Chapter 3.1 The Information of Life 1
Molecular biology in less than 50 words 1
Thermal motion and diffusion 2
All objects are in motion 2
Thermal energy is proportional to temperature 2
The thermal velocity of a typical molecule 2
Some molecules are slower and others faster than average 3
Figure Maxwell_Boltzmann 4
The three states of matter 5
Figure the3states 6
Diffusion and mixing 8
Diffusion as a random walk 8
The square root of N law 9
Walks in one, two, and three dimensions 9
Figure walk_123 11
Brownian motion 12
Albert Einstein’s early papers 12
Boltzmann 's tombstone 13
Atoms are elements 14
Molecules 17
Figure ribose_views 18
Chemical bonds 20
The covalent bond 20
The ionic bond 20
Water and the hydrogen bond 21
Figure water 22
A nucleotide base 24
Figure guanine 25
A nucleotide 26
Figure gmp 27
A polynucleotide 28
Figure actg 29
Figure dna_isdn 31
Chapter Summary 32

Chapter 3.1 The Information of Life

It's going to be more difficult to talk about information in living organisms than it was when we were talking about information transport through the Internet. One problem is due to the origin and nature of life. Living things are self-replicating, and their present form and structure is the result of evolution during many millions of cycles of replication. There are no manuals to read or engineers to interview in order to determine the purpose and motivations of the design. We must rely on deductive reasoning: observe what the living organism does, determine its composition, and try our best to deduce cause and effect.

Another complication in understanding information transfer and storage in living organism is the close relationship between the actual material, the molecules that the organism is made up of, and the information. In fact, the molecules contain the information, transfer the information, and are the result of the transfer of information. If this sounds rather recursive, it should; life is a cycle in many ways.

On the Internet you also don't need to be concerned with the meaning of messages in order to know how they are transmitted. When you study the actual use of the Internet you need to know something about formats for text, images, etc., but formats are not meaning. However, with life it is difficult to even appreciate that information exists without knowing something about its use and meaning.

How can a living cell contain, transmit, and use information? While this is the topic of the next hundred pages or so, you might find a preview useful, ever if it may seem a little vague. A preview can give you the motive to read a more complete description as well as alerting you to the concepts that are going to be important. Here is my attempt to summarize all of molecular biology in a few sentences without using technical words (for brevity I use the term molecular biology for all fields of modern biology that refer to molecules, e.g. biochemistry, biophysics, cell biology, etc.).

Molecular biology in less than 50 words

- Living cells contains many types of objects, molecules, in constant motion; these molecules frequently collide.
- Molecules with complementary shapes stick together to form aggregates.
- These aggregates have structural and mechanical functions or produce chemical reactions that create new molecules allowing the cell to function and grow.

Those of you in the know will immediately think of exceptions, or might question how much information is actually contained in these simple statements; I did the best I could with less than 50 words. Now all I have to do in the remaining sections is describe the objects in a living cell, why they are in motion, why some objects stick to each other, and how mechanical or chemical reactions and other useful events occur as a result of the aggregates.

Thermal motion and diffusion

All objects are in motion

A basic fact about the world we live in is that all objects, even those that appear stationary to the eye, are in motion. The smaller the object, the proportionally more rapid and extreme the motion. The objects we see with our eyes are composed of an immense number of very, very small molecules, and thus we usually observe molecules only indirectly. As an example, the indirect effect of molecular motion is very familiar, it is called heat, and thus molecular motion is also called thermal motion.

Thermal energy is proportional to temperature

The temperature of an object is just the average energy due to the motion of each of its molecules. Thus, a degree on a temperature scale is actually a kind of energy unit. In order to find the thermal energy of a molecule in "real" energy units, e.g. Joules, we need a conversion factor. This factor is named after Ludwig Boltzmann, an Austrian physicist who made fundamental contributions in thermodynamics in the 1870s. The Boltzmann constant is 1.38×10^{-23} Joules/degree Centigrade.

However, we need to remember that zero degrees on the Centigrade scale was never meant to be, and is not, absolute zero, the temperature at which thermal energy is zero. Zero degrees Centigrade is just the temperature of melting ice, picked as a convenient reference. Absolute zero is about -273 degrees Centigrade, and thus room temperature is about 300 degrees absolute (or degrees Kelvin). Now we can calculate the thermal energy of a molecule at room temperature. Boltzmann showed that molecules have an average energy of $kT/2$ for each dimension, where k is his constant and T is the absolute temperature. Thus for an object moving in three dimensional space the average energy is:

$$E = 3\,(kT/2) = 3 \times 1.38 \times 10^{-23} \times 300 \;/\; 2 = 6.2 \times 10^{-21} \text{ Joules}$$

The thermal velocity of a typical molecule

The energy associated with the movement of an object, its kinetic energy, is equal to:

$$E = m\,v^2 \;/\; 2$$

where E is the energy, m is the mass, and v is the velocity. Since we now know the average thermal energy of molecules at room temperature we can calculate the average velocity if we know the mass of the molecules. Let's use oxygen as an example, since it represents about 20 percent of the earth's atmosphere and is so vital to life.

Chemists first determined the masses of molecules and atoms before they had the tools to accurately measure their absolute values. They were thus forced to describe masses relative to the mass of the lightest atom, hydrogen, which was assigned a "molecular weight" of 1. The actual mass of a single hydrogen atom now know with great accuracy, it is approximately 1.6 X 10^{-27} Kg. However, the old molecular weight units are convenient, and are thus still in use.

The molecular weight of an oxygen molecule is 32, so the mass, m, is:

$$m = 32 \times 1.6 \times 10^{-27} = 5.3 \times 10^{-26} \text{ Kg}$$

Rearranging the equation that defines the kinetic energy to give v:

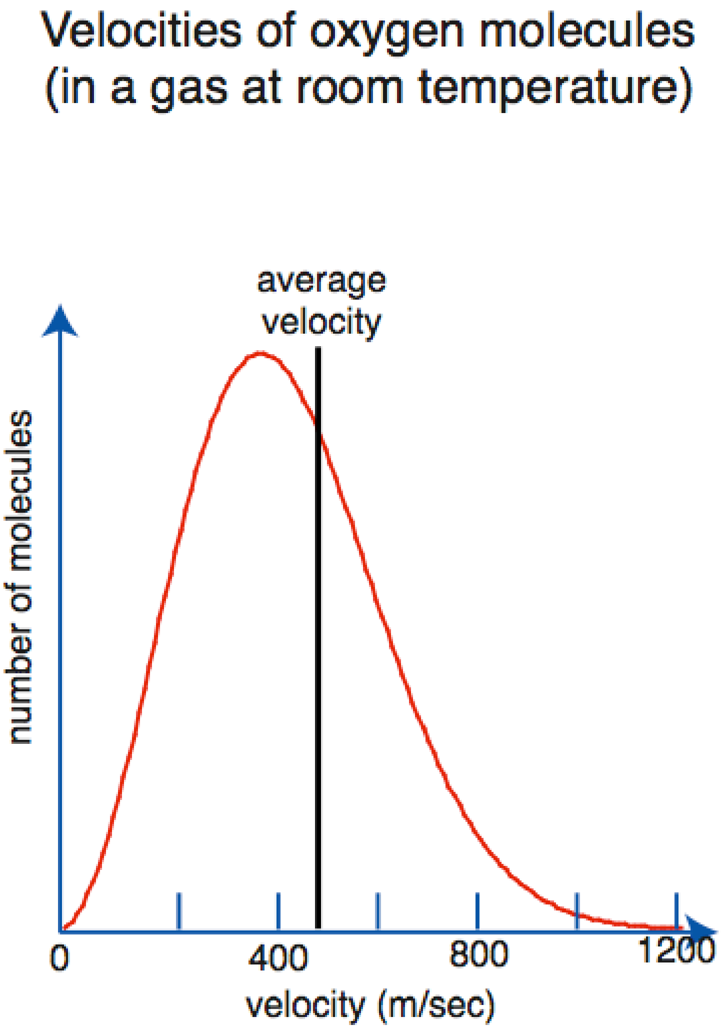
$$v = (2 E/m)^{1/2} = (2 \times 6.2 \times 10^{-21}/5.3 \times 10^{-26})^{1/2} = (2.3 \times 10^{5})^{1/2} = 480 \text{ m/s}$$

Does that seem fast? It’s close to the speed of sound in air, about 340 m/s, which is not just a coincidence; in air sound is transported by the thermal motion of molecules.

Some molecules are slower and others faster than average

I used the term average several times in the previous section when I talked about velocity due to thermal energy. In fact, molecules in a gas have a wide range of velocities. Boltzmann, along with Maxwell, (of electromagnetic fame) deduced the distribution of velocities in a collection of molecules at the same temperature. This Maxwell-Boltzmann velocity distribution, calculated for oxygen molecules at room temperature, is shown in Figure Maxwell_Boltzmann.

Figure Maxwell_Boltzmann



Maxwell_Boltzmann

Figure Maxwell_Boltzmann. The average velocity of molecules is determined by the temperature, but individual molecules can have very different velocities. This graph shows the distribution of velocities for oxygen molecules at a temperature of 20° C. It is important to notice that a few molecules have very high velocities (and thus energies). There are progressively fewer molecules with velocities greater than average, but there is no upper limit for the velocity that a molecule can have. Thus this graph doesn't show extremely energetic but correspondingly rare molecules.

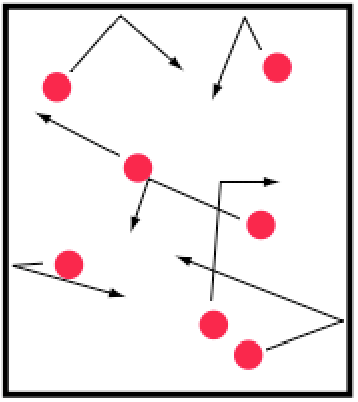
This curve can also be interpreted as the probability that a molecules will have the indicated velocity. The curve is asymmetrical about its maximum, with more molecules having very high velocities than very low. Thus, the average velocity of 480 m/s is greater than the most common velocity. The probability doesn't go to zero at the high end, 1200 m/s; the plot has just been arbitrarily clipped at this value. While the probability progressively decreases at higher velocities, it never goes to zero. The molecules that have high velocities, and thus high energy, are of great importance. They are the molecules that collide with other molecules to sometimes form new molecules. Molecules with lower energy also collide with each other, but they do not have the energy to get through the difficult (energy requiring) transition between new and old molecules.

The three states of matter

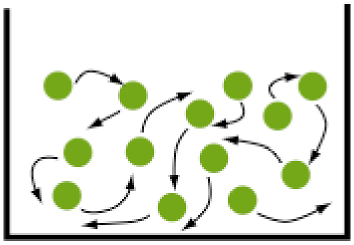
The nature of molecular motion is different in the three fundamental states of matter, gas, liquid, and solid, as diagramed in Figure the3states.

Figure the3states

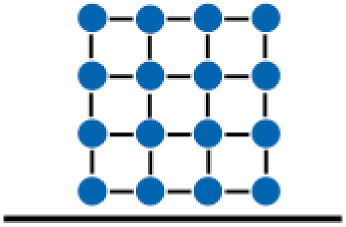
Three states of matter:
gas, liquid, solid



gas- no force between molecules until they collide; must be contained on all sides



liquid- molecules always feel each other but slide by neighbors; no difinite structure, so must be held in a vessel



solid- molecules held in fixed positions by strong forces; has a fixed shape that can be supported on a surface

the3states

Figure the3states. A collection of molecules can be in one of three states, a gas, a liquid, or a solid.

In a gas, molecules are not bound to each other and are typically separated by at least several molecular diameters. They are free to move in straight lines for some distance before colliding with the walls of the container or each other, after which they bounce off in different directions like billiard balls. Since gas molecules are not bound to each other, the gas must be completely contained or it will expand indefinitely. When two molecules collide and recoil, the total kinetic energy is redistributed between the two according to the relative speeds, point of contact, and angle between the two trajectories. Observation of a billiard game shows that sometimes all the energy is transmitted from one ball to the another while in other collisions each ball changes directions but keeps the same energy. This, effectively random, continuous energy redistribution generates a wide distribution of energies in the molecules.

In a gas at atmospheric pressure the distance between molecules is typically 10 - 20 molecular diameters. The only way one gas molecule can influence another is to hit it. Thus, the factor limiting the speed at which a disturbance, for example a sound wave, can move through a gas is the velocity of the molecules themselves. This explains the “coincidence” between the average speed of an oxygen molecule and the speed of sound in air (the other major constituent of air, nitrogen, has a similar molecular weight and thus a similar thermal velocity). High in the atmosphere the number of molecules per unit volume is much lower, and the molecules travel a greater distance between collisions. However, the temperature is also much lower, and thus the average molecular velocity is smaller. It is this lower velocity which causes the velocity of sound to be lower at high altitudes.

In a liquid, molecules are loosely bound to neighbors and are almost always in some sort of contact, but in their thermal dance they slip and slide past each other. Even though they don’t move in straight lines for any appreciable time, they are in constant motion. The attraction between liquid molecules is sufficient to keep most from leaving their neighbors and flying off into space, as gas molecules do. However, some liquid molecules do escape, a process called evaporation. In the absence of gravity a liquid will form a sphere, while under the influence of gravity it can be contained in a vessel, forming a horizontal surface in contact with the air. The liquid molecules may interact with the walls of the container ; a process called wetting.

In a solid, molecules have a permanent average position. Molecular movement consists mainly of vibration. Pure compounds can form crystals, in which each molecule and each atom in each molecule, has a fixed position defined by the symmetry of the crystal. Other solids are a somewhat random aggregate of micro-crystals and others may be almost completely amorphous.

While many materials can be easily classed as being in one of these three states, others have properties demonstrating a continuum of possible forms. An example is honey. At 100C it is a thick fluid, but below 0C it is almost a solid. The more slowly a fluid flows the more viscous it is said to be, thus honey becomes more viscous as it is cooled. Water also becomes more viscous as it is cooled, but it undergoes a sudden change at 0C and becomes a solid. Honey does not, rather the viscosity just continues to increase as it becomes colder. There is a special name for this type of material; a glass.

Diffusion and mixing

If several different chemical species are not uniformly distributed in a container or space the random thermal motion of molecules results produces mixing so that the final, equilibrium distribution is uniform. As an example, suppose we pour a solution of sugar, grape juice, and brandy into a glass and then very carefully and slowly add a layer of water over it, mixing the two layers as little as possible. The purpose of the sugar is to make the density of the lower layer greater than the pure water on top so that the system will be stable, the grape juice is to make the lower layer clearly distinguishable from the top, and the brandy is to give the concoction a little zing if you want to serve it as a cocktail. If you have been careful there will be a clean boundary between the two solutions. Since the heaver solution is on the bottom, this situation should be stable as long as the container is perfectly still. However, even under these calm conditions, as time passes the boundary between the two liquids will become less and less sharp. The movement of blue grape dyes may be the only visible change, but the other molecules in the lower layer will also move up into the top layer. The process of molecules moving from a region of high concentration to low is called diffusion. It is the major mechanism of molecular transport in the living cell.

Since thermal motion of molecules is random, a single sugar molecule at a certain distance above or below the boundary has an equal probability to move to the other side in a certain time interval. However, there are more sugar molecules in any giver volume on the high concentration side than in a corresponding volume on the low concentration side (that's what concentration means). Since the total number of molecules that move across the boundary in the time interval is equal to the product of the number of molecules and the probability that one will cross, the net flow will be from high to low concentration.

As with most real experiments, there are complications. To measure diffusion accurately you will have to make sure the temperatures of the glass, both fluids, and the air in the room are exactly the same. Any temperature difference will cause mixing because cool fluids are generally more dense than warm, and thus will sink toward the bottom of the glass. As an example, if the glass is colder than the liquid, the liquid right next to the glass will rapidly become colder than the liquid a short distance toward the center of the class, and thus it will sink and push the warmer inner fluid upward. This process is called thermal convection and is the main forces that generate the winds and thus our weather. You will also have to make sure there is no vibration or other movement that also causes mixing; and so on and so on.

Diffusion as a random walk

We can look at the same diffusion process in a different way by following a single molecule as it randomly bounces along between collisions with its neighbors. It will make the theory slightly easier to develop by imagining a model world in which the molecule moves by steps of equal length, but randomly left or right sequentially along each axis of its world (trust me that this approximation becomes arbitrarily close to the real world as the steps get smaller and smaller). In one dimension there is only the x axis and the molecule steps one unit to the left or to the right each tick of the clock. The distance to the right of the starting position at the end of N steps is just the number of steps to the right minus the number to the left.

This random process is analogous to tossing a coin N times. After each toss the coin will land heads or tails up with equal probability while after each step the walker will move to the right or left with equal probability. Luckily for us the probability of observing various numbers of heads or tails is one of the oldest problems in statistics and its solution is the familiar Gaussian probability distribution (sometimes called the normal distribution because it is observed in so many situations).

The square root of N law

When we take walks along the x axis of 100, 1,000, 10,000 ... steps the most likely final position is always the starting position, i.e. the most likely outcome is an equal number of steps to the right and left. However, as the total number of steps increases the probability increases that we will end our random walk at a greater and greater distance from the beginning. To be more precise, there is a 16 percent probability we will end to the left and an equal probability we will end to the right of a value called the standard deviation. The standard deviation is thus of some importance; in this example it is equal to the square root of N.

The absolute value of 16 percent is not important, but the fact that the standard deviation increases only as the square root on N is of central importance. This square root law means that when the time increases by 4 fold, the average distance the molecule has traveled by diffusion increases only the square root of 4, i.e. 2 fold. This means the movement of molecules by diffusion has no intrinsic velocity, at least not in the same way that the propagation of energy has. To send information by energy 10 km it will take 10 times as long as it did to send it one km. However, for molecules carrying information to move by diffusion 10 μm requires 100 times as long as it took for them to move 1 μm. Diffusion is not the Internet. The square root of N law is a major limitation on the size of independent biologic units, i.e. cells.

Walks in one, two, and three dimensions

It is easy to extend the model for the random walk in one dimension to walks in two and three dimensions. Just make a step along one axis, then a step along another axis, and so on. In a two dimensional world you make steps along the x and y axis, and at the end you have a 32 percent chance of being outside a circle with radius of the square root of the number of steps. In a three dimensional world you have an equal chance of being outside a sphere with a radius of the square root of the number of steps.

If you are relying on diffusion to transport a molecule from one position to another it makes a big difference if you are in a one, two, or three dimensional space. In one dimension the molecule only needs to end x steps along the x axis. In most situations, as long as the final position is greater than x, the molecule will have collided with the target at x in the process of moving to the more distant position, so the condition for collision is just that the final position is greater than x.

However, in two dimensions the molecule must both finish at the location x on the x–axis and the location y on the y-axis. The probability of landing at this location is the product of the probabilities of ending x steps along the x-axis and the probability of ending y steps along the y-axis, which is lower than either single probability. It will thus take a longer time to satisfy both requirements than either one alone. In addition, if the x position of the diffusing molecule exceeds the target x position it does not at all imply that the molecule has hit the target, because the y position of the molecule must equal the y position of the target at the same time for a hit to occur.

In three dimensions the molecule must end up or pass through x, y, z steps along each of the respective three axes. Now the probability of satisfying the three requirements simultaneously is the product of the three probabilities alone, and it will take even longer to get there.

Figure walk_123

Random walks covering an equal distance on a 1 D line, a 2 D plane, and in 3 D space

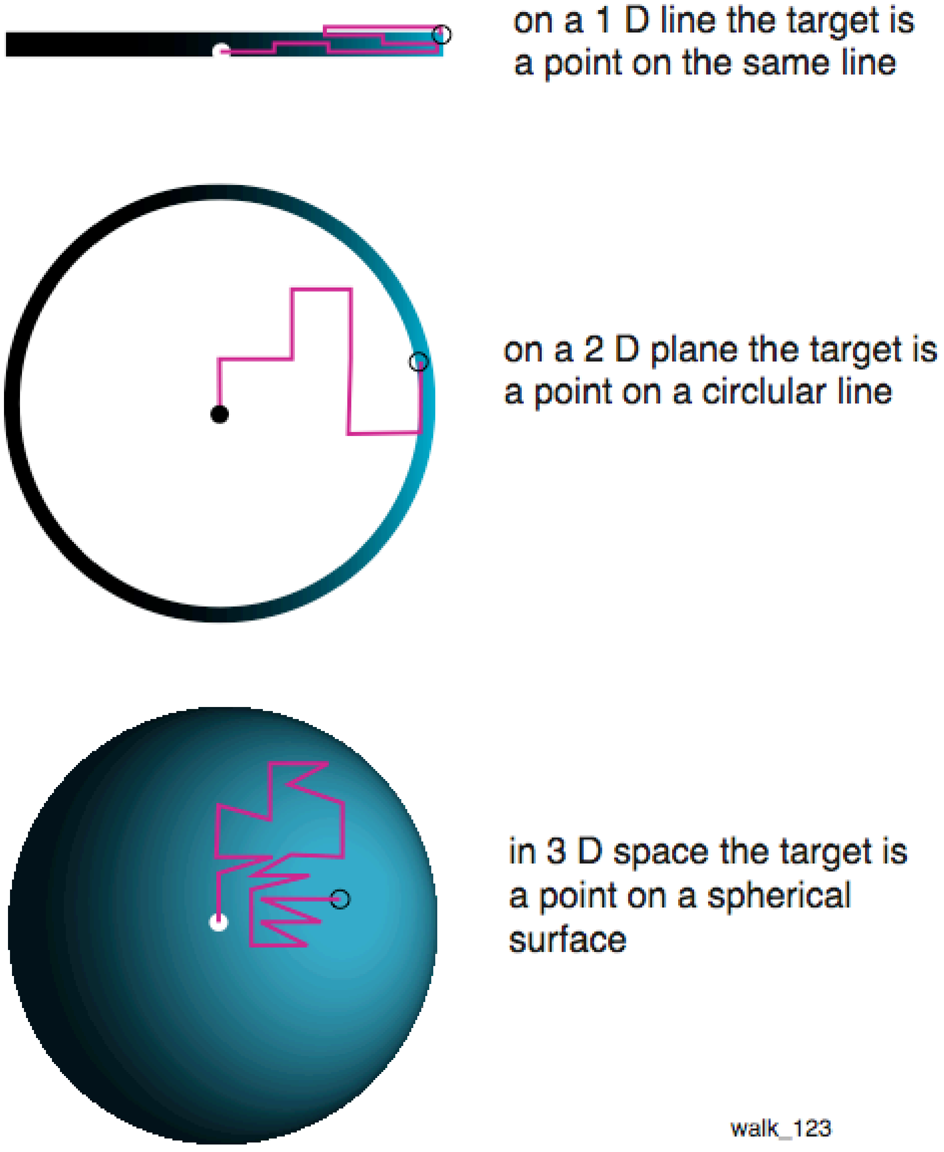


Figure walk_123. Random walks (diffusion) can occur on a one dimensional line, a two dimensional surface, or a three dimensional space. It takes the least time for a diffusing molecule to hit a fixed target on a one dimensional walk, more time in a two dimensional walk, and the most time in a three dimensional walk.

In Figure walk_123 we see diagrams of random walks in one, two, and three dimensions. The figure illustrates that it takes longer, or equivalently more steps, to hit a specific point on a two dimensional circle, and even more time or steps to hit a point on a three dimensional sphere as compared to just reaching a point along a one dimensional line. Since the world we live in is three dimensional, you might think the only motive for discussing diffusion in one and two dimensions is pedagogical, i.e. to derive equations for diffusion. Not so; in a living cell some molecules diffuse while loosely bound to a linear molecule (1 dimensional diffusion) or while confined to a membrane surface (2 dimensional diffusion). The speed of diffusion is more rapid in these constrained environments.

Brownian motion

I have claimed that a kind of mixing, diffusion, is the result of random thermal motion of molecules. However, you can't actually see molecules, and thus might assume this explanation must be accepted on faith or demonstrated with expensive, sophisticated instruments. However, if you have access to a simple microscope you can look at a liquid suspension of small particles, e.g. pollen, and you will see that they dance about in an apparently random way. This motion was described in 1827 by the Scottish botanist Robert Brown. Many explanations were proposed for "Brownian motion" over the decades. None were correct, and none hinted at the insight the true explanation of the motion would bring. The meaning of Brownian motion was discovered by Albert Einstein.

Albert Einstein's early papers

We associate the name Albert Einstein with insights into the fundamental nature of space and time. Distances are not absolute, they are influenced by, or relative to, the velocity of the observer. Among the many ramifications of this relation is the equation:

$$E = m c^2$$

between mass and energy. This relationship underlies the atomic bomb and nuclear power reactors. However, even before these studies Einstein published a body of work that would have ensured him a lasting reputation in the field of physics.

In a series of papers starting in 1905 Einstein proposed that the random motion of a small particle suspended in water is due to the thermal motion of molecules in the particle and collisions with moving water molecules. The first part of his model is a simple relationship between the square of the average distance the particle moves, the time, and the diffusion coefficient of the particle:

$$x^2 = 2 D t$$

where x is the distance, t is the time, and D is the diffusion coefficient. With a simple microscope you can follow a pollen grain for a minute and measure the distance it moves. Repeat this 100 times and you get a fairly good value for the average distance. Repeat this experiment for 100 four minute intervals, compute this average, and you will see that it moves about twice the distance in four times the time. Using the actual distance traveled by the particle you can calculate its diffusion coefficient.

The fact it takes four times as long to diffuse twice the distance has profound effect on the distances that molecules can move in reasonable times. If an organism has a high rate of metabolism, it can not use diffusion to transport material more than a few tens of microns. This limits the size of active organisms that do not have an active circulation system.

Einstein then proposed a formula for the diffusion coefficient of the particle that requires only the radius of the particle, the viscosity of the solvent, the temperature, and Boltzmann's constant:

$$D = k T / 6 \pi \eta r$$

where k is Boltzmann's constant, T is the absolute temperature, η is the viscosity of the solvent, and r is the radius of the particle. By 1905 fairly good values for k and the conversion of Centigrade to absolute temperature were available; the quantitative theory of Brownian motion was pretty much sewed up.

However, the greatest, but perhaps not so obvious insight obtained from these two equations is due to the fact that they hold for objects of any size; both molecules and pollen grains. You might have guessed this fact when we called D the diffusion coefficient, since the word diffusion is usually used to describe mixing due to motion of molecules, not the motion of visible pollen grains. Instead of watching pollen grains under a microscope you can make a concentration gradient in a sugar solution and measure how fast the sugar diffuses to generate a uniform concentration. Then using the simple equations that describe diffusion you can calculate the diffusion coefficient of sugar, and calculate the radius of a sugar molecule.

Boltzmann 's tombstone

Boltzmann produced a large body of important work on electromagnetic radiation and the kinetic theory of molecular motion in addition to determining the relation between the average energy of a molecule and its temperature, represented by the constant k used in the previous formula. As you remember from our earlier discussion of chemical reactions, Boltzmann along with Maxwell, predicted the distributions of energies a collection of molecules have at a given temperature (see Figure Maxwell_Boltzmann).

The fundamental equation that defines the Boltzmann constant is actually not the relation between energy and temperature of a molecule. Rather, it is an equation for entropy[1], previously only defined by the temperature dependence of chemical equilibrium concentrations of invisible molecules. Boltzmann's equation gives entropy as a function of the number of possible physical states molecules in a system can assume. It thus relates a molecular entropy, S, to the number of physical states, W, of the system. Like Einstein, Boltzmann was thus able to describe a property of the molecular world in terms of the larger physical world.

However, there were other influential scientists at the time that saw atomic models as artificial, imaginary constructs serving only as intellectual crutches to derive equations that must have deeper roots. Boltzmann suffered from periods of depression, and at least partially in response to negative comments on his work he committed suicide in 1906, just at the time Einstein's works on the relation between Brownian motion of individual particles and diffusion of molecules were being published. If only Boltzmann had waited. His formula for entropy is:

$$S = k \log W$$

This formula is engraved on his tombstone.

Atoms are elements

There are slightly more than 100 kinds of atoms, called chemical elements or just elements. The atom is composed of three types of smaller, sub-atomic particles[2]. There is a small, compact nucleus made of almost equal numbers of heavy neutrons and protons. The nucleus is surrounded by orbital electrons which are about 1/2000th as heavy as the particles in the nucleus. However, an electron does not behave as a compact object, like the moon orbiting around the earth, but is best imagined as a fuzzy cloud with a characteristic distribution around the central nucleus.

The number of protons in the nucleus is the atomic number, which defines the element and determines the basic chemistry of the atom. The sum of the number of protons and the number of neutrons in the nucleus of an atom is approximately the atomic weight of the atom. The number of electrons in the cloud around the nucleus is the same as the number of protons in the nucleus for the typical, neutral atom.

While the electrons do not behave as if they are in discrete orbits around the nucleus, there is a strong interaction between them that results in an internal organization or structure that is evident when you look at elements of increasing molecular weight. As an example, atoms with 2, 10, or 18 electrons have similar and very special chemical properties. These elements are helium, neon, and argon. They are called noble elements because they do not form bonds to other elements, and the

---

[1] As implied by this discussion, entropy is a measure of the number of possible states a system has. Thus a crystal at low temperature has a very low entropy, since the position of all the atoms is essentially fixed. A gas at high temperature has high entropy because the atoms can be almost everywhere and may be moving at almost any velocity.

[2] The physicist who studies the behavior of matter at very high energies describes these three particles in terms of other more fundamental particles. However, to understand basic chemistry it is sufficient to describe matter as composed of electrons, protons, and neutrons.

groups of 2, 10, or 18 electrons are called shells. This is a very important observation because the special nature of 2, 10, or 18 electrons is the basis of the strongest chemical bond between atoms, the covalent bond. The pattern of dependency of chemical properties with number of electrons is the basis of the Periodic Table of Elements.

There are many important elements in living organisms, although considerably less than 100. Some elements are bonded to each other to form the basic structure of biomolecules. Other elements are present as ions, charged forms of the element, which often diffuse throughout the fluid found in and out side cells. However, some ions are tightly bound to biomolecules at a specific location, and often play a direct role in chemical reactions with those biomolecules.

The principal elements that comprise the bonded structure of biomolecules

Element	Sym	A No	A Wt	Color*	Common function
hydrogen	H	1	2	white	present in almost all biomolecules; forms a single bond
carbon	C	6	12	green	present in almost all molecules; forms four bonds; basic skeleton of biomolecules
nitrogen	N	7	14	purple	present in half of biomolecules; forms three bonds with ionic character
oxygen	O	8	16	red	present in most biomolecules; forms two bonds with ionic character
phosphorus	P	15	31	violet	a metal found in half of biomolecules; bound to oxygen to make phosphate group
sulfur	S	16	32	yellow	minor but important element; often links polypeptide chains

* the standard color for this atom used in molecular diagrams

Elements present as charged ions in fluid or as complexes with biomolecules

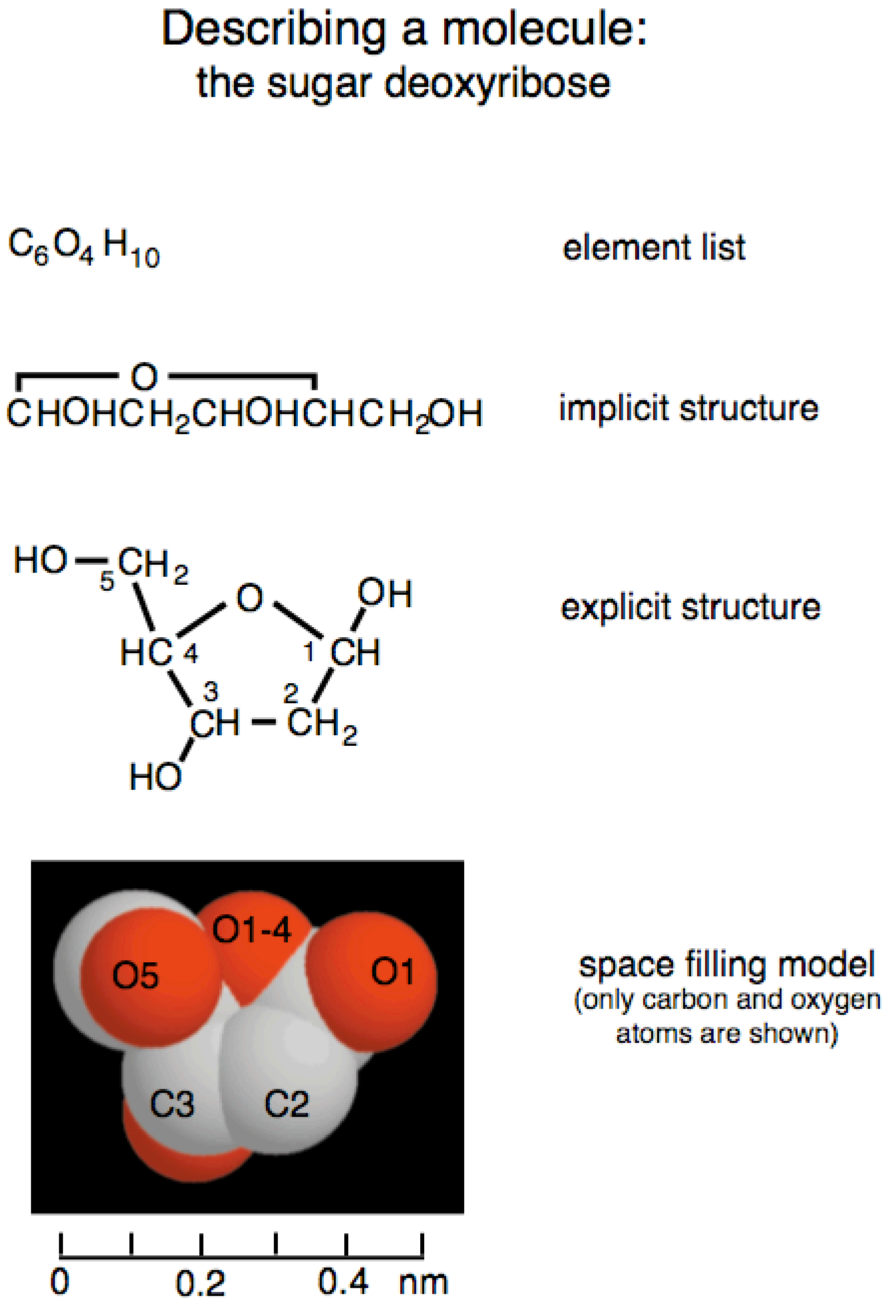
Element	Sym	A No	A Wt	charge	Common function or example
sodium	Na	11	23	+1	in all fluids
magnesium	Mg	12	24	+2	binds to phosphate
chloride	Cl	17	35	-1	in all fluids
potassium	K	19	39	+1	in all fluids; higher inside cells
calcium	Ca	20	40	+2	higher inside cells; signal molecule
manganese	Mn	25	55	+2	proteins involved in photosynthesis
iron	Fe	26	56	+2	hemoglobin
cobalt	Co	27	59	+1,+2,+3	vitamin B12
copper	Cu	29	64	+2	proteins involved in photosynthesis
zinc	Zn	30	65	+2	many DNA binding enzymes

Molecules

A molecule is a group of atoms that are linked together and behave as one unit or object. Atoms and molecules are the carriers of information in living organisms, as well as being the material from which the organisms are formed. There are many notations and types of diagrams that can be used to represent molecules. As an example of a molecule we use the sugar deoxyribose, since it is part of a giant molecule that caries the genetic information.

We just used one notation for this molecule, its name. However, there are several names for each chemical compound. Deoxyribose is called its trivial name, since you get few hints of the structure unless you already know what it is. Then there are a series of other names for the same molecule that are progressively longer and give correspondingly more explicit information about its structure. For example, the prefix "deoxy" means this sugar has one less oxygen than ribose, and we could have called the molecule 2–deoxyribose to indicate that it is the oxygen bound to carbon atom 2 of ribose which is missing. This suggests that there might be a sugar called 3-deoxyribose. There is, but it is not common in living organisms, and thus we don't need to use the longer name here. The series of possible names ends with the systematic name, which explicitly describes the position of every atom. The organic chemist who just synthesized a new compound never before described in the scientific literature or perhaps not even present on earth would need a systematic name. We will almost always use trivial names in our discussions. Four common ways to represent deoxyribose are shown in Figure ribose_views.

Figure ribose_views



ribose_views

Figure ribose_views. This Figure shows four formats for describing a specific sugar molecule. The basic problem is that a molecule is a three dimensional object, but it is rather inconvenient to draw a picture every time you want to refer to it.

The first notation in the Figure, the chemical formula, is just a list of the elements. The subscript indicates the number of atoms of each element present in the molecule, i.e. the ribose molecule contains six carbon atoms, four oxygen atoms and ten hydrogen atoms. It doesn't tell much more, although a biochemist would know that a molecule with that many oxygen atoms must be a sugar. If the sugar was ribose, it would have five oxygens, one for each carbon, and the formula could have then been written: $(CH_2O)_6$ or $C_6(H_2O)_6$. Since water is H_2O, this formula would suggests that the molecule is a polymer (a string units) of carbon linked to water, i.e. hydrated carbon. The molecule is not really constructed from water, since the carbon is between the H and OH groups, but sugars and polymers of sugars are still called carbohydrates.

The second notation in the Figure, the implicit structure, tells you the structure of deoxyribose if you know some chemistry and a few notational conventions. Assume the carbons are linked in a chain with the other atoms hanging off from them, and assume that "OH" is a group (the hydroxyl group). The oxygen at the top is evidently linking the first and forth carbons to make a ring.

The third notation in the Figure gives it all away, even to the novice. The molecule is a ring of five atoms and three of the atoms have hydroxyl groups extending from them. There is a standard numbering scheme for the backbone carbon atoms, and you can see that there is no OH group at C2. There is also no OH at C4, but that is because it effectively is being used to link to C1 and form the five-atom ring.

The forth and last representation of deoxyribose is called a space filling structure because the volume occupied by the electron clouds is indicated. The relative distances and the angles between the atoms are those of the actual molecule. It is thus the most realistic, but may not be the most useful. The electron clouds fill so much space that even some atoms of this small molecule are obscured, and it is difficult to see if atoms are attached by bonds or are just positioned next to each other. The hydrogen atoms have been omitted from this picture so you at least have a chance to follow the backbone of the molecule. Carbon atoms 2 and 3 are clearly visible, but carbons 4 and 5 are almost hidden by the oxygen bonded to carbon 5. However, C5 can rotate around the bond that connects it to C4, and thus this picture indicates only one possible configuration the molecule can have. Since each atom of the molecule has thermal energy, the atoms will rapidly move around and cause the molecule to assume all possible configurations. Thus, this particular model represents a snapshot at one time.

If we could rotate the space-filling model in space it would be clear that the five-atom ring is not quite flat. In fact, it assumes a continuous spectrum of puckered shapes, and some of these shapes have been given names, e.g. envelope, twist, C1-endo. As a single sugar molecule in solution the ring would be rapidly changing between most of these shapes in a random manner, although when the sugar is part of a larger molecule it may be forced to assume one specific ring configuration.

These four representations of deoxyribose, and other representations not shown here, each have advantages and disadvantages. We will mostly use combinations of the last two representations in this book, in addition to showing large molecules as “blobs” that capture only the major features of the molecular surface. When serious research makes it necessary to actually determine if molecules have specific shapes, e.g. to measure the points of contact when two molecules touch, solid plastic models or 3-D computer representations are used. Unless the molecules are small, physical models are prohibitively expensive and awkward to manipulate. Fortunately the cost of computer hardware and software has decreased sufficiently that even amateurs have access to computer models. Software is available free of change and we describe how to obtain and use it in the Software appendix.

Chemical bonds

Bonds hold atoms together to form molecules. These bonds are called chemical only for historical reasons; chemists study them. Chemical bonds are no less physical that the electrons that generate them, and many physicists also study chemical bonds. There are stereotypes for chemical bonds, but actual bonds are hybrids of the stereotypes. The force between two atoms is usually called a chemical bond only if it is sufficiently strong to create a stable link. Stability is relative, but for a group of atoms to be a molecule it should have a half life of at least hours or days. Some complexes of large molecules are stable because of the presence of many weak inter-atomic forces that individually might not be considered chemical bonds. In fact, most of the aggregates mentioned in "Molecular biology in less than 50 words" are held together by these weak bonds. Are these aggregates molecules, and are they held together by chemical bonds? The chemical vocabulary has some ambiguity, as it must, since the forces between atoms actually have a continuum of characteristics.

The covalent bond

The strongest chemical bond is the covalent bond that is created when two atoms share electrons. Since most of the molecules we will talk about contain a large fraction of carbon atoms, the carbon-carbon bond is not only important, but can be considered a stereotype bond for the biochemist. As mentioned before, the carbon atom has four electrons in its outer shell, while eight electrons is the most stable number. When a carbon atom shares an electron with another carbon atom, it creates a very stable bond. “Sharing an electron” means that the electron clouds of the two carbon atoms overlap, and the electron density around each atom becomes more similar to the electron density that would be expected if the carbon atom had one more electron in its outer shell. The most stable situation for a carbon atom is to be surrounded by four other carbon atoms with which it shares electrons, for then it is close to having eight electrons in its outer shell. A carbon atom also forms molecules with other atoms that can share electrons, e.g. hydrogen. Methane, CH_4, is the most simple example.

The ionic bond

The opposite of a covalent bond is one in which the two atoms play a completely asymmetrical role. An example is the bond between sodium and chlorine in a crystal of sodium chloride; table salt. Here each sodium atom has a charge of +1 while each chloride atom has a charge of -1. The electrostatic attraction between the two atom types is the force that holds the crystal together.

Water and the hydrogen bond

Most bonds are hybrids between covalent and ionic. A very important example for us is the hydrogen bond, as seen in Figure water.

Figure water

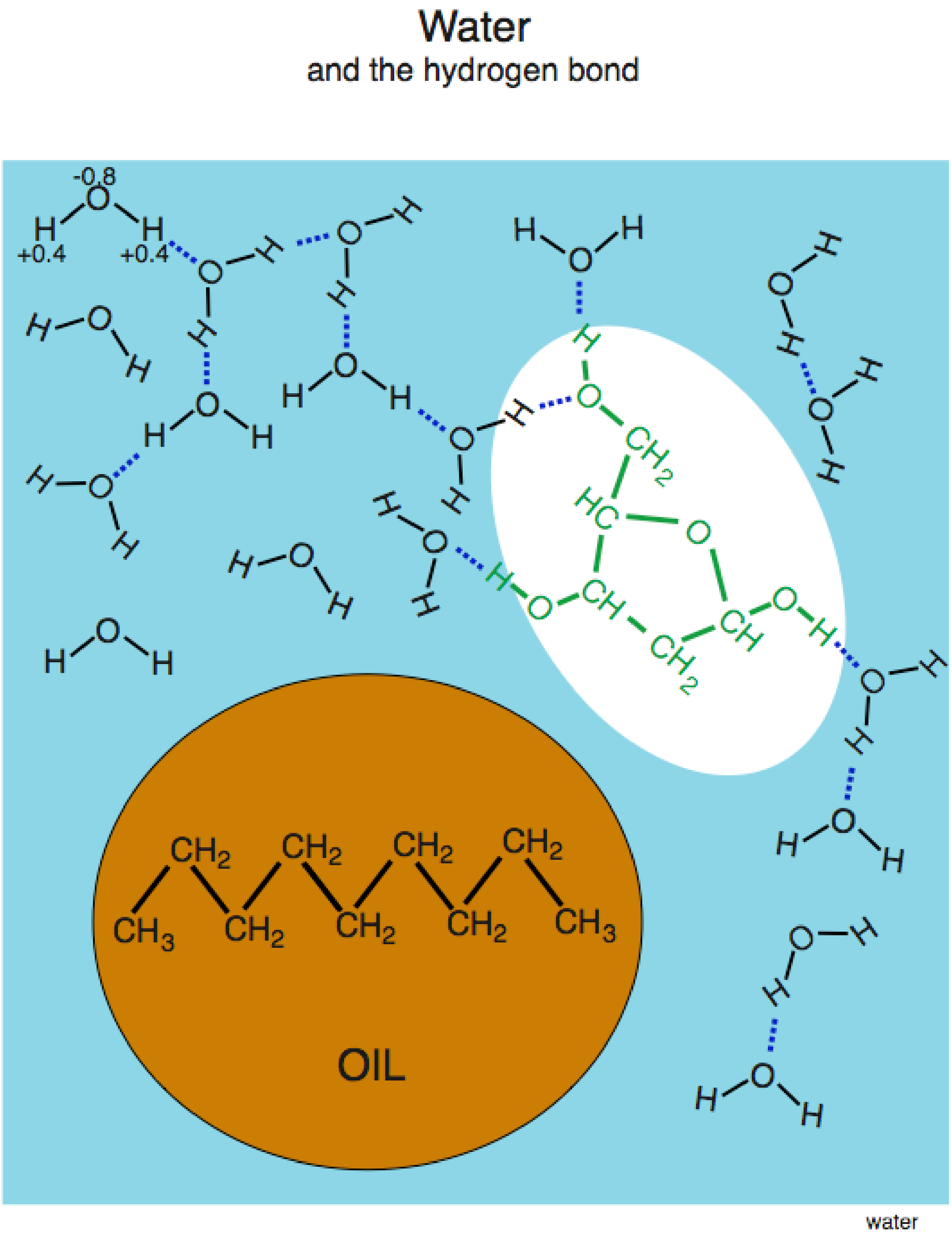


Figure water. Water, H_2O, has two hydrogen atoms bonded to an oxygen atom at an angle of 105 degrees. The hydrogen atoms in one water molecule can form weak bonds to the oxygen atom of another water molecule. However, these "hydrogen bonds" are transient, and thus the water molecules can move past each other to make water a liquid. Water can also form hydrogen bonds to OH groups in other molecules, such as sugar. These bonds allow sugar to dissolve in water. Water can not form hydrogen bonds with oils, thus oils do not dissolve in water.

In a water molecule the covalent bond between each hydrogen atom and the oxygen atom makes an angle of 105 degrees. The bonds have considerable ionic character, oxygen having an effective charge of -0.8 and each hydrogen atom a charge of +0.4. As seen in Figure water, the water molecules can form a variety of clusters in which an OH bond (called a dipole since it contains both a positive and negative electrostatic pole) aligns with the dipole of an entire water molecule. A single hydrogen bond has a modest energy, not sufficient to produce a stable, permanent molecule at room temperature. Thus the water aggregates seen in Figure water represent vary transient structures. Even though the hydrogen-bonded structures are not permanent, they result in a very significant time-average attractive force between water molecules. This attractive force makes water a unique solvent, with a high freezing and boiling temperature for its molecular weight, and a high heat capacity. The physiochemical properties of the molecules that make up a living cell are dependent on a water environment. In any other solvent they would have such different properties that they would not function together to sustain life. In a world with out water you might be able to have life, but it would have to be composed of a very different collection of molecules.

A molecule of deoxyribose is seen in the white zone of Figure water. The hydroxyl groups of the sugar make hydrogen bonds with water, and thus the presence of the sugar molecule does not decrease the total number of hydrogen bonds in the solution to a great extent. Sugars are soluble in water because they can make hydrogen bonds with water molecules. Other molecules with hydroxyl-groups, or other groups having ionic character on their surface are also soluble in water. Most biomolecules have polar groups on the outer surface which is in contact with water, while the non-polar groups are in the interior of the molecule where they can at least stick to each other. Molecules or groups that can form hydrogen bonds with water are called hydrophilic.

In contrast to sugar, an oil molecule, seen in the brown zone of the Figure has no polar groups to form hydrogen bonds with water. Since it occupies space in the middle of the water, it prevents water molecules from forming hydrogen bonds with each other. This is not a stable situation if there are other configurations that allow more water molecules to bond with each other. If oil molecules collide, they will coalesce to form oil droplets, because this will decrease the total surface area between oil and water and thus decrease the number of hydrogen bonds between water molecules that the oil prevents.

Since water molecules can not form hydrogen bonds with air, oil molecules on a water surface do not decrease the total number of possible hydrogen bonds. Thus, oil molecules dispersed in water that reach the surface remain and form a film. Molecules or groups that can not form hydrogen bonds with water are called hydrophobic.

A nucleotide base

In Figure guanine the molecule guanine, a nucleotide base, is shown in five views. A conventional bonded structure is seen in the upper left panel. The valences are all correct, hydrogen atoms have one bond, oxygen atoms have two bonds, nitrogen atoms have three bonds, and carbon atoms have four bonds. However, the right top panel is a far more realistic picture of the molecule. Most of the bonds in this molecule have significant ionic character, and thus are not well described as either single or double bonds since they have intermediate properties.

Figure guanine

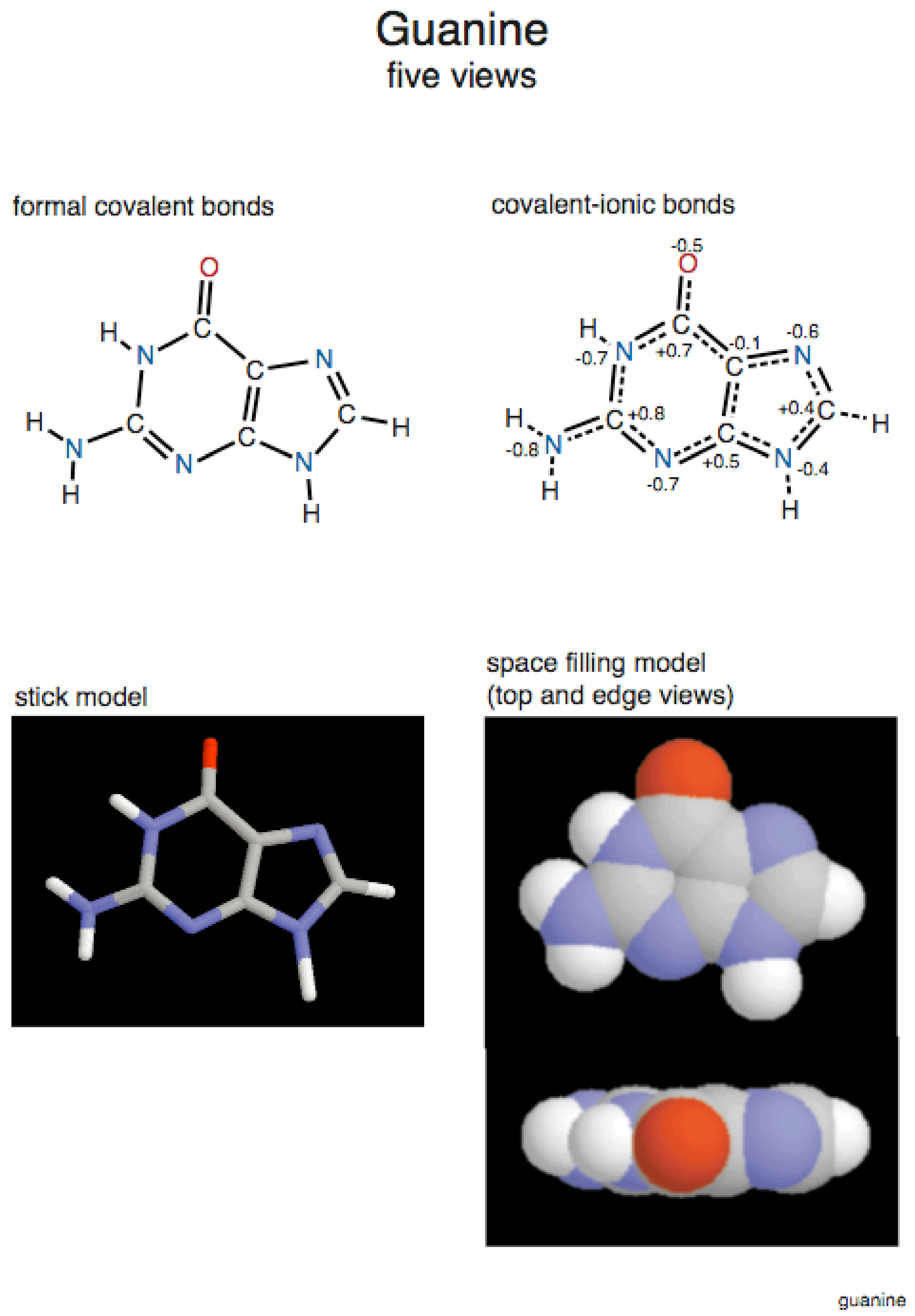


Figure guanine. The nucleotide base guanine is shown in four representations.

The representation in the lower left panel is called a stick model. Bond lengths and angles between bonds are realistic, but the electron clouds are missing. While this is not completely realistic, it allows you to see the structure of large molecules and thus is often preferred over a space filling model. Finally, in the lower right panel we see the space filling models of guanine, with hydrogen molecules included. It is clearly a flat structure, without an empty space in the middle of the rings (which one might imagine from the stick model).

A nucleotide

In Figure gmp we see a nucleotide, deoxyriboguanosine monophosphate. It is built from the guanine base of the previous section, the deoxyribose sugar of the section before that, and a phosphate group. Phosphate is a familiar chemical. It is the oxidized form of phosphorus, a highly reactive and dangerous metal that is typically found only in artillery shells or bombs intended to start fires. A mildly acidic form of phosphate helps give CocaCola its tart taste. A basic form, trisodiumphosphate (TSP), is used as a cleaning agent. It is a common component of fertilizers, since phosphate is often a limiting mineral in soil.

Figure gmp

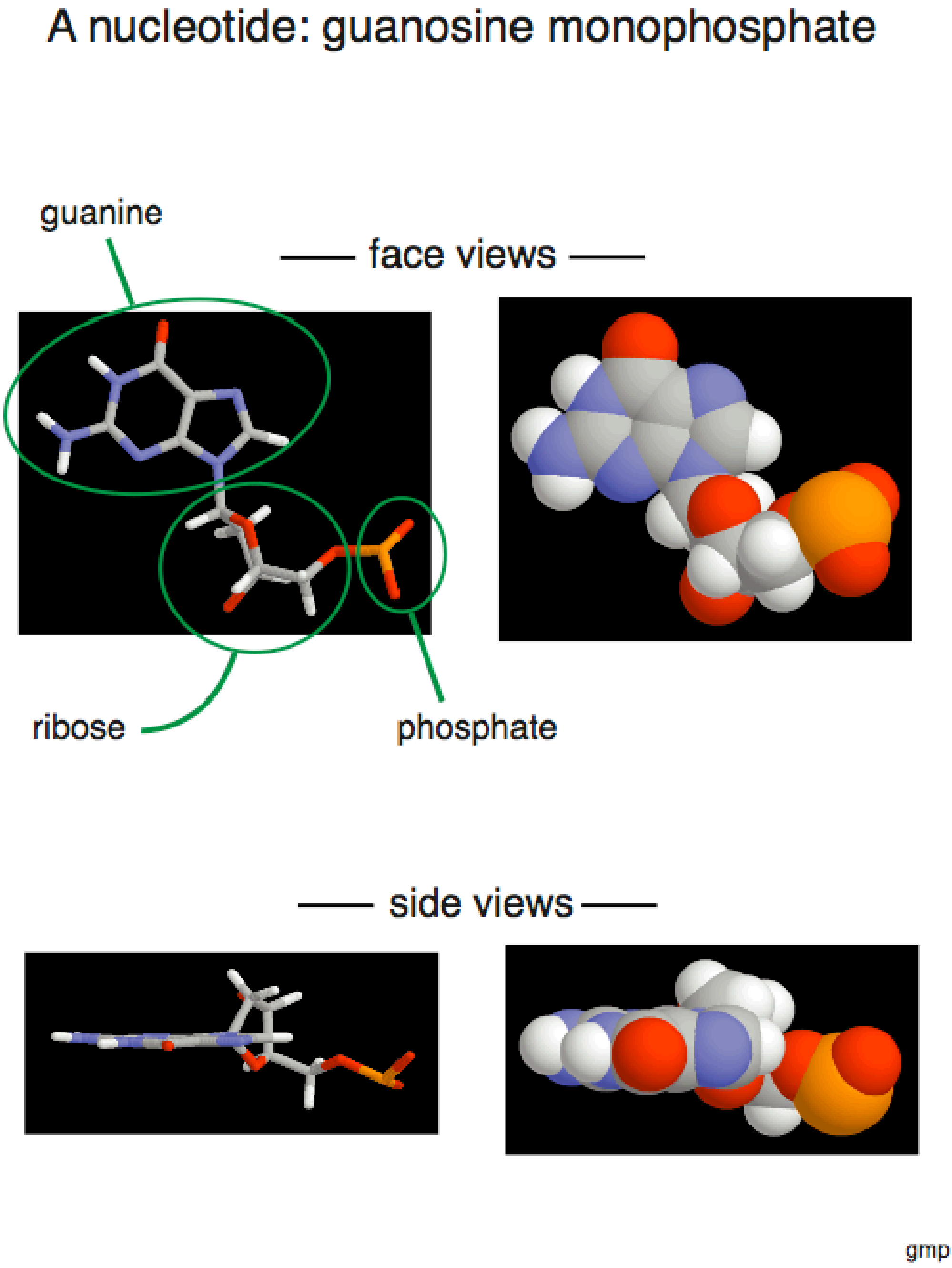

A nucleotide: guanosine monophosphate
guanine
face views
ribose
phosphate
side views
gmp

Figure gmp. This nucleotide is a combination of deoxyribose, guanine, and phosphate.

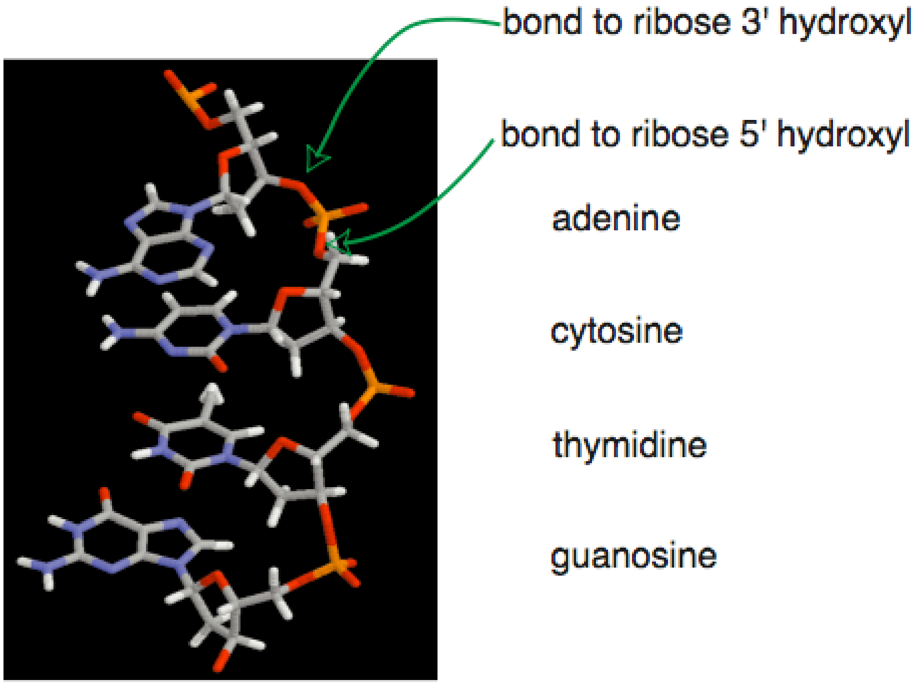
In the bottom panels of Figure gmp we again that the guanine ring is flat and the deoxyribose and phosphate are not in the same plane. Since the sugar and phosphate groups can rotate around the single bonds that connect them to each other and to the guanine base, the configuration shown here is just one of many possible ones.

A polynucleotide

In Figure actg we see guanosine and three other nucleotides, adenosine, cytosine, and thymine, linked together to form a polynucleotide. In these four nucleotides only the bases are different. The adenosine base, like guanine, is composed of two fused rings, but the polar groups on the six membered ring are different. The other two bases, cytosine and thymine, have only one ring, again with different polar groups on the rings. The nucleotides are joined by phosphate groups, or more precisely phosphodiester bonds. An ester is the result of an acid, phosphoric acid, reacting with a hydroxyl group, on the sugar. Since each phosphate forms an ester with two nucleotides, the linkage is called a phosphodiester link.

Figure actg

A polynucleotide: actg



actg

Figure actg. Guanosine and three other nucleotides are linked together by phosphodiester bonds. This polymer is a short, single stranded DNA molecule.

In a living cell there are several types of polynucleotides, and some are more than a million nucleotide units in length. Almost every possible sequence of the nucleotides is found in these long polymers, and the four are often not present in equal numbers. The nucleotide chains containing deoxyribose are localized in the nucleus of the cell. Since the phosphate groups in these polynucleotides make them an acid, they are called deoxyribonucleic acid or just DNA. A class of polynucleotides similar to DNA contain the sugar ribose, and these ribonucleic acids are thus called RNA.

The sequence of nucleotides in DNA represents the genetic information of the cell. Since there are four nucleotides, DNA is a two bit per nucleotide information string, similar to the two bit per pulse system used in the ISDN format. As seen in Figure dna_isdn, position on a DNA chain is equivalent to time for an ISDN signal. The information in the DNA nucleotide sequence, after several conversions, is actually read sequentially (but three nucleotides at a time). Thus the nucleotides in DNA are converted into a time sequence also. As we will see in the next chapter, most of the DNA in the cell is actually in the form of a double stranded helix. However, all of the information is contained in one strand.

Figure dna_isdn

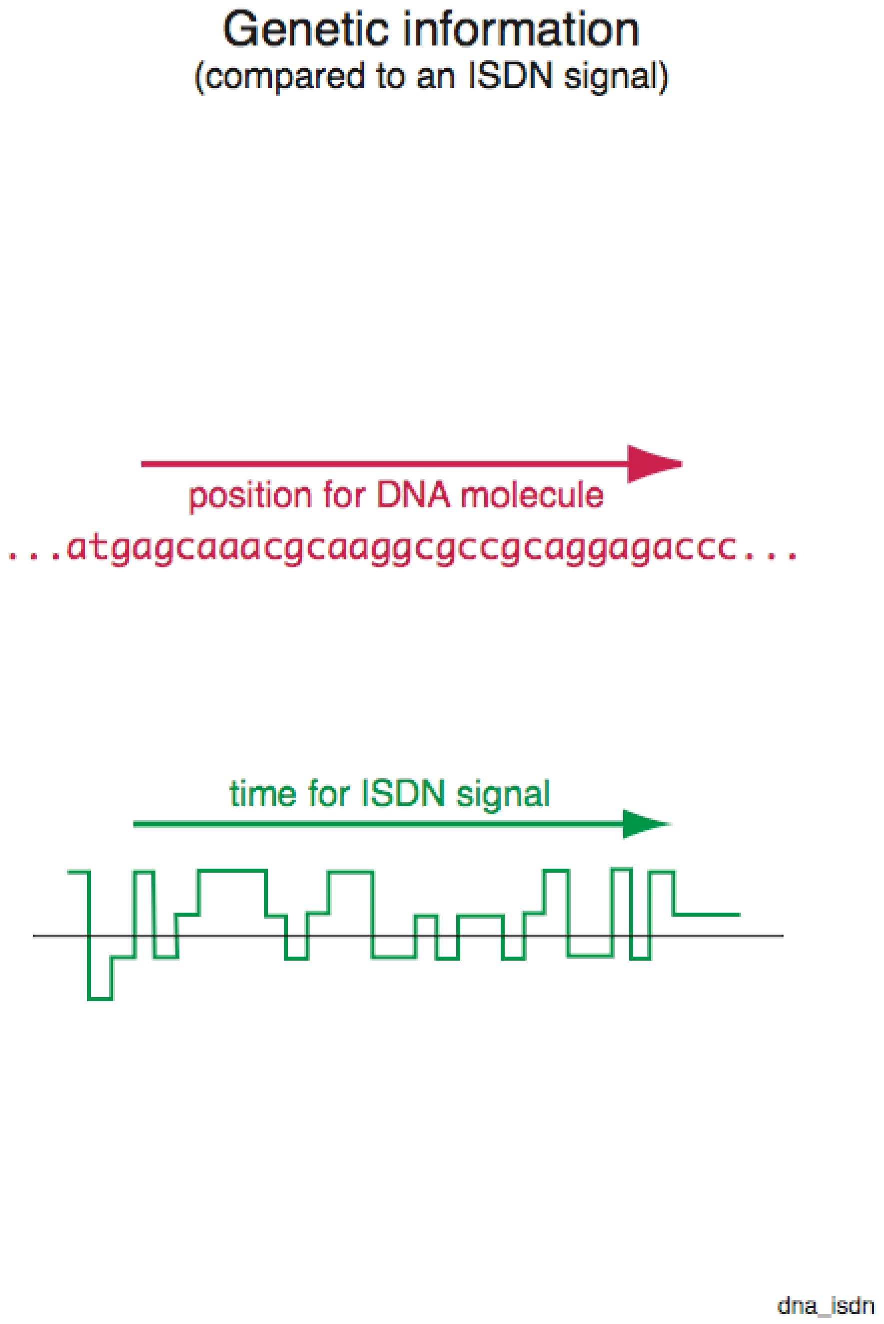


Figure dna_isdn. In terms of information transfer the four bases of DNA are equivalent to the four pulse levels of an ISDN signal.

The amount of information encoded in the DNA of a human cell is quite large. One human cell contains several DNA molecules with a total of 6 X 10^9 nucleotides[3], which equals 12 Gb or 1.5 GB. At the relatively slow speed of a PRI ISDN link, 128 kbps, it would take 12 X 10^9 / 128 X 10^3, about 10^5 seconds or 28 hours to transmit the entire genetic information of one human. Using a fast DSL line it would still take about 3 hours. However, it would only take two CDs to store the data; data storage has become very cheap.

All the characteristics transmitted from parents to offspring are encoded in the DNA. As is true in all information systems, including the Internet and the living cell, a message can only be read in a particular environment. In the Internet the environment includes hardware that recognizes the pulses as discrete bits, quads, etc. and software implementing codes, formats, and protocols that enable the bit stream to be decoded to regenerate the original message. In the living cell the environment includes molecules that directly make contact with the DNA in a nucleotide specific pattern, and a sequence of other molecular complexes, or machines, that translate the nucleotide message into meaning. Meaning for the living cell means proteins and RNA. Proteins and RNA, directly or indirectly, accomplish the tasks the cell requires to live and survive. Since all cells die, an essential task is to generate progeny, which requires making a copy of the DNA. The proteins that copy the DNA have been created using information contained in the DNA they copy. Life is a cycle.

Chapter Summary

Objects in our world are in constant motion, which increases with temperature. In gases and liquids thermal motion of molecules results in transport and mixing by diffusion, while in solids the atoms merely vibrate. Diffusion in liquids is sufficient for transport over distances of microns, but not meters. Constrained diffusion in a two dimensional membrane or along a one dimensional molecule is more rapid than free diffusion three dimensions.

There are more than 100 different kinds of chemical elements, or atoms. Atoms contain a central compact, positively charged nucleus surrounded by one or more negatively charged electrons. The electrons behave as if they are smeared out in a delocalized cloud surrounding the nucleus. Atoms are linked by bonds to form molecules. These bonds are the result of interaction between the electrons of the atoms. The bond may be completely ionic; a negatively charged atom is attracted to a positively charged atom. At the other extreme the bond can be completely covalent;

---

[3] This is actually the number of base pairs (next chapter) but since one nucleotide of each pair defines the other, the second is redundant as far as information is concerned. The number includes both copies of each chromosome, which are actually very similar in nucleotide sequence.

two atoms share electrons, with neither having a net charge. Most bonds are hybrids of these two types.

Water molecules transiently bond to each other by hydrogen bonds. This interaction defines many of the physical properties of water. Many biomolecules have hydroxyl groups that also form hydrogen bonds to water, which makes them soluble in water. Other important biomolecules can not form hydrogen bonds with water, and thus they coalesce into droplets or membranes. The partitioning of hydrophilic and hydrophobic molecules is a major organizing force in both individual biomolecules and in the cell.

Nucleotides, an important class of molecules in the cell, are composed of a base, a sugar, and phosphate. The deoxyribo-nucleotides contain one of four different bases and the sugar deoxyribose. The ribonucleotides contain the sugar ribose. Deoxyribonucleic acid, DNA, found in the nucleus of cells, is a chain, or polymer, of deoxyribo-nucleotides. The sequence of nucleotides in DNA encodes all the genetic information of the cell.