

Lecture 2a – Origin of Life and the transition from the RNA world to the DNA world

Eigen's theory is very useful in understanding the origin of life. The theory essentially shows that a selfreplicating molecule must be shorter (in terms of base pairs) than the reciprocal of the error rate for copying each base.

It is thought that the first self-replicating molecule was an RNA (or perhaps an RNA-like molecule). We will not discuss the important question of how the 4 bases that make up the RNA polymer came into existence, we will simply assume that they were reasonably abundant on the early earth, possibly at volcanic vents at the bottom of the oceans, possibly in ponds on land.

We know that these (and DNA) can undergo template-directed polymerization. Since the sequence of bases on the template directs the sequence of bases on the new polymer (by Crick-Watson base pairing), the new polymer is, in principle, copy of the old (actually of course it is a complementary copy). However, such selfreplication is extremely inefficient (for example, because the new and old copies have to completely separate before further selfreplication can occur), and also it is essentially random (because the initial copy has to form in the absence of a template).

So we think that "in the beginning" there must have been a catalyst for selfreplication. Possibly at first this catalyst was some type of inorganic molecule, but whatever it was, it would equally enhance the selfreplication of ALL sequences, and if $s = 1$ Darwinian evolution is not possible.

However (unlike DNA) RNA molecules can fold up on themselves (by internal Crick-Watson base pairing) to form complex 3-dimensional shapes that can function as catalysts. We call such RNA-catalysts ribozymes (compare to protein-catalysts which are called enzymes). So we think that after an initial phase of purely random synthesis of RNA, an RNA-sequence appeared that could fold up and act as a catalyst of RNA replication (either of synthesis or of separation or both).

So the critical question in the origin of life is, is this scenario plausible (we will probably never know the details)?

In other words, (1) could a sequence that can fold to form a selfreplicase have formed spontaneously (by random chemical processes) and (2) if it did arise could it catalyse its own replication with sufficient accuracy to commence Darwinian evolution?

The first question is merely one of time and opportunity, but the second question is the key to the origin of life. It is essentially the following question: are there RNA sequences long enough to fold to form replicases, but short enough to undergo Darwinian evolution (i.e. with length less than the reciprocal of the error rate with which they operate)?

This question has recently been addressed by David Bartel and colleagues (Science vol 292, p1319, 2001). They started with a ribozyme that, while not template-directed, had some ability to link bases together (a ligase) by extending an RNA primer. They added random tails to this ribozyme, and then enriched the contents of the starting mixture of random ligases using methods similar to PCR, ending up with a ribozyme that could direct polymerization of any RNA sequence terminating in the appropriate primer sequence. Eventually they were able to isolate a replicase that was only 165 bases long and that functioned with 98.5 % accuracy. From the Eigen theory, we know that the first replicase should only be $1/0.015 = 67$ bases long.

So it seems plausible that an even shorter replicase might exist, but there is as yet no direct proof. If one could screen all possible 67-base sequences and show that not one of them can function as a replicase with greater than 98.5% accuracy, one would have to conclude (provisionally) that life did NOT originate in this manner. The Bartels experiments only tested a very small subset of possible sequences, and yet they came very close to this “jackpot” sequence”, making it rather likely that this scenario is in fact correct.

[In addition, it should be noted that to undergo Darwinian evolution the replicase must be a general replicase, able to catalyse the synthesis of any RNA sequence; this condition was met in the Bartels experiments; if the replicase is general, for Darwinian evolution to occur there is the further requirement that synthesis take place in small isolated compartments, called “protocells”, otherwise the synthesis of all possible sequences would be equally enhanced by the replicase, eliminating Darwinian competition.]

From RNA to DNA/protein

Although RNA can function both as an information carrier (template for replication) and as a catalyst (replicase), it is not very good at either. This means that both e and $\ln s$ are unfavorable – a lot of errors are made in copying RNA (even using modern RNA polymerases) and copying is not very efficient. RNA is a jack of both trades (storing information and make copies of that information). Life as we know it started with a gradual shift from the limited RNA-world (where polynucleotides were necessarily very short and unable to catalyse complex functions) to the modern DNA/protein world).

The accuracy of DNA replication (catalysed by polymerases and aided by proofreading) is very high (error rates on the order of 10^{-9}).

Furthermore, the DNA does not itself act as a catalyst, but rather it encodes the information necessary to make protein-based catalysts. Indeed, one can regard every protein that an organism’s DNA encodes as promoting the reproduction of that organism (and if it did not, on average, aid reproduction, it would have been modified or eliminated).

This means that the polynucleotides of modern organisms can encode enormously complex sets of proteins, which are extraordinarily efficient in aiding the reproduction of modern organisms). Of course “modern” here means since the RNA world, i.e. 3.5 billion years ago. Also, the modern world is in many ways far

less conducive to selfreplication than was the early earth (we have to capture our own food, for example, unlike the first protocells, which just used nucleotides that were laying around.).

Each living (and every organism that has ever lived) organism is a sophisticated machine for making copies of its DNA: crows make copies of crow DNA, blue jays make copies of bluejay DNA etc. Only humans may be partly exempt from this iron logic, because of our large brains.