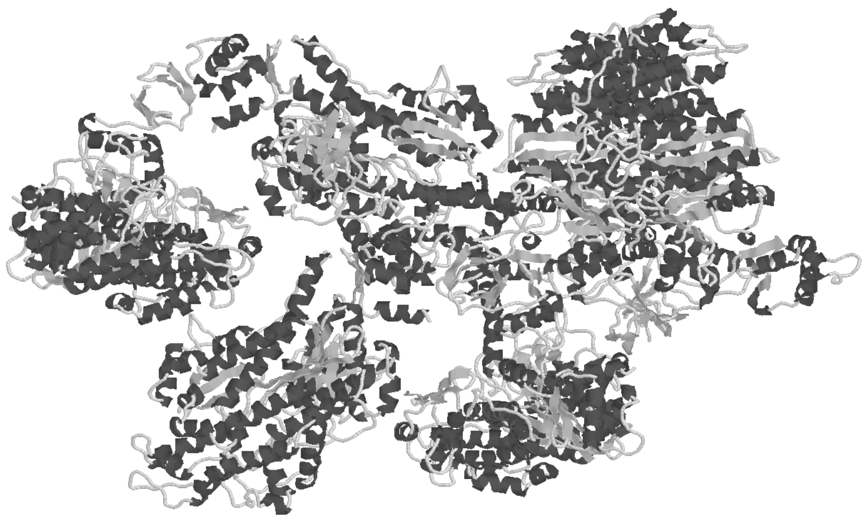
Part 2: Chemical Evolution



The figure is a cartoon representation of the protein Myosin. Myosin and actin are the two proteins responsible for muscle contraction.

Chapter 5:Information & Knowledge before the Genetic Code

To calculate the information and knowledge for insulin in the last chapter, the genetic code was used to assign a probability to each amino acid arising by chance. For this calculation to be meaningful, both the code and a method to turn the knowledge in DNA into proteins must already exist. So the calculations in chapter 4 assume that life already exists. What about before life exists? How would one calculate the information or knowledge in the very first protein? This is not an easy problem.

Several authors have used thermodynamics, but thermodynamics only applies when the system reaches equilibrium. The relevance of thermodynamic calculations is questionable as amino acids do not polymerize into peptides chains unless external conditions force them away from equilibrium.

This chapter will use information theory to solve the problem. Unlike thermodynamics, information theory can easily deal with non-equilibrium systems.

Information theory cannot normally be used to predict how chemicals will react because some chemicals react with each other readily, and others only react very slowly. Others do not react with each other at all. Thus, the likelihood of two chemicals joining together depends on both the quantity of the chemicals present and their chemical properties. Information theory can easily deal with the effects of quantity, but it has no way to deal with chemical properties.

This chapter will require several assumptions. Without these assumptions information theory cannot be applied to chemical reactions. Fortunately, these assumptions will improve the probability for creating a protein in the primordial soup.

Assumptions:

- The probability of a peptide bond forming between two amino acids only depends on how many of each amino acid is present in the system.
- The primordial soup only contains amino acids.
- Amino acids do not form non-proteinous bonds with each other. So for example, the carboxylic acid functional groups in aspartate and glutamate do not react with the n-terminus of other amino acids.

The first assumption allows all amino acids to be treated equally. While this assumption ignores the chemical properties of each amino acid, the assumption is not an unreasonable approximation because all amino acids must join together by forming a peptide bond. This assumption improves the odds because it ignores the finding that glycine and alanine are not only the most common amino acids but they are also the most likely to form the alpha peptide bonds required by proteins (Fox, 1972, page144 and 154). The second assumption also greatly improves the odds of creating a functional protein. By excluding chemicals that react quickly with amino acids, this assumption eliminates chemical reactions that can prematurely terminate a growing peptide chain. It also ensures that the amino acids will be available to interact with each other. The third assumption is not true, but it greatly simplifies the math, and at the same time, it improves the odds of creating a protein in the soup.

With these assumptions, information theory may be applied to the primordial soup. The first step is to estimate the number of each amino acid in the primordial soup. There are two methods. For 50 years, scientists have been trying to find better ways to synthesize amino acids under plausible prebiotic conditions. Many of the 20 amino acids used by life have been synthesized. Because these experiments are riddled with speculation about conditions on the primitive earth and investigator interference, the second method is preferable. This method relies on the amino acids found in meteorites.

Meteorites

Some meteorites contain organic carbon, and several of these have been analyzed for amino acids. This analysis has shown that the amino acids, glycine and alanine, are quite common in some meteorites. Most of the other amino acids used by life are rare, but some are present. In addition, many amino acids not used by life are present. More than 50 non-biological amino acids are found in the Murchison meteorite.

Meteorites are easily contaminated by biological amino acids. So samples are always taken from the meteorite interior. Unfortunately, contamination is still a major problem. Nevertheless, several generalizations are possible.

- The biologically relevant amino acids in meteorites are always predominantly glycine and alanine. Sometimes aspartate and glutamate run a close second, but in many cases, this appears to be the result of contamination. Serine and valine are sometimes present. The other amino acids used by life are absent.

- Non-biological amino acids are common in variety and in number. The most common non-biological amino acids are the many isomers of aminobutyric acid. The second most common non-biological amino acids are two forms of alanine that life does not use.

One comparison of four different meteorites that contain amino acids revealed that only 25% of the amino acids are biologically relevant.[4] If the primordial soup has a similar composition, then only 25% of the amino acids in the soup are biologically relevant, and even if a way is found to make the amino acids join together, the odds of a protein emerging are very small.

The average protein in the Swiss Prot database contains 362 amino acids, and most contain more than 150 amino acids. If the composition of amino acids in the soup is similar to that of meteorites, what is the probability of creating a peptide composed of 150 amino acids if all of the amino acids must be biological?

The knowledge required to build this peptide is simply the knowledge required to exclude all amino acids not used by life. Today, random amino acid sequences do not contain such knowledge because the machinery used by life to build proteins ensures that one of the 20 amino acids used by life will always be placed at each position in the growing chain. This is not true in the primordial soup. The term molecular knowledge in this book is reserved for useful information that conveys a selective advantage. A random sequence of biological amino acids that evolves in the soup will not possess molecular knowledge because the sequence will most likely have no function. Thus, the term information is preferred in this case. To avoid any possible confusion with terminology, this type of information will always be referred to as primordial information. Primordial information is the information needed to exclude non-biological chemicals found in the primordial soup from a growing peptide, RNA or DNA molecule. Since primordial information is a form of knowledge, it can be safely related to a probability. Furthermore, this calculation does not rely on human insight. Before self replication, natural selection cannot exist, so all events are guided by chance and chance alone.

Each addition to the growing chain has a 25% chance of being an amino acid used by life. So each amino acid added to the chain has a 1 in 4 chance of being correct. Thus, there are 4 possible outcomes and only one is desirable. Using equation 1 in chapter 1, $2^{\text{information}} = 4/1$, and because $2^2 = 4$, the information content for each amino acid added is 2 bits. So a random chain of 150 amino acids that emerges from the soup will contain 300 bits of information. The odds of this arising by chance are 1 time in 2^{300} tries or a 1 in 2×10^{90} chance. What do odds like 1 in 2×10^{90} really mean?

The number 10^{90} is so large that naturalistic explanations will always fail to explain any event whose odds are this poor. To understand why, assume that every single star in the universe has one planet composed entirely of amino acids. Further assume that every one of these amino acids exists as a 150 amino acid peptide chain. The highest estimate for the number of stars in the universe currently available is 7×10^{22}. If the planets orbiting these 7×10^{22} stars are about the same size as the earth, then on average each has a mass of 6×10^{24} Kg. A planet with this mass composed entirely of the amino acid glycine will be made from 5×10^{49} glycine molecules. If all the planets have the same number of amino acids, then there will be 3.5×10^{72} amino acids in the universe. Since every amino acid exists in a chain of 150, there will be 2.3×10^{70} peptide chains. The odds that 1 of these chains will contain only biologically relevant amino acids is only 1 in 8.6×10^{19}. So further assume, that all of these peptide chains break down each year only to reform, and that this process has been going on every year for 15 billion years. The odds improve to 1 in 6 billion. So while the odds are not zero, they might as well be. Nature simply cannot accumulate enough tries to overcome the poor odds.

One can certainly speculate that the first proteins used amino acids that are no longer used today or that these proteins were very short. Both assumptions improve the likelihood for evolution. Nevertheless, all readers need to realize that when a scientist in the lab mixes together pure amino acids that are only used by life, the scientist is adding so much information to the system that the experiment can no longer be considered representative of the conditions on the early earth. The starting point for such experiments is not plausible.

If the soup existed, then the first proteins evolved in a soup that contained many amino acids not used by life. The soup also contained a host of other chemicals like aldehydes that react readily with amino acids. These undesirable side reactions make the evolution of information in the primordial soup very difficult to explain. When a scientific experiment models evolution by excluding these other chemicals, the experiment no longer models the origin of life. Such experiments only model evolution in a test tube.

The Evolution of Primordial Knowledge

The odds of a random amino acid chain evolving in the soup are quite poor, but what about a protein? A protein is not a sequence of random amino acids. The order and type of amino acids in a protein determine how it folds, how it behaves, and its biological function. The sequences are not random. They contain knowledge. The odds of a functional protein evolving are certainly expected to be much less than that of a random sequence.

How Many Solutions?

One of the more important experiments concerning the origin of life was performed by Keefe and Szostak.[1] The authors of this paper in Nature searched six trillion random peptides each composed of 80 amino acids. They were looking for a sequence that could bind the chemical, ATP. They found four sequences in this large pool with ATP binding activity.

This allows for a direct computation of the molecular knowledge required for ATP binding. Using equation 2 in chapter 1, molecular knowledge = 3.32 x log (6 trillion/4) = 40 bits. Notice, that this is not 40 bits of information because the proteins that were selected only possessed minimal functionality. These proteins were subjected to several rounds of selection greatly improving their affinity for ATP.

This experiment provides a direct measurement of molecular knowledge. It also shows that there are very few solutions.

Binding a chemical like ATP is one of the functions that many enzymes possess. So while Keefe and Szostak did not actually find a useful enzyme, they did find a function than many enzymes require. The 40 bits calculated above are for evolution in a test tube. How many bits are required for evolution in the primordial soup?

The minimum possible primordial information in a random sequence of 80 amino acids is 160 bits (2 bits per amino acid). The odds of such a peptide evolving are one in 1.5×10^{48}. Given that the odds that a random sequence of 80 amino acids will bind ATP are only 4 in 6 trillion, the odds of finding a primitive peptide on the earth that can bind ATP are simply the product of the two numbers or one chance in 2.2×10^{60}. Alternatively, the 160 bits needed to construct an 80 amino acid peptide in the soup may be added to the 40 bits calculated above. The total knowledge is thus 200 bits, and the odds of this happening are 1 in 2^{200} or 1 in 2.2×10^{60}. In this calculation, the total knowledge required is simply the sum of the molecular knowledge and the primordial information. After life exists, primordial information always equals zero, and molecular knowledge always equals total knowledge.

Binding ATP is a simple function. Clay, a simple mineral, binds ATP. Furthermore, the function by itself does not confer a selective advantage. Thus, ATP binding is below the threshold of molecular knowledge. This function must be combined with another function before natural selection will preserve it. To create a functional enzyme that can be preserved by natural selection quite a bit more knowledge is required. Since it takes 200 bits to bind ATP, assume that it also takes 200 bits to bind another molecule. Thus, 400 bits is a more reasonable approximation for a functional enzyme, and the odds for such evolution are given by 1 time in 2^{400} tries or a 1 in 2.5×10^{120} chance.

The origin of the first enzyme just cannot be explained in this way. The odds are too poor.

Molecular Knowledge Before Life

This section will investigate how the composition of the soup influences knowledge. If the soup contains mostly glycine and alanine along with a host of other amino acids not used by life, then the probability of a useful protein emerging from it must be very low. Chapter 4 calculated the molecular knowledge for the protein insulin. This chapter will repeat this procedure assuming that insulin emerged in the soup before life. By this calculation this chapter does not mean to suggest that insulin originated in the soup. The calculation is for comparison only. Remember insulin was only chosen because it does not contain many amino acids, and this makes the calculations easier.

The Composition of the Soup

If meteorites are used to reconstruct the composition of the soup, then 14 of the 20 amino acids used by life will be absent. Only glycine, alanine, valine, serine, aspartate and glutamate would be available in the soup. The proteins used by life today require more than 6 amino acids. While this prediction of the soup's composition is probably the most accurate, it is an undesirable composition. So this chapter will assume a much more favorable composition.

Life uses 20 amino acids. Seventeen of these have been synthesized in the lab under conditions that might be similar to the conditions found on earth 4 billion years ago. Some amino acids are quite easy to synthesize, and others are very difficult. The amino acids that are easy to synthesize invariable are the primary product of these experiments. The other amino acids occur in various concentrations depending on the conditions chosen to carry out the experiment. Three amino acids, histidine, arginine, and lysine, have not been synthesized under plausible conditions.[2]

Because no single experiment has generated all of the amino acids, if the soup's composition is taken from the results of a single prebiotic experiment, then the composition will be unfavorable for protein evolution. Most proteins need 18 or 19 different amino acids to function. To construct a favorable composition for protein evolution, it is either necessary to combine many different prebiotic experiments or to just assume that the absent amino acids are present. This section will take the latter approach.

On page 87 of his book, Miller lists the results from one of the most successful prebiotic experiments.[3] The yields of ten amino acids are listed in this table.

As a reasonable starting point, assume the abundance of the amino acids in the primordial soup tracks Miller's table. Ten amino acids are not found in Miller's table. Seven of these have been synthesized under plausible prebiotic conditions. Assume that these seven are as abundant as threonine. Threonine is the least common amino acid listed in Miller's table. Three amino acids have not been synthesized in the lab. Assume that these are found in the soup at 1/10 the concentration of threonine. Finally, assume that the 20 amino acids that life uses comprise 1/4 of all amino acids present in the soup. Thus, the soup ratio of biological to non-biological amino acids is similar to the ratio found in meteorites.

These assumptions improve the odds that a protein will emerge in the soup. For example, one could easily assume that the ten proteins not found in Miller's table were also absent from the soup. With this single assumption, the information and molecular knowledge found in most proteins becomes infinite. Furthermore, the assumption to exclude chemicals like aldehydes and formic acid greatly improves the likelihood for protein evolution.

With these assumptions in place, labeling wooden blocks according to amino acid abundance yields table 5.1. The numbers in the second column are taken from Miller's table. The right column is based on what might have been given the constraints of the favorable assumptions discussed above.

Table 5.1: Wooden Blocks Used to represent Chemicals in the Soup

amino acid	number of blocks	amino acid	numbers of blocks
glycine*	440,000	tryptophan	400
alanine	395,000	tyrosine	400
valine	9,750	histidine	40
leucine	5,650	lysine	40
isoleucine	2,400	cysteine	400
proline	750	methionine	400
aspartate	17,000	phenylalanine	400
glutamate	3,850	arginine	40
serine	2500	asparagine	400
threonine	400	glutamine	400
Total number of blocks labeled with amino acids used by life		880220 (sum of column 2 and 4)	
Total number of blocks		4 x 880220 = 3,520,880	

* Most amino acids exist in two forms. The forms are mirror images of each other. Life only uses one image. Glycine is the only amino acid that does not have a mirror image. Thus, the number reported for glycine in table 5.1 corresponds to the concentration reported in Miller's table. The numbers associated with the other amino acids in this column are ½ the value reported in Miller's table.

The Evolution of a Functional Protein in the Primordial Soup

Because three million blocks cannot fit in a basket, the trapped scientist is now given a truck (figure 5.1). The blocks in the truck are determined by table 5.1. The scientist can draw blocks from a tube that connects his room to the back of the truck. How much information would insulin contain, if it evolves given these constraints.

Figure 5.1: Trapped Scientist with a Truck

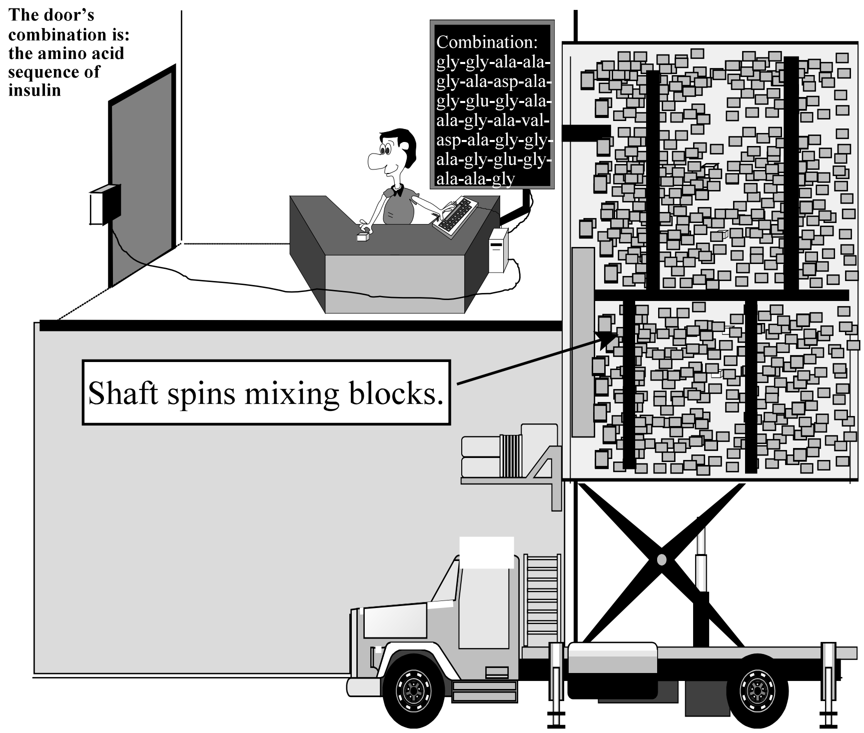


Table 5.2: Information in Insulin B Chain (Primordial Evolution)

pos	allowed amino acids	bits	pos	allowed amino acids	bits
2	phe, ala, leu, val	3.1	17	phe, tyr	12.1
3	val, ala, pro	3.1	18	leu	9.3
4	pro, lys, asn	11.5	19	val, ile	8.2
5	gln	13.1	20	cys	13.1
6	his, arg	15.4	21	gly	3.0
7	leu	9.3	22	asp, glu	7.4
8	cys	13.1	23	arg	16.4
9	gly	3.0	24	gly	3.0
10	ala, pro, ser	3.1	25	phe	13.1
11	his	16.4	26	phe, tyr	12.1
12	leu	9.3	27	tyr	13.1
13	val	8.5	28	thr, ser, asn	10.1
14	glu, asp	7.4	29	pro	12.2
15	ala	3.2	30	lys, arg	15.4
16	leu	9.3	31	ala, arg, thr, ser	3.1

Total bits = 280.5 bits

Example calculation: Phenylalanine, alanine, leucine and valine are possible at position one. There are 5,650 blocks labeled leu, 395,000 labeled alanine, 9,750 labeled valine, and 400 labeled phenylalanine in the truck. The total number of blocks is 3,520,880. So the probability that the scientist will pull a leucine, alanine, phenylalanine, or valine is (395,000+9,750+5,650+400) = 410,800 times in 3,520,880 tries. Thus the information at position one is calculated as follows: information = 3.32 x log (3,520,880/410,800) = 3.1 bits.

The total number of bits is 280.5. In chapter 4, the total for the B chain was only 108 bits. Intuitively, this is obvious because any proteins that emerge in the primordial soup will be composed of mostly alanine and glycine. Since real proteins do not follow this pattern, they are less likely to evolve in the primordial soup. The conclusion is that it is much harder for information and knowledge to evolve in the primordial soup.

The B chain of insulin contains 30 amino acids. So the average information contributed by each amino acid is equal to the total information divided by 30.

Information before life = 280.5/ 30 = 9.35 bits per amino acid

Information with the genetic code = 108/30 = 3.6 bits per amino acid

Because knowledge is defined in terms of information, it too must increase.

Molecular Knowledge in The Primordial Soup

Table 5.3 calculates the knowledge in the B chain of insulin assuming that the protein evolved in the primordial soup.

Table 5.3: Molecular Knowledge in Insulin B Chain

pos	allowed amino acids	bits	pos	allowed amino acids	bits
2	phe, ala, leu, val	2.0	17	phe, tyr ,(trp)	2.0
3	val, ala, pro	2.0	18	leu, (ile),(val), (ala), (met)	3.1
4	pro, lys, asn	2.0	19	val, ile, (ala), (leu), (met)	3.1
5	gln, (asn)	12.1	20	cys	13.1
6	his, arg, (lys)	14.8	21	gly	3.0
7	leu, (ile), (leu), (val), (met)	3.1	22	asp, glu	7.4
8	cys	13.1	23	arg, (lys), (his)	14.8
9	gly	3.0	24	gly	3.0
10	ala, pro, ser	2.0	25	phe, (tyr), (trp)	11.5
11	his, (lys), (arg)	14.8	26	phe, tyr, (trp)	11.5
12	leu, (ile), (val), (ala), (met)	3.1	27	tyr, (phe), (trp)	11.5
13	val, (ile), (leu), (ala), (met)	3.1	28	thr, ser, asn	2.0
14	glu, asp	7.4	29	pro	12.2
15	ala, (leu), (ile),(val), (met)	3.1	30	lys, arg, (his)	14.8
16	leu, (ala), (val),(Ile), (met)	3.1	31	ala, thr, ser, -	0*

Total = 211 bits.

* Any position with a gap does not need an amino acid and therefore the knowledge is set to 0 bits.

Example calculation: At position 7, only leucine is found in the alignment. Nevertheless, the technique to calculate knowledge assumes that the other amino acids in this group are allowed. The number of blocks labeled with the 5 amino acids belonging to group 1 in the truck (figure 5.1) is 413,200. There are 3,520,880 total blocks. So the knowledge is 3.32 x log (3,520,880/413,200) = 3.1 bits.

Notice that no position can ever contribute less than 2 bits. If all 20 amino acids are found at a particular position, the position still contributes 2 bits. This accounts for the amino acids not used by life found in the soup.

The average knowledge per amino acid in the soup is calculated as follows: knowledge = 211 total bits / 30 amino acids = 7 bits per amino acid. The average knowledge per amino acid with the genetic code is only 76 total bits /30 amino acids = 2.5 bits per amino acid. (Refer to pg. 76, table 4.4 for number of bits using the code).

Because of the nature of logarithms, the implications are dramatic. Suppose that one of the first proteins to evolve contains 100 amino acids, and that 30% of this protein shows a conservation pattern similar to insulin.

Knowledge today = molecular knowledge =
100 amino acid x 2.5 bits per amino acid x 30% = 75 bits

Odds of evolving today are 1 time in 2^{75} tries or 1 time in 4 x 10^{22} tries. This could happen with enough tries.

Knowledge soup = molecular knowledge + primodial information

Knowledge soup = 100 x 7 bits per amino acid x30%
+ 100 x 2 bits per amino acid x 70%
350 bits

Odds of evolving in the primordial soup are 1 time in 2^{350} tries or 1 time in 2.2 x 10^{105} tries. This can never happen.

Before life exists, chance will require an incredible number of tries to create knowledge, and the vastness of space, the number of atoms in the universe, and the incredible age of the universe do not make a dent in the problem. Nature simply cannot accumulate enough tries to overcome the poor odds.

Finally, this chapter had to make quite a few assumptions. Some readers may be concerned about these assumptions, but realize that almost every assumption was for the benefit of protein evolution. For example, this chapter assumed that the primordial soup did not contain aldehydes, carboxylic acids, and amines. This assumption is obviously false, but it greatly improves the chance for a protein evolving because it eliminates many side reactions. Also the amino acids not listed in Miller's table because they are not present in significant quantities are assumed to be in the soup at a very generous level. Allowing every star in the universe to have one planet is certainly a generous assumption, but perhaps the most generous assumption is to allow every single one of these planets to be composed entirely of peptide chains each containing 150 amino acids. This assumption is only rivaled by the next one that allows these peptide chains to break down and reform every year, and even with all of these generous assumptions, the probabilities do not budge from zero.

Any scientist who believes that nature can create molecular knowledge before life exists is relying on faith to justify his opinion because the math just does not support this belief.

References:

1) Keefe and Szostak, "Functional Proteins from a Random Sequence Library," Letters to Nature, 410: 715-718, 2000.

2) Miller, Which Organic Compounds Could have Occurred on the Prebiotic Earth?, Cold Spring Harbor Symposium of Quantitative Biology Volume L11, 17-25, 1987.

3) Miller, Orgel, The Origins of Life on Earth, Prentice Hall, 1974

4) Ehrenfreund, Glavin, Botta, Cooper, Bada, "Extraterrestrial amino acids in Orguil and Ivuna: Tracing the parent body of CI type carbonaceous chondrites," PNAS, 98: 2138-2141, 2001.

5) Fox, Dose, Molecular Evolution and the Origin of Life, 1972.

Chapter 6: Introduction to Chemistry and Entropy

This chapter will introduce chemistry, organic chemistry, quantum mechanics, and thermodynamics. The goal is to make sure that all readers understand how and why chemicals react with each other, and how and why the laws of thermodynamics influence these reactions.

The key concept of entropy will be introduced in this chapter. Entropy is often defined as disorder, but this definition is both misleading and incorrect. In classical thermodynamics, entropy is a mysterious concept. Entropy is difficult to define without considering quantum mechanics and micro-states. While these last two topics are usually only found in advanced chemistry and physics textbooks, they are absolutely necessary to understand entropy. Entropy is not a difficult concept. It is simply a measure of uncertainty that must always increase with time.

Entropy makes it very difficult for a self replicating molecule to exist because self replication decreases entropy. Life has many ways around this problem. The most common solution involves tapping plentiful energy sources to drive replication (chapter 7). Simple self replicating molecules cannot do this.

Chemicals and Atoms

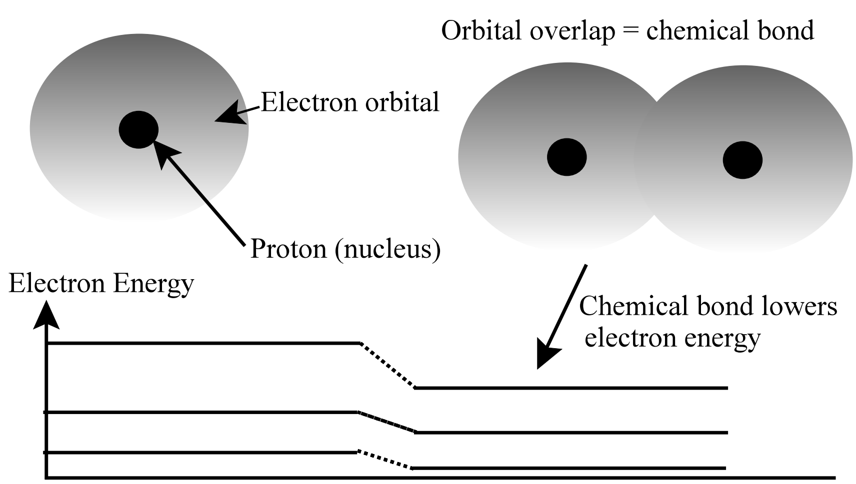
Chemicals make up everything that is a solid, liquid or gas. The earth's atmosphere today is composed predominantly of two chemicals, nitrogen and oxygen. Chemicals are made from atoms. A chemical composed entirely of the same type of atom is called an element. Oxygen is an element because it is composed entirely of oxygen atoms. Hydrogen is an element because it is composed of hydrogen atoms. Water is not an element because it is made from oxygen and hydrogen atoms. A molecule is a collection of two or more atoms. A water molecule consists of two hydrogen atoms and one oxygen atom.

The Hydrogen Atom

Hydrogen is the simplest atom. It contains one proton and one electron. The proton is the central core, the nucleus. Contrary to popular belief, electrons do not orbit the nucleus like the planets orbit the sun. When a small particle like an electron is confined to a small space, it no longer behaves like a particle. Its energy now must jump in discrete steps. This situation is analogous to a baseball that can only be thrown at 20, 50 or 90 mph. If a pitcher tries to throw it at any other speed, the baseball will still travel at 20, 50 or 90 mph. Furthermore, the electron can never be precisely located at any instant in time. The electron is always found somewhere in its orbital (figure 6.0).

When two electron orbitals merge to form a hydrogen molecule, the two electrons are free to occupy either orbital. Both hydrogen atoms share these two electrons. Because each electron can be found anywhere in this merged orbital, the electrons can find a lower energy state. This is why chemical bonds form.

Figure 6.0: Hydrogen Atom



Representing Chemicals with Symbols

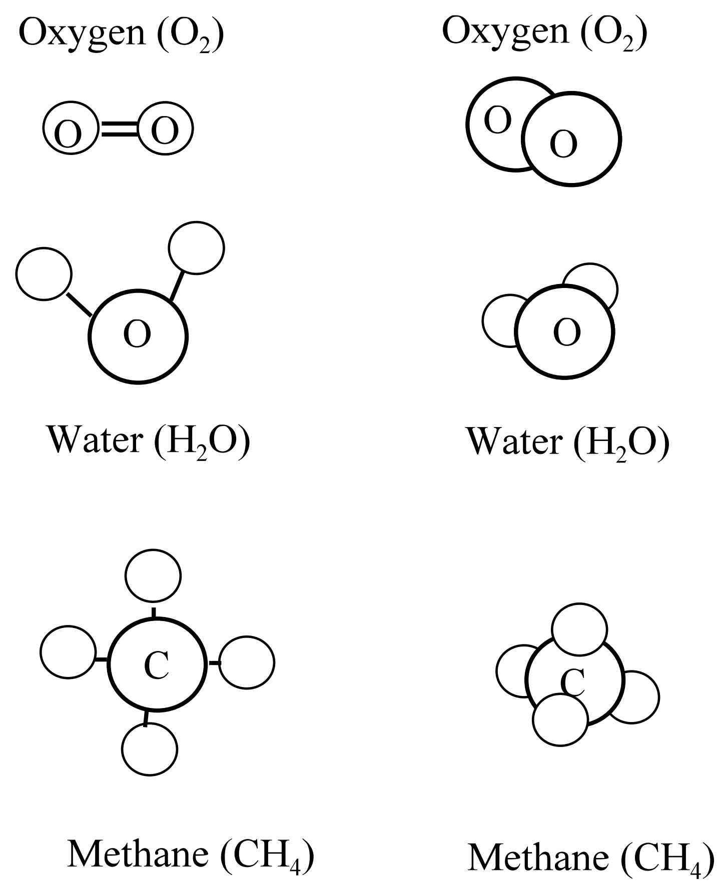
Every atom can be represented by a symbol. These symbols are usually the first letter in the name. For example, the symbol for hydrogen is H, the symbol for carbon is C, the symbol for oxygen is O, the symbol for nitrogen is N, and the symbol for sulfur is S.

Using symbols, a molecule of water is represented by H_2O where the subscript indicates that there are two hydrogen atoms in every water molecule. Figures 6.1 and 6.2 show how chemicals can be represented by names, symbols, balls and sticks, and spheres. The lines connecting the atoms are chemical bonds. Chemical bonds are the glue that hold molecules together.

In figure 6.1, three chemicals are represented, oxygen, water and methane. The chemical formulas for each are O_2, H_2O and CH_4. The ball and stick models make it easy to see what chemical looks like, but the ball and stick models do not show the electron orbitals. Atoms are really much bigger as shown in the space filling models of figure 6.1.

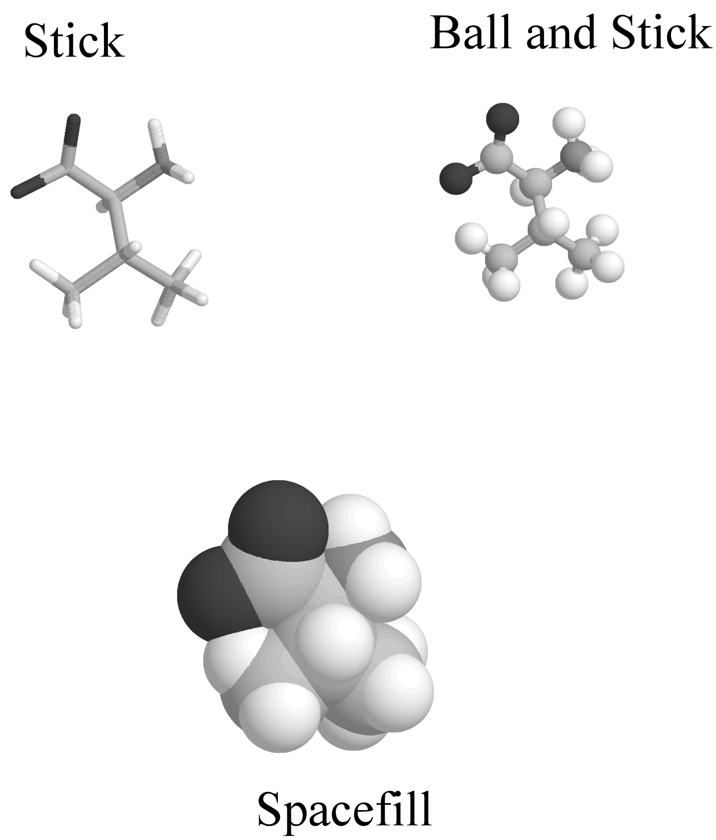
This book will also generate many images using a molecular visualization tool called Rasmol. The images of chemicals are more accurate with this tool. Figure 6.2 shows the amino acid, valine. Because the atoms are no longer labeled, the color must be used as an indication of atom type. In figure 6.2, black is oxygen, white is hydrogen, light gray is carbon, and dark gray is nitrogen. Rasmol can generate a space fill image, a ball and stick image, and a stick image as shown in figure 6.2. Rasmol also allows the image to be rotated by the user. While the space fill model in figure 6.2 is the only accurate representation of valine, it is too difficult to see how the atoms are connected using this model. So the stick and ball and stick are often preferred. In complex molecules, the cartoon view generated by Rasmol is often preferable (figures 3.13 and 3.14 on pages 58 and 59).

Figure 6.1: Ball and Stick and Space Filling Model



In both models, unlabeled small circles are hydrogen.

Figure 6.2: Rasmol Images of Valine



Chemical Bonds

Atoms are composed of three components, electrons, neutrons, and protons. The protons and neutrons form the central core, the nucleus, and electron orbitals interact with each other to form chemical bonds in such a way that each orbital is filled with electrons. In many chemicals, this can only be accomplished by sharing the available electrons.

To fill their orbitals, carbon atoms must share 4 electrons. To do this, they need to form 4 chemical bonds. Nitrogen atoms must share 3 electrons, so they typically form 3 chemical bonds. Hydrogen atoms must share 1 electron, so they typically form one chemical bond. Oxygen atoms must share 2 electrons, so they typically from two bonds.

By knowing how many chemical bonds each atom requires, it is possible to predict chemical structures. For example, every water molecule has 2 hydrogen atoms and 1 oxygen atom. The oxygen atom is satisfied because it shares 2 electrons, one with each hydrogen atom. Likewise, both hydrogen atoms are satisfied because each shares one electron with the single oxygen atom (figure 6.1).

Multiple Bonds

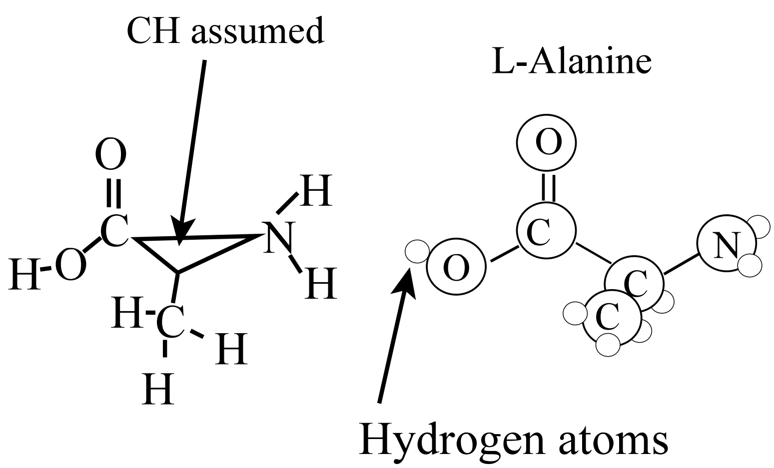
In many cases, if atoms cannot satisfy their requirements for sharing electrons, they will form double and triple bonds. For example, oxygen forms a double bond with itself. Since oxygen desires two electrons, by forming a double bond instead of a single, each oxygen atom in an oxygen molecule is satisfied (both have 2 bonds and share 2 electrons).

Chemical Symbols

It is not practical to draw the ball and stick model or the space filling model every time a chemical is mentioned. So instead chemists have developed various shorthand representations. The most condensed is to simply use symbols to represent chemicals. The symbol H_2O represents a water molecule.

This shorthand is not that useful for large molecules because it does not indicate how the various atoms are arranged. It is easier to visualize chemicals with the ball and stick model, but these are too cumbersome to draw. So the compromise is to replace the ball in the ball in stick model with a letter representing the atom. So hydrogen atoms are replaced by the letter H, and oxygen atoms are replaced by the letter O. Lines then connect the symbols to show chemical bonds. Since carbon often forms the backbone of large molecules, it is generally depicted by a line that bends. The hydrogen atoms connected to this assumed carbon are also assumed. This technique is shown for alanine in figure 6.3.

Figure 6.3: Alanine



Matter, Energy, Heat, and Temperature

Matter is composed of atoms. Matter takes three forms, solid, liquid and gas.

Temperature is a measure of how fast the atoms in matter are moving. If a room is hot, then the oxygen and nitrogen atoms in the room are moving very fast. If the room is cold, the atoms are moving slowly. The same is true in water. Even in a solid rock, the atoms are free to vibrate, and this vibration is a measure of the temperature.

Heat is a measure of energy transfer. Heat always flows from hot objects to cold ones. Fast moving atoms impart some of their energy to slower ones when they collide. This slows down the fast atoms and speeds up the slow ones. Energy is thus transferred from hot objects to cold ones.

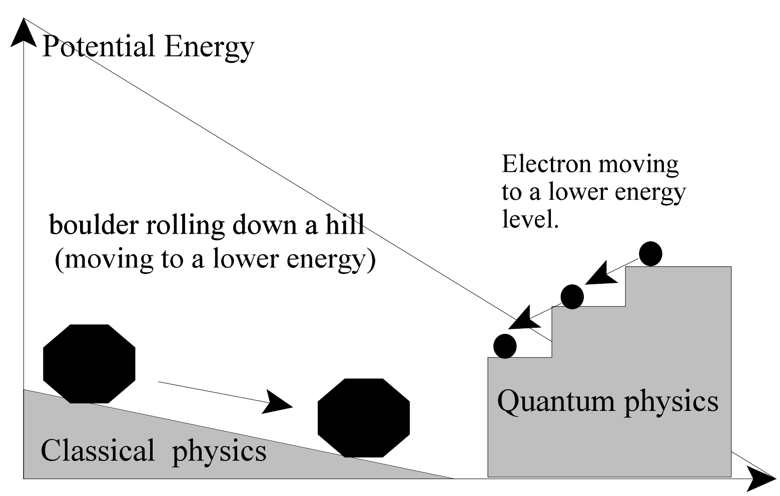
Energy is the ability to do work. A boulder sitting on top of a hill is said to have potential energy. When it starts rolling down the hill, this potential energy is converted to kinetic energy. If this boulder is tied to a rope attached to a cart on the other side of the hill, then it can lift the cart up the other side as it rolls. Lifting the cart is work. So in this example, the boulder accomplishes work as it rolls.

As the boulder rolls some of its potential energy will be converted to heat. This will raise the temperature of the boulder, the hill, the air, and the cart. This means that some of the work done in lifting the boulder to the top of the hill cannot be recovered when the boulder rolls down the hill.

Quantum Mechanics

As mentioned earlier when a small particle like an electron is confined to a small space it no longer behaves like a particle. Its energy becomes quantized. This means that it can only take on discrete energy levels. The best way to illustrate this is to envision a boulder rolling down a hillside. In classical physics, the hillside is gently sloped allowing the boulder to be anywhere on the hill (figure 6.4). It gradually gains speed as it progresses down the hill. The loss in potential energy is continuous. Some of this energy becomes kinetic energy (the energy associated with the moving boulder) and some of it becomes heat. Quantum physics tells a different story. The hill is like a series of steep cliffs separated by flat areas. Atoms and electrons can only reside on the plateaus. Each plateau is a quantum energy level. For atoms and molecules, this quantum energy level is similar to kinetic energy. That is faster moving atoms occupy a higher energy state. Atoms can rotate, vibrate, and move through space. All of these energy states are quantized.

Figure 6.4: Quantum Physics vs. Classical Physics



Micro-states and Entropy

Consider two very cold chambers that are completely empty except for a few oxygen atoms (figure 6.5 and 6.6).

Chamber 1: 4 atoms, 5 quantum states, total energy = 8 units
Chamber 2: 3 atoms, 5 quantum states, total energy =6 units

There are 5 quantum energy states available to these atoms. The lowest has an energy of 2 units, and the highest has an energy level of 10 units. The atoms in both chambers are only free to occupy the lowest energy state. If one occupies a higher energy state then the total energy of the system will be too high. In this case, there is only one possible way to arrange the atoms in each chamber. The ways that the atoms are arranged to fill the quantum energy states are called micro-states. Figures 6.5 and 6.6 illustrate this principle.

Figure 6.5: Micro-state for Chamber 1, Total Energy = 8 Units

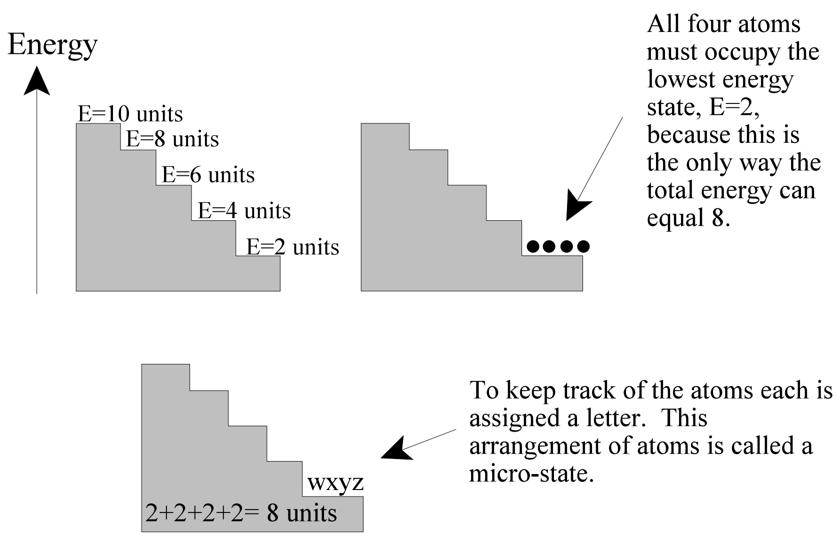
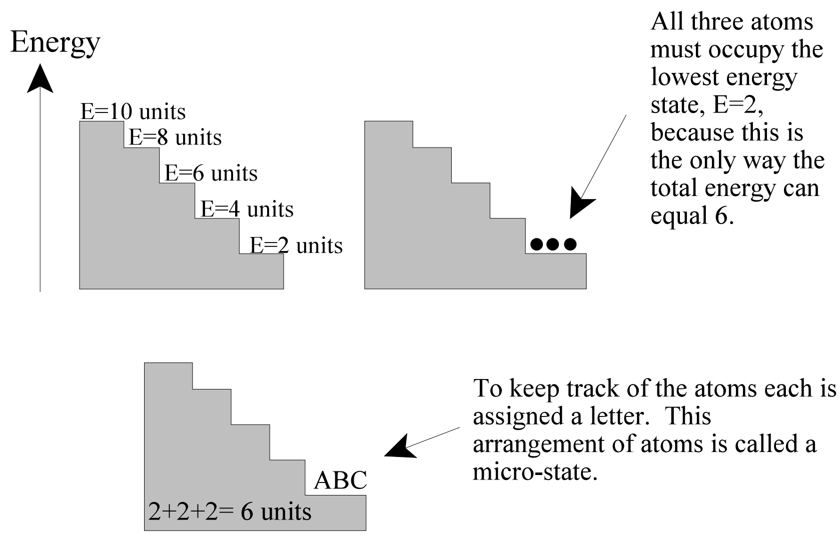


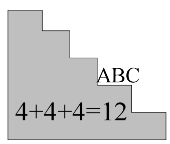
Figure 6.6: Micro-states for Chamber 2, Total Energy = 6 Units



If chamber 2 is heated, the atoms will move faster. Assume the total energy is now 12 units. Figure 6.7 shows that the number of available micro-states is greatly increased.

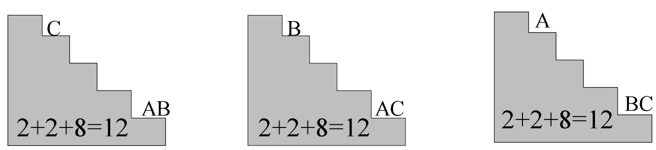
Figure 6.7: Ten Micro-states Available After Heating

Distribution 1: all atoms have 4 energy units



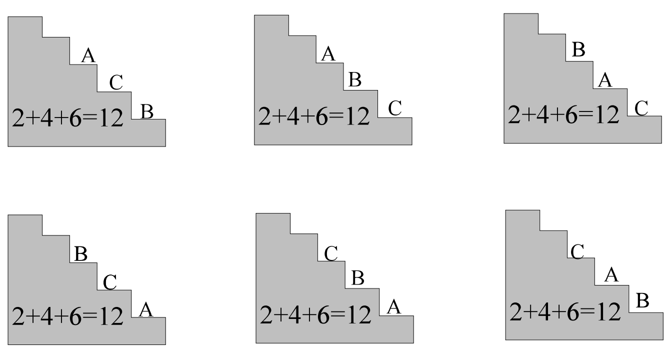
1 microstate

Distribution 2: one atom has 8 units and two have 2 units



3 microstates

Distribution 3: one atom has 6 units, one has 4 and one has 2



6 microstates

Entropy is often defined as a measure of disorder, but this definition is not only misleading it is wrong. Entropy is a measure of available micro-states. So in this example, the oxygen atoms that are heated have more entropy (10 micro-states available vs. 1 micro-state available).

Entropy can also be defined as a measure of uncertainty. Because as more micro-states become available to the system, the state of the particles becomes more uncertain. Entropy has nothing to do with disorder.

From this example, it should be clear that there are several ways to increase the entropy of a system. Increasing the temperature is one. Suppose 5 quantum energy levels are added so that all of the odd energies are represented, E=1, E=3, E=5, E=7, and E=9. The oxygen atoms will be able to distribute themselves in many more ways and thus find more micro-states.

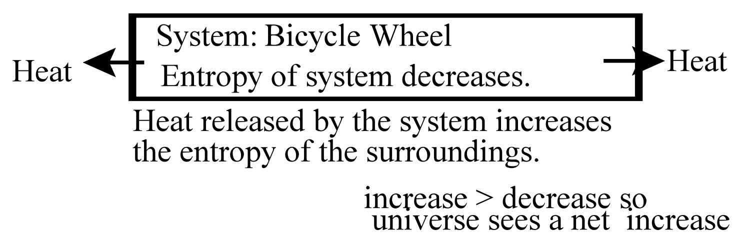
The Second Law

The second law of thermodynamics states that in any spontaneous process the entropy of the universe will increase. What does this mean? It means that all spontaneous processes must increase the number of available micro-states. The number of available micro-states after any event will always be greater than the number of available micro-states before the event. Because atoms form large objects like boulders, large objects must also obey the second law.

The second law can be stated in a very intuitive way. The uncertainty of the universe increases with time. This is why it is more difficult to predict what will happen far in the future. Weather is a great example. The weatherman may be able to forecast rain tomorrow, but he cannot forecast rain a month in advance.

Consider a bicycle that is turned upside down. The back wheel is spun until it is moving very fast. The second law explains why the wheel will not turn for long. The atoms that make up the wheel are moving very fast, so these atoms have lots of energy. As the wheel spins some of this energy is transferred as heat to the air around the wheel and to the frame that holds the wheel. This increases the air temperature which in turn increases the number of micro-states available to the air molecules. The entropy of the air molecules increases. Since the frame also heats up, its atoms are free to occupy more micro-states. Eventually all of the energy in the wheel will be dissipated as heat. The wheel's entropy decreases as it slows. The entropy of its surroundings increase. The increase is more than enough to offset the decrease. Thus, the entropy of the universe as a whole increases.

Figure 6.8: System vs. Surroundings



The equation that must always be satisfied is as follows:

entropy change of the system + entropy change of the surroundings = entropy change of the universe > 0.

The entropy of any system can decrease as long as the entropy of the surroundings increases, and the increase is greater than the decrease.

Heat Flows From Hot Objects to Cold Ones

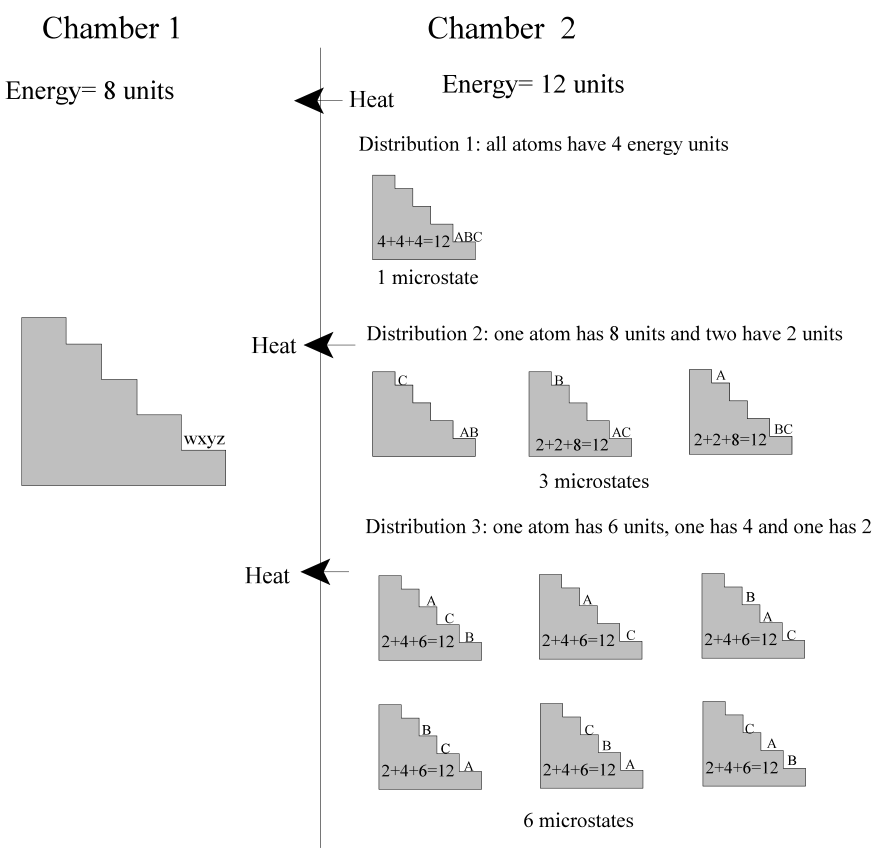
Suppose that chamber 1 (figure 6.5) and chamber 2 (figure 6.7) are brought together so that their walls touch. Heat can now be transferred between the systems, but the atoms are confined to their respective chambers. Because chamber 2 is hot and chamber 1 is cold, heat should flow out of chamber 2 and into chamber 1. Does such a flow increase the number of available micro-states? Figure 6.9 and 6.10 show that this process increases the number of available micro-states. The steady state (maximum number of micro-states) is reached when energy is distributed equally among both chambers.

This example in meant to convey an intuitive feel for the second law, what it means, and how it works. Keeping track of how micro-states change in real processes is not practical. There are too many atoms and too many quantum states. The number of available micro-states in most systems is far greater than the number stars in the universe. Even the most powerful computer cannot keep track of this many micro-states.

Fortunately, when the number of atoms is increased to 10,000 or more, one distribution dominates. In figure 6.7, distribution three is the most probable. The system will spend 60% of its time in one of the micro-states belonging to this distribution. As the number of atoms is increased, the dominance of the most probable distribution also increases. With 100,000 or more atoms, the system will spend all of its time in the most probable distribution because the most probable distribution is always the one with the most available micro-states.

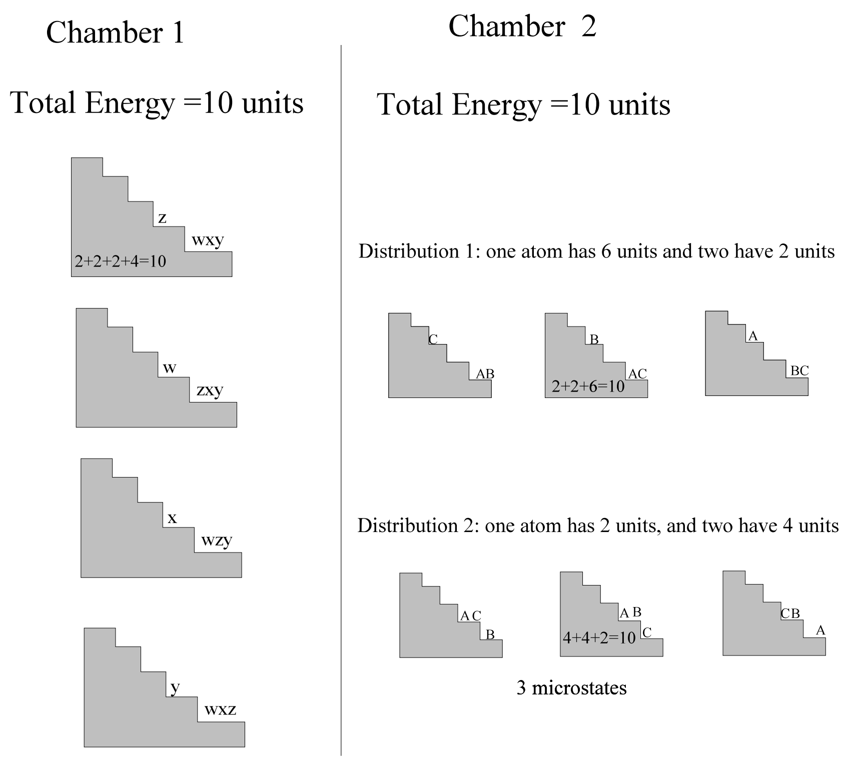
Entropy is now easy to understand. As chemicals or atoms react and move around, they try to find their most probable distribution. Since this is the distribution that maximizes the number of available micro-states, it is also the distribution that maximizes entropy. Thus, the entropy of the universe always increases.

Figure 6.9: Initial Distribution of Micro-states



Mathematically, if the universe is defined as both chambers, then there are only 10 micro-states (not 11). No matter which micro-state chamber 2 chooses, chamber one is always in the same micro-state. Thus the total number of micro-states is given by multiplication not addition and 10 x1 =10.

Figure 6.10: Final Distribution of Micro-states



Mathematically, if the universe is defined as both chambers, then there are only 24 micro-states (not 10). For each micro-state chamber 2 chooses, chamber one can choose 4. Thus the total number of available micro-states is given by multiplication not addition and 4x6 =24. Thus, when heat flows from a hot object to a cold one, the number of microstates and hence the entropy of the universe is increased.

Entropy and Chemical Reactions

Chemicals interact with each other in chemical reactions. Chemical reactions break existing chemical bonds and form new ones. The reaction is represented symbolically by an arrow. Consider the following chemical reaction: $2H_2 + O_2 \text{–>} 2H_2O$. In this reaction two molecules of hydrogen combine with one molecule of oxygen to yield two molecules of water (figure 6.11).

The second law determines whether or not this reaction will happen. Let the chemicals be the system. The electrons in water have less energy than those in hydrogen and oxygen. This decrease in electron energy releases heat as the reaction takes place. This heat increases the entropy of the surroundings, figure 6.12. This heat drives the reaction forward because it ensures that the entropy of the universe increases. Thus, this chemical reaction is said to be spontaneous. This means that given time, the reaction will happen. It does not mean that the reaction will happen quickly.

Figure 6.11: A Simple Chemical Reaction

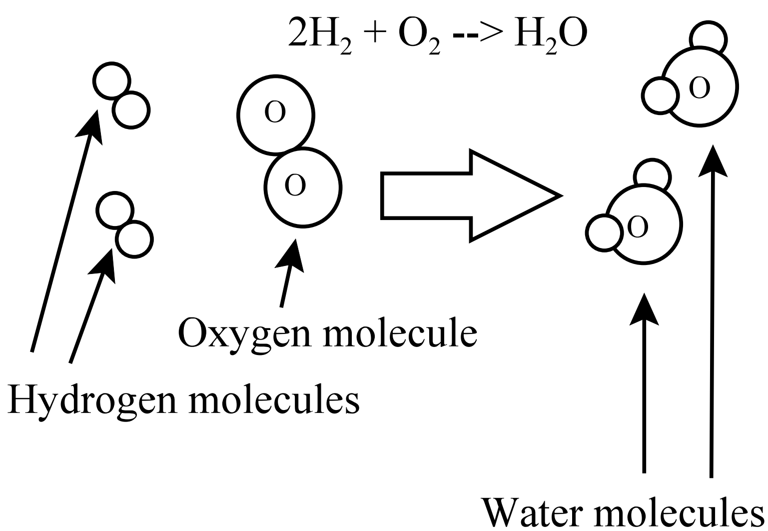
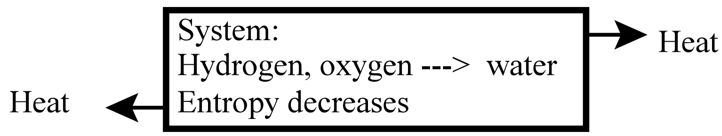


Figure 6.12: System Diagram for Above Reaction



Heat released by system increases the entropy of the surroundings.

increases > decrease so the universe realizes a net increase

Notice that the entropy decreases for the isolated system because when the system exists as oxygen and hydrogen more micro-states are available than when the system exists as water, but just like the bicycle example, the heat released increases the entropy of the surroundings. This allows the entropy of the universe to increase. Therefore, the reaction is spontaneous.

Chemical Kinetics

When hydrogen and oxygen are mixed together in a chamber at room temperature, nothing happens. There is no chemical reaction, but if a match is lit in the chamber, the chemical reaction happens so fast that it is explosive.

The reaction of hydrogen and oxygen to form water is spontaneous, but it will not happen unless some energy is put into the system. The energy required is called the activation energy. The chemical equation

$2H_2 + O_2 \rightarrow 2H_2O$ can be written as

$2H_2 + O_2 \rightarrow$ very high energy intermediate state $\rightarrow 2H_2O$

to indicate that for the reaction to happen the chemicals involved must transition through a short lived high energy intermediate state.

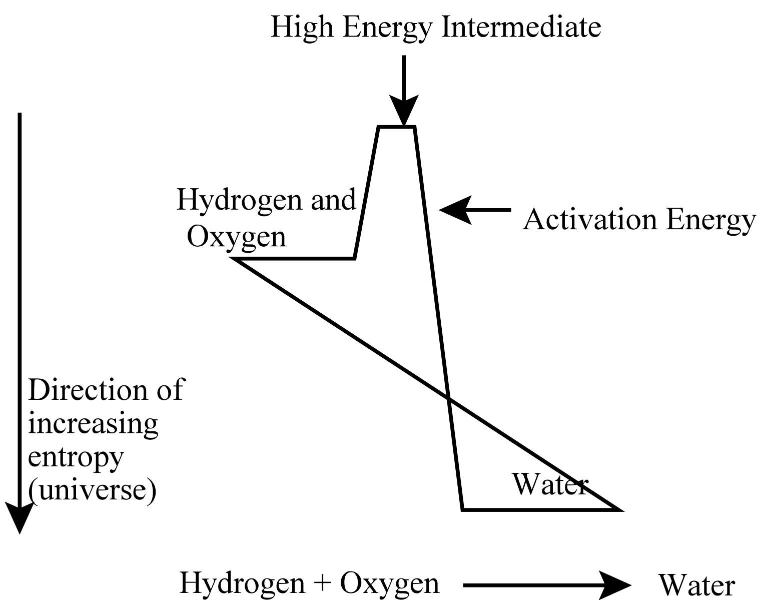
Because this state has more energy than the initial and final states, energy must be put into the chemicals to allow them to reach the intermediate state. In this case, the match allows some hydrogen and oxygen atoms to transition to the intermediate state. They then form water and release heat. The heat release causes more hydrogen and oxygen atoms to transition to the intermediate state. The reaction builds on itself in this manner until almost all of the hydrogen and oxygen are used up and only water remains.

Activation energy is very important for life. It allows chemicals to exist in a state indefinitely even if a change in state may increase the entropy of the universe. Figure 6.13 conveys these key concepts. The reaction to create water will increase the entropy of the universe as indicated by the arrow on the left side of the figure. Therefore, the reaction is said to be spontaneous. Nevertheless, it is not spontaneous unless the hydrogen and oxygen are provided with enough energy to cross through the high energy intermediate stage.

It is the activation energy that allows life to exist. Because of this barrier, chemicals that are not thermodynamically favored can exist for many hundreds of years. If a process increases the entropy of the universe, then the second law defines the process as spontaneous, but it does not have to happen right away. The process may take years to complete. The speed of a chemical reaction depends on the activation energy. The second law does not determine how fast a chemical reaction will happen. Hydrogen and oxygen can coexist in a chamber for a thousand years if no energy source is present to start the reaction.

Notice that the direction of increasing entropy is drawn downward in figure 6.13 to indicate that water is the preferred state as it maximizes the entropy of the universe.

Figure 6.13: Activation Energy



Chemical Equilibria

The reaction of hydrogen and oxygen to form water only happens in one direction. Figure 6.13 illustrates why. It is almost impossible for water to cross the activation barrier. Many chemical reactions happen in both directions. Figure 6.14 shows the reaction of a glycine-glycine molecule with water to yield two glycine molecules. Figure 6.15 shows how this reaction affects entropy. Notice that the entropy change is very small. The small change in entropy means that the reaction happens in both directions.

Figure 6.14: The formation of a glycine-glycine

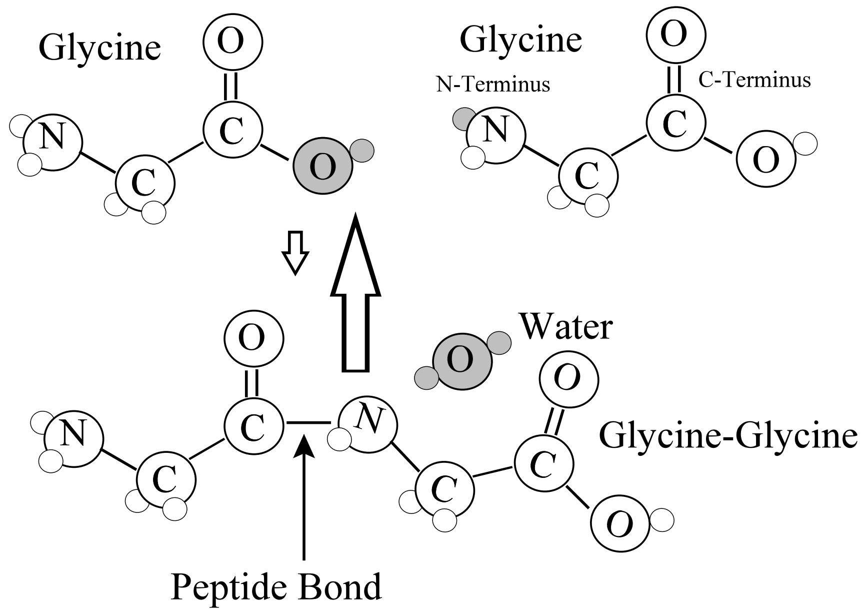
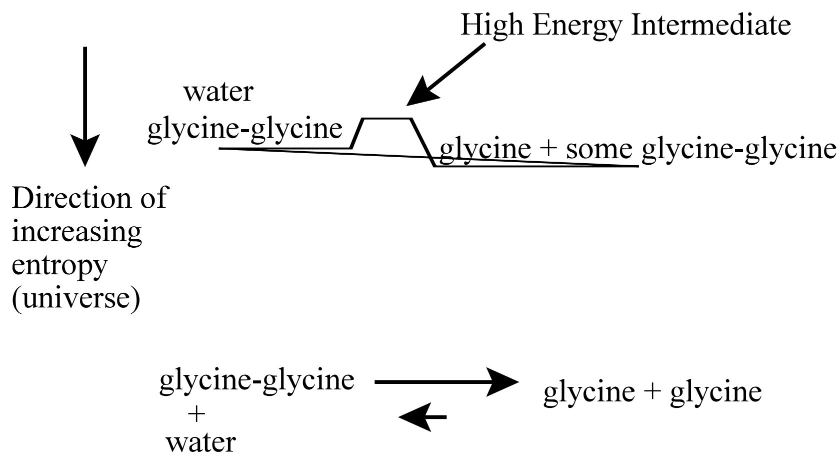


Figure 6.15: Entropy Change Associated with a Peptide Bond



In this case, the energy barrier favors the formation of glycine + glycine, but the reverse direction also happens as indicated by the reverse arrow in figure 6.14 and 6.15. The number of available micro-states in this reaction is maximized when there is some glycine-glycine and quite a bit of free glycine. To satisfy the second law, this chemical reaction will find the point that maximizes the available micro-states.

The optimal mixture is the one that maximizes the available micro-states and hence the entropy. Figure 6.15 should be compared to figure 6.13. The change in entropy is so great in figure 6.13, that the number of available micro-states is maximized when the universe exists only as water.

Notice that this chapter uses the term *available micro-states* as opposed to *micro-states*. The number of micro-states is a property of a system and its surroundings, and as such, in many reactions the total number of micro-states does not change, but as unavailable micro-states become accessible to more atoms and electrons, the number of available micro-states increases.

At equilibrium the concentration of the chemicals in a system no longer changes. That is in figure 6.15, the concentration of glycine and glycine-glycine remains constant once the system reaches chemical equilibrium. The forward and reverse reactions still take place, but they cancel each other. Thus, no net change is observed.

Closed vs. Open Systems

Chemical equilibrium is a hard state to maintain. Almost anything that changes will alter equilibrium. If the temperature rises or falls, a new equilibrium will have to be found. If energy is put into the system, the temperature will rise. Any chemicals that enter or leave the system will also change equilibrium. Because of these factors, only closed systems reach equilibrium. A closed system is one that is completely isolated from its surroundings. Chemicals in the system are not allowed to leave. New chemicals are not allowed to enter, and no heat can be transferred to or absorbed from the surroundings.

The earth is an open system. The sun continually transfers energy into the earth's system, and the amount of energy varies with the time of year. On a smaller scale, the earth's oceans are open systems. They continually receive new water and chemicals from rivers, and lose water to evaporation. Lakes and ponds are also open systems. Closed systems are very rare.

The implication is that most chemicals do not ever reach chemical equilibrium. This is why the second law is often stated as follows: all spontaneous processes tend to increase the entropy of the universe. The second law does not state that all spontaneous processes must instantly maximize the entropy of the universe. This is very fortunate for life. It allows the chemicals in life to exist in a state very far from equilibrium. This is why life is possible. This topic will be discussed in the next chapter.

References:

Vemulapalli, Physical Chemistry, Prentice Hall 1993
Energy Flow in Biology, Morowitz, Ox Bow Press, 1979.