

Chapter 10: Self Replicating Molecules and Systems

This chapter will attempt to quantify the amount of molecular knowledge needed for self replication. Both proteins and RNA will be considered. While many researchers have theorized that one of these molecules emerged as the first self replicator, origin theories stand a much better chance if both are involved. While RNA can perform some of the functions normally performed by proteins, proteins are much more efficient. Amino acids have many functional groups available in their side chains, and these functional groups impart to proteins a versatility that RNA cannot possibly possess. To understand why a system comprised of both is better, consider how numbers and letters are used in the following two sentences.

- The number is 4,900,555,015 dollars.
- The number is four billion nine hundred million five hundred fifty five thousand and fifteen dollars.

Often numbers communicate numerical concepts better than words. The first sentence is much easier to understand. Forcing RNA to do the job of a protein is clumsy. It is analogous to writing out a very large number using words to represent the numbers. Just because it is possible, does not mean that it is the easiest or best way to accomplish the task. RNA is good at storing information. Proteins are good at regulating chemical reactions. The first system of replicating molecules was probably a combination of both, and a good model for such a system is alive and well today in the simplest bacteria. Nevertheless, because chemical evolution does not explain the spontaneous emergence of bacteria from the primordial soup something simpler needs to be considered. The goal of this chapter is to show that something simpler does not work because simple systems cannot self replicate.

A Self Replicating Peptide

In 1996, an article was published in Nature in which David Lee reports to have found a self replicating peptide.¹ The title of the article is appropriately “A Self Replicating Peptide.” Unfortunately, the investigator interference required for self replication is perhaps the most extreme in the history of origins research.

The peptide of interest contains 32 amino acids. The sequence is as follows:

arg-met-lys-gln-lys-glu-glu-lys-val-tyr-glu-lys-lys-ser-lys-val-ala-
cys-leu-glu-tyr-glu-val-ala-arg-leu-lys-lys-leu-val-gly-glu.

The peptide does not self replicate using amino acids. Instead it uses a pool of two peptides, one is 17 amino acids long and the other is 15 amino acids long. The amino acid sequences of these two peptides are shown below. Notice that if a peptide bond forms between ala (last amino acid on right in the peptide with 17 amino acids) and cys (first amino acid on left in the peptide with 15 amino acids) then a replica of the self replicating peptide results.

arg-met-lys-gln-lys-glu-glu-lys-val-tyr-glu-lys-lys-ser-lys-val-ala

cys-leu-glu-tyr-glu-val-ala-arg-leu-lys-lys-leu-val-gly-glu

Because the peptide with 32 amino acids facilitates the formation of this single peptide bond, Lee claims that this peptide can self replicate. But is this really true? To self replicate, this peptide requires a pool of two peptides. One of these peptides has the same amino acid sequence as the first 15 amino acids in the self replicating peptide, and the other has the same amino acid sequence as the next 17 amino acids. Where do these peptides come from? In this case, they are supplied by the investigator.

Chapter 9 discussed the difficulties of creating peptide chains under plausible prebiotic conditions. Due to the difficulties, peptides with more than six amino acids are expected to be very rare chemicals. Peptides composed of 15 to 17 amino acids will be much more scarce. Yet to self replicate, this peptide requires an abundant supply of both, and not just any peptide. One of these peptides must be identical to the first half of the self replicator, and the other peptide must be identical the second half of the self replicator.

This last requirement is particularly troublesome. Suppose the self replicator comes into contact with two random peptide chains. One is 15 amino acids long and the other is 17. How often will the two smaller peptides be an exact replica of the self replicator? Answer 1 time in every 4×10^{41} tries (assuming that every amino acid has a 1 in 20 chance of occurring at each position). Given the low concentration of peptides in the primordial soup, the probability for such an encounter is zero.

The interference does not stop here. It is critical that the first amino acid in the peptide with 15 amino acids be a cysteine. Cysteine has chemical properties that facilitate peptide bond formation, and to make sure that the interference sets the record for the most ever, the alanine (last amino acid on right in the peptide with 17 amino acids) must be chemically altered to make it much more susceptible to attack by cysteine.

Finally, the self replicating peptide contains eight lysines. Lysine is instrumental in its self replication as its charge plays a role in aligning the two small peptides. Lysine is one of the amino acids that has yet to be synthesized under plausible prebiotic conditions. So even if lysine was present in the soup, its concentration would have been negligible.

Every possible strategy of interference is employed by this investigator to promote replication. This mixture of peptides has almost no chance of existing on the primitive earth. Even if it did, as soon as the supply of 15 and 17 amino acid peptides runs out, the replication stops. Despite all of this interference, the claim of self replication is not valid. Self replication involves a system that can duplicate all of its components. In this system, the self replicating peptide is supplied with one peptide containing 15 amino acids and one with 17 amino acids. A true self replicating molecule could generate these two smaller peptides from the amino acids in the primordial soup.

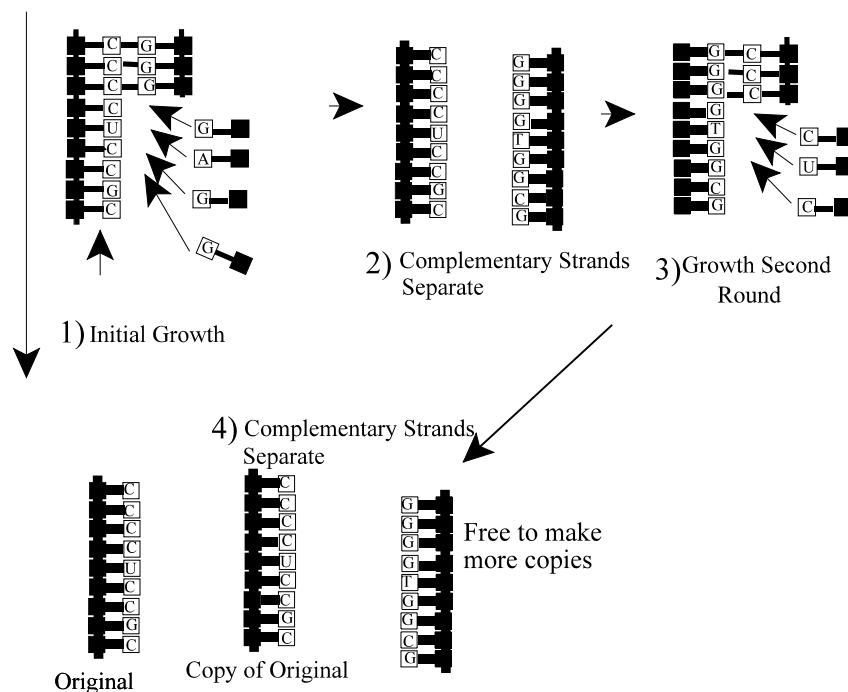
The authors of this paper tried to use dynamite to blow up the door in figure 9.4, but the door withstood the blast and did not open. So the authors just claimed that it opened.

Proteins do not self replicate, and this explains why most scientists rejected the self replicating protein hypothesis in favor of the self replicating RNA hypothesis.

RNA Self Replication

Conceptually RNA should be able to self replicate without the help of proteins. This is shown in figure 10.1. The original strand serves as a template. New base pairs arrive and form weak bonds with their complement. Adenine can form a bond with Uracil, and Guanine can form a bond with Cytosine. After one replication, two complementary strands exist. Another round of replication is necessary to duplicate the original strand. The complement to the original strand is also free to make more copies.

Figure 10.1: Conceptual Model for RNA Self Replication



On paper, this model is great. Nevertheless, it does not work in the lab. The problems were described by Joyce and Orgel as follows:

- Most strands of RNA are unsuitable templates. The original RNA molecule that serves as the template must contain a very high concentration of cytosine to make process 1 in figure 10.1 viable.^{2,3} This situation is unlikely to be met because as discussed earlier cytosine has no plausible prebiotic synthesis pathway and it decays rapidly. Nevertheless, the original strand depicted in figure 10.1 meets the high C requirement.
- The chain will not grow correctly unless a very specific activation agent is used to activate the nucleotides. The activation agent of choice is not ATP (GTP, CTP or UTP). While life uses these, if these activation agents are used without proteins the phosphate groups usually attach to the wrong carbon atom in ribose.^{2,3} ImpA, impG, impU and impC are the activation agents of choice. These activation agents contain the same side group as the amino acid histidine, which is one of the three amino acids that have not been synthesized in prebiotic experiments. Thus, it is unlikely that these activation agents were present in the primordial soup.
- The complement of the original chain will have a high G content. This is inevitable due to the requirement for high C in the original chain. This is problematic because RNA with a high concentration of guanine tends to fold up in such a way that it cannot be an effective template for replication.^{2,3} Thus, the second round of growth in figure 10.1 does not happen.
- If different isomers of ribose are present, these isomers will terminate the growing chain.^{2,3}

Joyce and Orgel comment that “In light of the available evidence, it seems unlikely that a pair of complementary sequences can be found each of which facilitates the synthesis of the other . . . ”³

Just to add to the difficulties, if too many steps form in the replication ladder (complementary bonds between base pairs), then the strands will never separate.⁴ Furthermore, figure 10.1 is oversimplified in that it does not show that in order for the RNA strands to grow, an RNA enzyme is required to catalyze the reaction. Because a growing chain cannot catalyze its own replication, two identical RNA molecules must arise simultaneously in the soup. Each capable of replicating the other.

A pattern is beginning to emerge for the RNA world. The RNA world is a speculative world without proteins where RNA is the most important molecule. RNA regulates all chemical reactions and contains all of the molecular knowledge for life. The pattern that is emerging is that perhaps this world is too speculative in that it may have never existed.

Again Joyce and Orgel put it best: “Scientists interested in the origins of life seem to be divided neatly into two classes. The first, usually but not always molecular biologists, believe that RNA must have been the first replicating molecule and that chemists are exaggerating the difficulties of nucleotide synthesis . . . The second group of scientists are much more pessimistic. They believe that the de nova appearance of oligonucleotides on the primitive earth would have been a near miracle. The author’s subscribe to this latter view. Time will tell which is correct.”³

One last point, RNA replication in the lab makes use of extensive investigator interference. Chemicals like amino acids, aldehydes, and sugars (other than ribose) are arbitrarily excluded. Very specific activation agents are used to encourage replication (ImpA for adenine, ImpG for guanine, ImpC for cytosine, and ImpU for uracil). The concentration of the chemicals (especially cytosine and ribose) is billions and billions of orders of magnitude higher than what one would expect under plausible prebiotic conditions.

Dynamite is being used to blow the door open in figure 9.4, and the door is just too solid. It remains closed and the scientist remains trapped. Fortunately, many scientists understand this, and they no longer claim that the door is open.

How Much Knowledge is Required to Create a Ribozyme

RNA molecules capable of facilitating chemical reactions do exist. Because such RNA molecules perform a role traditionally carried out only by protein enzymes, they are called ribozymes. Ribozymes have been shown to facilitate the creation of both peptide bonds in proteins, and the bonds between phosphate and ribose in RNA. This discovery is very significant in that it means RNA can both store and implement knowledge. It also explains the popularity of RNA as the first living molecule.

Bartel carried out a very relevant experiment. In this experiment, 65 ribozymes were isolated from a pool of 1×10^{15} RNA molecules. All ribozymes isolated contained 200 bases. This result allows for a direct calculation of the knowledge in ribozymes. If 65 sequences have some minimal enzymatic activity out of a pool containing 10^{15} random sequences, then one in every 15 trillion sequences is a ribozyme. Thus the molecular knowledge is as follows: knowledge = $3.32 \times \log(15 \text{ trillion})$ or 44 bits. Note that knowledge and not information is used because the 65 ribozymes were not yet optimized. The experiment also subjected the ribozymes to several rounds of selection in which only the best were chosen. Selection dramatically improved their catalytic efficiency. Thus, Bartel's experiment proves that both information and knowledge can evolve under the guidance of natural selection.

Given the extreme difficulties associated with synthesizing an RNA molecule containing 200 or more bases, it is unlikely that even one such molecule ever existed on the primitive earth, and 15 trillion are needed to just get 65 functional ribozymes. Furthermore, ribozymes are not self replicators. The knowledge required for self replication is certainly many orders of magnitude more than the 44 bits required for a marginally functional ribozyme. Finally, the 44 bits calculated above is in a test tube where all competing side reactions are eliminated. If the real primordial soup contains free amino acids, aldehydes, and undesirable isomers of ribose, then the 44 bits will increase by a factor similar to the increase seen for the protein insulin in chapter 5. Taking this last factor into account, the 44 bits is at least one order of magnitude too small.

Molecular Knowledge in the Primordial Soup

In chapter 5, the difficulties with creating a functional protein in the primordial soup were explored. A similar analysis will now be undertaken for RNA. Because of the scarcity of the RNA subunits (especially ribose and cytosine), the information content of any RNA molecule that evolves in the soup is expected to be very high.

If the soup existed, its exact composition is unknown. Nevertheless, several generalizations are possible. Ribose and cytosine should be extremely rare (see chapter 9). Furthermore, ribose will react with any free amino acids in the soup forming an insoluble polymer. Adenine can be synthesized in the lab, but not under plausible conditions with high yield. Even phosphate will be scarce if inorganic salt is present in the soup.⁷

While the concentration of cytosine and ribose in the soup is probably zero, applying information theory to this situation is not productive because infinite information, implies zero chance for success. So instead this section will make some very favorable assumptions concerning the composition of the soup. The assumptions are not realistic. They are made for educational purposes only.

Favorable Assumptions:

1) All phosphate, sugar and base molecules in the soup exist only as activated nucleotides. That is any adenine in the soup is assumed to be attached to ribose or another sugar. All sugars either have a high energy phosphate group attached or they are attached to some other activating agent.

2) No amino acids are found in the soup. While these are easily synthesized in prebiotic experiments, they must be excluded as they react quickly with ribose and other aldehydes, removing ribose from solution and preventing more ribose from forming. Amino acids and ribose cannot coexist in the soup.

3) No aldehydes exist in the soup. While these are required for the synthesis of ribose and other sugars, they cannot be allowed to persist. Aldehydes react with the four biological bases. These reactions will interfere with the formation of RNA.

Given this starting point, what is the probability that an RNA molecule will emerge from the soup?

- Every time an activated nucleotide attacks a ribose, it has a 50% chance of attacking the wrong carbon atom. This results in premature chain termination.^{2,3}
- Half of the ribose present is the wrong isomer, this also results in premature chain termination.^{2,3}

- 3/4 of the bases attached to the ribose are not biological. That is adenine, guanine, cytosine, and uracil are only used in 1/4 of the activated nucleotides. The most common base is likely ammonia or some other simple amine.
- 3/4 of the activated nucleotides use a sugar other than ribose or deoxyribose. This also results in premature chain termination. Given that ribose is usually only a minor product in any prebiotic experiment that synthesizes simple sugars, this is a very generous assumption.

Even with these most favorable assumptions that ignore all competing side reactions, every nucleotide added to the RNA chain still contributes a minimum of 6 bits of primordial information (for every 64 nucleotides added to the chain, only 1 is expected to be biologically relevant, and this corresponds to $3.32 \times \log(64/1) = 6$ bits of primordial information). This is three times the value calculated for amino acids in chapter 5.

Thus, a 200 base pair random RNA sequence contains $6 \times 200 = 1200$ bits of primordial information, and as explained in chapter 5, this information can be related to a probability because it is a form of knowledge - the knowledge to exclude chemicals found in the soup that are not used by life today.

Thus, a 200 base pair random RNA sequence has a 1 in 2^{1200} chance of emerging in the primordial soup. Given that only 65 out of 15 trillion will exhibit any ribozyme functionality, the odds are staggering - 1 time in 3.9×10^{372} tries. Furthermore, this calculation is only for a ribozyme capable of regulating a simple chemical reaction. The odds of a self replicating ribozyme emerging are certainly much smaller.

In summary, the probability of creating a 200 base ribozyme is extremely small because so few random sequences contain the required knowledge, but given that no 200 base RNA molecules existed on the primitive earth, the odds are no longer almost zero, but instead almost zero multiplied by zero.

Finally, as noted in chapter 5, using information theory to calculate the odds has some drawbacks. Information theory only takes into account the concentration of the various chemicals. It does not have the ability to deal with chemical properties that may make certain reactions more probably, and this can skew the results in favor of evolution or against it. In the case of RNA, a very strong argument can be made that the skewing is strongly in favor of evolution. This is because the above calculation excluded amino acids and aldehydes from the soup. Thus, the information calculated above represents RNA that evolves in a test tube, not the real world.

Self Replication and Perpetual Motion

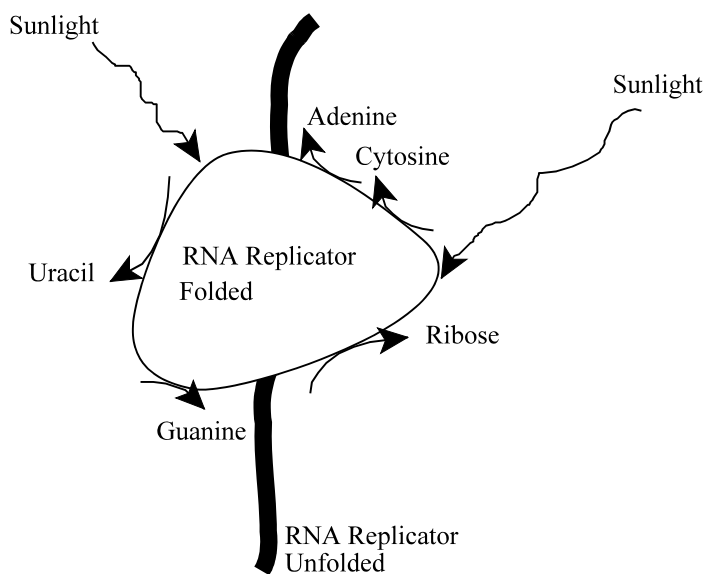
Researchers today are actively seeking and finding new ribozymes. Many are artificially engineered and others arise from random sequences. Many of these researchers believe that in time they will find a self replicating RNA molecule. Others like Joyce and Orgel who are at the forefront of the research disagree.

In chapter 7, several techniques used by life to circumvent the second law of thermodynamics were discussed. Unless a self replicating RNA molecule has the capability to implement some of these same techniques, its existence can be ruled out on purely theoretical grounds.

Based on fundamental laws of physics, science can state with certainty that if a self replicating RNA molecule is found, the molecule will only be able to replicate in a test tube. It will require a continuous supply of activated nucleotides to drive its replication. While this might work in the test tube, it will certainly not work in the primordial soup. Activated nucleotides in the soup will not last for more than a few days. Given that their decay will dominate any conceivable path for prebiotic synthesis, the soup will only contain at most a very dilute supply of activated nucleotides.

Given the difficulty associated with the prebiotic synthesis of ribose, adenine, and cytosine, the concentration of these critical molecules in the soup will also be extremely low. This means that the first successful self replicating RNA molecule must be able to direct the synthesis of adenine, cytosine, ribose, uracil and guanine. If it cannot do this, it will not be able to replicate in the soup. Furthermore, it must be able to activate all of the nucleotides. So this special RNA molecule must know how to tap a plentiful energy source and use it to drive many different chemical reactions. If it cannot perform all of these functions, then it is a perpetual motion machine, and its very existence is limited to biology textbooks.

Figure 10.2: A Self Replicating RNA Molecule



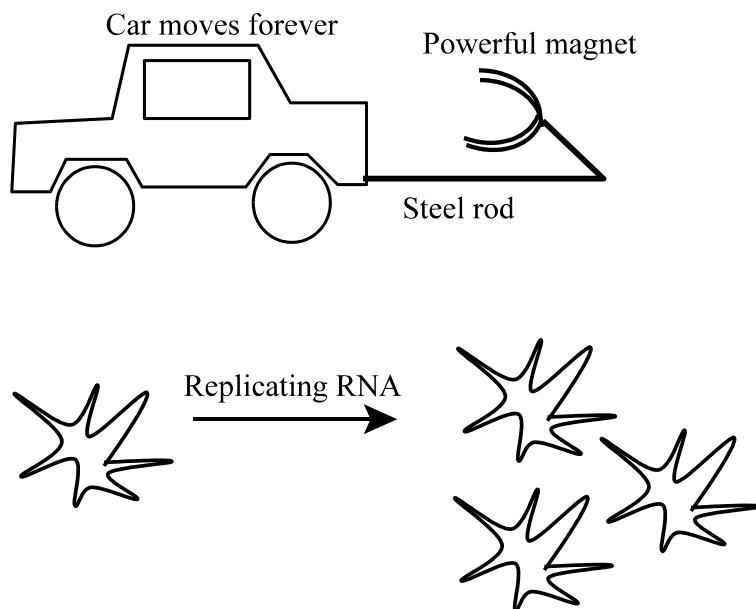
In figure 10.2, the RNA molecule can exist in two states, folded and unfolded. When folded, it catalyzed RNA replication, and the formation of adenine, ribose, cytosine, uracil, and guanine. It also must create activated nucleotides (not shown). When unfolded, it serves as a template for replication. The folded version must also know how to replicate the unfolded version.

This particular ribozyme taps into sunlight as an energy source using a primitive form of photosynthesis. Other self replicating RNA molecules could potentially oxidize a chemical like methane, hydrogen, or sulfur to generate the required energy.

Figure 10.2 is what is required of a “living molecule.” Anything less is not alive. This figure was constructed with due consideration to the second law. Any RNA molecule that does not possess all of the capabilities shown in figure 10.2 is a perpetual motion machine. It may replicate in the lab as long as it is supplied with activated nucleotides, but it will not replicate in the soup. Thus, it only exists in textbooks, and there is no need to wait to see if researchers can locate it.

Inventors have been trying to invent perpetual motion machines for at least 2000 years. They have all failed. Nevertheless, many have been issued patents by various governments throughout the world. Two examples of perpetual motion are shown in figure 10.3. Both examples are equally absurd. While many scientists apparently only recognize the absurdity of the first picture, nature can recognize both, and it does not allow either to exist.

Figure 10.3: Perpetual Motion Machines



The first picture in figure 10.3 is a clear violation of energy conservation. It does not work because the force that the magnet exerts on the car is exactly cancelled by the force that the car exerts on the magnet. The magnet does not cause the car to move. The second violation is more subtle only because it violates a different law of nature. When a self replicating molecule replicates, the replication decreases the entropy of the universe. The second law is violated. To get around this problem, any real self replicator must know how and be able to couple its replication to a plentiful energy source. If it is unable to do this, then it is a special type of perpetual motion machine, and it only exists on paper and in the imagination of researchers.

References:

- 1) Lee, Granja, Martinez, Severin, Ghadiri, "A Self Replicating Peptide," Letters to Nature, 382:525-528, 1996.
- 2) Joyce, Visser, Boeckel, Boom, Orgel, Westrenen, "Chiral Selection in Poly (C) Directed Synthesis of Oligo (G)," Letters to Nature, 310: 602-604, 1984.
- 3) Joyce and Orgel, The RNA World, Gesteland, Cech, Atkins, Cold Spring Harbor, "Origin of the RNA World," 1999.
- 4) Bartel, The RNA World, Gesteland, Cech, Atkins, Cold Spring Harbor, "Recreating an RNA Replicase," 1999.
- 5) Eklund, Szostak, Bartel, "Structurally Complex and Highly Active RNA Ligase Derived from Random RNA Sequences," Science, 1995.
- 6) Bartel and Szostak, "Isolation of New Ribozymes from a Large Pool of Random Sequences," Science, 261:1411-1418, 1993.
- 7) Thaxton, Bradley, Olsen, The Mystery of Life's Origin: Reassessing Current Theories, Philosophical Libraries, 1984.
- 8) Orgel, Self-organizing Biochemical Cycles, Salk Institute of Biological Studies, 99:12503-12507, 2000.
- 9) Green, Szostak, "Selection of Ribozyme that Functions as a Superior Template in Self Copying Reaction," Science, 258:1910-1915, 1992.

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