Lies, Damned Lies, Statistics, and Probability of Abiogenesis Calculations

[Last Update: December 21, 1998]

Contents

Introduction

Every so often someone comes up with the statement "the formation of any enzyme by chance is nearly impossible, therefore abiogensis is impossible". Often they cite an impressive looking calculation from the astrophysicist Fred Hoyle, and trot out something called "Borel's Law" to prove that life is statistically impossible. These people, including Fred, have committed one or more of the following errors.

Problems with the creationist "it's so improbable" calculations

1) They calculate the probability of the formation of a "modern" protein, or even a complete bacterium with all "modern" proteins, by random events. This is not the abiogenesis theory at all.

2) They assume that there is a fixed number of proteins, with fixed sequences for each protein, that are required for life.

3) They calculate the probability of sequential trials, rather than simultaneous trials.

4) They misunderstand what is meant by a probability calculation.

5) They seriously underestimate the number of functional enzymes/ribozymes present in a group of random sequences.

I will try and walk people through these various errors, and show why its not possible to do a "probability of abiogenesis" calculation in any meaningful way.

Glossary

Acyl transferase:
An enzyme or ribozyme that synthesizes peptides.

Ligase:
An enzyme or ribozyme that adds a monomer to a polymer, or links two shorter polymers together.

Monomer:
Any single subunit of a polymer. An amino acid is a monomer of a peptide or protein, a nucleotide is a monomer of an oligonucleotide or polynucleotide.

Nucleotide:
Adenine, Guanine, Cytosine and Uracil. These are the monomers that make up oligo or polynucleotides such as RNA.

Oligonucleotide:

A primordial protoplasmic globule

So the calculation goes that the probability of forming a given 300 amino acid long protein (say an enzyme like carboxypeptidase) randomly is $(1/20)^{300}$ or 1 chance in 2.04 x 10^{390}, which is astoundingly, mind-beggaringly improbable. This is then cranked up by adding on the probabilities of generating 400 or so similar enzymes until a figure is

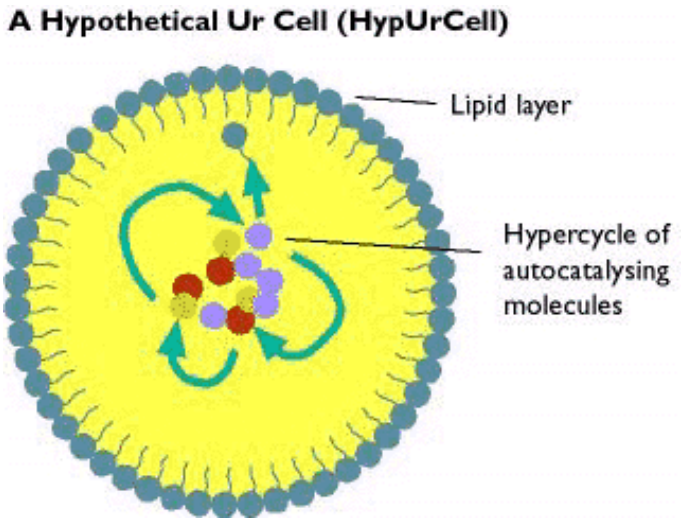
reached that is so huge that merely contemplating it causes your brain to dribble out your ears. This gives the impression that the formation of even the smallest organism seems totally impossible. However, this is completely incorrect.

A short polymer of nucleotide subunits.

Polymerase:
A enzyme or ribozyme that makes a polymer out of monomers. For example RNA polymerase makes RNA out of single nucleotides.

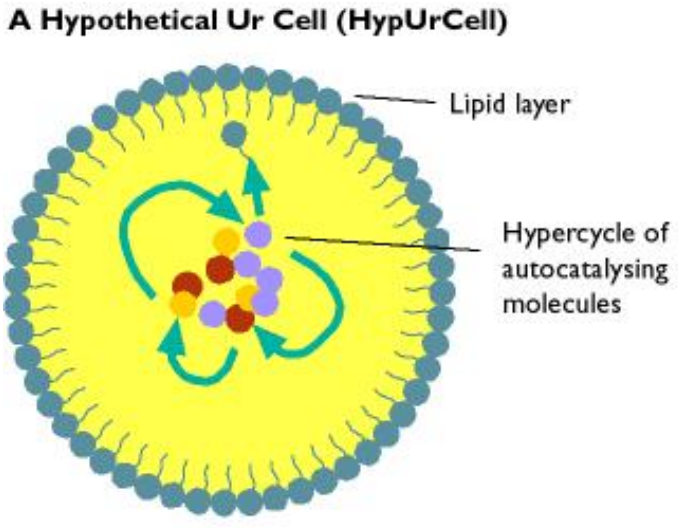
Ribozyme:
A biological catalyst made from RNA.

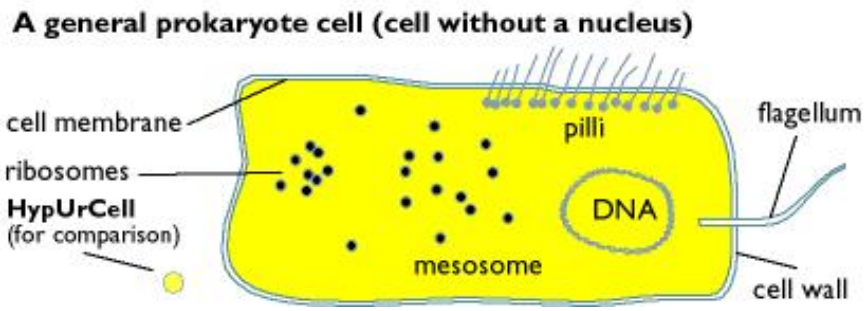
Self-replicator:
A molecule which can make an identical, or near identical copy of itself from smaller subunits. At least 4 self-replicators are known.



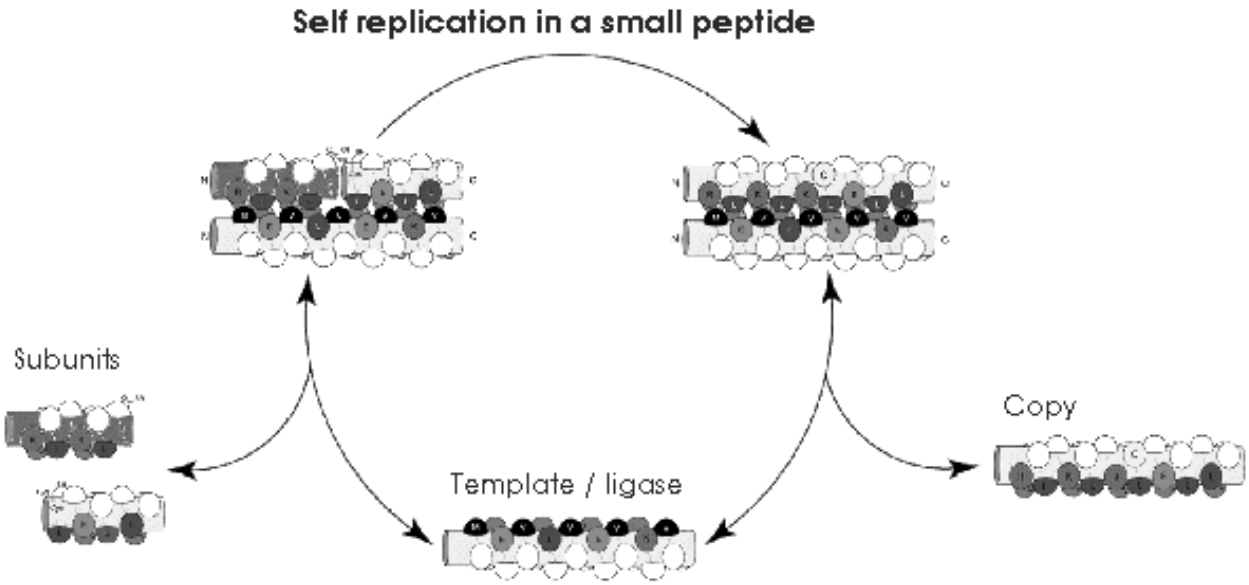
Firstly, the formation of biological polymers from monomers is a function of the laws of chemistry and biochemistry, and these are decidedly *not* random.

Secondly, the entire premise is incorrect to start off with, because in modern abiogenesis theories the first *"living things"* would be much simpler, not even a protobacteria, or a preprobacteria (what Oparin called a protobiont [8] and Woese calls a progenote [4]), but one or more simple molecules probably not more than 30-40 subunits long. These simple molecules then slowly evolved into more co-operative self replicating systems, then finally into simple organisms [2,5,10,15,28]. An illustration comparing a hypothetical probiont and a modern bacteria is given below.





The first *"living things"* could have been a *single* self replicating molecule, similar to the "self-replicating" peptide from the Ghadiri group [7,17], or the self replicating hexanucleotide[10], or possibly an RNA polymerase that acts on itself [12].

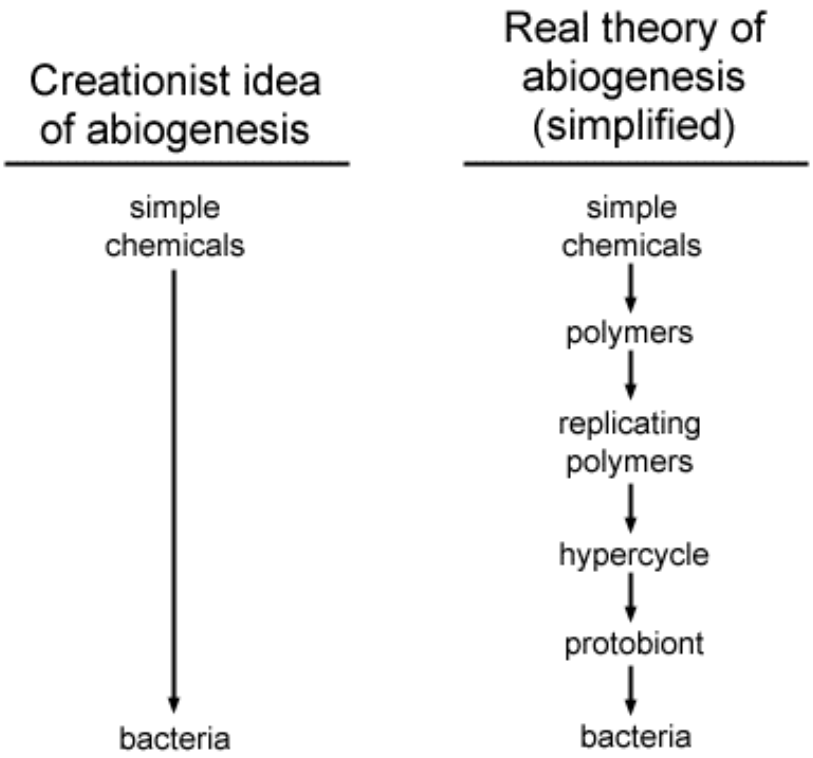


Self replication in the Lee et al peptide, two subunits bind to the original peptide then are ligated and a new copy released.
modified from Lee et al, Curr Opinion Chem Biol, 1, 491-496, 1997

Another view is the first self-replicators were groups of catalysts, either protein enzymes or RNA ribozymes, that regenerated themselves as a catalytic cycle [3,5,15,26,28]. An example is the SunY three subunit self-replicator [24]. These catalytic cycles could be limited in a small pond or lagoon, or be a catalytic complex adsorbed to either clay or lipid material on clay. Given that there are many catalytic sequences in a group of random peptides or polynucleotides (see below) it's not unlikely that a small catalytic complex could be formed.

These two models are not mutually exclusive. The Ghadiri peptide can mutate and form catalytic cylces [9].

No matter whether the first self-replicators were single molecules, or complexes of small molecules, this model is nothing like Hoyle's "tornado in a junkyard making a 747". Just to hammer this home, here is a simple comparison of the theory criticised by creationists, and the actual theory of abiogenesis.



Note that the real theory has a number of small steps, and in fact I've left out some steps (especially between the

hypercycle-protobiont stage) for simplicity. Each step is associated with a small increase in organisation and complexity, and the chemicals slowly climb towards organism-hood, rather than making one big leap [4,10,15,28].

Where the creationist idea that modern organisms form spontaneously comes from is not certain, the first modern abiogenesis formulation, the Oparin/Haldane hypothesis from the 20's, starts with simple proteins/proteinoids developing slowly into cells. Even the ideas circulating in the 1850's were not "spontaneous" theories. The nearest I can come to is *Lamarks* original ideas from 1803![8]

Given that the creationists are criticising a theory over 150 years out of date, and held by no modern evolutionary biologist, why go further? Because there are some fundamental problems in statistics and biochemistry that turn up in these mistaken "refutations".

The myth of the "Life sequence"

Another claim often heard is that there is a "Life sequence" of 400 proteins, and that the amino acid sequences of these proteins cannot be changed, for organisms to be alive.

This however is nonsense. The 400 protein claim seems to come from the protein coding genome of *Mycobacterium genetalium*, which has the smallest genome currently known of any modern organism [20]. However, inspection of the genome suggests that this could be reduced further to a minimal gene set of 256 proteins [20]. Note again that this is a *modern* organism. The first protobiont/progenote would have been smaller still [4], and preceded by even simpler chemical systems [3,10,11,15].

As to the claim that the sequences of proteins cannot be changed, again this is nonsense. There are in most proteins regions where almost any amino acid can be substituted, and other regions where conservative substitutions (where charged amino acids can be swapped with other charged amino acids, neutral for other neutral amino acids and hydrophobic amino acids for other hydrophobic amino acids). Some functionally equivalent molecules can have between 30 - 50% of their amino acids different. In fact it is possible to substitute structurally non-identical bacterial proteins for yeast proteins, and worm proteins for human proteins, and the organisms live quite happily.

The "Life Sequence" is a myth.

Coin tossing for beginners and macromolecular assembly

So lets play the creationist game and look at forming a peptide by random addition of amino acids. This certainly is not the way peptides formed on the early Earth, but it will be instructive.

I will use as an example the "self-replicating" peptide from the Ghadiri group mentioned above [7]. I could use other examples, such as the hexanucleotide self-replicator [10], the SunY self-replicator [24] or the RNA polymerase described by the Eckland group [12], but for historical continuity with creationist claims a small peptide is ideal. This peptide is 32 amino acid long with a sequence of RMKQLEEKVYELLSKVACLEYEVARLKKVGE and is an enzyme, a peptide ligase that makes a copy of itself from two 16 amino acid long subunits. It is also of a size and composition that is ideally suited to be formed by abiotic peptide synthesis. The fact that it is a self replicator is an added irony.

The probablilty of generating this in successive random trials is $(1/20)^{32}$ or 1 chance in 4.29×10^{40} this is much, much more probable than the 1 in 2.04×10^{390} of the standard creationist "generating carboxypeptidase by chance" scenario, but still seems absurdly low.

However, there is another side to these probability estimates, and it hinges on the fact that most of us don't have a feeling for statistics. When some one tells us that some event has a one in a million chance of occuring, many of us expect that 1 million trials must be undergone *before* the said event turns up, but this is wrong.

Here is a experiment you can do your self, take a coin, flip it 4 times, write down the results, do it again. How many times would you think you had to repeat this procedure (trial) before you get 4 heads in a row?

Now the probability of 4 heads in a row is is $(1/2)^4$ or 1 chance in 16, do we have to do 16 trials to get 4 heads (HHHH)? No, in successive experiments I got, 11, 10, 6, 16, 1, 5, and 3 trials before HHHH turned up. The figure 1 in 16 (or 1 in a million or 1 in 10^{40}) gives the likelihood of an event in a given trial, but doesn't say *where* it will occur in a series. You can flip HHHH on your very *first* trial (I did). Even at 1 chance in 4.29×10^{40}, a self-replicator could have turned up suprisingly early. But there is more.

1 chance in 4.29×10^{40} is still orgulously, gobsmackingly unlikely, it's hard to cope with this number. Even with the argument above (you could get it on your very first trial) most people would say "surely it would still take more time than earth existed to make this replicator by random methods". Not really, in the above examples we were examining sequential trials, as if there was only one protein/DNA/proto-replicator being assembled per trial. In fact there would be billions of *simultaneous* trials as the billions of building block molecules interacted in the oceans, or on the thousands of kilometers of shorelines that could provide catalytic surfaces or templates. [2,15]

Let's go back to our example with the coins. Let's say it takes a minute to toss the coins 4 times, to generate HHHH would take on average 8 minutes. Now get 16 friends, each with a coin, you all flip the coin simultaneously 4 times, the average time to generate HHHH is now 1 minute. Now try to flip 6 heads in a row, has a probability of $(1/2)^6$ or 1 in 64, this would take half an hour on average, go out and recruit 64 people, and you can flip it in a minute. If you want to flip a sequence with a chance of 1 in a billion, just recruit the population of china to flip coins for you, you will have that sequence in no time flat.

So, if on our prebiotic earth we have a billion peptides growing simultaneously, that reduces the time taken to generate our replicator significantly.

Okay, you are looking at that number again, 1 chance in 4.29×10^{40}, thats a *big* number, and although a billion starting molecules is a lot of molecules, could we ever get enough molecules to randomly assemble our first replicator in under half a billion years?

Yes, *1 kilo* of the amino acid arginine has 2.85×10^{24} molecules in it (that's well over a billion billion), a tonne of arginine has 2.85×10^{27} molecules. If you took a semi-trailer load of each amino acid and dumped it into a medium size lake you would have enough molecules to generate our particular replicator in a few tens of years, given that you can make 55 amino acid long proteins in 1 to 2 weeks.[14,16]

So how does this shape up with the prebiotic Earth? On the early Earth it is likely that the ocean had a volume of 1×10^{24} litres. Given an amino acid concentration of 1×10^{-6} M (a moderately dilute soup, see Chyba and Sagan 1992[23]), then there is roughly 1×10^{50} potential starting chains, so that a fair number of efficent peptide ligases (about 1×10^{31}) could be produced in a under a *year* let alone a million years, the synthesis of primitive self-replicators could happen relatively rapidly, even given a probability of 1 chance in 4.29×10^{40}. (and remember, our replicator could be sythesized on the very first trial).

Assuming that it takes a week to generate a sequence [14,16], Then the Ghadiri ligase could be generated in one week, and any cytochrome C sequence could be generated in a bit over a million years (along with about half of all possible 101 peptide sequences, a large proportion of which will be functional protiens of some sort).

Although I have used the Ghadiri ligase as an example, as I mentioned above the same calculations can be performed for the SunY self replicator, or the Ekland RNA polymerase. I leave this as an exercise for the reader, but the general conclusion (you can make scads of the things in a short time) is the same for these oligonucleotides.

Search spaces, or how many needles in the haystack?

So I've shown that generating a *given* small enzyme is not as mind bogglingly difficult as creationists (and Fred Hoyle) suggest. Another misunderstanding is that most people feel that the number of enzymes/ribozymes, let alone the ribozymal RNA polymerases or any form of self-replicator, represent a very unlikely configuration and that the chance of a single enzyme/ribozyme forming, let alone a number of them, from random addition of amino acids/nucleotides is mind-beggaringly small.

However, an analysis by Ekland suggests that in the sequence space of 220 nucleotide long RNA sequences, a staggering 2.5×10^{112} sequences are efficent ligases [12]. Not bad for a compound previously thought to be only structural. Going back to our primitive ocean of 1×10^{24} litres and assuming a nucleotide concentration of 1×10^{-7} M (23), then there is roughly 1×10^{49} potential nucleotide chains, so that a fair number of efficent RNA ligases (about 1×10^{34}) could be produced in a *year* let alone a million years. The potential number of RNA polymerases is high also, about 1 in every 10^{20} sequences is an RNA polymerase [12]. Similar considerations apply for ribosomal acyl transferases, (about 1 in every 10^{15} sequences), and ribozymal nucleotide synthesis [1,6,13].

Similarly, of the 1×10^{130} possible 100 unit proteins, 3.8×10^{61} represent cytochrome C alone!![29]. There's lots of functional enyzmes in the peptide/nucleotide search space, so it would seem likely that a functioning ensemble of enzymes could be brewed up in an early Earths prebiotic soup.

So, even with more realistic (if somewhat mind beggaring) figures, random assemblage of amino acids into "life-supporting" systems (whether you go for protein enzyme based hypercycles [10], RNA world systems [18], RNA ribozyme-protein enzyme coevolution [11,25]) would seem to be entirely feasible, even with pessimistic figures for the original monomer concentrations [23] and synthesis times.

Conclusions

The very premise of creationists' probability calculations is incorrect in the first place as it aims at the wrong theory. Furthermore, this argument is often buttressed with statistical and biological fallacies.

At the moment, we have no idea how probable life is, it's virtually impossible to assign any meaningful probabilities to any of the steps to life except the first two (monomers to polymers p=1.0, formation of catalytic polymers p=1.0). For the replicating polymers to hypercycle transition, the probability may well be 1.0 if Kauffman is right about catalytic closure and his phase transition models, but this requires real chemistry and more detailed modelling to confirm. For the hypercycle->protobiont transition, the probability here is dependent on theoretical concepts still being developed, and is unknown.

However, in the end life's feasibility depends on chemistry and biochemistry that we are still studying, not coin flipping.

References

[1] Unrau PJ, and Bartel DP, RNA-catalysed nucleotide synthesis. Nature, **395**: 260-3, 1998

[2] Orgel LE, Polymerization on the rocks: theoretical introduction. Orig Life Evol Biosph, **28**: 227-34, 1998

[3] Otsuka J and Nozawa Y. Self-reproducing system can behave as maxwell's demon: theoretical illustration under prebiotic conditions. J Theor Biol, 194, 205-221, 1998

[4] Woese, C, The universal ancestor, Proc Natl Acad Sci USA, 95: 6854-6859.

[5] Varetto L, Studying artificial life with a molecular automaton. J Theor Biol, 193: 257-85, 1998

[6] Wiegand TW, Janssen RC, and Eaton BE, Selection of RNA amide synthases. Chem Biol, 4: 675-83, 1997

[7] Severin K, Lee DH, Kennan AJ, and Ghadiri MR, A synthetic peptide ligase. Nature, 389: 706-9, 1997

[8] Ruse M, The origin of life, philosophical perspectives. J Theor Biol, 187: 473-482.

[9] Lee DH, Severin K, Yokobayashi Y, and Ghadiri MR, Emergence of symbiosis in peptide self-replication through a hypercyclic network. Nature, 390: 591-4, 1997

[10] Lee DH, Severin K, and Ghadri MR. Autocatalytic networks: the transition from molecular self-replication to molecular ecosystems. Curr Opinion Chem Biol, 1, 491-496, 1997

[11] Di Giulio M, On the RNA world: evidence in favor of an early ribonucleopeptide world. J Mol Evol, 45: 571-8, 1997

[12] Ekland EH, and Bartel DP, RNA-catalysed RNA polymerization using nucleoside triphosphates. Nature, 383: 192, 1996

[13] Lohse PA, and Szostak JW, Ribozyme-catalysed amino-acid transfer reactions. Nature, 381: 442-4, 1996

[14] Ferris JP, Hill AR Jr, Liu R, and Orgel LE, Synthesis of long prebiotic oligomers on mineral surfaces [see comments]. Nature, 381: 59-61, 1996

[15] Lazcano A, and Miller SL, The origin and early evolution of life: prebiotic chemistry, the pre- RNA world, and time. Cell, 85: 793-8, 1996

[16] Ertem G, and Ferris JP, Synthesis of RNA oligomers on heterogeneous templates. Nature, 379: 238-40, 1996

[17] Lee DH, Granja JR, Martinez JA, Severin K, and Ghadri MR, A self-replicating peptide. Nature, 382: 525-8, 1996

[18] Joyce GF, Building the RNA world. Ribozymes. Curr Biol, 6: 965-7, 1996

[19] Ishizaka M, Ohshima Y, and Tani T, Isolation of active ribozymes from an RNA pool of random sequences using an anchored substrate RNA. Biochem Biophys Res Commun, 214: 403-9, 1995

[20] Mushegian AR and Koonin, EV, A minimal gene set for cellular life derived by comparison of complete bacterial genomes. Proc. Natl. Acad. Sci. USA, 93: 102688-102773.

[21] Ekland EH, Szostak JW, and Bartel DP, Structurally complex and highly active RNA ligases derived from random RNA sequences. Science, 269: 364-70, 1995

[22] Breaker RR, and Joyce GF, Emergence of a replicating species from an in vitro RNA evolution reaction. Proc Natl Acad Sci U S A, 91: 6093-7, 1994

[23] Chyba C, and Sagan C, Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. Nature, 355: -32., 1992

[24] Doudna JA, Couture S, and Szostak JW, A multisubunit ribozyme that is a catalyst of and template for complementary strand RNA synthesis. Science, 251: 1605-8, 1991

[25] Lahav N, Prebiotic co-evolution of self-replication and translation or RNA world? J Theor Biol, 151: 531-9, 1991

[26] Stadler PF, Dynamics of autocatalytic reaction networks. IV: Inhomogeneous replicator networks. Biosystems, 26: 1-19, 1991

[27] Eigen M, Gardiner W, Schuster P, and Winkler-Oswatitsch R, The origin of genetic information. Sci Am, 244: 88-92, 96, et passim, 1981

[28] Eigen M, and Schuster P, The hypercycle. A principle of natural self-organization. Springer-Verlag, isbn 3-540-09293,

1979

[29] Yockey HP, On the information content of cytochrome c. J Theor Biol, 67: 345-76, 1977

Useful books

Statistics at Square One, T.D.V. Swinscow, 8th Edition Paperback, Published by Amer College of Physicians, 1983, ISBN: 0727901753

Evolution from Space, F Hoyle and Wickramasinghe, JM Dent and sons, London, 1981

Vital Dust : Life As a Cosmic Imperative, by Christian De Duve, Basic Books 1995, ISBN: 0465090451

The Major Transitions in Evolution, Maynard Smith J & Szathmary E, 1995, WH Freeman, ISBN: 0716745259

The Origins of Order: Self Organization and Selection in Evolution. By Stuart Kauffman, S. A. (1993) Oxford University Press, NY, ISBN: 0195079515.

At Home in the Universe. By Stuart Kauffman, 1995) Oxford University Press, NY.

Links

- Creation Column: Evolutionary Improbabilities. A creationist page which uses Hoyles calculation.
- Chandra Wickramasinghe's Testimony in Arkansas, 1981. transcribed by Brig Klyce.
- The Ghadiri home page. Go to publications and download the Current Opinions PDF for a more detailed explanation.
- A plain language description of the properties of the Lee peptide.
- Another description, with comments by Stuart Kauffman.
- Some other self-replicating molecules
- An overview of the RNA world, with comments on hypercycles, several of the suggested experiments in the conclusion have been performed, with results compatible with the RNA world.
- An American Scientist article the on the origin of life by C. de Duve. This account was written before the ribozymal polymerases were described, and a number of other issues resolved so is slightly more pessimistic than needs be.
- A discovery article on Deamers work on protocells. From the Discover site, go to the Archives, search on November 1995 and click on the First Cell link.

Acknowlegements

Thanks to John Wilkins and Jthomford for helpful suggestions and discussions. Thanks to John for some nifty GIFs and JPEGS as well.

Home | Browse | Search | Feedback | Links
The FAQ | Must-Read Files | Index | Creationism | Evolution | Age of the Earth | Flood Geology | Catastrophism | Debates