Organic Macromolecules and the Genetic Code

- A cell is mostly water.
  - The rest of the cell consists mostly of carbon-based molecules.
  - Organic chemistry is the study of carbon compounds.

- Carbon can use its bonds to
  - Attach to other carbons.
  - Form an endless diversity of carbon skeletons.

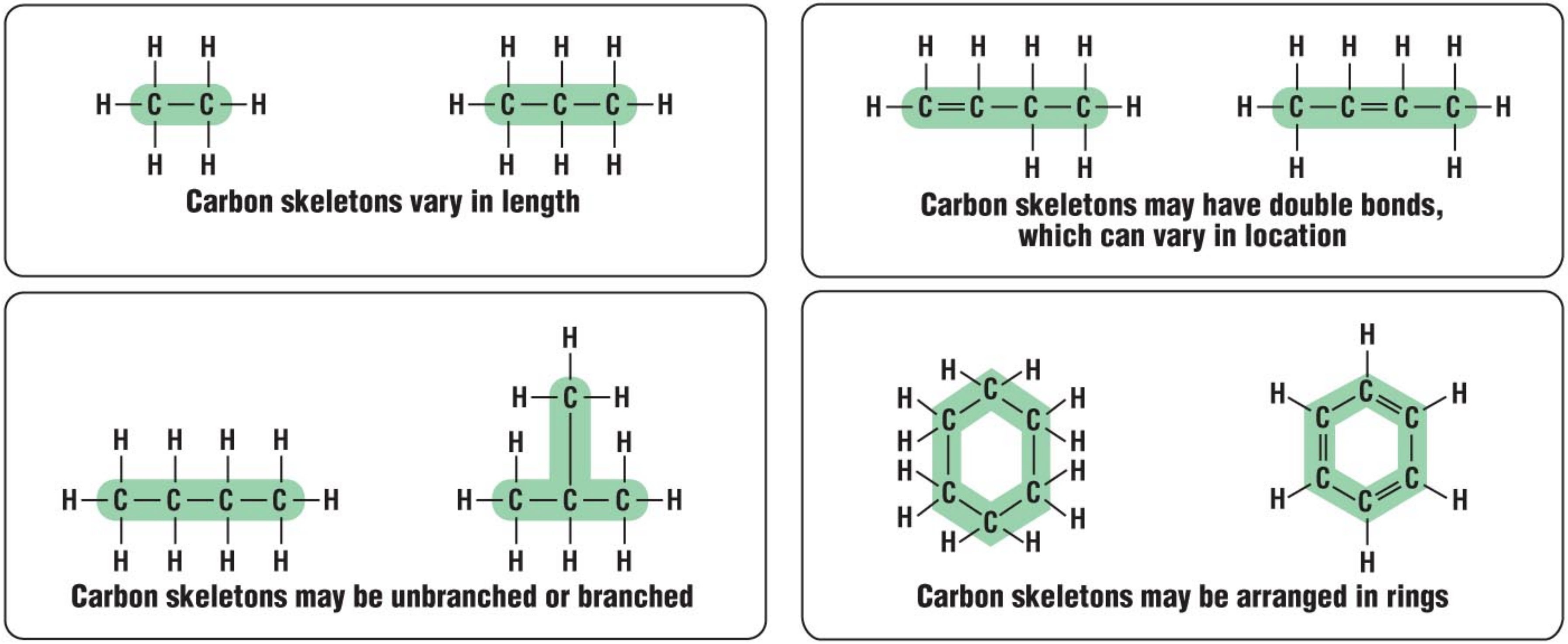

Carbon skeletons vary in length
Carbon skeletons may have double bonds, which can vary in location
Carbon skeletons may be unbranched or branched
Carbon skeletons may be arranged in rings

Figure 3.2

- Each type of organic molecule has a unique three-dimensional shape that defines its function in an organism.
  - The molecules of your body recognize one another based on their shapes.

Giant Molecules from Smaller Building Blocks

- On a molecular scale, many of life's molecules are gigantic.
  - Biologists call them macromolecules.
  - Examples: DNA, proteins

Biological Molecules

- There are four categories of large molecules in cells:
  - Carbohydrates
  - Lipids
  - Proteins
  - Nucleic acids

Proteins

- A protein is a polymer constructed from amino acid building blocks.
- Proteins perform most of the tasks the body needs to function. Behind most cell functions, there is a protein.

PLAY Structural Proteins

PLAY Storage Proteins

PLAY Contractile Proteins

PLAY Transport Proteins

PLAY Defensive Proteins

PLAY Receptor Proteins

PLAY Enzymes

PLAY Hormonal Proteins

PLAY Sensory Proteins

PLAY Gene Regulatory Proteins

The Monomers: Amino Acids

- All proteins are constructed from a the same set of 20 kinds of amino acids.

- Each amino acid consists of
  - A central carbon atom bonded to four groups.
  - Three of the groups are identical in all 20 amino acids, and each has a distinct 4^{th} group.

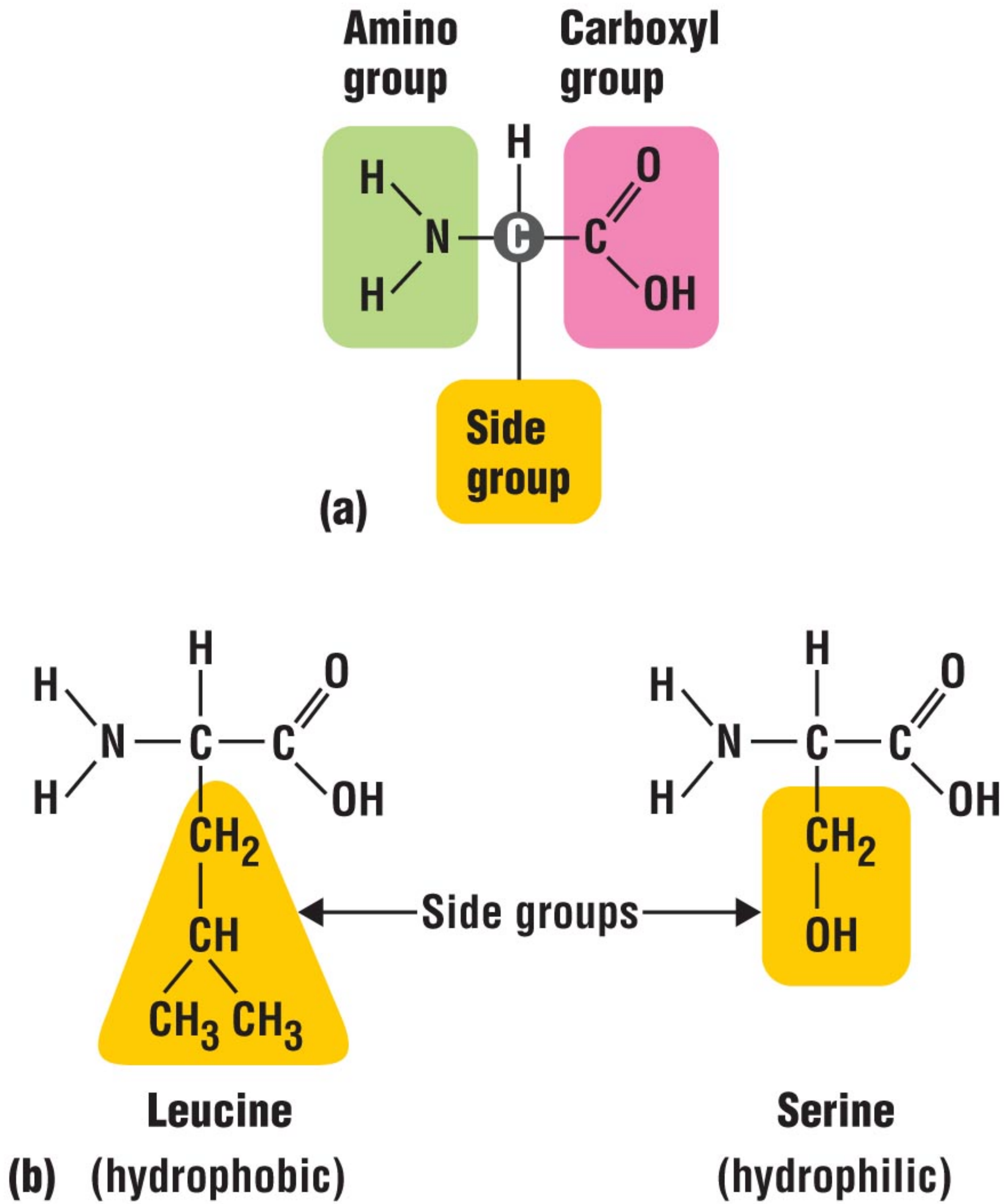
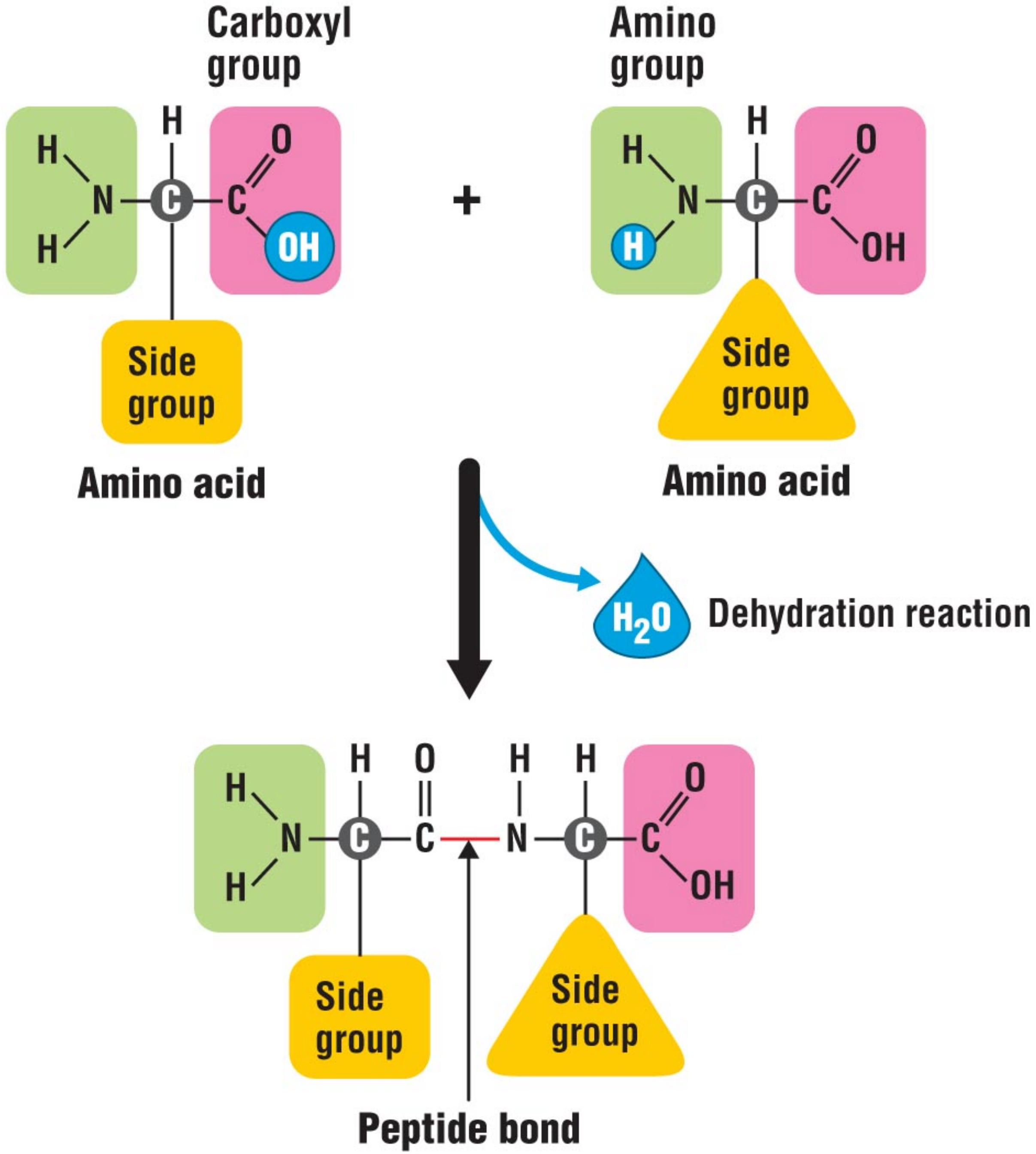

Amino group
Carboxyl group
Side group
(a)
Side groups
Leucine
(b) (hydrophobic)
Serine
(hydrophilic)

Figure 3.20

Proteins as chains

- A linked chain of amino acids

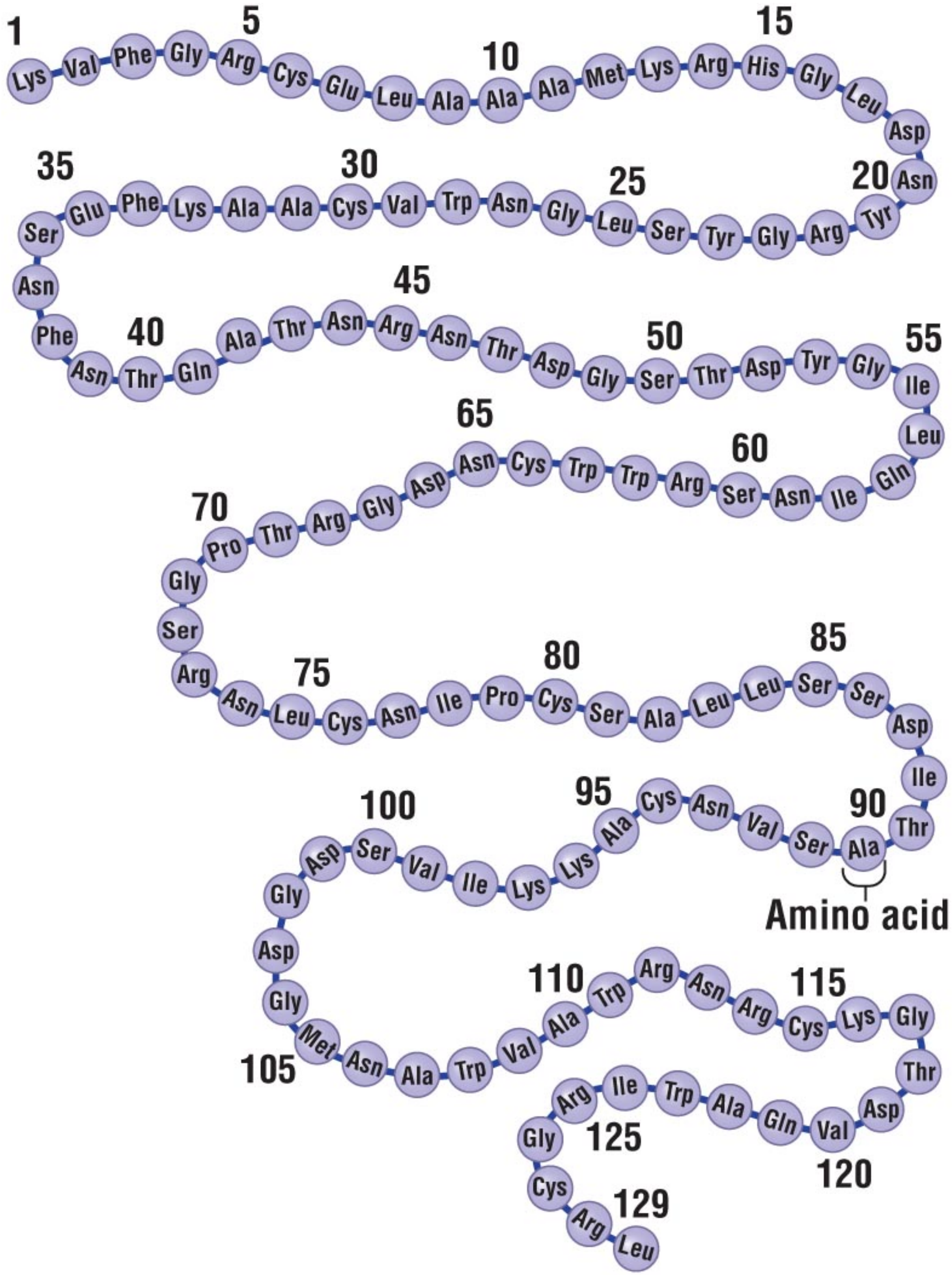


Carboxyl group
Amino group
H
H
N
C
C
O
OH
+
H
H
N
C
C
O
OH
H
Side group
Side group
Amino acid
Amino acid
H_2O
Dehydration reaction
H
H
N
C
H
O
C
H
N
C
H
C
O
OH
Side group
Side group
Peptide bond

Figure 3.21

- Your body has tens of thousands of different kinds of proteins.
  - The arrangement of amino acids of any given protein makes it different from any other protein.

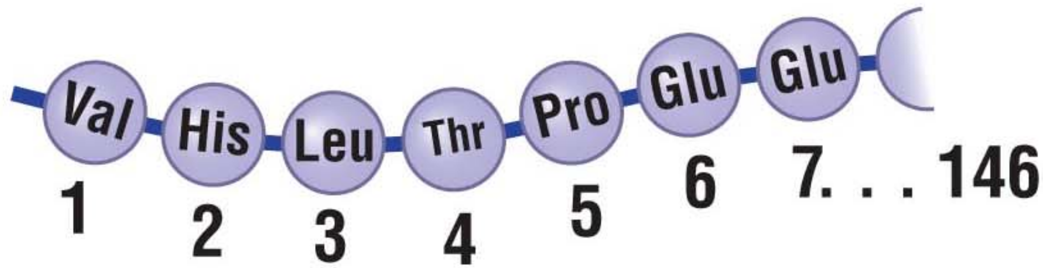
- Primary structure
  - The specific sequence of amino acids in a protein



1
Lys
Val
Phe
Gly
5
Arg
Cys
Glu
Leu
Ala
10
Ala
Ala
Met
Lys
Arg
15
His
Gly
Leu
Asp
20
Asn
Tyr
Arg
Gly
Tyr
Ser
25
Leu
Gly
Asn
Trp
Val
30
Cys
Ala
Ala
Lys
Phe
Glu
35
Ser
Asn
Phe
Asn
40
Thr
Gln
Ala
Thr
Asn
45
Arg
Asn
Thr
Asp
Gly
50
Ser
Thr
Asp
Tyr
Gly
55
Ile
Leu
Gln
Ile
Asn
60
Ser
Arg
Trp
Trp
Cys
65
Asn
Asp
Gly
Arg
Thr
70
Pro
Gly
Ser
Arg
Asn
75
Leu
Cys
Asn
Ile
Pro
80
Cys
Ser
Ala
Leu
Leu
85
Ser
Ser
Asp
Ile
Thr
90
Ala
Ser
Val
Asn
Cys
95
Ala
Lys
Lys
Ile
Val
100
Ser
Asp
Gly
Asp
Gly
105
Met
Asn
Ala
Trp
Val
110
Ala
Trp
Arg
Asn
Arg
115
Cys
Lys
Gly
Thr
Asp
120
Val
Gln
Ala
Trp
Ile
125
Arg
Gly
Cys
Arg
129
Leu
Amino acid

Figure 3.22

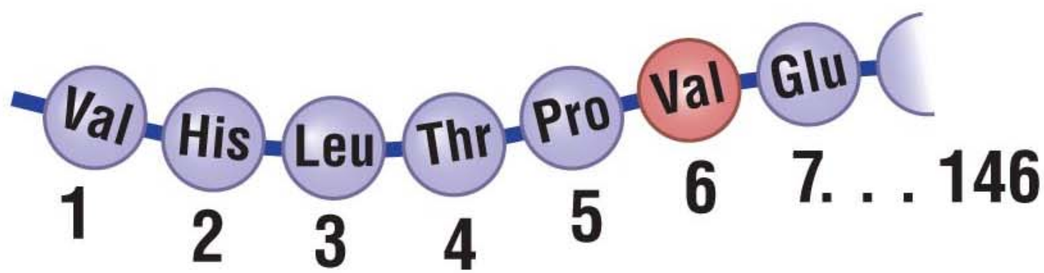
- A slight change in the primary structure of a protein affects its ability to function.
  - The substitution of one amino acid for another in hemoglobin causes sickle-cell disease.



Normal red blood cell

Normal hemoglobin

(a)



Sickled red blood cell

Sickle-cell hemoglobin

(b)

Figure 3.23

Nucleic Acids

- Nucleic acids are information storage molecules.
  - They provide the directions for building proteins.

- There are two types of nucleic acids:
  - DNA, deoxyribonucleic acid
  - RNA, ribonucleic acid

- The genetic instructions in DNA
  - Must be translated from “nucleic acid language” to “protein language.”

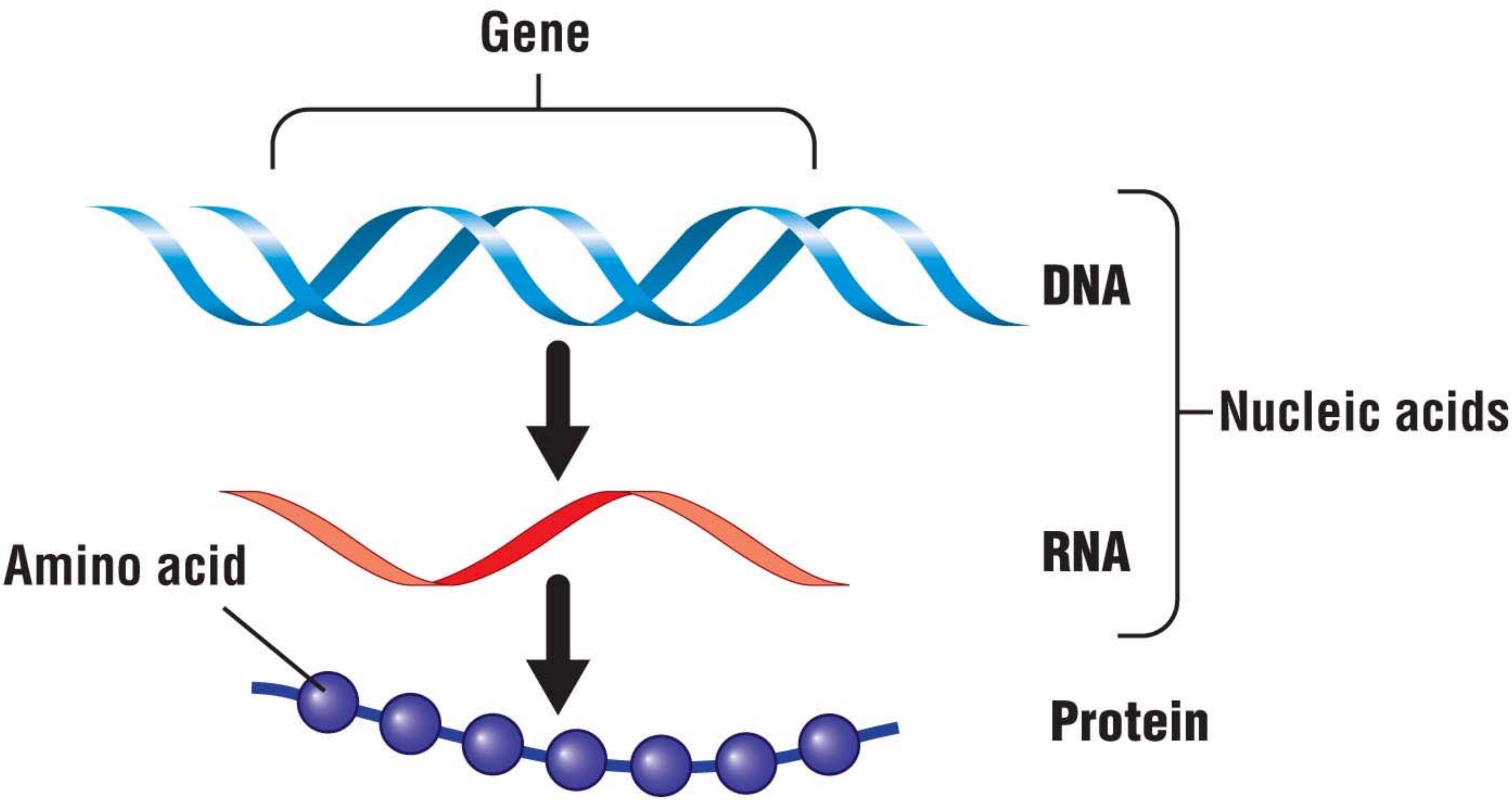
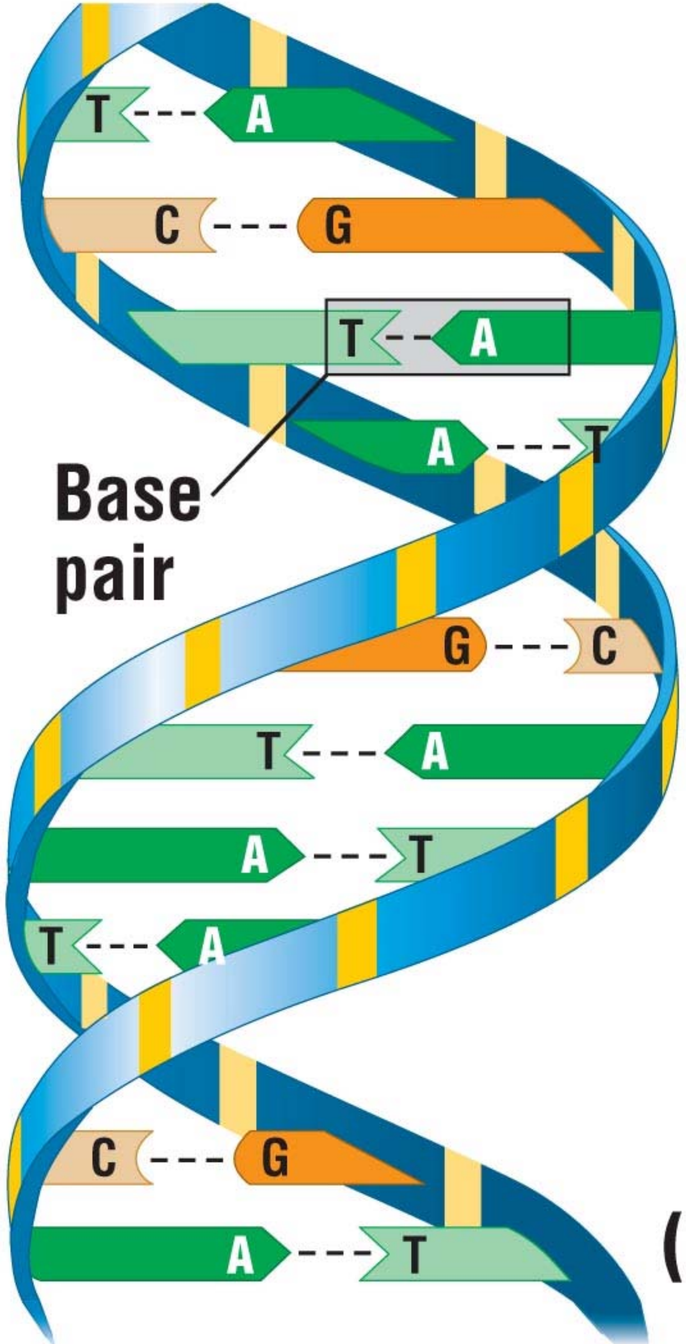


Figure 3.25

- Nucleic acids are chains of 4 nucleotides:
  - Guanine (G)
  - Thymine (T)
  - Cytosine (C)
  - Adenine (A)



(b) Double helix

Figure 3.28b

Now, here is one of the central problems of biology. Much of what is going on in a cell, as we have seen, is carried out by proteins. There are all kinds of proteins, depending on the sequence of their **20** amino acids:

Ala-Ala-Gly-Leu-Ser-. . .

Will be entirely different from

Gly-Ala-Leu-Ser-Ala- . . .

The information for the length and sequence of each protein must be contained in the genes, which are made of DNA. Here again, a gene is made up of a chain of **4** building blocks, called nucleotides:

A-G-C-T-C- . . .

C-G-T-A-C- . . .

This sequence of 4 building blocs must somehow code for sequence of 20 different amino acids in proteins. How can that be done?

To grasp nature's solution, let's examine a simpler code first

Imagine that you take part in a game show. You and your partner are competing with another pair of contestants. Your task is to invent a code which will allow you, when the show start, to communicate with your partner by relying on the code and by using 4 colors to make a series of dots on paper. How will you go about this?

How do you express 26 letters in 4 colors?

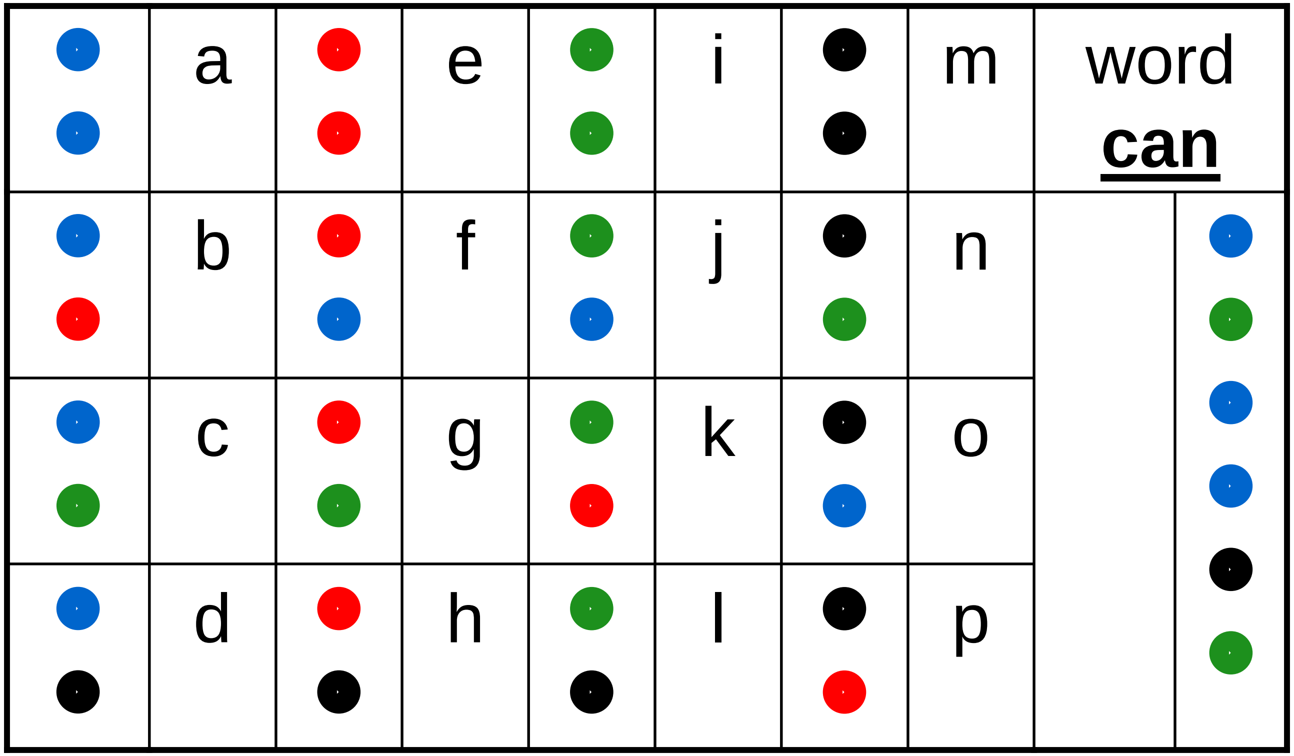
A one-for-one code will not work:

Let's try 1 color coding for a single letter: In this code, <u>dad</u> is: ⚫ 🔴 ⚫

Obviously, this code can't go very far: With just one color per letter, we can only specify 4 letters—but we need at least 26!

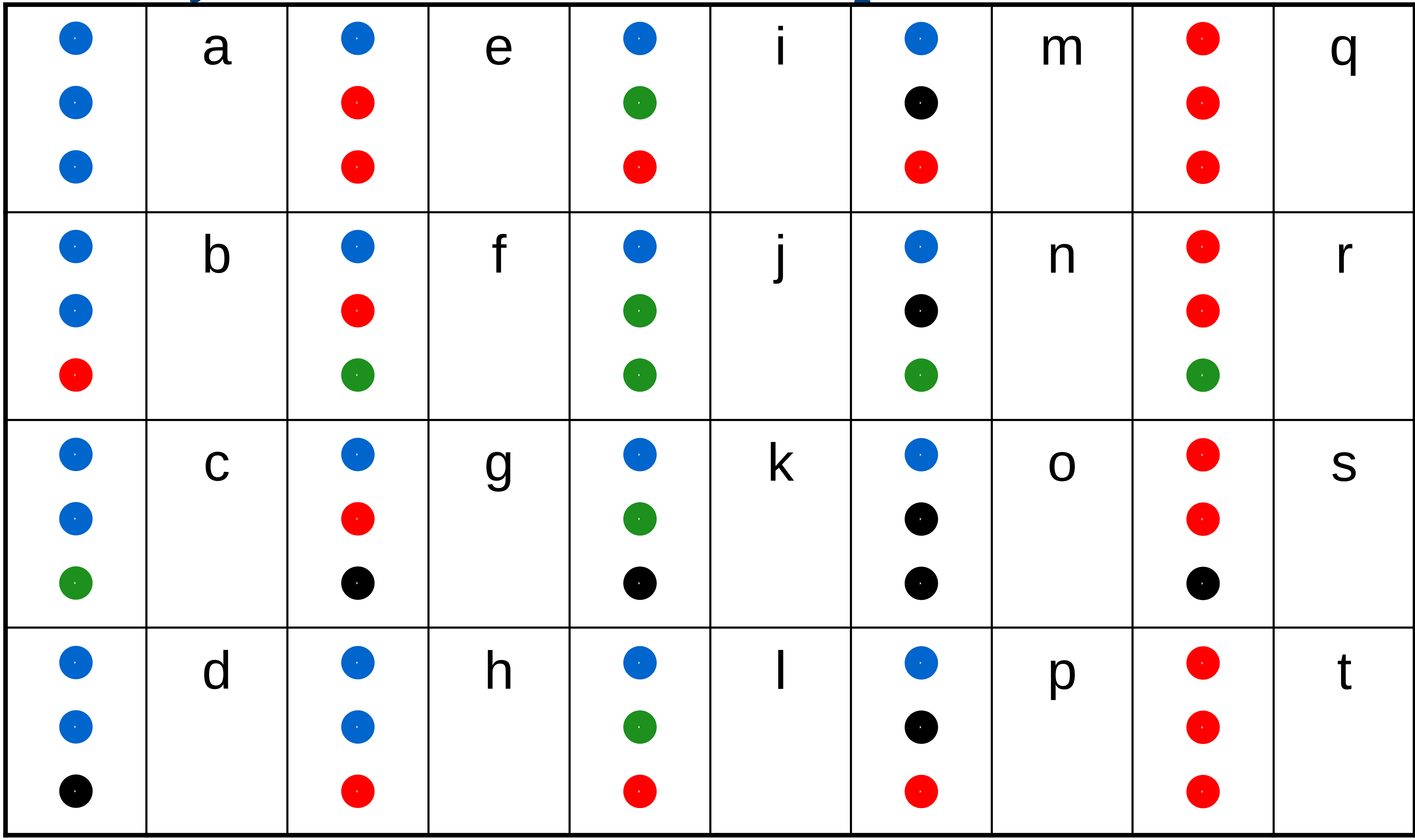
🔴	a
🟢	b
🔵	c
⚫	d

Will any combination of 2 colors be enough for 26 letters? When we try, we find out that only 16 permutations are possible?



So, 2 colors per letter: not enough. How about 3?

This is a PARTIAL table. Can you see how many color combinations are possible?



With color triplets then, I can write any message in English, coding even for caps and punctuation marks!

Now, life is written by a similar code. DNA is made of a chain of 4 links or “colors” (nucleotides) AGCT. Protein is made of a chain of 20 different links—amino acids. You can specify any amino acid you want, by using triplets like AAA, AAG, AAG, AAG, AGG, . . .

- So the idea is: Triplets of DNA bases which
  - Specify all the amino acids.
  - Are called codons.

The Genetic Code

- The genetic code is the set of rules relating nucleotide sequence to amino acid sequence.
- Here it is: one of the greatest scientific breakthroughs of the 20th century:

Second base

First base	U	C	A	G	Third base
U	UUU, UUC: Phenylalanine (Phe) UUA, UUG: Leucine (Leu)	UCU, UCC, UCA, UCG: Serine (Ser)	UAU, UAC: Tyrosine (Tyr) UAA Stop UAG Stop	UGU, UGC: Cysteine (Cys) UGA Stop UGG Tryptophan (Trp)	U C A G
C	CUU, CUC, CUA, CUG: Leucine (Leu)	CCU, CCC, CCA, CCG: Proline (Pro)	CAU, CAC: Histidine (His) CAA, CAG: Glutamine (Gln)	CGU, CGC, CGA, CGG: Arginine (Arg)	U C A G
A	AUU, AUC, AUA: Isoleucine (Ile) AUG Met or start	ACU, ACC, ACA, ACG: Threonine (Thr)	AAU, AAC: Asparagine (Asn) AAA, AAG: Lysine (Lys)	AGU, AGC: Serine (Ser) AGA, AGG: Arginine (Arg)	U C A G
G	GUU, GUC, GUA, GUG: Valine (Val)	GCU, GCC, GCA, GCG: Alanine (Ala)	GAU, GAC: Aspartic acid (Asp) GAA, GAG: Glutamic acid (Glu)	GGU, GGC, GGA, GGG: Glycine (Gly)	U C A G

Figure 10.11

- The genetic code is shared by virtually all organisms.