

## **Chapter 13: Nucleic Acid Synthesis: Adenine**

---

Science has yet to find a solution for chemical evolution or the origin of life, and many scientists realize that this is the case. It seems that the puzzle just gets harder as the years pass. In the 1950s, most of the scientists investigating the origin of life probably thought that they would solve the problem. Today, it is clear that this is not going to happen. The self replicating RNA molecule theory persists only because nobody has proposed a better hypothesis. Science is trapped because by definition, science is only allowed to consider naturalistic explanations. Science may never find the solution.

While most scientists realize that the origin of life is a rather formidable barrier, few realize that the barrier associated with the creation of the first genes and proteins is just as high. The next two chapters will explore this topic.

### **Nucleic Acids are Critical to Life**

The first living organism either needed to find adenine and ATP in its surroundings or synthesize these two chemicals from more abundant chemicals found in the soup. If the first living molecule used impA instead of ATP, then it would have to either synthesize impA or find impA. A similar argument can be made for the other activated nucleotides required by life to replicate (GTP, UTP, CTP, TTP, impG, impU or impC).

In the laboratory, chemists can create adenine in an enclosed vessel from a concentrated solution of hydrogen cyanide and ammonia. The first form of life would be destroyed by these conditions, and because there is simply no way to concentrate volatile gases in an open environment (like the primitive earth), the conditions required to create large yields of adenine in the lab do not model the conditions on the primitive earth. So finding enough adenine in the primordial soup to support life seems very unlikely.

Miller has suggested that adenine was synthesized from hydrogen cyanide at very cold temperatures as water freezes, melts and re-freezes. While this process may have produced some adenine, if the entire body of water ever melts, then the adenine will again be too dilute. Furthermore, the freezing cycles described above are limited to small ponds because the salts present in the ocean interfere with adenine synthesis. Despite these concerns, adenine has been found in very small concentrations in some meteorites. So nature can make it in small quantities; as a result, some adenine was probably dissolved in the primitive ocean.

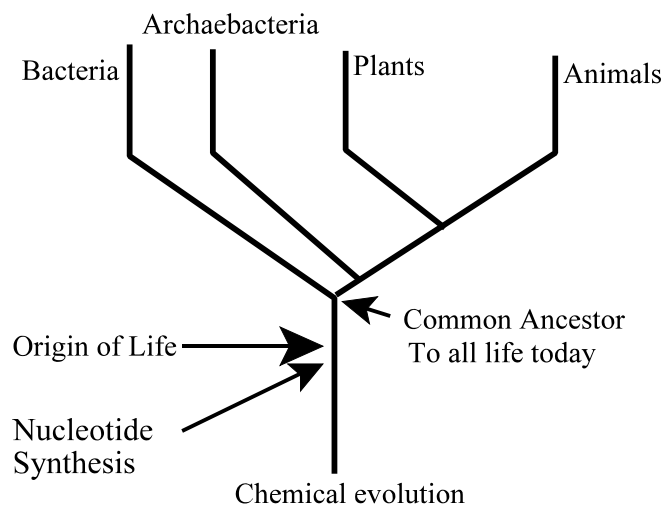
The problem is one of dilution. If the first living molecule does not have the ability to synthesize adenine and ATP, then it will have to search the ocean for decades, collecting these chemicals before it can replicate. Since the first living molecule must be simple, it would be more reasonable to speculate that the necessary chemicals float by the living molecule, and it simply collects them. Such a molecule may have to wait several thousand years to accumulate the ATP that it needs to replicate just once. This is not a reasonable model for a living system. The lifetime of the chemicals that make up the living molecule (assuming it is an RNA molecule) is very short (see chapter 9). So such a molecule will not exist long enough to replicate once.

Suppose instead that the primordial soup exists locally in a small pond or puddle. In theory, with the appropriate concentration mechanisms in place, such a puddle may have a concentration of nucleic acids many orders of magnitude higher than that of the primitive ocean. If life evolves in the puddle, it will quickly deplete the supply of free nucleic acids as it replicates. It will then run into the dilution problem outlined in the previous paragraph.

This argument and several others like it suggest that life was never as simple as many scientists have theorized. The enzymes that synthesize adenine and the other bases (guanine, uracil, cytosine, and thymine) almost certainly had to exist either before or coincidentally with the origin of life. The same can be said for the enzymes that synthesize ATP and the other activated nucleotides. The difficulties with chemical evolution demand that this be the case.

Furthermore, there is quite a bit of evidence that suggests that these enzymes did exist. Every single living thing (with the exception of a few parasites who have lost the genes) shares the same genes that encode the enzymes responsible for making adenine. This means that the genes responsible for the synthesis of adenine and the other bases existed in the common ancestor to all living things, 2.5 to 4 billion years ago.

Figure 13.1: The Tree of Life



Because life on the primitive earth was not possible without the genes responsible for making nucleic acids, these genes may have actually preceded life (figure 13.1). Even if life originated first, the problem remains unsolvable. The metabolic pathway that will be discussed next is perhaps the best example of an irreducibly complex system that can be found in life. The synthesis of adenine requires 11 enzymes. If a single enzyme is missing, the yield of adenine is zero. Therefore, all 11 enzymes must evolve together and become at least marginally functional before natural selection can preserve and optimize the system. The implication is that the very first step required to create the molecular knowledge is insurmountable. Chance and natural selection cannot explain the evolution of this metabolic pathway.

## Adenine Synthesis in Life

Life synthesizes adenine directly on the ribose ring. Eleven enzymes are required to make this happen. If a single enzyme is not functional, then no adenine can be synthesized. It is an irreducibly complex system. Before natural selection can help, all eleven enzymes must be marginally functional. The enzyme names and number of amino acids per enzyme are listed in table 13.1. The enzyme names are not important for this discussion and are given only for reference.

Table 13.1 Enzymes Required for Adenine Syntheses

| Enzyme Name  | Amino Acids |
|--|-------------|
| amidophosphoribosyl transferase  | 474         |
| phosphoribosylglycinamide synthase   | 416         |
| phosphoribosylglycinamide formyltransferase  | 195         |
| phosphoribosylformylglycinamide synthase   | 1295        |
| phosphoribosylaminoimidazole synthase  | 346         |
| phosphoribosylaminoimidazole carboxylase   | 162         |
| phosphoribosylaminoimidazole - succinocarboxamide synthase   | 239         |
| adenylosuccinate lyase   | 456         |
| phosphoribosylaminoimidazolecarboxamide & formyltransferase and IMP cyclohydrolase ( 2 enzymes in one protein) | 512         |
| adenylosuccinate synthase  | 432         |

Total amino acids = 4,527

### **Preliminary Calculation of the Information and Knowledge**

The analysis undertaken in chapters 4 and 5 could be repeated for all of these enzymes, but this would take several weeks to analyze the data and then another week to format the data in a presentable fashion.

Is there a simpler way? In chapter 5, the analysis revealed that the B chain of insulin contains 280.5 bits of information and 211 bits of knowledge. Since this chain contains 30 amino acids, the average information per amino acid is  $280.5/30 = 9.35$  bits per amino acid and  $211/30 = 7$  bits of knowledge per amino acid.

To find an answer quickly without any long drawn out mathematical analysis, the molecular knowledge required for life to synthesize adenine is given by: 7 bits per amino acid x 4527 amino acids = 31,689 bits. Refer to table 13.1 for the total amino acids.

Notice that the above calculation uses molecular knowledge instead of information. Nevertheless, the above calculation assumes that the average knowledge of a small highly conserved protein like insulin is representative of large enzymes. This assumption is incorrect, and so the number of bits calculated above is too large.

In a large enzyme, patches of amino acids that are located near each other on the three-dimensional structure of the protein tend to either be highly conserved or highly variable. The highly conserved patches will likely show a conservation pattern similar to insulin. Thus, the average knowledge per amino acid (based on the analysis of insulin) can only be safely applied to these highly conserved patches.

### 3-D Structure of Several Key Enzymes

The protein data bank houses a database which contains the 3-D structures of many proteins. All of the proteins involved in adenine synthesis can be found in this online database.

Another program called CONSURF will accept a protein database entry as input and generate a script that can be used to color the amino acids by their variability. The default of this program is to color highly conserved amino acids as shades of purple and weakly conserved amino acids as blue and white. The script so generated was modified so that all of the purple atoms appear as a shade of gray. Blue and off white were also modified to appear as white. The modifications were necessary so that the images would print correctly in black and white. The modified script was used to color several of the enzymes used in the synthesis of adenine (figures 13.2 - 13-5).

The black and gray regions in the following figures are regions of high conservation. These regions contain most of the information. Between twenty to thirty percent of the amino acids found in these proteins belong to a highly conserved region. In these regions, the knowledge per amino acid should be similar to that of insulin.

In chapter 5, because of the presence of non-biological amino acids in the primordial soup, 2 bits of primordial information are assigned to every amino acid that is used by life. So if these proteins evolved in the primordial soup, the white regions in the following figures must contribute at least 2 bits of information, and in this special case, because primordial information is a form of knowledge, it must be added to molecular knowledge to calculate the total knowledge, and this total can be related to a probability (see chapter 5).

If on the other hand, these proteins evolved after the genetic code was in place, then the white regions would not contribute much to knowledge. This scenario seems unlikely because life needs adenine to replicate, and chemical evolution does not create large quantities of adenine.

Figure 13.2: Amidophosphoribosyl Transferase (474 amino acids)

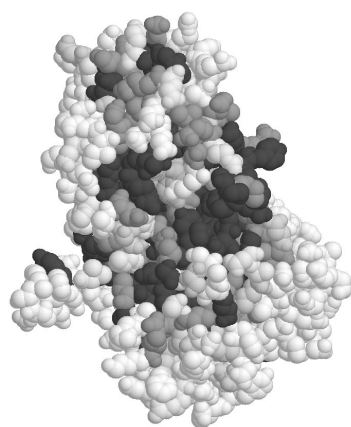


Figure 13.3: Phosphoribosylaminoimidazole carboxylase (162 amino acids)

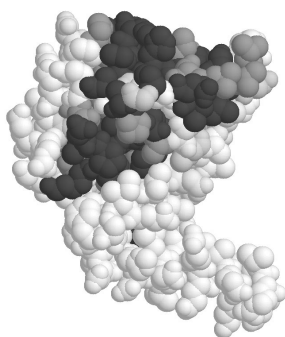


Figure 13.4: Adenylosuccinate Lyase (456 amino acids)

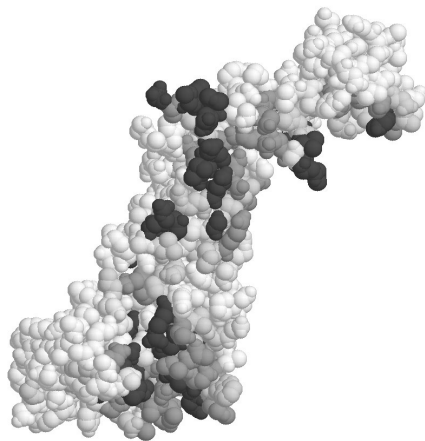
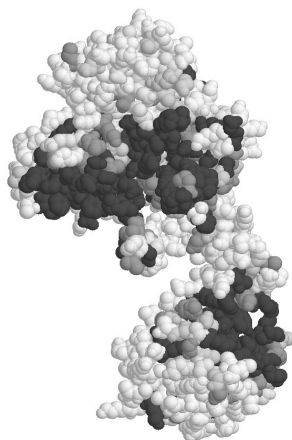


Figure 13.5: Phosphoribosylaminoimidazolecarboamide and IMP Cyclohydrolase (2 enzymes in one - 512 amino acids)



## Calculation of the Knowledge and Information

The total number of amino acids found in these enzymes is not nearly as important as the number of higher conserved amino acids. From figures 13.2 to 13.5, it appears that approximated 20-30% of the amino acids in these proteins belong to highly conserved patches.

So of the 4527 amino acids found in these enzymes, assume that only 20% contribute to molecular knowledge. With this assumption, the numbers found in the preliminary calculation now need to be multiplied by 20%.

Molecular Knowledge =  $4527 \times .2 \times 7$  bits per amino acid = 6,338 bits  
(black and gray regions)

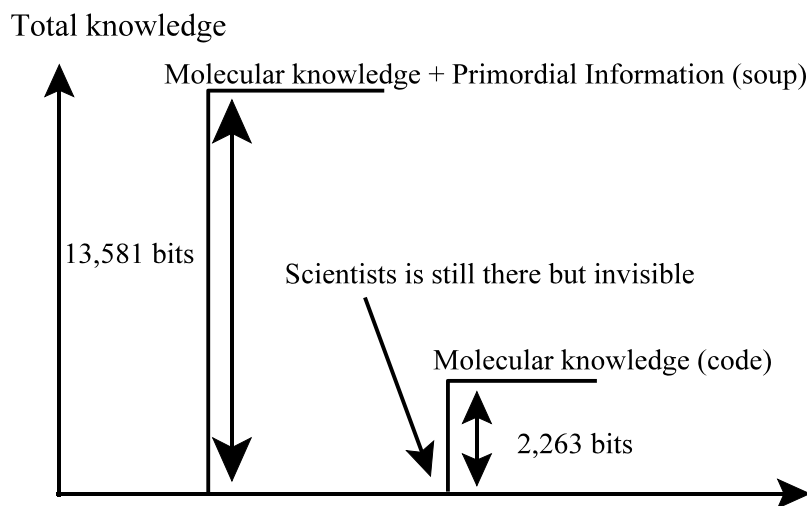
Primordial  
Information =  $4527 \times .8 \times 2$  bits per amino acid = 7,243 bits  
(white regions)

Total number of bits: 13,581 bits.  
Odds for success: 1 in  $2^{13581}$  of 1 in  $2 \times 10^{4088}$

In figure 13.6, the total number of bits calculated above is shown as the first step. In this special case, information and knowledge are additive. This can be done because the information must evolve before self replication is possible, so natural selection is not active. Furthermore, in this case, the information is a form of knowledge - the knowledge to use only the twenty biological amino acids in the soup.

The above calculation assumes that the 11 enzymes responsible for synthesizing adenine evolved before or coincidentally with the origin of life. Consider the case where the opposite is true. Suppose that the first living molecule obtained adenine from the soup. Could these 11 enzymes evolve after the genetic code is already in place? The number of bits is now as follows: knowledge = 4527 amino acids \* 20% \* 2.5 bits per amino acid = 2,263.5 bits. Refer to page 102.

Figure 13.6: The Evolution of Molecular Knowledge Required to Synthesize Adenine in the Soup and with the Genetic Code



At this scale, the scientist is so small that he is not visible in the picture. There is no way he can climb either wall.

Adenine synthesis is perhaps the best example of an irreducibly complex system that can be found in life, but there are many other metabolic pathways that need all of their enzymes to at least be partially functional before they can have any function at all.

Many proteins are complexes of smaller peptides, and many of these proteins are themselves irreducibly complex. This will be the subject of the next chapter.

## **Did Life Arise All at Once?**

Researchers have attempted to reconstruct the tree of life by comparing the RNA sequences found in ribosomes (ribosomal RNA) in many diverse species of bacteria, archaeobacteria, fungi, protozoans, animals and plants. Others have done the same by comparing the amino acid sequences of ancient proteins that are found in all living things. The results have been ambiguous.

While few researchers have been bold enough to suggest that life arose all at once, the evidence from the reconstruction experiments suggest that it did. The tree as it is conventionally drawn in figure 13.1 suggests that bacteria diverged first, and then archaeobacteria second. This means that archaeobacteria are more closely related to man than bacteria. The drawing in figure 13.1 is based on the analysis of ribosomal RNA, but this is not consistent with the protein comparisons that show that archaeobacteria and bacteria share many of the same genes that are not found in higher forms of life.<sup>1</sup>

“If these two prokaryotic groups span the primary phylogenic divide and their genes are vertically (genealogically) inherited, then the universal ancestor must have had all of these genes, these many functions: This distribution of genes would make the ancestor a prototroph with a complete tricarboxylic acid cycle, polysaccharide metabolism, both sulfur oxidation and reduction, and nitrogen fixation; it was motile by means of flagella; it had a regulated cell cycle and more. This is not the simple ancestor, limited in metabolic capabilities, that biologist originally intuited. That ancestor can explain neither the the broad distribution of diverse metabolic functions nor the origin of early autotrophy implied by this distribution. The ancestor that this broad spectrum of metabolic genes demands is totipotent, a genetically rich and complex entity, as rich and complex as any modern cell - seemingly more so.” - Carl Woese, The Universal Ancestor.

“For instance, transcription, translation and splicing machineries of the archaeobacteria resemble those of the eukaryotes, while the majority of the functional genes, coding primarily for metabolic enzymes, transport systems and enzymes of cell wall biogenesis, resemble the eubacterial ones. Microbiologists have reviewed a number of possible explanations for this mosaic, but none of them seems to be, at the present time, particularly convincing.” - Mayr, “Two empires of Three,”PNAS, 1998.

The ribosomal RNA data does not agree with the protein data. The ribosomal data suggest that the archaeobacteria are more closely related to man than bacteria. Yet the proteins tell another story. Woese resolved the dilemma by suggesting that the common ancestor was not a single cell, but that instead a collection of many different very simple cells. None of these simple cells had the ability to live alone, but together by sharing and transferring genes, they managed.

This is an interesting idea, but it does not change the probability. One of the nice features about information is that it is additive. So 1 complex protein containing 500 bits of information has the same chance of evolving as 10 smaller proteins containing 50 bits each. It is far more likely that the common ancestor was a complete cell similar to life today.

Mayr perhaps offered a better explanation by suggesting that life is composed of only two kingdoms not three. This of course still implies that the common ancestor to all life was not simple, but possessed a multitude of genes. It very well may have been more complex than any living cell found today.

The difficulties associated with chemical evolution almost demand that the first living thing be robust and complex. No organism whose genetic structure is based on RNA or DNA can replicate itself unless it can synthesize the bases required to do so (adenine, cytosine, guanine, thymine and uracil). And if ribose was used as the backbone in the first DNA/RNA, then it too must be synthesized. This implies photosynthesis and the Calvin cycle must be present. Furthermore, activation of nucleotides to create molecules like ATP requires that this first living molecule tap a plentiful energy source. The simplest living organisms today are parasitic. They obtain nourishment by absorbing nutrients from their host, and this process allows them the luxury of not having many genes that would otherwise be necessary. Given the dilute concentration of biological precursors in the soup (if it existed), the first living thing would not be able to rely on the soup for nourishment. The first form of life was a complete living cell with many if not all of the capabilities found in life today.

#### References:

- 1) Woese, "Interpreting the Universal Phylogenetic Tree," PNAS, 97: 8392-9396, 2000.
- 2) Woese, "The Universal Ancestor," PNAS, 95:6854-68-59, 1998.
- 3) Mayr, "Two Empires or Three," PNAS, 95:9720-9723, 1998.
- 4) Meyer, "The Origin of Biological Information and the Higher Taxonomic Categories," Proc. Of the Biological society of Washington, 117:213-239, 2004.
- 5) Glaser F./ Pupko T., Paz I., Bell R.E., Becher D., Martz E., Ben-Tal N., Consurf: Identification of Functional Regions in Proteins by Surface Mapping of Phylogentic Information, Bioinformatics, Vol. 19, no, 1, 2003 pp163-164.

<http://www.theory-of-evolution.net>