Pohorille, Andrew

MATHEMATICAL MODELING OF THE ORIGINS OF LIFE

Andrew Pohorille
NASA-Ames Research Center

The emergence of early metabolism – a network of catalyzed chemical reactions that supported self-maintenance, growth, reproduction and evolution of the ancestors of contemporary cells (protocells) was a critical, but still very poorly understood step on the path from inanimate to animate matter. Here, it is proposed and tested through mathematical modeling of biochemically plausible systems that the emergence of metabolism and its initial evolution towards higher complexity preceded the emergence of a genome. Even though the formation of protocellular metabolism was driven by non-genomic, highly stochastic processes the outcome was largely deterministic, strongly constrained by laws of chemistry. It is shown that such concepts as speciation and fitness to the environment, developed in the context of genomic evolution, also held in the absence of a genome.

The hypothesis is directly relevant to a long-standing debate about the nature of processes that led to the emergence of life. One view holds that the origin of life is an event governed by chance, and the result of so many random events is unpredictable. This view was eloquently expressed by Monod in his book "Chance or Necessity", in which he argued that life is a product of "nature's roulette." In an alternative view, the origin of life is considered a deterministic event. Its details need not be deterministic in every respect, but the overall behavior is predictable. An elegant exposition of this view can be found in Morowitz's book "Beginnings of Cellular Life." Our hypothesis bridges these two apparently disparate views. The processes underlying the emergence of life are stochastic and, therefore, can be described only in probabilistic terms. However, their outcome is predictable, although not in full detail.

We postulate that metabolic reactions were initially catalyzed by means of ancestral proteins (or their short-length analogs – peptides). Although proteins do not replicate, stochastic mechanisms still exist that can lead to reproduction and improvement of their catalytic functions *in a population*. In fact, self-replication of macromolecules was not required at the early stages of evolution; the reproduction of cellular functions alone was sufficient for self-maintenance of protocells. The precise transfer of information between successive generations of the earliest protocells was unnecessary and could have impeded the "discovery" of cellular metabolism. Many proteins with potentially unrelated sequences could have performed eac h biochemical function at an evolutionarily acceptable level. As evolution progressed, however, proteins must have performed their functions with increasing efficiency and specificity. These proteins constituted a progressively decreasing fraction of all proteins and, at some point, the likelihood of generating them through a non-coded synthesis was so small that further evolution was not possible without storing information about their sequences. Beyond this point, further evolution required the coupling between proteins and informational polymers that is characteristic to all known forms of life.

The starting point for the proposed mechanism of protocellular reproduction and evolution is the emergence of protoenzymes that catalyze the formation of peptide bonds (ligases) and, therefore, generate new peptides. It is known that even some dipeptides exhibit ligating activity. Although most of the peptides synthesized with the aid of these poor catalysts would be non-functional or only weakly functional, a

few of them would be better ligases than the peptides that generated them. They, in turn, would ligate even more peptide bonds and, by doing so, increase the repertoire of peptides in the protocellular system. As a consequence, the likelihood of finding an even better ligase increases, as does the likelihood of finding proteins catalyzing other, possible chemical reactions. It is often assumed that the set of possible reactions is prohibitively large. In reality, however, carbon-based chemistry is a subject of strong thermodynamic (free energy and redox potential) and kinetic constraints, especially in the absence of powerful, highly evolved enzymes. Once these constraints are accounted for, possible chemical reactions can be considered in a highly systematicway. These sets of rules are incorporated into our boichemical model.

The computational model used is an extension of a previously developed model that included only two functions: synthesis and degradation of peptides. The identities of the amino acids forming peptides are not considered. Instead, the key quantity is the probability distribution of finding a peptide with a given efficiency of catalyzing a desired function, irrespective of its sequence. In agreement with biochemical intuition, the catalytic properties of a new peptide are chosen from a probability distribution contingent upon the properties of the substrates from which it is formed. Biochemically plausible rules allow for both the emergence of novelty and increases in efficiency and specificity. Using these probabilistic rules, time evolution of a population of protocells can be simulated numerically using the recently developed Next Reaction Method.

The results of these simulations should help answer such questions as:

1. Can we observe self-organized pathways and autocatalytic cycles and what is the degree of complexity of the systems in which they emerge? Are these pathways and cycles a consequence of universal chemistry? How do they relate to contemporary biochemical pathways?
2. Do different types of networks, which can be interpreted as different "species", persist in the population (i.e. are "collectively inherited") even though they are not inherited at the level of individual protocells?
3. How broad are the probability distributions of solutions under specified conditions and, therefore, to what extent can the outcomes be interpreted as deterministic?
4. How robust are the results with respect to changes of the initial inventory of reactants and the parameters in the model? How do populations of protocells respond to parameter changes that represent changes in the environment (changes in fitness)?

The results influence views on simple forms of life that might exist beyond Earth. If the constraints of chemistry operate as we propose, these alien forms of life will turn out to be surprisingly (and perhaps disappointingly) similar to our own. This does not mean that they will be endowed with the same form of genetic material, genetic code, or the suite of amino acids, but the underlying organizational principles of metabolism will be similar.