A Thermodynamic Model for Prebiotic RNA-Protein Co-evolution

by Erkan Tuzel and Ayse Erzan

Review by Paul Welander
498BIN Assignment #1
October 15, 2001

Since the discovery of reverse transcription – a process whereby an RNA molecule can encode a DNA strand – it has been hypothesized that perhaps RNA was the first molecule of life. The means of evolution that transformed these simple molecules into the variety of organisms we observe today is then quite complicated. The first step in this journey was the selection of different strands for survival and reproduction. The catalyst for this selection process is anybody's guess. However, we do know that the primary role of RNA is the synthesis of proteins. Might these daughter molecules have played a role in the evolutionary course of their parent RNA? A new theory suggests such a method.

In their recent paper, *A Thermodynamic Model for Prebiotic RNA-Protein Co-evolution*[1], authors Erkan Tuzel and Ayse Erzan explore a thermodynamic model for the selection and evolution of proteins and RNA molecules in the pre-biotic world. The folding and unfolding of the protein molecules is treated as a heat pump whose efficiency depends on the entropy gap between states and the folding rate. Those proteins that are most efficient, along with the RNA that encoded them, are then favored in the evolutionary process. This method, the authors believe, is partly responsible for the selection of proteins observed today.

The refrigeration cycle begins with primordial *soup* contained in a porous rock. The rock acts not only as a heat reservoir for the cycle, but its pores loosely resemble cells with the molecules of life contained within. Contents of the soup include RNA and protein molecules, the amino acids needed for protein construction, and water. For purposes of studying thermodynamic selection, the presence of other organic and inorganic materials in the mixture is neglected.

The first step of the cycle is the production of a protein by an RNA molecule. Once separated from the RNA, the protein remains in an unfolded state due to the high temperature of the soup. However, the existence of both polar and hydrophobic residues (residue = amino acid) causes the protein to seek a surface to attach to. By concealing the hydrophobic residues from the water in the soup, the protein can reduce the total free energy of the system. Once attached to the rock, the surface acts as a guide to assist in the folding of the protein. During this process, the rock absorbs heat from the protein. The folding occurs such that the hydrophobic residues are restricted to the interior of the protein. When it is no longer energetically favorable, the protein would detach from the rock and return to swim amongst the soup. In reaching equilibrium with the soup, the protein will unfold and absorb heat, lowering the temperature in its vicinity.

The cooling power of this cycle depends on two parameters: entropy gaps and the rate of the folding/unfolding transition. The existence of a unique, low-energy ground state leads to a large entropy gap for the protein. The time required for small proteins to fold into their secondary structure has been observed to be less than a millisecond in some cases. The larger the entropy gap and the faster the transition, the more efficient the refrigeration cycle.

How is this important to the selection and evolution of RNA and protein molecules? Well, if the soup is cool enough, proteins can remain folded after detaching from the rock. Efficient proteins can thereby create an environment where they can exist in folded form and perform some function. The RNA molecules that synthesize these fast folding proteins might have then replicated more and become dominant in nature due to the lowering of the temperature in their vicinity by the help of their proteins. Better replication rates for RNA molecules may have been achieved at lower temperatures by the presence of folded proteins, a few of which could have acted as catalysts in the replication process.

The model for such an evolutionary method can be briefly described as follows. Two variables, σ_i and τ_i, are introduced at the ith node, where σ_i represents a folding value and τ_i a sequence value. At each node σ_i can take on several values, but only a value of σ_i* will allow the protein to fold correctly. Similarly, τ_i can assume several different values at each node, but only if the amino acid at that node is the correct one as coded by the RNA will $\tau_i = \tau_i$*. For simplicity, the model of Tuzel and Erzan assumes that there are only two residue types: hydrophobic and polar. Two state variables can then be defined as follows:

$$\psi_i = \delta_{\sigma i, \sigma i^*} \qquad \theta_i = \delta_{\tau i, \tau i^*}$$

Now the value of ψ can be either 0 or 1, depending on whether the fold at the ith node is correct. Similarly, θ may equal 0 or 1 depending on the survival, or correct synthesis, of the corresponding amino acid residue. By further defining the products,

$$\Psi_i = \psi_1 \ldots \psi_i \qquad \Theta_i = \theta_1 \ldots \theta_i$$

one can construct the following Hamiltonian for the protein

$$-H = \lambda\varepsilon \sum_{i=1..N} \Psi_i + (1-\lambda)\varepsilon\Psi_N N\Theta_N + J\Theta_N N$$

where ε is the energy loss due to each folding event, and J the energy loss due to each amino acid ordering event. The parameter λ provides for the absence or existence of intermediate folding states. For $\lambda = 0$, the protein has only the unfolded and native states. $\lambda = 1$ indicates that there are intermediate states whereby a protein may pass during a folding or unfolding transition. From this Hamiltonian, the authors proceed to calculate transition temperatures, a partition function, thermodynamic variables, and order parameters.

Within the description of this thermodynamic model are a few shortcomings. First, the authors assume that all amino acids can be grouped into the polar and hydrophobic groups, and all other discrepancies can be ignored. This effectively reduces the number of amino acids from twenty to two. They have neglected differences in size, type of side chain, and charge that result in twenty unique amino acids. In making this assumption, the authors obviously hoped to reduce the complexity of the problem. While their results shed light on the possibility of thermodynamic selection of RNA and proteins, one has to wonder what effect this number reduction might have. In making the problem easier, have the authors really solved anything of biological importance, or is the model too oversimplified to be applicable.

The second and perhaps more serious flaw in the authors' reasoning is the statement of a unique ground state for the protein folding. As we have seen in class for the folding of RNA molecules, protein folding is a dynamic process with a very complex free-energy landscape. The details of this landscape provide for the existence of several stable folded states. The possibility of multiple protein configurations introduces more complexity to the state variable ψ_i. Remember that ψ_i was defined according to whether or not the fold at the ith node was correct, and that there was one and only one correct fold. The possibility of multiple configurations introduces more correct folds at each node; σ_i^* is now multi-valued. Since the protein can take on one of many different configurations, the entropy of the folded state is also increased, and hence the entropy gap between the folded and unfolded state is reduced. One then must ask how the proteins we observe today could exist assuming the model of Tuzel and Erzan is correct. Perhaps it is enough to assume that proteins today have maximal entropy gaps. Another possibility is that the next step in the evolutionary process selected proteins that could assume additional configurations.

Regardless of the steps taken to simplify the problem at hand, the authors do provide arguments for the possibility of thermodynamic selection. The implication of a natural refrigeration cycle is the selection of fast-folding, large entropy gap proteins and the RNA strands that encode them. Perhaps this was one of the many processes that nature underwent in its quest to develop life.

---

[1] E. Tuzel and A. Erzan, cond-mat/0107315 v1, 14 July 2001.