Chapter 1

The Facts of Life: Chemistry Is the Logic of Biological Phenomena

Chapter Outline

- Properties of living systems
  - Highly organized: Cells > organelles > macromolecular complexes > macromolecules (proteins, nucleic acids, polysaccharides)
  - Structure/function correlation: Biological structures serve functional purposes
  - Energy transduction: ATP and NADPH—energized molecules
  - Steady state maintained by energy flow: Steady state not equilibrium
  - Self-replication with high, yet not perfect, fidelity
- Biomolecules
  - Elements: Hydrogen, oxygen, carbon, nitrogen (lightest elements of the periodic table capable of forming a variety of strong covalent bonds)
    - Carbon: 4 bonds, nitrogen: 3 bonds, oxygen: 2 bonds, hydrogen: 1 bond
  - Compounds: Carbon-based compounds—versatile
  - Phosphorus- and sulfur-containing compounds play important roles
- Biomolecular hierarchy
  - Simple compounds: H_2O, CO_2, NH_4^+, NO_3^-, N_2
  - Metabolites: Used to synthesize building block molecules
  - Building blocks: Amino acids, nucleotides, monosaccharides, fatty acids, glycerol
  - Macromolecules: Proteins, nucleic acids, polysaccharides, lipids
  - Supramolecular complexes: Ribosomes, chromosomes, cytoskeleton
- Membranes: Lipid bilayers with membrane proteins
  - Define boundaries of cells and organelles
  - Hydrophobic interactions maintain structures
- Organelles: Mitochondria, chloroplasts, nuclei, endoplasmic reticulum, Golgi, etc.
- Cells: Fundamental units of life
  - Living state: Growth, metabolism, stimulus response, and replication
- Properties of biomolecules
  - Directionality or structural polarity
    - Proteins: N-terminus and C-terminus
    - Nucleic acids: 5′- and 3′-ends
    - Polysaccharides: Reducing and non-reducing ends
  - Information content: Sequence of monomer building blocks and 3-dimensional architecture
- Three-dimensional architecture and intermolecular interactions (via complementary surfaces) of macromolecules are based on weak forces
  - van der Waals interactions (London dispersion forces)
    - Induced electric interactions that occur when atoms are close together
    - Significant when many contacts form complementary surfaces
  - Hydrogen bonding
    - Donor and acceptor pair: Direction dependence
    - Donor is hydrogen covalently bonded to electronegative O or N
    - Acceptor is lone pair on O or N

- Ionic interactions
  - Stronger than H bonds
  - Not directional
  - Strength influenced by solvent properties
- Hydrophobic interactions: Occur when non-polar groups are added to water
  - Water molecules hydrogen bond
  - Non-polar groups interfere with water H-bonding, and non-polar groups aggregate to minimize this

- Life restricted to narrow range of conditions (temperature, pH, salt concentration, etc.) because of dependence on weak forces
  - Denaturation: Loss of structural order in a macromolecule
- Enzymes: Biological catalysts capable of being regulated
- Cell types
  - Prokaryotes: Bacteria and archaea: Plasma membrane but no internal membrane-defined compartments
    - Archaea include thermoacidophiles, halophiles, and methanogens
  - Eukaryotes: Internal membrane-defined compartments: nuclei, endoplasmic reticulum, Golgi, mitochondria, chloroplasts, vacuoles, peroxisomes
  - Viruses and bacteriophages: Incomplete genetic systems

Chapter Objectives

Understand the Basic Chemistry of H, O, N ,and C

H forms a single covalent bond. When bound to an electronegative element, like O or N, the electron pair forming the covalent bond is not equally shared, giving rise to a partial positive charge on the hydrogen (this is the basis of H bonds which will be covered in the next chapter). In extreme cases, the H can be lost as a free proton.

O forms two covalent bonds and has two lone pairs of electrons. It is an electronegative element and when bound to hydrogen it will cause H to be partially positively charged. O is highly reactive due to its high electronegativity.

N forms up to three covalent bonds and has a single lone pair of electrons. It is an electronegative element and will create a partial positive charge on a hydrogen bonded to it.

C forms four covalent bonds. With four single bonds, tetrahedral geometry is predominant. With one double bond, carbon shows trigonal planar geometry, with an additional pair of electrons participating in a pi bond.

Macromolecules and Subunits

Proteins are formed from amino acids composed of C, H, O, N, and in some instances S.

Nucleic acids are formed from nucleotides that are composed of phosphate, sugar, and nitrogenous base components (nucleosides lack phosphate).

Polysaccharides are made of carbohydrates or sugar molecules.

Lipids are a class of mostly non-polar, mostly hydrocarbon molecules.

Macromolecular Structures

Macromolecular structures are composed of complexes of macromolecules (i.e., proteins, nucleic acids, polysaccharides, and lipids). The ribosome, made up of protein and ribonucleic acid, is a prime example.

Organelles

Organelles are subcellular compartments defined by lipid bilayer membranes.

Cell Types

There are two fundamental cell types: eukaryotic, having organelles and a defined nuclear region; and prokaryotic, lacking organelles and a membrane-enclosed region of genetic material. The archaea and bacteria comprise the prokaryotes.

Problems and Solutions

1. The nutritional requirements of* Escherichia coli *cells are far simpler than those of humans, yet the macromolecules found in bacteria are about as complex as those of animals. Because bacteria can make all their essential biomolecules while subsisting on a simpler diet, do you think bacteria may have more biosynthetic capacity and hence more metabolic complexity than animals? Organize your thoughts on this question, pro and con, into a rational argument.

Answer: Although it is true that *Escherichia coli* are capable of producing all of their essential biomolecules (e.g., there is no minimum daily requirement for vitamins in the world of wild-type *E. coli*), they are rather simple, single-cell organisms capable of a limited set of responses. They are self-sufficient, yet they are incapable of interactions leading to levels of organization such as multicellular tissues. Multicellular organisms have the metabolic complexity to produce a number of specialized cell types and to coordinate interactions among them.

2. Without consulting figures in this chapter, sketch the characteristic prokaryotic and eukaryotic cell types and label their pertinent organelle and membrane systems.

Answer: Prokaryotic cells lack the compartmentation characteristic of eukaryotic cells and are devoid of membrane-bound organelles such as mitochondria, chloroplasts, endoplasmic reticulum, Golgi apparatus, nuclei, peroxisomes, and vacuoles. Both cell types are delimited by membranes and contain ribosomes.

3.* Escherichia coli *cells are about 2 μm (microns) long and 0.8 μm in diameter.
a. How many* E. coli *cells laid end to end would fit across the diameter of a pinhead? (Assume a pinhead diameter of 0.5 mm.)
b. What is the volume of an* E. coli *cell? (Assume it is a cylinder, with the volume of a cylinder given by $V = \pi r^2 h$, where $\pi = 3.14$.)
c. What is the surface area of an* E. coli *cell? What is the surface-to-volume ratio of an* E. coli *cell?
d. Glucose, a major energy-yielding nutrient, is present in bacterial cells at a concentration of about 1 mM. What is the concentration of glucose, expressed as mg/mL? How many glucose molecules are contained in a typical* E. coli *cell? (Recall that Avogadro's number = 6.023×10^{23}.)
e. A number of regulatory proteins are present in* E. coli *at only 1 or 2 molecules per cell. If we assume that an E. coli contains just 1 molecule of a particular protein, what is the molar concentration of this protein in the cell? If the molecular weight of this protein is 40 kDa, what is its concentration, expressed as mg/mL?
f. An* E. coli *cell contains about 15 000 ribosomes, which carry out protein synthesis. Assuming ribosomes are spherical and have a diameter of 20 nm (nanometers), what fraction of the* E. coli *cell volume is occupied by ribosomes?
g. The* E. coli *chromosome is a single DNA molecule whose mass is about 3.0×10^9 Da. This macromolecule is actually a circular array of nucleotide pairs. The average molecular weight of a nucleotide pair is 660 Mr and each pair imparts 0.34 nm to the length of the DNA molecule. What is the total length of the* E. coli *chromosome? How does this length compare with the overall dimensions of an* E. coli *cell? How many nucleotide pairs does this DNA contain? The average* E. coli *protein is a linear chain of 360 amino acids. If 3 nucleotide pairs in a gene encode 1 amino acid in a protein, how many different proteins can the* E. coli *chromosome encode? (The answer to this question is a reasonable approximation of the maximum number of different kinds of proteins that can be expected in bacteria.)

Answer: a.

$$E.\,coli \text{ per pinhead} = \frac{\dfrac{0.5 \text{ mm}}{\text{dia.}}}{\dfrac{2\ \mu\text{m}}{E.\,coli}} = \frac{\dfrac{0.5 \times 10^{-3}\text{m}}{\text{dia.}}}{\dfrac{2 \times 10^{-6}\text{m}}{E.\,coli}}$$

$$= 250\ E.\,coli \text{ per pinhead}$$

b.

$$V = \pi \times r^2 \times h$$

$$= 3.14 \times \left(\frac{0.8\ \mu\text{m}}{2} \right)^2 \times 2\ \mu\text{m}$$

$$= 3.14 \times \left(0.4 \times 10^{-6}\text{m} \right)^2 \times 2 \times 10^{-6}\text{m}$$

$$= 1 \times 10^{-18}\text{m}^3$$

But $1\ \text{m}^3 = (100\ \text{cm})^3 = 10^6\text{cm}^3 = 10^6\text{mL} = 10^3\text{L}$

$V = 1 \times 10^{-18}\text{m}^3 = 1 \times 10^{-15}\text{L} = 1\ \text{fL}$ (femtolitre)

c.

$$\text{Surface Area} = 2 \times \pi \times r^2 + \pi \times d \times h$$

$$\text{Surface Area} = 2 \times 3.14 \times \left(0.4 \times 10^{-6}\text{m} \right)^2 + 3.14 \times \left(0.8 \times 10^{-6}\text{m} \right) \times \left(2 \times 10^{-6}\text{m} \right)$$

$$\text{Surface Area} = 6.03 \times 10^{-12}\text{m}^2$$

$$\frac{\text{Surface Area}}{\text{Volume}} = \frac{6 \times 10^{-12}\text{m}^2}{1 \times 10^{-18}\text{m}^3\ \text{(from b)}}$$

$$\text{Surface Area per volume} = 6 \times 10^6\text{m}^{-1}$$

d.

$$\left[\text{Glucose}\right] = 1\ \text{mM} = 1 \times 10^{-3}\ \frac{\text{mol}}{\text{L}}$$

$\text{Glucose} = C_6H_{12}O_6$

$M_r = 6 \times 12 + 12 \times 1.0 + 6 \times 16$

$M_r = 180$

$$\left[\text{Glucose}\right] = 1 \times 10^{-3}\ \frac{\text{mol}}{\text{L}} \times 180\ \frac{\text{g}}{\text{mol}}$$

$$\left[\text{Glucose}\right] = 0.18\ \frac{\text{g}}{\text{L}} = 0.18\ \frac{\text{mg}}{\text{mL}}$$

$$\text{moles of glucose} = \text{concentration} \times \text{volume}$$

$$\text{moles of glucose} = 1 \times 10^{-3}\ \frac{\text{mol}}{\text{L}} \times 1 \times 10^{-15}\text{L (from b)}$$

$$\text{moles of glucose} = 1 \times 10^{-18}$$

$$\#\ \text{molecules} = 1 \times 10^{-18}\text{mol} \times 6.023 \times 10^{23}\ \frac{\text{molecules}}{\text{mol}}$$

$$\#\ \text{molecules} = 6 \times 10^5\text{molecules}$$

e.

$$\frac{1\ \text{molecule}}{6.023 \times 10^{23}\ \frac{\text{molecules}}{\text{mol}}} = 1.66 \times 10^{-24}\text{mol}$$

$$\text{Molar Concentration} = \frac{\text{moles}}{\text{volume (in litres)}}$$

$$\text{Molar Concentration} = \frac{1.66 \times 10^{-24}\text{mol}}{1 \times 10^{-15}\text{L (from b)}}$$

$$\text{Molar Concentration} = 1.66 \times 10^{-9}\text{M} = 1.7\ \text{nM}$$

$$\left[\text{Protein}\right] = 1.66 \times 10^{-9}\ \frac{\text{mol}}{\text{L}} \times 40\,000\ \frac{\text{g}}{\text{mol}}$$

$$\left[\text{Protein}\right] = 6.6 \times 10^{-5}\ \frac{\text{g}}{\text{L}} = 6.6 \times 10^{-5}\ \frac{\text{mg}}{\text{mL}}\ \text{or}\ 66\ \frac{\mu\text{g}}{\text{L}}\ \text{or}\ 66\ \frac{\text{ng}}{\text{mL}}$$

f.

$$\text{Volume of 1 ribosome} = \frac{4}{3} \times \pi \times r^3$$

$$\text{Volume of 1 ribosome} = \frac{4}{3} \times 3.14 \times \left(\frac{20 \times 10^{-9}\text{m}}{2} \right)^3$$

$$\text{Volume of 1 ribosome} = 4.2 \times 10^{-24}\text{m}^3$$

$$\text{Volume of 15 000 ribosomes} = 15\,000 \times 4.2 \times 10^{-24}\text{m}^3 = 6.3 \times 10^{-20}\text{m}^3$$

$$\text{Fractional volume} = \frac{\text{Volume ribosomes}}{\text{Volume cell}}$$

$$\text{Fractional volume} = \frac{6.3 \times 10^{-20}\text{m}^3}{1 \times 10^{-18}\text{m}^3 \text{ (from b)}}$$

$$\text{Fractional volume} = 0.063 \text{ or } 6.3\%$$

g. The number of moles of base pairs in 3.0×10^9 Da dsDNA is given by

$$= \frac{3.0 \times 10^9 \dfrac{\text{g}}{\text{mol dsDNA}}}{660 \dfrac{\text{g}}{\text{mol bp}}}$$

$$= 4.55 \times 10^6 \frac{\text{mol bp}}{\text{mol dsDNA}}$$

$$\text{Length} = 4.55 \times 10^6 \frac{\text{mol bp}}{\text{mol dsDNA}} \times 0.34 \frac{\text{nm}}{\text{bp}}$$

$$\text{Length} = 4.55 \times 10^6 \times 0.34 \times 10^{-9}\text{m}$$

$$\text{Length} = 1.55 \times 10^{-3}\text{m} = 1.55 \text{ mm} = 1550\ \mu\text{m}$$

$$\text{Length } \textit{E. coli} = 2\ \mu\text{m}$$

$$\frac{\text{Length DNA}}{\text{Length } \textit{E. coli}} = \frac{1550\ \mu\text{m}}{2\ \mu\text{m}} = 775$$

To calculate the number of different proteins that would be encoded by the *E. coli* chromosome:

$$360 \frac{\text{aa}}{\text{protein}} \times 3 \frac{\text{bp}}{\text{aa}} = 1080 \frac{\text{bp}}{\text{protein}}$$

$$\text{\# different proteins} = \frac{4.55 \times 10^6 \text{bp}}{1080 \dfrac{\text{bp}}{\text{protein}}} = 4213 \text{ proteins}$$

The exact number can be found at NCBI (http://www.ncbi.nlm.nih.gov/). The genomes of a number of strains of *E. coli* have been sequenced, but the first one was K-12 strain MG1655. At NCBI, search for MG1655 and view hits in the genome database. There should be 11 of them and NC_000913 should be one of them. Activate this link (or search for NC_000913 directly and then activate it). The returned page should indicate that this strain of *E. coli* has 4132 protein-coding genes.

4. Assume that mitochondria are cylinders 1.5 μm in length and 0.6 μm in diameter.
a. What is the volume of a single mitochondrion?
b. Oxaloacetate is an intermediate in the citric acid cycle, an important metabolic pathway localized in the mitochondria of eukaryotic cells. The concentration of oxaloacetate in mitochondria is about 0.03 μM. How many molecules of oxaloacetate are in a single mitochondrion?

Answer: a. $V = \pi \times r^2 \times h$
$V = 3.14 \times (3 \times 10^{-7}\text{ m})^2 \times (1.5 \times 10^{-6}\text{ m})$
$V = 4.24 \times 10^{-19}\text{ m}^3$

But 1 $m^3 = 10^3$ L

$$V = 4.24\times10^{-19}\text{m}^3 \times \frac{10^3\text{L}}{\text{m}^3} = 4.24\times10^{-16}\text{L} = 0.424\text{ fL}$$

b.

$$\#\text{ molecules} = \text{Molar concentration}\times\text{volume}\times 6.023\times10^{23}\,\frac{\text{molecules}}{\text{mol}}$$

$$\#\text{ molecules} = 0.03\times10^{-6}\,\frac{\text{mol}}{\text{L}}\times 4.24\times10^{-16}\text{L (from a)}\times 6.023\times10^{23}\,\frac{\text{molecules}}{\text{mol}}$$

$$\#\text{ molecules} = 7.66\text{ molecules (fewer than 8 molecules)}$$

5. Assume that liver cells are cuboidal in shape, 20 μm on a side.
a. How many liver cells laid end to end would fit across the diameter of a pinhead? (Assume a pinhead diameter of 0.5 mm.)
b. What is the volume of a liver cell? (Assume it is a cube.)
c. What is the surface area of a liver cell? What is the surface-to-volume ratio of a liver cell? How does this compare to the surface-to-volume ratio of an* E. coli *cell? (Compare this answer to that of problem 3c.) What problems must cells with low surface-to-volume ratios confront that do not occur in cells with high surface-to-volume ratios?
d. A human liver cell contains 2 sets of 23 chromosomes, each set being roughly equivalent in information content. The total mass of DNA contained in these 46 enormous DNA molecules is 4 × 10¹² Da. Because each nucleotide pair contributes 660 Da to the mass of DNA and 0.34 nm to the length of DNA, what is the total number of nucleotide pairs and the complete length of the DNA in a liver cell? How does this length compare with the overall dimensions of a liver cell? The maximal information in each set of liver cell chromosomes should be related to the number of nucleotide pairs in the chromosome set's DNA. This number can be obtained by dividing the total number of nucleotide pairs calculated above by 2. What is this value? If this information is expressed in proteins that average 400 amino acids in length and three nucleotide pairs encode one amino acid in a protein, how many different kinds of proteins might a liver cell be able to produce? (In reality, liver cells express at most about 30 000 different proteins. Thus, a large discrepancy exists between the theoretical information content of DNA in liver cells and the amount of information actually expressed.)

Answer: a.

$$\#\text{ liver cells} = \frac{0.5\,\frac{\text{mm}}{\text{pinhead}}}{20\,\frac{\mu\text{m}}{\text{cell}}}$$

$$\#\text{ liver cells} = \frac{0.5\times10^{-3}\,\frac{\text{m}}{\text{pinhead}}}{20\times10^{-6}\,\frac{\text{m}}{\text{cell}}}$$

$$\#\text{ liver cells} = 25\,\frac{\text{cells}}{\text{pinhead}}$$

b.

$$\text{Volume of cubic liver cell} = \text{length}^3 = \left(20\times10^{-6}\text{m}\right)^3$$

$$\text{Volume of cubic liver cell} = 8\times10^{-15}\text{m}^3\times\left(\frac{100\text{ cm}}{\text{m}}\right)^3\times\left(\frac{1\text{ L}}{1000\text{ cm}^3}\right)$$

$$\text{Volume of cubic liver cell} = 8\times10^{-12}\text{L} = 8\text{ pL}$$

c.

$$\text{Surface Area} = 6\times\left(20\times10^{-6}\text{m}\right)\times\left(20\times10^{-6}\text{m}\right) = 2.4\times10^{-9}\text{m}^2$$

$$\frac{\text{Surface Area}}{\text{Volume}} = \frac{2.4\times10^{-9}\text{m}^2}{8\times10^{-15}\text{m}^3\text{(from b)}}$$

$$\frac{\text{Surface Area}}{\text{Volume}} = 3.0 \times 10^5 \text{m}^{-1}$$

The surface-to-volume ratio of liver to that of *E. coli* is given by:

$$\frac{3.0 \times 10^5 \text{m}^{-1}}{6 \times 10^6 \text{m}^{-1} \text{ (from 3c)}} = 0.05 \ (1/20^{\text{th}})$$

The volume of a cell sets or determines the cell's maximum metabolic activity, while the surface area defines the surface across which nutrients and metabolic waste products must pass to meet the metabolic needs of the cell. Cells with a low surface-to-volume ratio have a high metabolic capacity relative to the surface area for exchange.

d.

$$\text{\# base pairs} = \frac{4.0 \times 10^{12} \text{Da}}{660 \dfrac{\text{Da}}{\text{base pair}}}$$

$$\text{\# base pairs} = 6.1 \times 10^9 \text{bp}$$

$$\text{Length} = 0.34 \frac{\text{nm}}{\text{bp}} \times 6.1 \times 10^9 \text{bp}$$

$$\text{Length} = 2.06 \text{ m}$$

$$\text{Length relative to liver cell} = \frac{2.06 \text{ m}}{20 \ \mu\text{m}} = \frac{2.06 \text{ m}}{20 \times 10^{-6} \text{ m}}$$

$$\text{Length relative to liver cell} = 1.03 \times 10^5 \text{ or about } 100\,000 \text{ times greater!}$$

The maximal information is 3.0×10^9 bp.

$$\text{\# proteins} = 400 \frac{\text{aa}}{\text{protein}} \times 3 \frac{\text{bp}}{\text{aa}} = 1{,}200 \frac{\text{bp}}{\text{protein}}$$

$$\text{\# proteins} = \frac{3.0 \times 10^9 \text{bp}}{1200 \dfrac{\text{bp}}{\text{protein}}} = 2.5 \times 10^6 \text{proteins}$$

6. Biomolecules interact with one another through molecular surfaces that are structurally complementary. How can various proteins interact with molecules as different as simple ions, hydrophobic lipids, polar but uncharged carbohydrates, and even nucleic acids?

Answer: The amino acid side chains of proteins can participate in a number of interactions through hydrogen bonding, ionic bonding, hydrophobic interactions, and van der Waals interactions. For example, the polar amino acids, acidic amino acids and their amides, and the basic amino acids all have groups that can participate in hydrogen bonding. Those amino acid side chains that have net charge can form ionic bonds. The hydrophobic amino acids can interact with non-polar, hydrophobic surfaces of molecules. Thus, amino acids are capable of participating in a variety of interactions. A protein can be folded in three dimensions to organize amino acids into surfaces with a range of properties.

7. What structural features allow biological polymers to be informational macromolecules? Is it possible for polysaccharides to be informational macromolecules?

Answer: Biopolymers, like proteins and nucleic acids, are informational molecules because they are vectorial molecules, composed of a variety of building blocks. For example, proteins are linear chains of some 20 amino acids joined head-to-tail to produce a polymer with distinct ends. The information content is the sequence of amino acids along the polymer. Nucleic acids (DNA and RNA) are also informational molecules for the same reason. Here, the biopolymer is made up of 4 kinds of nucleotides. Monosaccharides can be linked to form polymers. When a polymer is formed from only one kind of monosaccharide—as for example in glycogen, starch, and cellulose—even though the molecule is vectorial (i.e., it has distinct ends), there is little information content. There are, however, a variety of monosaccharides and monosaccharide derivatives that are used to form polysaccharides. Furthermore, monosaccharides can be joined in a variety of ways to form branch structures. Branched polysaccharides composed of a number of different monosaccharides are rich in information.

8. Why is it important that weak forces, not strong forces, mediate biomolecular recognition?

Answer: Life is a dynamic process characterized by continually changing interactions. Complementary interactions based on covalent bonding would of necessity produce static structures that would be difficult to change and slow to respond to outside stimuli.

9. What is the distance between the centres of 2 carbon atoms (their limit of approach) that are interacting through van der Waals forces? What is the distance between the centres of 2 carbon atoms joined in a covalent bond? (See Table 1.4.)

Answer: The limit of approach of two atoms is determined by the sum of their van der Waals radii, which are given in Table 1.4. For 2 carbon atoms, the limit of approach is (0.17 nm + 0.17 nm) 0.34 nm. The distance between the centres of 2 carbon atoms joined in a covalent bond is the sum of the covalent radii of the two carbons or (0.077 nm + 0.077 nm) 0.154 nm. Clearly, two carbons sharing electrons in a covalent bond are closer together than are two carbons interacting through van der Waals forces.

10. Why does the central role of weak forces in biomolecular interactions restrict living systems to a narrow range of environmental conditions?

Answer: The weak forces such as hydrogen bonds, ionic bonds, hydrophobic interactions, and van der Waals interactions can be easily overcome by low amounts of energy. Slightly elevated temperatures are sufficient to break hydrogen bonds. Changes in ionic strength, pH, concentration of particular ions, etc., all potentially have profound effects on macromolecular structures dependent on the weak forces.

11. Describe what is meant by the phrase "cells are steady-state systems."

Answer: Life is characterized as a system through which both energy and matter flow. The consequence of energy flow in this case is order, the order of monomeric units in biopolymers, which in turn produce macromolecular structures that function together as a living cell.

12. The genome of the* Mycoplasma genitalium *consists of 523 genes, encoding 484 proteins, in just 580 074 base pairs (Table 1.6). What fraction of the* M. genitalium *genes encode proteins? What do you think the other genes encode? If the fraction of base pairs devoted to protein-coding genes is the same as the fraction of the total genes that they represent, what is the average number of base pairs per protein-coding gene? If it takes 3 base pairs to specify an amino acid in a protein, how many amino acids are found in the average* M. genitalium *protein? If each amino acid contributes on average 120 daltons to the mass of a protein, what is the mass of an average* M. genitalium *protein?

Answer:

$$f_{protein} = \frac{484}{523} = 0.925 \text{ or } (92.5\%)$$

The other genes likely code for ribosomal RNAs and transfer RNAs. To make a functional ribosome it takes at least three ribosomal RNAs, a small subunit rRNA, a large subunit rRNA, and a 5S rRNA. To decode 61 triplet codons requires a minimum of 32* tRNAs. So, a minimum set of tRNAs and rRNAs is 35 (32 + 3). Of the 523 genes, 484 are proteins leaving 39 genes to code for RNAs.

Of the few RNAs that we are missing by this accounting one is the RNA portion of RNase P, a ribonuclease involved in tRNA processing. Another is the so-called 10Sa RNA, a tRNA-like RNA that is involved in decoding faulty mRNAs. The 4.5S RNA of the signal recognition particle, a complex involved in synthesis of membrane and secreted proteins, is also coded in the genome. This leaves perhaps one or two RNAs unaccounted for whose functions are still unknown.

A complete listing of genes for *M. genitalium* may be found by doing a search at the NCBI website (http://www.ncbi.nlm.nih.gov/) for this organism. You can either restrict your search to "Genome" using the pull-down search menu or do a search on all databases and then inspect hits for the genome database. Information for *M. genitalium* G37 is in NC_000908.

* Essentially 2 tRNAs for each XXN triplet set except for TAN, which only requires one. This is because TAA and TAG are stop codons that require proteins for recognition. This would give 31 tRNAs, but an extra one should be included for initiation of protein synthesis. In bacteria, a methionine codon starts a protein-coding region and it is decoded by a special initiator tRNA, which is different from the one used at internal methionine codons.

Assuming no overlap of genes:

$$\text{Amount of genome devoted to proteins} = \frac{484}{523} \times 580\,074 = 536\,818 \text{ bp}$$

The average number of base pairs per protein-coding gene is found by dividing this number by the number of protein genes. Thus,

$$\text{Average size of gene coding for protein} = \frac{536\,818}{484} = 1109 \ \frac{\text{bp}}{\text{protein}}$$

Note: This number is simply the genome size divided by the total number of genes.

$$\text{Average number of amino acids} = \frac{1109}{3} = 370 \ \frac{\text{amino acids}}{\text{protein}}$$

To calculate the actual average number of amino acids in *M. genitalium* proteins, visit NC_000908 at the NCBI website (www.ncbi.nlm.nih.gov/). When you search "All Databases" with NC_000908, you come up with a list of resources. Click on "Genome" to show 5 different "Mycoplasma genitalium" strain genome-sequencing projects. Under "Organism," find "Mycoplasma genitalium G37," the most representative of the species, and click "475" under "Protein." This gets you a list of all the *M. genitalium* G37 proteins and a protein length histogram. Here, you will find the "average" protein length to be 369 amino acids.

$$\text{Average protein size} = 370 \times 120 = 44\,280 \text{ daltons}$$

Note: The actual average of the weight of 20 amino acids is 138 daltons, but a weighted average (to consider that some amino acids are more common than others) gives an average of 128 daltons. In addition, both these numbers include H_2O, which should be removed when calculating the size of a polypeptide, since linking 2 amino acids involve the loss of H_2O. Hence, 120 or 110 daltons should be used for calculations.

13. Studies of existing cells to determine the minimum number of genes for a living cell have suggested that 206 genes are sufficient. If the ratio of protein-coding genes to non–protein-coding genes is the same in this minimal organism as the genes of* Mycoplasma genitalium*, how many proteins are represented in these 206 genes? How many base pairs would be required to form the genome of this minimal organism if the genes are the same size as* M. genitalium *genes?

Answer: For *M. genitalium* we determined in Problem 12 that 92.5% of the genes of this organism are protein-coding genes. Assuming the same percentage applies to a minimum set of genes, then 191 of the 206 genes are protein-coding genes.

$$\text{Protein-coding genes} = 0.925 \times 206 = 190.6 = 191$$

In Problem 12, we were told that 580 074 base pairs code for 523 genes. The genome size required to code for 206 genes is calculated as follows:

$$\frac{580\,074}{523} = \frac{x}{206}$$

$$x = 206 \times \frac{580\,074}{523} = 228\,480$$

Note: This calculation assumes that genes essentially do not overlap. A smaller genome size could be possible by allowing overlapping, but this would constrain the protein sequences.

14. Virus genomes range in size from approximately 3500 nucleotides to approximately 280,000 base pairs. If viral genes are about the same size as* M. genitalium *genes, what is the minimum and maximum number of genes in viruses?

Answer: In Problem 12, we determined that the average gene size in *M. genitalium* is 1109 (the genome size [580 074] divided by the number of genes [523]). Applying this average gene size to viral genomes, we find:

$$\text{Minimum number of viral genes} = \frac{3\,500}{1\,109} = 3.15 \approx 3 \text{ genes}$$

$$\text{Maximum number of viral genes} = \frac{280\,000}{1\,109} = 252 \text{ genes}$$

15. The endoplasmic reticulum (ER) is a site of protein synthesis. Proteins made by ribosomes associated with the ER may pass into the ER membrane or enter the lumen of the ER. Devise a pathway by which

a. a plasma membrane protein may reach the plasma membrane

b. a secreted protein may be deposited outside the cell

Answer: Protein synthesis starts out on ribosomes located in the cytoplasm of cells. Proteins destined to be excreted or to become membrane proteins are synthesized with a signal sequence located near the N-terminus of the protein. (Protein synthesis begins at the N-terminus.) This signal sequence directs the ribosome to the endoplasmic reticulum where the ribosome docks with the reticular membrane. Endoplasmic reticulum studded with ribosomes is called rough endoplasmic reticulum. The signal-sequence-containing protein is synthesized by rough endoplasmic reticulum-bound ribosomes that synthesize the protein and simultaneously export it into the lumen of the endoplasmic reticulum. For a protein to be transported to the plasma membrane, it must be packaged into membrane vesicles in the endoplasmic reticulum, since the reticular membrane is separate from the plasma membrane. Vesicles from the endoplasmic reticulum containing membrane proteins do not, however, move directly to the plasma membrane. Rather, they are routed to the Golgi apparatus where a variety of post-translational modifications occur. Once proteins move through the Golgi, they are repackaged into vesicles that are directed to the plasma membrane. Secreted proteins follow the same pathway. Both membrane proteins and excreted proteins contain signal sequences that get them into the endoplasmic reticulum. Membrane proteins contain an additional domain or domains that are hydrophobic in nature and anchor the proteins into the reticular membrane.

Additional Problems

1. Silicon is located below carbon in the periodic chart. It is capable of forming a wide range of bonds similar to carbon, yet life is based on carbon chemistry. Why are biomolecules made of silicon unlikely?

2. Identify the following characters of the Greek alphabet: α, β, γ, δ, Δ, ε, ζ, θ, κ, λ, μ, ν, π, ρ, σ, Σ, τ, χ, ϕ, ψ, and ω.

3. Give a common example of each of the weak forces at work.

4. On a hot dry day, leafy plants may begin to wilt. Why?

Abbreviated Answers

1. Covalent silicon bonds are not quite as strong as carbon covalent bonds because the bonding electrons of silicon are shielded from the nucleus by an additional layer of electrons. In addition, silicon is over twice the weight of carbon. Also, silicon oxides (rocks, glass) are extremely stable and not as reactive as carbon.

2. These Greek letters are commonly used in biochemistry, though this set is not the complete Greek alphabet: alpha (α), beta (β), gamma (γ), delta (δ), capital delta (Δ), epsilon (ε), zeta (ζ), theta (θ), kappa (κ), lambda (λ), mu (μ), nu (ν), pi (π), rho (ρ), sigma (σ), capital sigma (Σ), tau (τ), chi (χ), phi (ϕ), psi (ψ), and omega (ω), the last letter of the Greek alphabet.

3. Ice is an example of a structure held together by hydrogen bonds. Sodium and chloride ions are joined by ionic bonds in table salt crystals. A stick of butter is a solid at room temperature because of van der Waals forces. The energetically unfavorable interactions between water and oil molecules cause the oil to coalesce.

4. The tonoplast loses water and begins to shrink, causing the plant cell membrane to exert less pressure on the cell wall.

Summary

The chapter begins with an outline of the fundamental properties of living systems: complexity and organization, biological structure and function, energy transduction, and self-replication. What are the underlying chemical principles responsible for these properties? The elemental composition of biomolecules is dominated by hydrogen, carbon, nitrogen, and oxygen. These are the lightest elements capable of forming strong covalent bonds. In particular, carbon plays a key role serving as the backbone element of all biomolecules. It can participate in as many as four covalent bonds arranged in tetrahedral geometry and can produce a variety of structures including linear, branched, and cyclic compounds.

The four elements are incorporated into biomolecules from precursor compounds: CO_2, NH_4^+, NO_3^-, and N_2. These precursors are used to construct more complex compounds such as amino acids, sugars, and nucleotides, which serve as building blocks for the biopolymers; and proteins, polysaccharides, and nucleic acids, as well as fatty acids and glycerol, which are the building blocks of lipids. These complex macromolecules are organized into supramolecular complexes such as membranes and ribosomes that are components of cells, the fundamental units of life.

Proteins, nucleic acids, and polysaccharides are biopolymers with structural polarity due to a head-to-tail arrangements of asymmetric building block molecules. In these biopolymers, the building blocks are held together by covalent bonds, but they assume an elaborate architecture due to weak, non-covalent forces such as van der Waals interactions, hydrogen bonds, ionic bonds, and hydrophobic interactions. The three-dimensional shape is important for biological function, especially for proteins. At extreme conditions such as high temperature, high pressure, high salt concentrations, extremes of pH, and so on, the weak forces may be disrupted, resulting in loss of both shape and function in a process known as denaturation. Thus, life is confined to a narrow range of conditions.

Life demands a flow of energy during which energy transductions occur in the organized, orderly, small, manageable steps of metabolism, each step catalyzed by enzymes.

The fundamental unit of life is the cell. There are two types: eukaryotic cells with a nucleus and prokaryotic cells without a nucleus. Prokaryotes are divided into two groups, eubacteria and archaea. All cells contain ribosomes, which are responsible for protein synthesis; however, prokaryotic cells contain little else in the way of subcellular structures. Eukaryotic cells, found in plants, animals, fungi, and single cell organisms (protista), contain an array of membrane-bound compartments or organelles, including a nucleus, mitochondria, chloroplasts, endoplasmic reticulum, Golgi apparatus, vacuoles, lysosomes, and perixosomes. Organelles are internal compartments in which particular metabolic processes are carried out.

