

# 1

## Atoms and Molecules

### 1.1 Atoms

The most important elements of life are hydrogen H, carbon C, nitrogen N, and oxygen O. These four elements together with calcium Ca and phosphorus P make up 99 percent of the mass of the human body. Potassium K, sulfur S, chlorine Cl, sodium Na, and magnesium Mg constitute another 3/4 of 1 percent. Trace amounts of manganese Mn, iron, Fe, cobalt Co, copper Cu, zinc Zn, iodine I, selenium Se, bromine Br, and molybdenum Mo make up the remaining 1/4 of 1 percent of the mass of the human body. These 17 elements are among the most abundant elements on the surface of the Earth. Elements heavier than lithium Li were made in stars that later exploded.

### 1.2 Bonds between atoms

Electrostatic interactions form four kinds of bonds between atoms: ionic, hydrogen, van der Waals, and covalent.

The simplest is the **ionic bond** which occurs when atoms have opposite charges. The potential energy between two atoms carrying charges  $q$  and  $q'$  and separated by a distance  $r$  in the vacuum is

$$V(r) = \frac{q q'}{4\pi\epsilon_0 r} \quad (1.1)$$

in which  $\epsilon_0$  is the **electric constant**  $\epsilon_0 = 8.854 \times 10^{-12}$  F/m. The resulting force

$$\mathbf{F} = -\nabla V(r) \quad (1.2)$$

is attractive when the charges have opposite signs and repulsive when they

have the same sign. Many crystals are arrangements of ions of opposite charge in solid lattices. Table salt, NaCl, is a common example.

A typical ionic bond in vacuum is 0.25 nm long and has an energy of 3.47 eV or 80 kcal/mol. This energy is the kinetic energy an electron gets from a potential difference of 3.47 V; it also is the energy that when given to Avogadro's number ( $N_A = 6.022 \times 10^{23}$ ) of water molecules would raise the temperature of a liter of water by 80°C.

**Example 1.1** (Units of energy ). An **electron Volt eV** is the energy that an electron gets when falling through a potential difference of one Volt. Since the charge of an electron is  $e = 1.60217662 \times 10^{-19}$  of a Coulomb C this energy is  $1.602 \times 10^{-19}$  of a Joule J so  $1 \text{ eV} = 1.602 \times 10^{-19} \text{ J}$ . The **kilocalorie, kcal** is the energy ( 4,184 J ) that raises 1 kilogram kg of water or 1 liter of water by 1 degree centigrade or Kelvin at a pressure of 1 atmosphere. A **mol** of a given molecule is **Avogadro's number**  $6.0221367 \times 10^{23}$  of those molecules. So  $1 \text{ eV} = 6.0221367 \times 10^{23} \times 1.602 \times 10^{-19} / 4184 = 23.06 \text{ kcal/mol} = 6.0221367 \times 10^{23} \times 1.602 \times 10^{-19} = 96.485 \text{ kJ/mol}$ . A Dalton (Da) is one gram per mole or  $1.66 \times 10^{-27} \text{ kg}$ . An atom of oxygen has a mass of 16 Da.  $\square$

Biology does not normally take place in a vacuum. The potential energy of two atoms of charges  $q$  and  $q'$  separated by a distance  $r$  in water is

$$V(r) = \frac{qq'}{4\pi\epsilon r} \quad (1.3)$$

in which the **permittivity**  $\epsilon$  of water is  $\epsilon = 80.1\epsilon_0$  at 20°C and  $74.16\epsilon_0$  at 37°C. This drop of permittivity with temperature means that the electrostatic interactions become stronger as the temperature rises, which may partly explain why cells are so sensitive to temperature. Because ionic bonds are so short, their energies in water drop from their vacuum values by a factor of 27 rather than by a factor of 80, to about 3 kcal/mol.

A hydrogen atom that is bonded to an atom of nitrogen or oxygen loses part of its electron cloud to the nitrogen or oxygen atom. The hydrogen atom therefore carries a small positive charge, and the atom of nitrogen or oxygen a small negative charge. The resulting ionic bond between a positive hydrogen atom bonded to an atom of nitrogen or oxygen and a negative nitrogen or an oxygen atom is called a **hydrogen bond**. Hydrogen bonds are about 0.3 nm long and have an energy of 4 kcal/mol in vacuum and 1 kcal/mol in water.

The **van der Waals bond** is about 0.35 nm long and has an energy of about 0.1 kcal/mol in vacuum and in water. It arises because two neutral

atoms in their ground states attract each other weakly. This electrostatic attraction is simplest in the case of two hydrogen atoms.

**Example 1.2** (The van der Waals bond). Let  $R$  be the distance between two nuclei one at the origin and the other at  $R\hat{z}$  so that the electrons are at  $\mathbf{r}_1$  and at  $R\hat{z} + \mathbf{r}_2$ . Then the hamiltonian  $H$  for the two atoms is  $H = H_{01} + H_{02} + V$  in which the electrostatic potential is

$$V = \frac{e^2}{4\pi\epsilon_0} \left[ \frac{1}{R} + \frac{1}{\sqrt{(R\hat{z} + \mathbf{r}_2 - \mathbf{r}_1)^2}} - \frac{1}{\sqrt{(\mathbf{r}_1 - R\hat{z})^2}} - \frac{1}{\sqrt{(R\hat{z} + \mathbf{r}_2)^2}} \right] \quad (1.4)$$

in SI units. If the distance  $R$  between the atoms is much bigger than the distances  $r_1$  and  $r_2$  of the electrons from their nuclei, then one can expand the square-roots in powers of  $1/R$  and approximate (exercise 1.1) the potential as

$$V \approx \frac{e^2}{4\pi\epsilon_0 R^3} (x_1 x_2 + y_1 y_2 - 2z_1 z_2) \quad (1.5)$$

Since hydrogen atoms are spherically symmetric in their ground states, the mean value of this potential vanishes. But second-order perturbation theory gives the change in the energy of the two atoms as a sum over all the states of the two atoms except the ground state  $|0\rangle$

$$\Delta E(R) = - \sum_{n>0} \frac{|\langle n|V|0\rangle|^2}{E_n - E_0}. \quad (1.6)$$

Since the potential  $V$  goes as  $1/R^3$ , we see that the energy  $\Delta E(R)$  varies with the distance between the atoms as  $1/R^6$ . Since  $E_n - E_0 \geq E_2 - E_0$  and since  $\langle 0|V|0\rangle = 0$ , we have

$$\begin{aligned} \Delta E(R) &\leq - \frac{1}{E_2 - E_0} \sum_n |\langle n|V|0\rangle|^2 \\ &\leq - \frac{1}{E_2 - E_0} \sum_n \langle 0|V|n\rangle \langle n|V|0\rangle \\ &\leq - \frac{\langle 0|V^2|0\rangle}{E_2 - E_0}. \end{aligned} \quad (1.7)$$

A detailed computation shows that the van der Waals potential between two hydrogen atoms is

$$\Delta E(R) \approx -6.5 \frac{e^2 a_0^5}{4\pi\epsilon_0 R^6} \quad (1.8)$$

in which  $a_0 = 4\pi\epsilon_0 \hbar^2 / m e^2 = 5.292 \times 10^{-11}$  m is the Bohr radius of the hydrogen atom (Schiff, 1968).  $\square$

As two neutral atoms approach each other more closely than a few nanometers, the van der Waals interaction can turn into a **covalent bond**. Electrons are several thousand times lighter than nuclei, and so they quickly go where the attractive force of the nuclei of the approaching atoms pulls them. They go between the two nuclei and form a negative charge that attracts the nuclei toward them. This attraction of the electrons to the nuclei and of the nuclei to the electrons holds the atoms in a tight covalent bond about 0.15 nm apart. The energy of covalent bonds is of the order of a few electron volts or 90 kcal/mol and is the same in vacuum as in water.

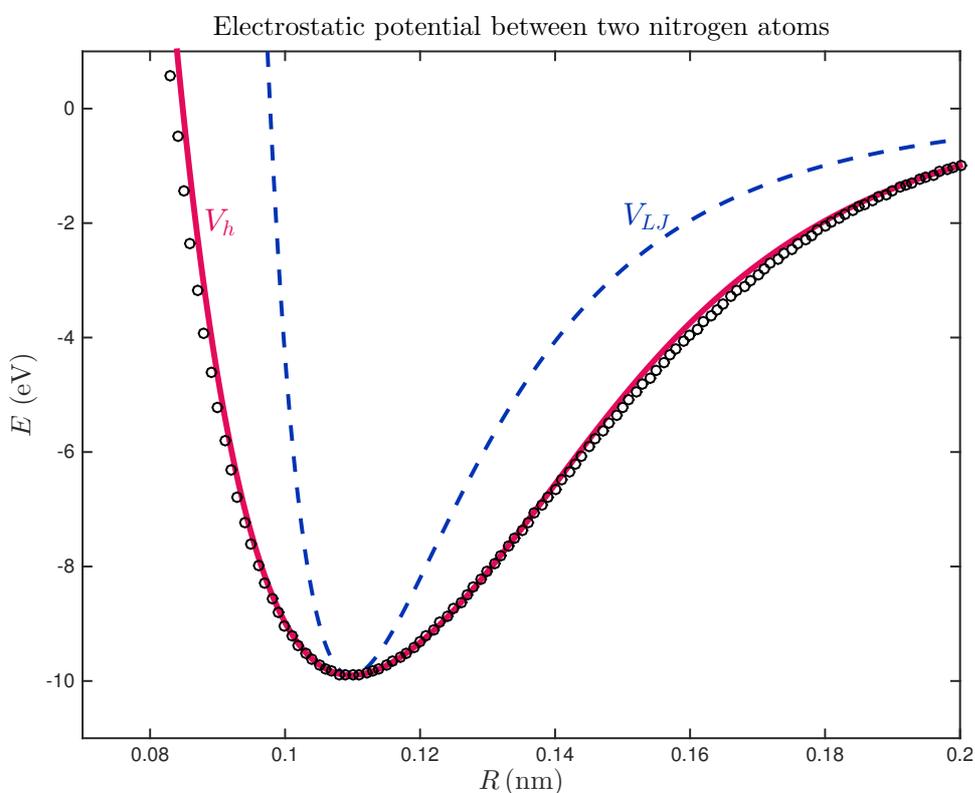


Figure 1.1 The hybrid potential  $V_h(R)$  (1.9, red solid line) fits the potential between two nitrogen atoms obtained from spectroscopy (black circles) better than the Lennard-Jones potential  $V_{LJ}(R)$  (1.10, blue dashes). The binding of  $N_2$  is unusually strong because of its triple bond  $N \equiv N$ , which is why nitrogen is used in explosives.

One can approximate the electrostatic potential between two neutral atoms

by the hybrid potential

$$V_h(R) = ae^{-bR}(1 - cR) - \frac{C_6}{R^6 + dR^{-6}} \quad (1.9)$$

which combines the Rydberg formula of spectroscopy with the London formula for pairs of atoms. Formulas exist that relate the parameters  $a$ ,  $b$ ,  $c$ , and  $d$  to the internuclear separation  $R_0$ , the depth  $E_0$ , curvature  $k$  of the potential at its minimum, and the van der Waals coefficient  $C_6$  of the pair (Cahill, 2007). In 1924, John Lennard-Jones introduced as a teaching device the simple potential

$$V_{LJ} = E_0 \left[ \left( \frac{R_0}{R} \right)^{12} - 2 \left( \frac{R_0}{R} \right)^6 \right] \quad (1.10)$$

in which  $R_0$  is the distance between the nuclei at which the potential energy assumes its minimum value  $-E_0$  (Lennard-Jones, 1924). The hybrid potential is more accurate than the Lennard-Jones potential as illustrated by the figure (1.1) for the case of two nitrogen atoms.

### 1.3 Oxidation and reduction

Atoms bind electrons more tightly are said to be more **electronegative**. On Linus Pauling's scale, the electronegativities of the main atoms that occur in the human body are listed in the table (1.1). Among them, oxygen

Table 1.1 *Electronegativities*

K	Na	Ca	Mg	P	H	C	S	N	Cl	O
0.82	0.93	1.00	1.31	2.19	2.20	2.55	2.58	3.04	3.16	3.44

is the most electronegative with an electronegativity of 3.44. Because it is more electronegative than most other atoms, oxygen steals electrons from them, raising their charges in a process called **oxidation**. More generally, any process that raises the electric charge of an atom or molecule is an oxidation. **Reduction** is the gain of electrons or the lowering of the charge on an atom or molecule.

Hydrogen is less electronegative than C, S, N, Cl, and O. It lends its electron to these atoms. So when one of them loses a hydrogen atom, it also loses the electron it was borrowing from the hydrogen atom, and its charge rises. Thus the loss of a hydrogen atom is an oxidation, and the gain of a hydrogen atom is a reduction. When methane  $\text{CH}_4$  is burned in the reaction  $\text{CH}_4 + 2\text{O}_2 \rightarrow \text{CO}_2 + 2\text{H}_2\text{O}$ , the carbon atom, which had borrowed

electrons from four hydrogen atoms, is oxidized as it loses them to two oxygen atoms. The oxygen atoms of  $O_2$ , which had shared their electrons with each other, are reduced as they gain electrons from the carbon atom and from the hydrogen atoms.

Electronegative implications: Because the electronegativity of oxygen 3.44 exceeds that of hydrogen 2.20 by 1.24, the electrons in a water molecule stay close to the oxygen atom and give water its huge dipole moment  $1.855 \text{ D} = 6.188 \times 10^{-30} \text{ C}\cdot\text{m}$ . Water is a liquid of dipoles in which the positively charged hydrogen atoms form hydrogen bonds with the negatively charged oxygen atoms of nearby water molecules. Similar hydrogen bonds form in proteins between hydrogen atoms drained of their electrons by oxygen or nitrogen atoms and nearby negatively charged oxygen or nitrogen atoms. Because the electronegativity of carbon 2.55 exceeds that of hydrogen 2.20 by only 0.35, the electrons of hydrocarbons arrange themselves fairly uniformly among the carbon and hydrogen atoms. Thus hydrocarbons have no intrinsic electric dipole moments and are **nonpolar**.

#### 1.4 Molecules in cells

Water molecules make up about 70 percent of the mass of a living cell. This molecule  $H_2O$  is an electric dipole: the two hydrogen atoms carry a small positive charge equal to one-half of the negative charge on the oxygen atom. Its dipole moment is  $d = 6.2 \times 10^{-30} \text{ mC}$  or 1.85 D, a unit named after Debye ( $1 \text{ D} = 3.34 \times 10^{-30} \text{ mC}$ ). It is this huge electric dipole moment that increases the permittivity of water from that of vacuum by a factor of about 80.

Macromolecules make up about 26 percent of the mass of a cell. These are mainly proteins, nucleic acids, and polysaccharides.

Sugars, amino acids, nucleotides, fatty acids, their precursors, and other small molecules make up the remaining 4 percent of the mass of a living cell.

#### 1.5 Sugars

Sugars are carbohydrates, that is, they are molecules made from carbon and water.

**Example 1.3** (Sugar and sulfuric acid). If you pour some concentrated sulfuric acid  $H_2SO_4$  into a beaker of sugar and stir, the acid will draw the water from the sugar, and a column of black carbon will rise from the beaker

in a cloud of steam and acid. Do this experiment outside, not in a closed classroom.  $\square$

Simple monosaccharides  $(\text{CH}_2\text{O})_n$  are chains of  $n \geq 3$  carbon atoms, including one **carbonyl**  $\text{C}=\text{O}$  group,  $n - 1$  **hydroxyl**  $\text{OH}$  groups, and  $n + 1$  hydrogen atoms. A carbonyl group at one end of the chain forms an **aldehyde**  $\text{H}-\text{C}=\text{O}$  group  $\text{H}_2\text{C}=\text{O}$ , and the sugar is an **aldose**. A carbonyl group in the middle of the chain is a **ketone** group  $>\text{C}=\text{O}$ , and the sugar is a **ketose**. The simplest sugars are **glyceraldehyde** and **dihydroxyacetone**, both  $(\text{CH}_2\text{O})_3$ . Two important aldoses are **ribose**  $\text{C}_5\text{H}_{10}\text{O}_5$ , which has 5 carbon atoms and 5 water molecules, and **glucose**  $\text{C}_6\text{H}_{12}\text{O}_6$ . **Fructose** is an  $n = 6$  ketose  $\text{C}_6\text{H}_{12}\text{O}_6$ .

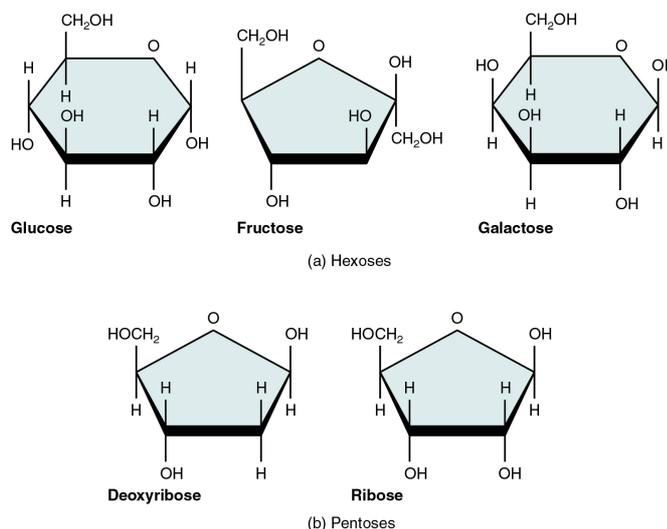


Figure 1.2 In these diagrams of five of the most important simple sugars, the unlabeled vertices denote carbon atoms. Thus ribose has five carbon atoms and glucose has six. From Anatomy & Physiology, Connexions Web site.

A monosaccharide's linear chain of carbon atoms can form a ring when its carbonyl group reacts with one of the hydroxyl groups so that its carbonyl carbon atom binds both to its oxygen and to the oxygen of the hydroxyl group, as illustrated the figure (1.2). The carbon atom thus bound to two oxygen atoms can then react with a hydroxyl group on a second sugar expelling a water molecule and forming a **glycosidic bond**  $\text{C}-\text{O}-\text{C}$  between

the two sugars. This joining of two sugars to form a **disaccharide** is called a **condensation reaction**.

Monosaccharides  $(\text{CH}_2\text{O})_n$  with the same number  $n$  of carbon-water units differ by where the carbonyl group is and in their ring forms by where the hydroxyl groups are. Some of these **isomers** are related by spatial reflexion. In cells right-handed sugars like D-glucose are more common than their left-handed isomers such as L-glucose. Two D-glucoses can form 11 different disaccharides.

Any number of monosaccharides can condense forming linear or branched polymers. Sugars can form huge numbers of different branched and unbranched polymers. A sugar polymer with only a few sugars is called an **oligosaccharide**. A sugar polymer with many sugars is a **polysaccharide**. **Glycogen** and starch are polysaccharides of glucose that cells use as sources of energy. **Glycosaminoglycans (GAGs)** are long unbranched chains of the same disaccharide, typically an **amino** sugar (a sugar in which an amine group replaces its terminal hydroxyl group,  $\text{C}-\text{OH} \rightarrow \text{C}-\text{NH}_2$ ) and a **uronic** sugar (a sugar in which a carboxyl group replaces a hydroxyl group,  $\text{C}-\text{OH} \rightarrow \text{C}-\text{COOH}$ ). **Cellulose** is an unbranched chain of glucose used in the cell walls of plants; **chitin** is a linear polysaccharide used in the exoskeletons of insects. Sugars bound to proteins are **glycoproteins**; sugars bound to fats are **glycolipids**.

### 1.6 Hydrocarbons, fatty acids, and fats

**Saturated hydrocarbons** are mainly polymers of  $\text{CH}_2$  groups. The shortest ones are methane  $\text{CH}_4$ , ethane  $\text{H}_3\text{C}-\text{CH}_3$ , propane  $\text{H}_3\text{C}-\text{CH}_2-\text{CH}_3$ , and butane  $\text{C}_4\text{H}_{10}$ . They are gases at room temperature. Longer polymers from pentane  $\text{C}_5\text{H}_{12}$  to pristane  $\text{C}_{19}\text{H}_{40}$  are liquids at room temperature. Hydrocarbon polymers as long as eicosane  $\text{C}_{20}\text{H}_{42}$  or longer are solids at room temperature, like wax. A saturated hydrocarbon has single bonds between adjacent carbon atoms and a maximal number of hydrogen atoms in each molecule; it is saturated with hydrogen. These molecules have a zigzag shape, but they are basically straight, like several v's **v v v v v v**.

**Unsaturated hydrocarbons** have double bonds between some carbon atoms. These molecules are crooked like the fatty acids of the unsaturated fat in figure 1.5.

Hydrocarbons are a source of energy for cells as for industry.

An organic acid has a **carboxylic group**  $\text{O}=\text{C}-\text{OH}$  which in pH 7 water

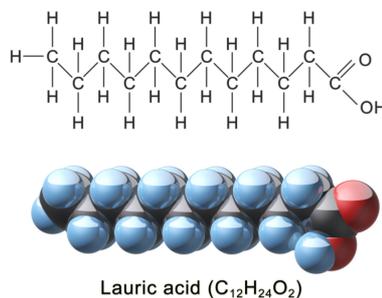


Figure 1.3 Lauric acid is a saturated fatty acid. (Courtesy of Bruce Blaus.)

loses a proton and becomes  $O=C-O^-$ . A **fatty acid** is a hydrocarbon with a carboxylic-acid group  $COOH$  instead of a methyl  $CH_3$  group at one end ( $-CH_2-CH_3 \rightarrow -CH_2-COOH$ ). Figure 1.3 displays Lauric acid, a saturated fatty acid.

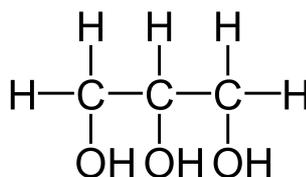
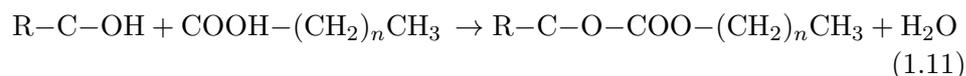


Figure 1.4 *Enzymes attach fatty acids and phosphate groups to the three hydroxyl groups of this glycerol or glycerin molecule to make triacylglycerides and phospholipids. (Courtesy of NEUROtiker.)*

A molecule  $X-O-H$  is an **oxoacid** if the group  $X$  is sufficiently **electronegative** as to attract electrons from the oxygen atom releasing a proton when in water,  $X-O-H \rightarrow X-O^- + H^+$ . An **acyl** group (aka, an **alkanoyl** group) is what's left when the hydroxyl group  $OH$  is removed from an oxoacid. An acyl group formed from a carboxylic acid  $R-COOH$  has a double-bonded oxygen atom  $R-C=O$  and a carbon atom that can form a single bond with something else.

Cells store fatty acids by connecting them in groups of three to **glycerol**  $HO-CH_2-CHOH-CH_2-OH$  (figure 1.4) through **ester linkages**



forming **triacylglycerides** also known as **triglycerides** and more simply as fats. In the image of the very unsaturated triacylglyceride in figure 1.5, the vertices represent carbon atoms, the hydrogen atoms are omitted (except for the methyl  $\text{CH}_3$  groups), the glycerol molecule is black, and the three unsaturated fatty acids are red, blue, and green. The saturated fatty acids of saturated fats are nearly straight, and so the fats can pack themselves into hard, dense arrays that stick to dinner plates and to the insides of one's arteries. Unsaturated fatty acids are crooked and so are softer and have lower melting points than saturated fats.

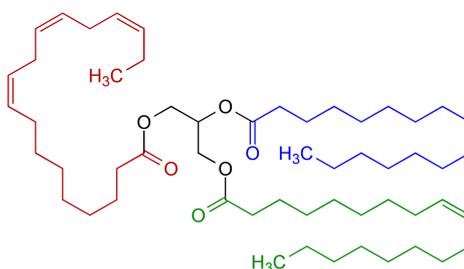


Figure 1.5 In this image of a very unsaturated triacylglyceride, the vertices represent carbon atoms, the hydrogen atoms are omitted (except for the methyl  $\text{CH}_3$  groups), the glycerol molecule is black, and the three unsaturated fatty acids are red, blue, and green. (Courtesy of Jü.)

## 1.7 Nucleotides

A **nucleotide** is a molecule made up of a **purine or pyrimidine base** attached by a glycosidic bond to a five-carbon sugar, itself bonded to one or more negatively charged phosphate  $\text{PO}_4$  groups as in figure 1.6. The bases are single or double rings of carbon and nitrogen atoms. The single rings are **cytosine C**, **thymine T**, and **uracil U**; all three are derived from a six-atom, **pyrimidine** ring of four carbon atoms and two nitrogen atoms. The double rings are the **purines adenine A** and **guanine G**. The pyrimidines C, T, and U and the purines A and G are called *bases* because their amine groups absorb protons under acidic conditions. In **ribonucleotides**, the five-carbon sugar is ribose, the bottom-right sugar in figure 1.2. In **deoxyribonucleotides**, the five-carbon sugar is deoxyribose, or ribose with a missing oxygen atom, the bottom-left sugar in that figure.



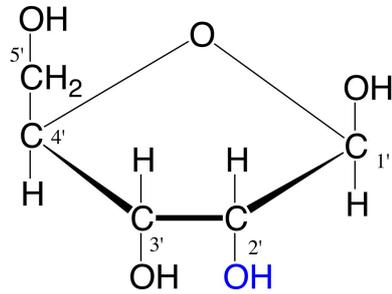


Figure 1.7 Ribose with the carbon atoms labeled. In deoxyribose, the 2' OH is replaced by H. (Courtesy of chronoxphy and flickr.)

the 3' carbon atom of the next ribose by another phosphodiester bond as in figure 1.8. (It is an acid because **phosphoric acid**,  $(\text{HO})_3\text{P}=\text{O}$ , loses

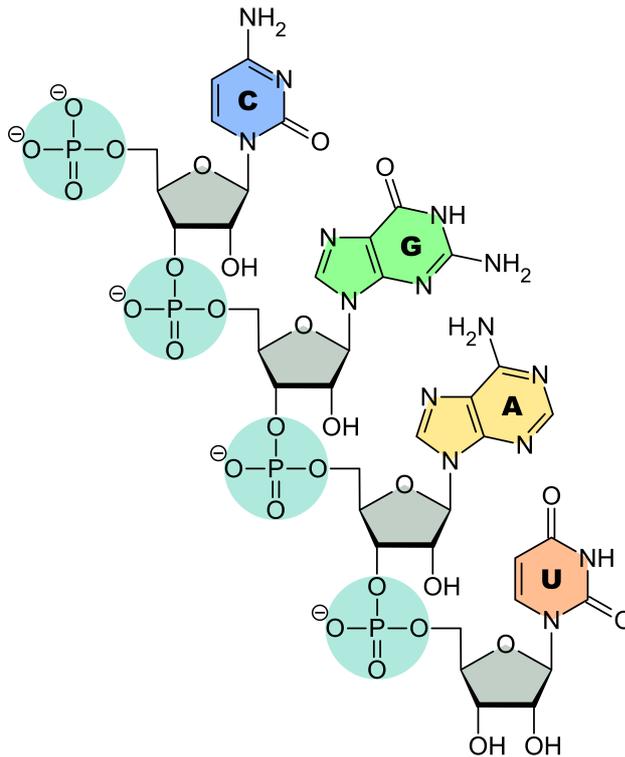


Figure 1.8 A very short strand of RNA which is a linear chain of CMP, GMP, AMP, and UMP. (Courtesy of Sponk and Wikimedia.)

two protons when it joins with a ribose and a base to form a nucleotide.) RNA is made of the monophosphate ribonucleotides CMP, GMP, AMP, and UMP. **DNA**, deoxyribonucleic acid, is made of the monophosphate deoxyribonucleotides CMP, GMP, AMP, and TMP.

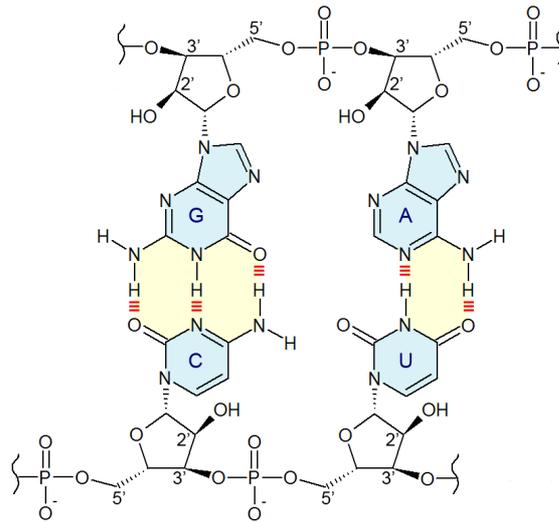


Figure 1.9 In RNA, guanine G forms three hydrogen bonds with cytosine C, and adenine A forms two with uracil U. (Courtesy of Wikipedia.)

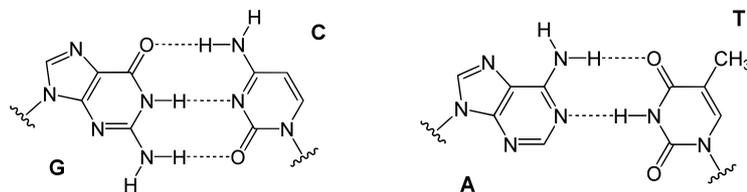


Figure 1.10 In DNA, guanine G forms three hydrogen bonds with cytosine C, and adenine A forms two with thymine T. (Courtesy of Wikipedia.)

The bases have oxygen and nitrogen atoms and amine and hydroxyl groups that can form hydrogen bonds (section ??) between pairs of bases. As shown in figure 1.9, guanine forms three hydrogen bonds with cytosine C, while adenine A forms only two with uracil in RNA and with thymine T in DNA. These Crick-Watson base pairs contribute to the structure of molecules of RNA and DNA.

Cells use RNA to do many things. Long strands of **messenger RNA**, mRNA, carry the protein recipes as sequences of **codons** (triplets of nu-

cleotides) each codon denoting a specific amino acid. A **ribosome** is a 2.5 MDa molecular machine that is one-third protein and two-thirds **ribosomal RNA**, rRNA, and that reads mRNA and links the correct amino acids in a chain to make the protein specified by the mRNA. There are about 20,000 ribosomes in an *E. coli* cell and as many as 10 million in a mammalian cell.

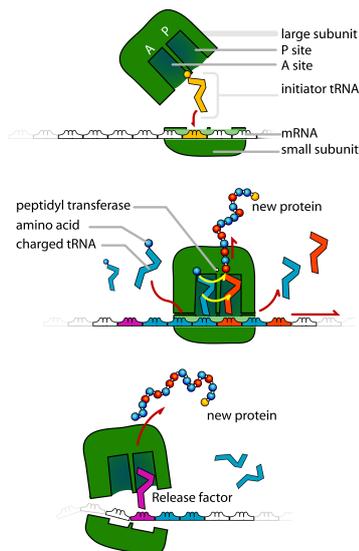


Figure 1.11 Cartoon of a ribosome using mRNA and tRNA's to make a protein. (Courtesy of LadyoffHats and Wikimedia.)

**Transfer RNA**, tRNA, are small molecules of RNA that embody the genetic code of figure 1.16. Figure 1.12 shows a bacterial tRNA without its amino acid. Each tRNA molecule has an amino acid at one end and the anticodon that specifically binds to the codon of mRNA that asks for that amino acid. For instance, suppose RNA polymerase reads the codon in DNA  $5'TTC3'$  and translates it to the mRNA codon  $5'UUC3'$ . Then a tRNA molecule with the anticodon  $3'AAG5'$  will bind to this mRNA codon  $5'UUC3'$ . That tRNA molecule will carry at its  $3'$  end, attached to an adenine by its carboxylic-acid group, the amino acid F, Phe, Phenylalanine, as required by the genetic code (figure 1.16). Most cells have 20 **synthetases**, one for each amino acid, that attach the right amino acid to the tRNA molecule that has its anticodon. An *E. coli* cell has some 200,000 tRNAs.

A **micro RNA**, miRNA, is a short (21–26 nucleotide) strand of RNA that controls how many protein molecules the cells ribosomes will make from the

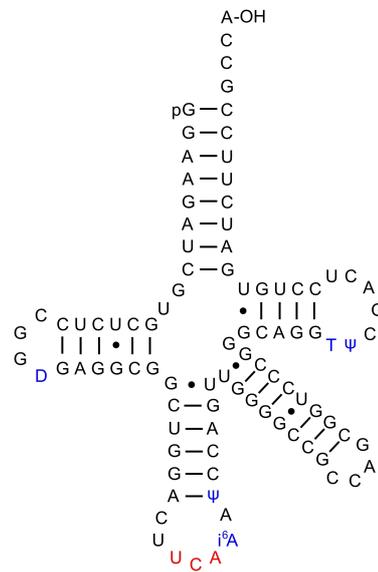


Figure 1.12 A tRNA of the bacterium *E. Coli* without its amino acid. (Courtesy of Yikrazuul and Wikimedia.)

mRNA for that protein. **Small nuclear RNAs (snRNAs)** help cells to process mRNA in the nucleus, and **small cytoplasmic RNAs (scRNAs)** play other roles in the cytoplasm. Cells surely use RNA for other purposes as yet undiscovered.

Cells use the Crick-Watson base pairs shown in figure 1.10 to store genetic information in DNA. DNA uses thymine instead of uracil because cytosine sometimes loses its amine group, gains an oxygen, and turns into uracil,  $C + H_2O \rightarrow U + NH_3$ . So by using thymine instead of uracil in DNA, cells can identify uracil in DNA as a mistake due to the decay of cytosine, and then correct that error by replacing the uracil with cytosine. When a cell divides into two daughter cells, the enzyme **DNA polymerase** duplicates the DNA of the cell. Its error rate is less than  $10^{-7}$  per base pair. Repair enzymes correct 99% of these errors for a net error rate of less than  $10^{-9}$  per base pair (McCulloch and Kunkel, 2008).

To make a protein, the enzyme **RNA polymerase II** copies the 5' to 3' strand of DNA onto a 5' to 3' strand of mRNA which is then exported from the nucleus of the cell. The **ribosome**, which is made of rRNA and protein, draws the strand of mRNA through it, starting at the 5' end and proceeding to the 3' end. (Carl Friedrich Gauss 1777-1855, Francis Crick 1916-2004, James Watson 1928-)

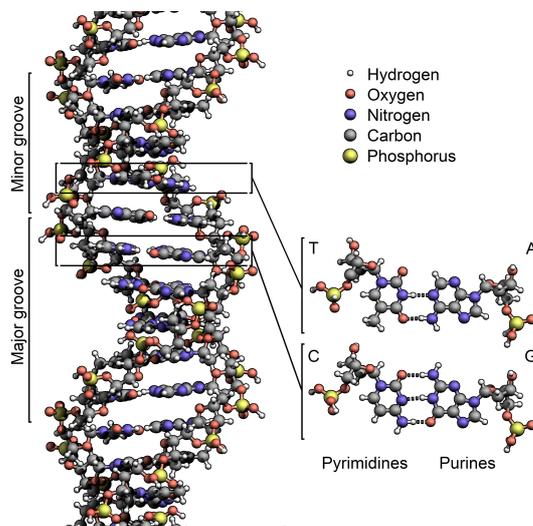


Figure 1.13 DNA.

### 1.8 Amino acids

An amino acid has an N–C–C backbone. In vacuum, the nitrogen atom is bonded to two hydrogen atoms forming an **amine group**  $\text{NH}_2$  and to the first carbon atom, called the **alpha carbon**, as illustrated by the left side of figure 1.14. The alpha carbon is bonded to that amine group, to a hydrogen atom, to the second carbon atom, which is part of a **carboxylic-acid group**  $\text{HO}-\text{C}=\text{O}$ , and to a side group represented as R in the figure (1.14). The 20 different amino acids differ by their side groups R. The second carbon atom is bound to the alpha carbon, to a hydroxyl group, and by a double bond to an oxygen atom, forming the carboxylic-acid group. Under physiological conditions, the carboxylic acid group of an amino acid loses a proton and the amine group gains a proton, as shown on the right side of the figure (1.14).

What are physiological conditions? The hydrogen-ion concentration of blood is  $c_{\text{H}^+} \approx 10^{-7.4}$  moles/liter (M). Its negative logarithm to the base 10 is the **pH**, so the pH of blood is about  $7.4 = -\ln_{10} c_{\text{H}^+}$ . The term **physiological conditions** means a pH near 7.4 and normal concentrations of **electrolytes** such as  $\text{Ca}^{2+}$ ,  $\text{Cl}^-$ ,  $\text{Mg}^{2+}$ ,  $\text{K}^+$ ,  $\text{Na}^+$ , and so forth.

The **genetic code** uses 64 codons, each of three nucleotides, to redun-

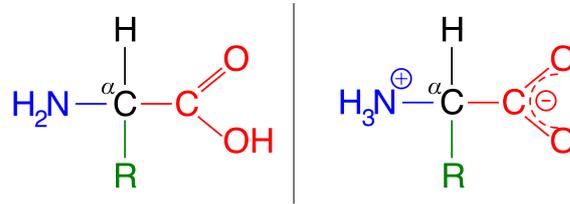


Figure 1.14 On the left is an amino acid with a side chain R in vacuum. On the right is an amino acid in water at pH 7.4. Its amine group has added a proton, and its carboxylic acid group has lost a proton, leaving it with an extra electron which its two oxygen atoms share (Courtesy of Yassine Mrabet.)

dantly specify the 20 amino acids and the stop signal as beautifully illustrated in figure 1.16.

Eight of the 20 amino acids have hydrocarbon side chains that are **nonpolar**, that is, they are not electric dipoles unless polarized by an electric field. Because water molecules are permanent electric dipoles, they attract each other like magnets and shove aside nonpolar molecules. Amino acids with nonpolar side chains are said to be **hydrophobic** or “water fearing.” The 8 hydrophobic amino acids, shown in yellow in the figure (1.16), are alanine Ala A, isoleucine Ile I, leucine Leu L, methionine Met M, phenylalanine Phe F, tryptophan Trp W, proline Pro P, and valine Val V. The side chain of proline loops around and forms a covalent bond with the amine group of its backbone.

Five of the 20 amino acids are **charged**. Because an amine group can add a proton, arginine Arg R, lysine Lys K, and histidine His H carry positive charges, that of H being smaller. When they take protons from a solution, they raise its pH. The side chains of these three positive amino acids therefore behave as bases and are blue in the figure (1.16).

Because a carboxylic group can lose the proton of its hydroxyl group leaving its electron stuck to the oxygen atom, aspartic acid Asp D and glutamic acid Glu E carry negative charges. They add protons to a solution. The side chains of these two negative amino acids therefore are acids are shown in rose in the figure. The five charged amino acids strongly attract and are attracted by water molecules. They tend to be on the outside of rather than on the inside of globular proteins.

The remaining 7 amino acids, shown in green in the figure, are serine Ser S, threonine Thr T, tyrosine Tyr Y, asparagine Asn N, and glutamine Gln Q, cysteine Cys C, and glycine Gly G. They are **polar** molecules, that is, they are permanent electric dipoles. Serine, threonine, and tyrosine have

hydroxyl groups to which **kinases** can attach **phosphate groups**  $\text{PO}_4^{3-}$  and to which other enzymes can attach sugars through a glycosidic bond. Asparagine and glutamine have **amide**  $\text{H}_2\text{N}-\text{C}=\text{O}$  groups. Glycine with only a hydrogen atom as its side chain carries the intrinsic polarity of any amino acid in a protein: the hydrogen atom of its amine group is slightly positive and the oxygen atom of its carboxylic group is slightly negative.

The seventh polar amino acid cysteine has a side chain that ends in a **thiol group**  $\text{SH}$ , which is similar to a hydroxyl group. Two cysteines, located anywhere in the chain of amino acids that constitutes a protein, can form a disulfide bond as illustrated in figure 1.15. The disulfide bond locks the cysteines together and turns the linear chain into a loop that joins parts of the protein that otherwise would be widely separated.

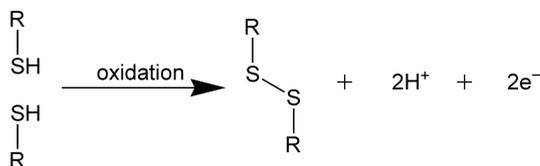


Figure 1.15 The thiol groups of two cysteines can form a disulfide bond releasing two protons and two electrons in an **oxidation** reaction.

As explained in section 1.3, an atom or molecule is **oxidized** if its charge increases; it is **reduced** if its charge decreases. Oxygen atoms tend to grab electrons from other atoms and molecules; they are highly **electronegative**. When an oxygen atom grabs electrons from a molecule, that molecule is said to be **oxidized**. The charge on the molecule becomes more positive. The successive oxidations of a carbon atom from methane  $\text{CH}_4$  to methanol  $\text{CH}_3\text{OH}$  to formaldehyde  $\text{CH}_2=\text{O}$  to formic acid  $\text{HO}-\text{CH}=\text{O}$  to carbon dioxide  $\text{O}=\text{C}=\text{O}$ , transfer negative charge from the carbon atom to oxygen and hydrogen atoms. The charge on the carbon atom becomes more positive. The disulfide reaction  $2\text{SH} \rightarrow \text{S}-\text{S} + 2\text{H}$  is an **oxidation** of the sulfur atoms because the electron clouds that the hydrogen atoms had shared with the sulfur atoms leave them and go off with the two hydrogen atoms or as two protons and two electrons. The charge on the sulfur atoms becomes more positive. The opposite reactions from  $\text{CO}_2$  to  $\text{CH}_4$  add negative charge to the carbon atom and so **reduce** its positive charge. Similarly, the breaking of a disulfide bond  $\text{S}-\text{S} + 2\text{H} \rightarrow 2\text{SH}$  is a **reduction**.

## 1.9 Proteins

Ribosomes join amino acids together by means of **peptide bonds** as in the figure 1.17.

A chain of a few amino acids is an **oligopeptide**; a chain of several amino acids is a **polypeptide**; a long chain is a **protein**. Proteins are most of a cell's dry mass. They do most of the things cells need done. They form the filaments that give a cell its structure. They catalyze its huge numbers of chemical reactions. They form the channels and pumps in the cell's plasma membrane that control the flow of small molecules into and out of the cell. They relay signals across the plasma membrane. They pull membrane-enclosed sacs of molecules along microtubules. They untie tangles of DNA. They function as antibodies, hormones, toxins, ropes, and so forth.

Most proteins are globular. Under physiological conditions, a protein assumes its **native state** of minimal **Gibbs free energy**,  $G(p, T) = U + pV - TS$  in which  $U$  is the internal energy and  $S$  is the entropy (explained in chapter 4). In its native state, a globular protein has a density nearly as high as that of a crystal and a specific form that permits it to do its work. One kind of globular protein, an **enzyme**, makes a specific chemical reaction run much faster than it otherwise would. Another kind, a **kinase**, adds a phosphate group to another protein. Another kind, a **phosphatase**, removes a phosphate group from another molecule. **Synthases** bind two molecules together. **Proteases** break up proteins by hydrolyzing their peptide bonds.

The smallest side group is a single hydrogen atom. That amino acid is called glycine abbreviated as Gly or G. Because the side group is a single hydrogen atom, G is very flexible—it can rotate freely about the N-C bond and about the C-C bond.

### Exercises

- 1.1 Derive the dipole approximation (1.5) from the quadrupole potential (1.4).

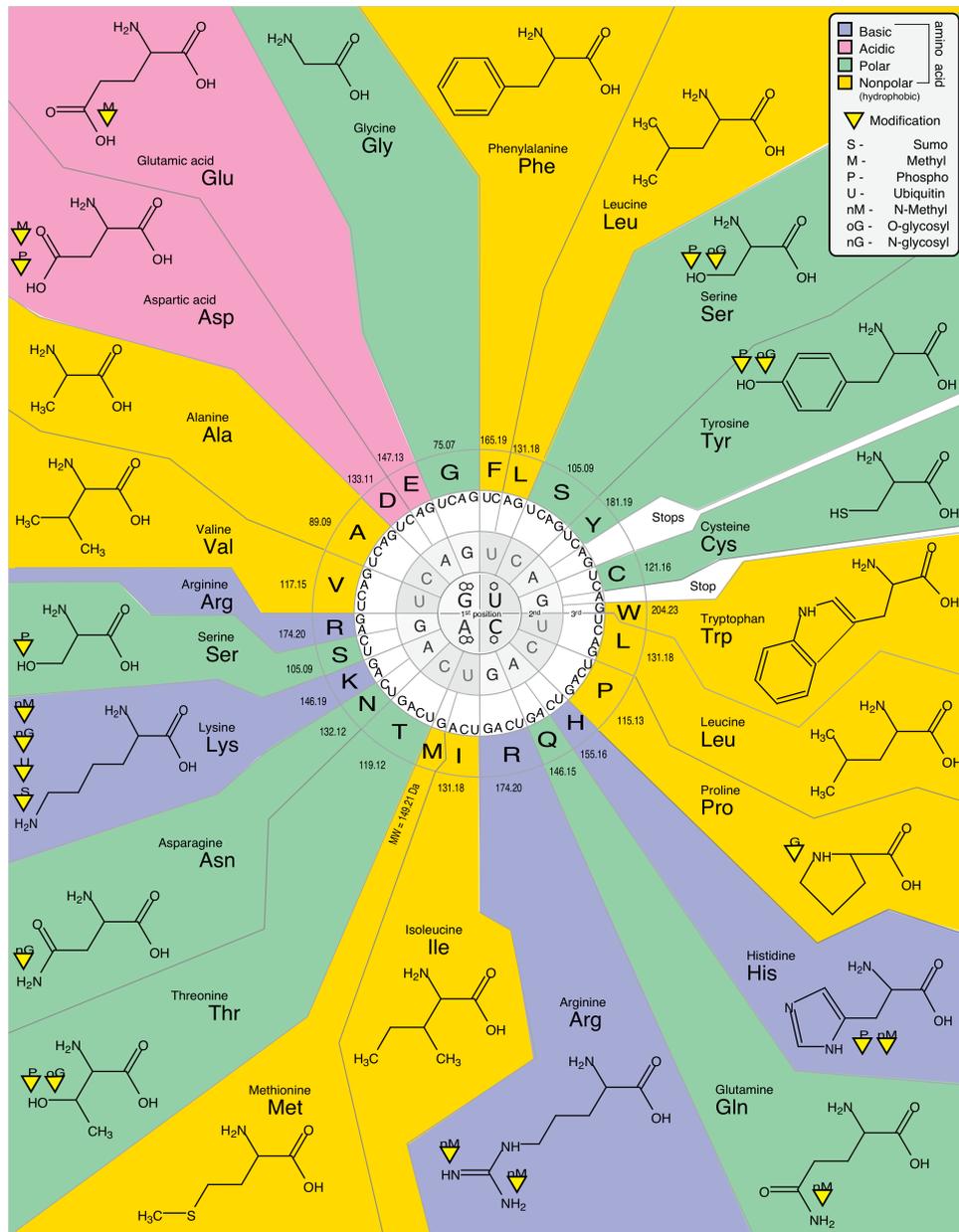


Figure 1.16 A codon is a triplet of the four nucleotides A, C, G, and T of DNA or of the four nucleotides A, C, G, and U of RNA. The 64 different codons redundantly specify 20 amino acids and 3 stop codons. For instance, the codon AUG specifies methionine, which also serves as a start codon. (Courtesy of Seth Miller and Kosi Gramatikoff.)

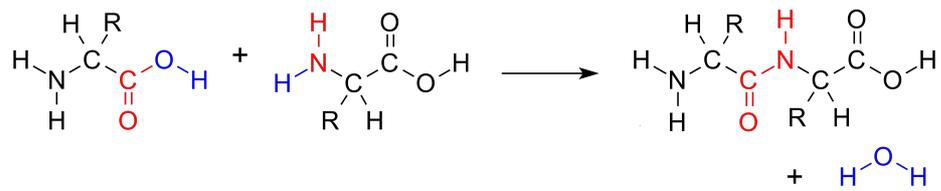


Figure 1.17 A peptide bond joins the carboxylic carbon of one amino acid to the nitrogen of the amino group of another amino acid. The four atoms O=C-N-H (red) lie in the same plane, whereas rotations occur around the N-C<sub>α</sub> and C<sub>α</sub>-C bonds. (Courtesy of V8rik at en.wikipedia.)