

# Molecules of Life

The Breakthrough Science Society is going to observe the three years from 2007 to 2009 as “Darwin Years”. The year 2007 will mark the 125<sup>th</sup> year of his death and the year 2009 marks the 200<sup>th</sup> year of his birth as well as the 150<sup>th</sup> year of the publication of his seminal book “Origin of Species”. Darwin’s theory of evolution brought about a radical change in the outlook of the contemporary man. Its impact was felt not only in the fields of the biological sciences but also in all the fields of knowledge involving history. Even today his ideas and arguments appear as instruments of struggle in the scientific campaign against irrationalities of all sorts. So, during the next three years, we will study and cultivate the life and works of Darwin through various kinds of science programmes. On this occasion, we intend to publish a series of articles dealing with the theory of evolution as well as the evolution of plants and animals upto man in the form of translation from the second volume of the Bengali booklet “*Vivartan Yuge Yuge*” (Evolution through Ages) published by the BSS in 1999. In this issue the third chapter, dealing with the biochemical basis of life, is presented. – *Editorial Board, Breakthrough*

We saw in the previous chapter (published in the last issue) that the origin of self replicating molecules marked the advent of life on this earth. And now it is also known how evolution of these molecules started on the earth, which itself was in the early phases of its own evolutionary course. But the living organisms that we see around us at the present time – elephants, horses, trees, snakes, frogs – they are not just molecules. How could this vast, diversified living world come into existence? How could they be linked to that self replicating molecule? Put in a different way, the question can be rephrased as: “Was the possibility of producing such diversity inherent in the common laws of chemistry?” In this chapter we shall try to find an answer to this question.

Everything around us is composed of atoms of elements. Of the 92 natural elements (there are 11 more artificial ones) some are quite rare. A few are abundant in nature (e.g. nickel, helium) but have hardly any role to play in the system of life. So we can exclude them from our discussion. Again certain elements are present only in some

particular animals and plants, but are not found in all life forms. So, in our search for the essential ingredients of life, these elements can also be excluded from the list.

Proceeding this way, we would find that only six elements are essential for life. They are carbon, hydrogen, oxygen, nitrogen, sulfur and phosphorus.

Doesn’t it sound strange? The entire living world is mainly composed of only six elements. Yet, how complex and diverse it is! The beautiful colour of a rose, the venom in a snake, the aroma of onion – everything in ultimate consideration is rather a jig-saw combination of a few specific molecules. That means a large variety of molecules is required for such huge variations in appearance and activities that we meet with in the living world. And mind that, another huge portion of the organic world is no longer living on this earth. They are extinct, as the dinosaurs are. Logically, for such an apparently infinite variation, we would need an infinite number of different molecules. But how is it possible to have an infinite number of different molecules composed of just six elements?

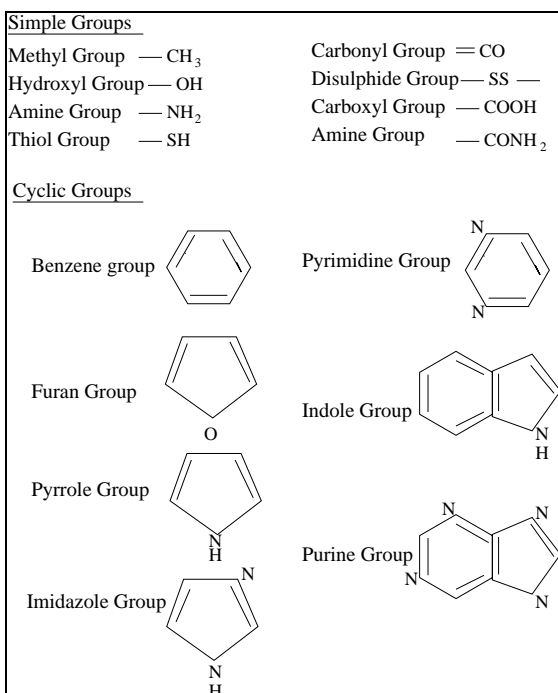


Fig. 1. Simple and Cyclic organic groups that make up almost the whole of the living world. Here the cyclic groups are represented by their geometric shapes. One has to assume that there is a CH combination in each vertex. If there is some atom other than hydrogen in a corner, it is shown explicitly.

Yes, it is possible; thanks to a unique property of the carbon atom. Carbon atoms can link together into a chain to form long stable molecules. That is why it is important to understand the structure of such complex chainlike molecules.

To illustrate such molecules, it is necessary to pictorially represent their structures showing the position of the atoms. However, it is almost impossible to draw each and every atom in such giant chains. Scientists have developed a mid-way course to represent such molecules in figures or diagrams. It has been found that a few atoms tend to form specific combinations that occur repeatedly in a long chain. These are called *groups* or radicals. But these groups are not self-sufficient, i.e., they

have one or two unsatisfied bonds through which they bind themselves with other atoms. Besides, these groups may also contain chainlike molecules of closed rings like the benzene ring.

It has been experimentally found that the millions and millions of molecules found in the living world are composed of only fifteen such groups. And these groups in their turn are composed of just five different atoms (see Fig. 1).

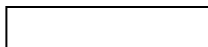
Readers may find it strange that just 15 groups, each comprising of only 5 different atoms, can arrange themselves in various ways to form the enormously complex living world. But that is the truth and in that sense the chemistry of life does not present anything that cannot be understood.

### From the small to the large

The groups or atomic arrangements shown in Fig. 1 are not very complicated. But if we examine any molecule from any part of our body we will find them too large – million times larger than a molecule of water.

To analyze such large molecules, scientists usually split up them into smaller segments. The method is simple. Just heat the material in presence of concentrated sulfuric acid or nitric acid for a few minutes. By analyzing these fragments, we have come to know that all large organic molecules are made of combinations of only a few smaller molecules.

The process of formation of large molecules based on combination of smaller molecules is known as polymerization. Starch – the principal constituent of rice and bread – when heated with acid produces only glucose (a saccharide molecule). It proves that large starch molecules are nothing but polymers of small glucose molecules. Cellulose, that makes paper and cotton,



also breaks down to give glucose. This means cellulose is also a polymer of glucose. But the pattern of the polymer chain is different in cellulose and in starch, and that makes their properties so different.

Nowadays scientists can synthesize polymers by combining different molecular units in laboratory. The polythene sheet we use in our day-to-day life is nothing but a polymer of a gaseous molecule known as ethylene. The automobile tyre, nylon rope, terylene cloth or plastic bucket – all are synthetic polymers of various smaller organic molecules.

Molecules of starch, cellulose or polythene may be large, but structurally they are simple. As we see they are just repetitions of same type of units. But organic molecules found in living system are not only large; they are also very complex in nature.

To understand the nature of these molecules scientists followed the same process – they simply split up the large molecules and examined the smaller fragments. They found that however complex the structure of large molecules be, by composition they belong to only two categories. Most of such complex macromolecules in a living body belong to a class known as *protein*. Millions and millions of protein molecules found in different cells starting from hair to nail have their unique functions to keep our body machine run smoothly. The hair on our head is nothing but a pure protein known as keratin. The haemoglobin found in red blood cells is also a protein by nature. Enzymes that help digest food are again protein molecules. Hormones that regulate our body temperature, blood pressure, and glucose level in blood are also proteins. Muscle, skin, eye, nose, ear, tongue – proteins are the main constituents of all the tissues of these organs or parts of body. One can guess

that this single type of molecule has the possibility of producing almost endless variations.

Another kind of complex molecule can be found in every cell, which is called *nucleic acid*. The nucleic acids present in all cells of an individual are identical. But between different individuals there are variations in the structure of these molecules.

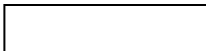
Analyzing these nucleic acids and protein molecules, scientists came to know that these molecules are basically polymers of smaller molecules known as *nucleotides* and *amino acids* respectively. Both nucleotides and amino acids are smaller molecules and quite simpler in structure.

Interestingly, millions and millions of proteins are composed of just 22 amino acids. And in nucleic acid the numbers of nucleotides are even lesser – only four.

### **Amino acids and proteins**

The general structure of amino acids is depicted in Fig. 2. Examination of this structure reveals that in one half of an amino acid, there is one amino group ( $-NH_2$ ), and on the other, a carboxylic group ( $-COOH$ ). In between them, there is a carbon atom attached to a hydrogen atom with one bond at one end and to another group (shown as R) with another bond at the other. Except this R-part, the rest of the structure is same in all amino acids. If there is only a hydrogen atom in place of R, the resulting amino acid is known as *glycine*. If R is  $-CH_3$  the amino acid is *alanine*. When R is  $-CH_2OH$  the amino acid becomes *serine*. The group in the side chain can also be of a ring structure. In Table 1 one can see the structures of R in the twenty-two amino acids.

In certain conditions when two amino acid molecules come close to each other, they can form a bond between them by



losing one molecule of water – known as the *peptide bond* (see Fig. 3).

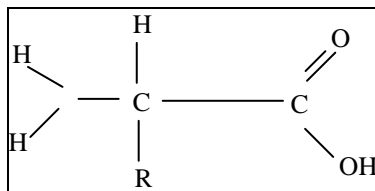


Fig. 2. The structure of amino acid.

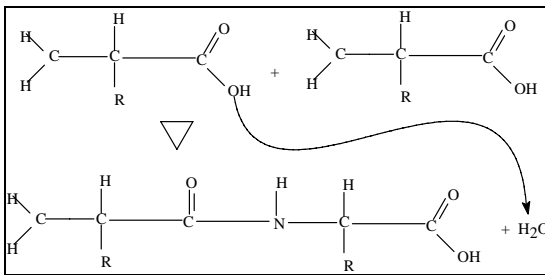


Fig. 3. Two amino acid molecules can join together forming *peptide bond*.

This way not only two, a number of amino acid molecules can get attached side by side in a chain-like structure to form a polymer. The backbone of this polymer is made up of the peptide linkage from which different side chains (i.e., different R groups) project outwards. Fig 4 depicts a simplified diagram of such kind of a polymer. This chain-polymer of amino acids are called *Protein*.

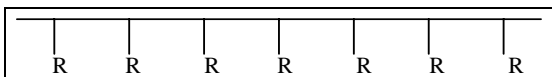


Fig. 4. Chain polymer of amino acid.

Such molecules that have all identical side chains are obviously monotonous. If you start making words with only one letter, say A, you will get A, AA, AAA, AAAA etc. But if you use two letters say A and B – they can be arranged in different ways like AB, AAB, ABB, AABB, ABAB etc. – i.e., more variations are possible. If we take a third letter, possible variations become still greater. And if there are 22

letters – an entire language system can be generated. In fact in the English language, the total number of letters used are surprisingly just four more than the number of “letters” in amino acids.

There is no limit to the number of amino acid molecules that can be present in a protein molecule. Millions of proteins can be generated by arranging different amino acids in different ways. This endless diversity of the protein molecules is the basis of the chemistry of life. Such variations have made it possible to evolve different protein molecules for different activities of our body. Moreover, there is obviously no limit to the generation of new variety every day.

The function of different proteins inside a living body depends mainly on their three dimensional structure. Some are straight, like silk (a polymer of glycine and alanine). Some are bent. In a protein chain, if there is an accumulation of excess positive charge in any part and excess negative charge in another, these two parts will try to come closer. In doing so, some chains become tangled within itself. Moreover, in some proteins there are more than one chain. This is due to the presence of cystine, a sulphur-containing amino acid, in these chains. The sulphur atoms of both chains form a disulphide linkage, which holds the two chains together (see Fig.5).

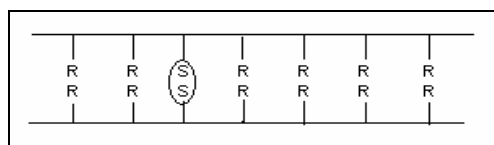


Fig. 5. Joining two chains by a disulphide bonding.

Let us examine the structure of an average protein molecule, say, haemoglobin. This is the basic constituent of red blood cell. In each haemoglobin molecule there are four protein chains, each consisting of 472



amino acids. These four chains are bound together by S-S linkages. In consequence they are winded together in such a way that the entire 3-dimensional structure of this molecule looks much like a thorny bush.

The rate of formation of such complex molecules is surprisingly high, about 25 millions per second. Moreover, in each copy there is an incredible degree of fidelity regarding the sequence of amino acids. Their three dimensional structures are also identical. It is not just a coincidence. If you twist a chain and release it, it recoils just like a spring to its original configuration. It is possible because of the excess positive and negative charges in specific locations along the chains.

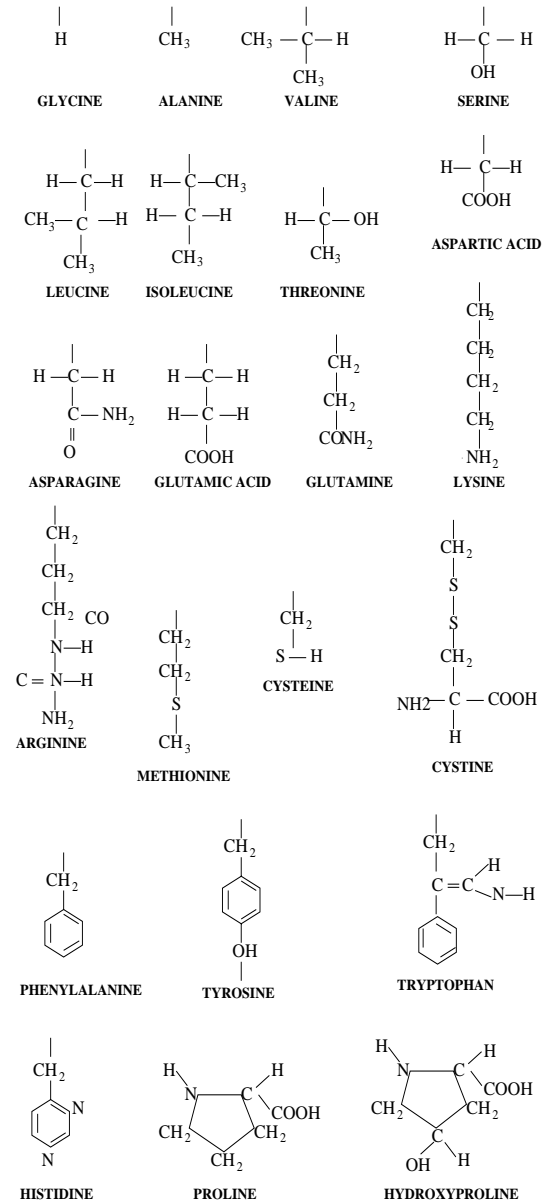
The natural question is, how can such molecules be generated in such large numbers inside a living body? Just as no engineer can construct a house without a proper blue-print, without a proper template inside a cell it is not possible to generate billions and billions of replicas of a single molecule. And this template is the *nucleic acid*.

### Nucleic Acid

On examining the structures of nucleic acids, scientists found them to be constituted of smaller molecules called nucleotides. The nucleotides are in turn composed of three types of molecules.

One of these three molecules is a carbohydrate molecule – composed of carbon, hydrogen, and oxygen. Like glucose (grape sugar), sucrose (cane sugar) or lactose (milk sugar), there are two other types of carbohydrate molecules – ribose and deoxyribose. They differ from each other by the presence of only one oxygen atom more in the former (see Fig. 6).

Table – 1: The structure of the side-chains in the 22 amino acids. In the last two, the side-chain takes a ring shape by combining with the amino group in the other side. That is why the whole molecule is shown, instead of showing only the R-part.



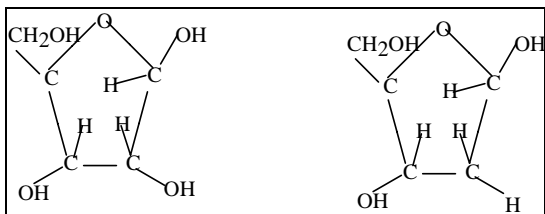


Fig. 6. Ribose and deoxyribose carbohydrate molecules.

The second molecule is inorganic – it is a phosphoric acid molecule; so herein is introduced compounds of phosphorus for their role in molecules of life. And that role is in the formation of nucleic acid. Phosphorus has five bonds. In phosphoric acid, all these valencies are satisfied (Fig 7). But if we remove one H atom, one bond becomes unsatisfied. It then can combine with one other group. When two H atoms are removed, two valencies become unsatisfied, and the molecule can combine with two other groups. What does it attach to? We shall come to that soon.

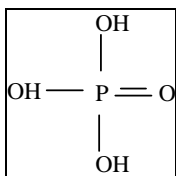


Fig. 7. Structure of phosphoric acid.

The third type of molecule in the nucleotide is called *organic base*. If we go back to Fig. 1, we will find two cyclic groups, purine and pyrimidine. The organic bases are just variations of purine and pyrimidine rings. Practically most of the nitrogen-containing organic compounds are like pyrimidine and purine rings. So they are also called nitrogen bases. There are three types of purine (thymine, uracil and cytosine) and two types of pyrimidine (adenine and guanine), as shown in Fig 8.

Now let us suppose that one ribose sugar and thymine base are combining with phosphoric acid. In that case, by giving up two molecules of water, this

combination will form a nucleotide molecule (Fig. 9). If two hydrogen atoms are removed from a phosphoric acid molecule, the two open bonds can bind with two nucleotides at the two sides. Thus the phosphoric acid molecules can act like joints to form long chains of nucleotides (Fig. 10). Such polymers of nucleotides are called *nucleic acid*.

Nucleic acids can be of different types. As mentioned before, pentose sugar is of two types- ribose and deoxyribose. Therefore, nucleic acid is also of two types. When nucleic acid contains ribose sugar, it is called *ribonucleic acid* (RNA). When it contains deoxyribose sugar, it is called *deoxyribo nucleic acid* (DNA).

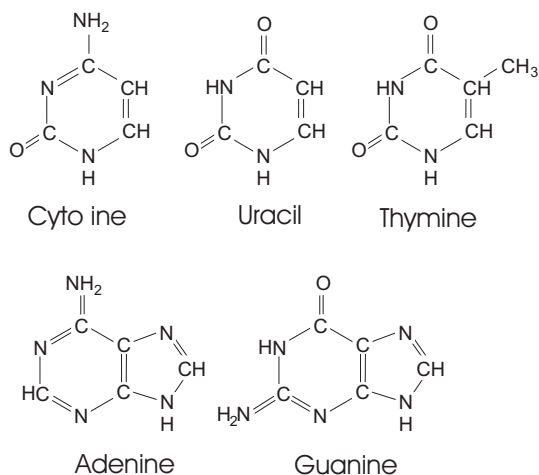


Fig. 8. Different types of purines and pyrimidines.

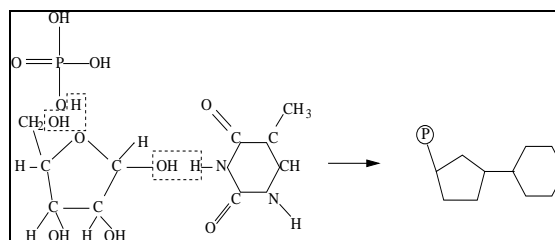


Fig. 9. Phosphoric acid, ribose and thymine join together to form nucleotides. In the figure to the right, the three components are shown as P, a pentagon, and a hexagon.

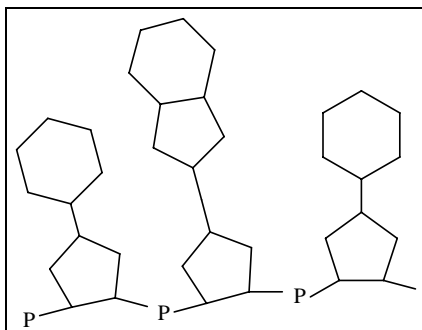


Fig. 10. Chain structure of Nucleic acid.

Again, in both of these nucleic acid chains, the nitrogen bases may be of different types (as shown in Fig 8). It has been found that DNA contains four types of nitrogen bases: adenine, thymine, guanine and cytosine. These are often abbreviated as A, T, C, and G. In RNA, uracil (U) is found in place of thymine. These four or five types of bases attach themselves to a backbone of phosphoric acid and sugar to give rise to the chain-like structure of DNA or RNA.

In the year 1953 Watson and Crick discovered the three dimensional structure of a nucleic acid molecule which consists of two helical chains of nucleic acid winding around each other. It is called a double helix. Due to the typical structure of the A and T parts, they form two hydrogen bonds between themselves whenever they come close in a double helix molecule, and get attached to each other. Similarly G and C tend to join hands with three hydrogen bonds between themselves. No other bonding is possible between these bases. Due to such structural compulsions, the bases in the opposite chains are just complementary to each other: the T-base always occurs opposite to A, and G occurs opposite to C.

The chains may contain any arrangement of the four bases described above. Moreover, there is no limit as far as the length of the chain is concerned. That explains why there are innumerable types of arrangements possible in DNA and RNA

molecules. Another important point should be noted. We know that in computers, all data are stored by arranging just two binary digits 0 and 1 in different sequences. Just like that, a DNA molecule can store information by arranging the ATGC bases in various possible sequences. The sequence of ATGC bases in the DNA molecule is called the *DNA code*. Information regarding the organization of an organism's body — what are really the hereditary traits — is transferred from one generation to the next with the help of these DNA codes. Obviously, it is the DNA code system that makes a crocodile a crocodile, a rhino a rhino, and so on. And it is due to the transmission of the same code system that a crocodile can produce only crocodile and no turtle.

(Translated by Avijit Garai)