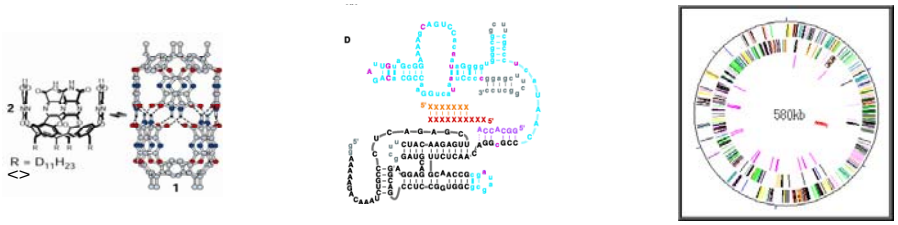
DRAFT

Threshold for Life

The Threshold Energy and Complexity for Self-Replicating Systems

Introduction

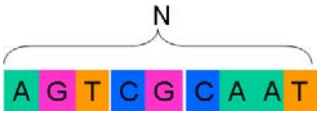
There is currently great interest in creating systems which have a property which is routine in biology but which we as engineers have yet to master, namely self-replication or more simply life. [Venter, Bartels, Reebeck,Griffiths, Lipson]. In this manuscript we wish to identify a number of fundamental attributes of such self-replicating systems and in particular point out thresholds in energy consumption and complexity required by such systems.



Minimal Self Replicating Systems (a. Autocatalytic chemistry b. Ribozymal polymerase c. Minimal organism)

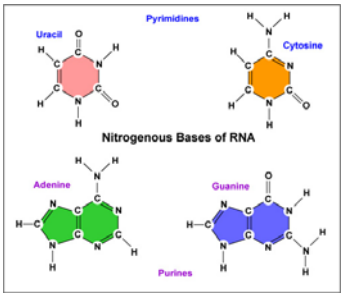
Biology and chemistry offer an excellent test bed for ideas in self replication. Investigation of such systems exists at three separate levels, the organism in which the goal is to find the minimal gene set which is fully self-replicating [Venter], information rich self replication in which a string of DNA which carries at least some amount of arbitrary code may be self replicated by the molecule itself [Bartels] and information poor self-replication or self catalyzing chemistries in which the input parts are complex but the the number of possible outputs is small [Reebeck].

I] Definition of Information Rich (Non-Trivial) Self replication:



Define the complexity for an *N*-mer or *M* Types as : $C_{string} = \ln M^N$

The building blaocks themselves (e.g. nucleotides) can be viewed as strings themselves



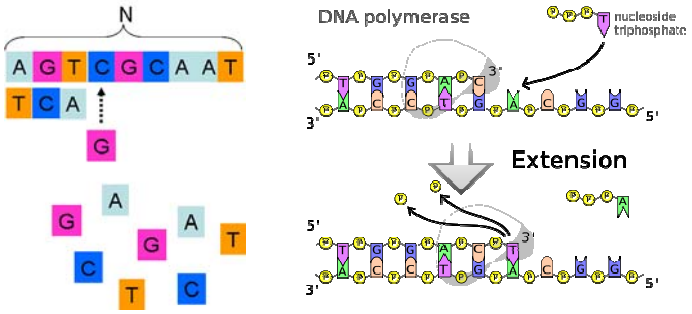
In which their complexity can be calculated as: $C_{string} = \ln M^N$. As an example: RNA. The nucleotide pyrimidine or purine bases have a configuration space of ~ 8 and 10 atoms , configured in either 1 or 2 rings respectively. If we constrain ourselves to substituent atoms of C,N or O then the complexity for the nucleotide bases is ~ 10 ln 3 ~ 11

The definition of Information Rich self replication is then given as:

$$C_{String} > C_{BuildingBlock}$$

Therefore, for Information Rich self replication we require an RNA string of about 8 bases (such that 8 ln 4 ~ 11) of arbitrary sequence to be replicated.

II] Threshold Energy to Replicate a String N at Temperature T



In order for a faithful copy to be made, a putative nucleotide base from solution (e.g. G) must interrogate its complimentary base to ascertain whether the new base is the correctly matching complimentary base or not. Let us postulate that that interrogation is carried out by a communication of some type (e.g. optical) using photons of energy E. Let us assume further that the entire system sits at a temperature T. If E is too small we may mistake the thermal background for communication. The question which we wish to ask is how much communication energy is required to accurately replicate an N-mer?

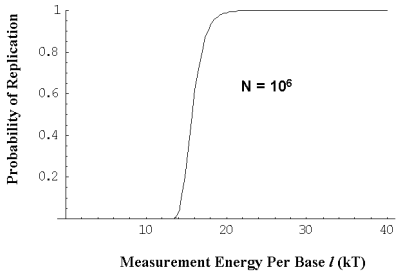
The Maxwell-Boltzmann distribution gives the energy density as:

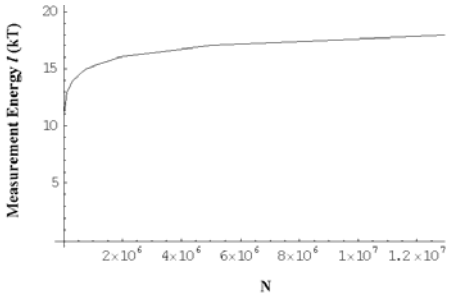
$$f(E)dE = 2\sqrt{\frac{E}{\pi(kT)^3}}e^{-E/kT}dE$$

The probability of correctly identifying the complimentary nucleotide for communication energy $E = lkT$ is:

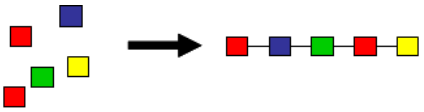
$$p(l) = \int_0^{lkT} f(E)dE = \frac{e^{-l}\left(-2\sqrt{l} + e^n\sqrt{\pi}Erf[\sqrt{l}]\right)}{\sqrt{\pi}}$$

The probability of building an N-mer is given as: $p(l)^N$





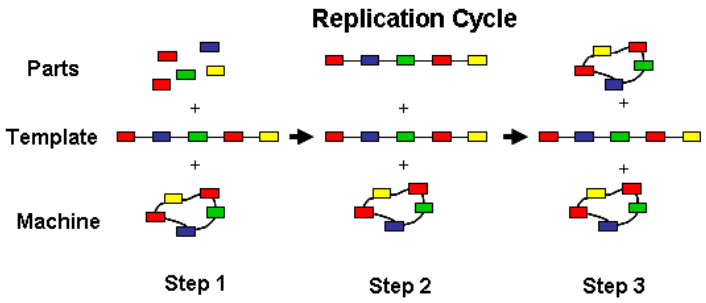
Comparison to experiment. The conversion of ATP to ADP yields ~ 0.3 eV. DNA replication is driven by the conversion of two Phosphate bonds yielding ~ 0.6 eV. From the above considerations, for the error correcting functionality alone, ~ 20 kT = 0.5eV is required per yielded base.



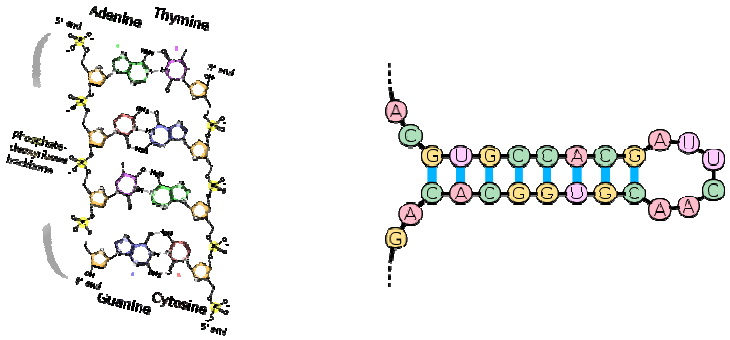
$$\Delta S = k\ln M^N,\ E = C \bullet N\ \therefore \frac{dS}{dE} = \frac{k\ln M}{C} = \frac{1}{T}\therefore C = kT\ln M$$

III] Complexity Threshold for Universal Self Replication

As we have seen from above, if we have a piece of external machinery (e..g. a protein based polymerase) which is carrying out the replication then, on completely general grounds as per above, the only threshold which we impose is a threshold error correction or measurement energy per bond addition. However it is known that linear polymers such as RNA can form secondary structures which are a function of their sequence and which have enzymatic function. If such enzymatic secondary structures have the function that they copy the linear sequence itself then we have a universal replicator. In this section we wish to calculate the size entity which comprises a universal self replicator – e.g. What is the complexity threshold for a universal self replicator?



As we had above, the key function which the replicator must carry out in order to produce a faithful copy of itself is error correction. The secondary structure of the polymer carries out this function by making a measurement of the canonical base pairing in DNA using a molecular ruler. If the ruler finds the wring base pairing it excises the added base and tries again.



In a thermal bath at kT what is the probability that the molecular ruler is sufficiently accurate to make the necessary measurement?

Assume that the probability that a single bond is open is ε. Therefore the probability that *N* bonds are open is given by ε^N and the per step yield is $p = 1 - \varepsilon^N$. We may now calculate the probability that our molecular ruler stay intact for the duration of the replication of the entire *N*-mer as:

$$P = p^N = (1 - \varepsilon^N)^N$$

Example DNA: From DNA melt curve data [Zewail] we have that a 10-mer (of balanced C-G content) spends 50% of its time in a dissociated (melted) state at ambient temperature. Therefore $\varepsilon_{DNA-CG0.5} = .933$.