. Because reverse transcriptase (copies RNA to DNA) is used in the first step, this is called *RT-PCR*.



5′
3′
3′
5′
denature
add excess
oligonucleotides
Sequences between
primers are amplified
5′
3′
5′
5′
3′
5′
tac polymerase
elongates primer using
template
heat to denature
cool to anneal
new primers
new cycle of
synthesis using
DNA from the first
cycle of synthesis
repeat cycle
repeat cycle again
and again
+

←**Figure 6-7 The Polymerase Chain Reaction**
**PCR** is used to amplify (synthesize) specific DNA sequences that lie between a 5′ primer and a 3′ primer. The primers are annealed to the appropriate DNA strand and are lengthened (5′ to 3′) by adding deoxynucleotides, using DNA polymerase and the longer DNA strand as a template. The newly synthesized DNA is denatured by heating, cooled to allow more primer to anneal to the newly synthesized strands, and the cycle of synthesis, melting, and annealing new primer is repeated over and over. Each cycle increases the amount of DNA by twofold. Note that with increasing numbers of cycles the sequences between the two primers are amplified more than sequences outside the primers.