Models of protocellular structures, functions and evolution

Andrew Pohorille and Michael H. New

NASA Center for Computational Astrobiology, NASA-Ames Research Center

1 Introduction

The central step in the origin of life was the emergence of organized structures from organic molecules available on the early earth. These predecessors to modern cells, called "protocells," were simple, membrane-bounded structures able to maintain themselves, grow, divide and evolve. Since there is no fossil record of these earliest of life forms, it is a scientific challenge to discover plausible mechanisms for how these entities formed and functioned. To meet this challenge, it is essential to create laboratory models of protocells that capture the main attributes associated with living systems, while remaining consistent with known, or inferred, protobiological conditions. These model protocells could be considered as fundamental "proofs of concept" for our understanding of the origin of life.

Unfortunately, a comprehensive scenario describing the transition from inanimate matter into the simplest protocell is still lacking. However, several progressively more sophisticated projects have been initiated to probe protocellular life. Their ultimate goal is to create a "minimal protocell" [1], which can be defined as a structure bound by membrane-like material, capable of capturing energy and chemical substrates from the environment and using them to support the chemical reactions (metabolism) needed for self-maintenance and growth. A community of such model protocells should be capable of evolving towards higher complexity.

Currently, most projects aimed at understanding the emergence of protocells focus on self-replication. Nucleic acids [2, 3], proteins [4, 5], membranes [6] and non-biological organic molecules [7] have been used as self-replicating units. In some instances, the evolutionary potential of the self-replicating system was also examined. Even though none of these projects will yield a true protocell, they still involve highly complicated processes. To accomplish these processes, it is usually necessary to use molecular components that, though simple by biological standards, are too complex to have existed in a protobiological milieu. While this type of approximation is undesirable, the results of these studies are still of great value because they shed light on principles underlying the fundamental processes involved in the emergence of life.

In this brief overview, we report on progress in another project aimed at understanding how protocells worked and evolved. This project is a collaborative effort between scientists from NASA-Ames Research Center and the laboratories of J. Szostak (Harvard Medical School), J. Lanyi (University of California, Irvine) and D. Deamer (University of California, Santa Cruz). It is distinguished from other projects by its focus on protocellular metabolism and the coupling of metabolism to energy transduction, rather than on self-replication. This research is founded on the assumption that the emergence of systems endowed with genomes and capable of Darwinian evolution was preceded by a pre-genomic phase, in which protocells functioned and evolved using mostly proteins, without self-replicating nucleic acids (*e.g.*, RNA). This view is motivated by findings that, while the building blocks of proteins and simple cell walls are relatively abundant in space and can be readily generated in experiments simulating conditions on the pre-biotic earth, the building blocks of nucleic acids are quite rare under these conditions. Furthermore, RNA cannot readily perform such essential cellular functions as energy capture and transport. Finally, since there is no relationship between the function of a catalytic RNA and the function, if any, of the protein for which it can code, there is no clear path from the "RNA World" to today's world of protein catalysis and nucleic acid information storage.

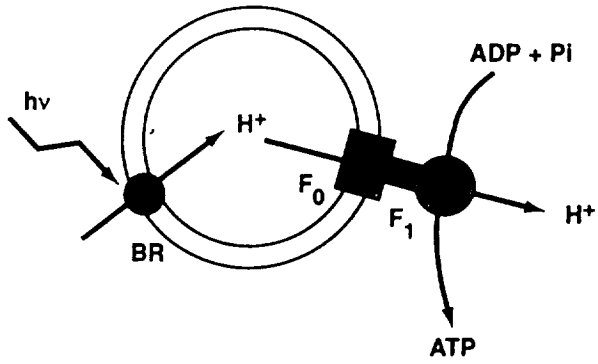

Figure 1: Schematic of the coupled BR-ATPase energy transduction system.

2 Protocellular Membranes

It is hypothesized that first protocells were encapsulated in vesicles — closed, spheroidal assemblies of amphiphilic molecules enclosing an aqueous medium. Contemporary cell membranes and most model vesicles (liposomes) are typically composed of phospholipids. However, in the absence of primitive biosynthetic pathways yielding phospholipids, it is unlikely that these molecules were present on the early earth. Instead, the most likely candidates for protocell-forming material were monocarboxylic acids and alcohols with chain lengths sufficient to stabilize self-assembled structures. Recent studies have shown that under narrowly defined conditions of pH, temperature and concentration, a monocarboxylic acid as short as nine carbons in length was able to assemble into vesicular membranes. However, the addition of small amounts of an alcohol, nonanol, markedly stabilized the bilayers, so that vesicles were present at any pH above 7 and at lower concentrations of carboxylic acid (D. Deamer, unpublished results). Significantly, monocarboxylic acids and alcohols are also the main components of the amphiphilic material extracted from the Murchison meteorite and shown to form vesicles [8]. We conclude that membranous vesicles produced by mixed short-chain monocarboxylic acids and alcohols are plausible models for early membranes.

3 Energy Transduction

One of the essential functions of protocells was to capture energy from the environment and utilize it to synthesize the high-energy compounds that served as the source of energy to drive other protocellular reactions. The initial source of environmental energy may have been light or, more likely, chemical energy (for a recent review see [9]). All known modern organisms transduce environmental energy by creating a transmembrane proton gradient which is coupled to the production of adenosine triphosphate (ATP) from adenosine diphosphate (ADP) and inorganic phosphate. The synthesis of ATP is performed by membrane proteins called ATP synthases (ATPases). Although phylogenetic studies indicate that ATPases must have emerged before the last common ancestor [10], these molecules are too complex to be a part of an early protocell, and the nature of the earliest energy transduction system is unknown.

To begin the study of the coupling of energy transduction to metabolism in protocells, a simple, model transduction system was constructed. It consists of two proteins, bacteriorhodopsin (BR) and the F_0F_1 ATPase from the thermophile *Bacillus PS3*, embedded in liposomes. BR is a light-driven proton pump, which generates the transmembrane proton gradient needed for the ATP-making activity of the ATPase. Although this system is most likely not protobiological, it is known to have high turnover and stability [11].

To construct the system, purple membranes containing wild type BR were solubilized and incubated with preformed liposomes in the presence of detergents. When illuminated, the proteoliposomes demonstrated very active proton transport directed inward, implying that the preferential orientation of the pumping species was inside-out. Alkalinization of unbuffered medium as large as 1 pH unit was observed. Next, BR was co-reconstituted with a thermostable ATPase from *Bacillus PS3* into liposomes using a similar procedure as for the BR-only liposomes. The ATPase incorporated into the membrane with its F_1 unit pointing outward, making it possible to utilize the pH gradient created by BR upon illumination (see Fig. 1). The system was stable under illumination for several hours (J. Lanyi and L. Brown, unpublished results).

It was further demonstrated that the production of ATP in liposomes could be coupled to a model, metabolic chemical reaction. As the model reaction, the production of acetyl-coenzyme A (acetyl CoA) from acetate and coenzyme A was chosen. This reaction is catalyzed by a thermophilic acetyl-CoA synthetase from *Pyrococcus furiosus*, which utilizes ATP, but not ADP to make acetyl-CoA. After adding this enzyme and appropriate substrates to liposomes, net synthesis of acetyl-CoA upon illumination was observed (J. Lanyi and L. Brown, unpublished results).

The next step will be to use the same enzyme, which is known to have broad specificity, to form different long-chain acyl-CoAs using fatty acids as substrates. Then, acyl-CoA transferase will add the fatty acid chain to lyso-phosphatylcholine (a lipid with only one hydrocarbon tail), and produce a new lipid. In a similar experiment performed previously, oleyl-CoA was coupled to lyso-phosphatylcholine, and production and growth of liposomes was observed (D. Deamer, unpublished results). Since higher concentrations of the lyso-lipid act as a detergent, a dynamic balance between liposome formation and destruction will be achieved if a mixture of different hydrocarbon chains is included. This, in turn, will introduce "natural selection" for the best components that can self-assemble into liposomes.

The results, described so far, demonstrate how simple metabolic reactions can be coupled to external sources of energy. The next objective is to simplify the components of the system and direct them to match better our ultimate goal — the creation of a protein-based model of primitive metabolism.

4 A Simplified Proton Pump

The task of creating a simple proton pump built of molecules that could have existed in the protobiological milieu may be helped by an unlikely model: the M_2 protein from human influenza virus. Four identical fragments of this protein, each only 25 amino acids long, fold into helices and aggregate to form a small, membrane-spanning channel. In response to an applied voltage, protons are conducted through a pore in the middle of the channel at remarkably high rates [12]. It has been postulated that protons travel along the network of water molecules filling the pore of the channel. This mechanism, however, must involve an additional step because the channel contains four histidine amino acid residues, one from each of the helices, which are sufficiently large to occlude the pore and interrupt the water network. Thus, the histidine residues form a gate, which controls transport of ions. Two mechanisms of gating have been proposed. In one mechanism, all four histidines acquire an additional proton and, due to repulsion between their positive charges, move away from one another, thus opening the channel [13]. In the alternative mechanism, one proton is captured on one side of a histidine residue while another proton is released from the opposite side. Then, the channel returns to the initial state through tautomerization [15].

Large-scale, atomic-level molecular dynamics simulations of the channel with the histidine residues in different protonation states [16] revealed that all intermediate states of the the tautomerization mechanism are structurally stable (see Fig. 2, right), and the arrangement of water molecules in the channel is conducive to the proton transport. In contrast, in the four-protonated state, postulated to exist in the gate-opening mechanism, the electrostatic repulsion between the histidine residues appears to be so large that the channel loses its structural integrity (see Fig. 2, left). This result indicates that such a mechanism of proton

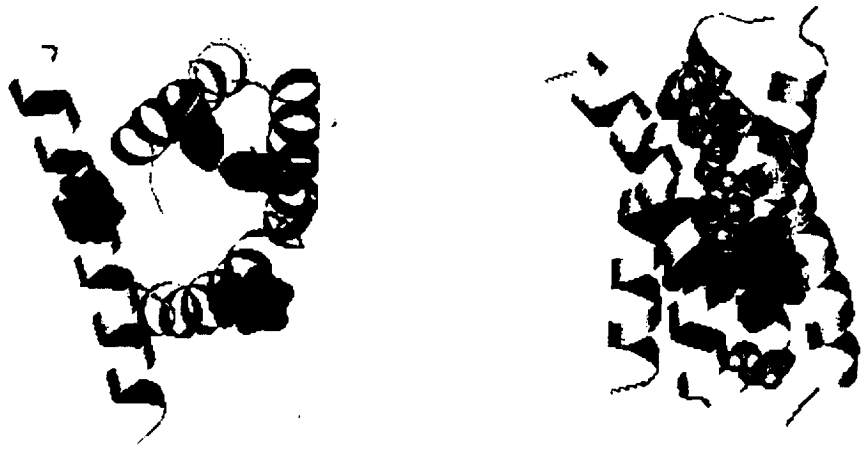

Figure 2: Richardson diagram of intermediate states of two mechanisms for proton translocation through the M_2 protein of human Influenza A. Histidines and water wire shown as dark and light CPK models, respectively. Solvating bilayer and waters omitted. Figure made with MOLSCRIPT [14]

transport is unlikely.

The results also suggest how the channel can be re-engineered to become a simple proton pump. First, a proton source must be placed inside the channel. Among many possible sources of protons, probably the simplest are polycyclic aromatic hydrocarbons [17] — a class of molecules readily available on the early earth. Once the channel is coupled with a proton source, the last thing that remains to be done is to ensure that protons transferred through the channel to the aqueous medium are unlikely to be used to regenerate the proton source. Since the proposed proton pump is asymmetrical, this should be possible to accomplish by appropriate genetic modification of the channel.

5 *In vitro* Evolution of Novel Peptide Enzymes

In considering the other functions performed by protocells, no suitable models from modern organisms are readily available. Previous attempts to design small, functional proteins from first principles have been less than successful. Instead, a novel method for the *in vitro* evolution of protein enzymes towards arbitrary catalytic targets will be employed. A similar approach has already been developed for nucleic acids: First, a very large population of candidate molecules is generated using a random synthetic approach. Next, the small number of molecules that can accomplish the desired task are selected. These molecules are then vastly multiplied using the polymerase chain reaction. A mutagenic approach, in which the sequences of selected molecules are randomly altered, can yield further improvements in performance or alterations of specificities. Unfortunately, unlike nucleic acids, proteins cannot be directly amplified. In the new technique, this problem is circumvented by covalently linking each protein of the initial, diverse, pool to the RNA sequence that codes for it. Then, selection is performed on the proteins. The selected protein-nucleic acid fusion molecules are then cleaved and the nucleic acids amplified and, if desired, mutated. These nucleic acids are then translated into proteins, fused to these proteins and another round of selection is performed [18].

Using this technique, we will generate models for several important proto-enzymes involved in the creation of new proteins. The formation of a peptide bond between amino acids requires the input of energy. Thus the first target of the new *in vitro* evolution proce-

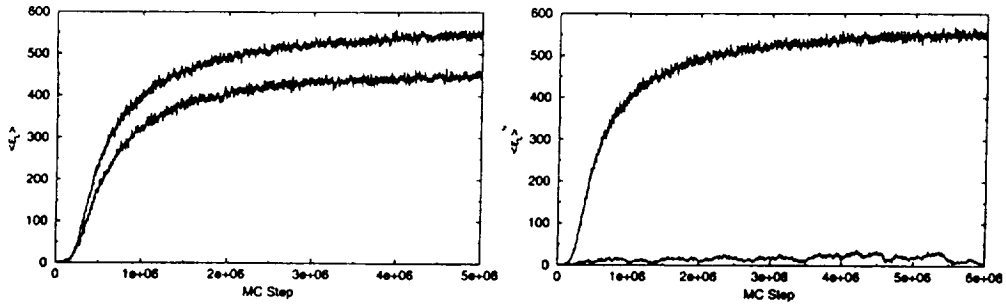

Figure 3: Average catalytic efficiency of peptides in a test protocell simulated by the Inherited Efficiencies Model. Left panel shows the effect of decreasing ease of forming ligases (lower line) relative to a reference model (upper line). Right panel shows effect of decreasing preference for cutting bonds in unstructured peptides (lower line) relative to the same reference model (upper line).

dure will be an enzyme that can use the ATP generated by the transduction system to phosphorylate amino acids, thereby preparing them to be used to form new proteins. The second target for *in vitro* evolution will be an enzyme that can catalyze the formation of a peptide bond in the absence of a nucleic acid template, a peptide ligase. In modern biology, there are very few examples of such enzymes [19], and all are quite large and complex.

6 Evolution without a genome

Protocells must have been able not only to maintain themselves but also to evolve. Darwinian evolution occurs through the accumulation of a series of small alterations to functional molecules whose identities are stored. Protocells, however, may have been incapable of such storage. We hypothesize that under such conditions, the replication of functions and their inter-relationships, rather than the precise identities of the functional molecules, is sufficient for survival and evolution. This process is called non-genomic evolution.

Central to this concept is the emergence of weak ligases, that would produce a diverse pool of peptides. A few of the peptides so generated could have been better catalysts of peptide bond formation than the proto-enzymes that formed them, thereby generating even more peptides and increasing the chances of producing functional ones. Some of these functional peptides would be proteases, proto-enzymes that cut peptide bonds. Since proteases cleave unstructured peptides more rapidly than structured ones, and since functional peptides have some degree of ordered structure, proteases would preferentially destroy non-functional peptides. Occasionally, the newly produced peptides would be capable of performing novel functions. If these novel peptides integrated into the protocellular metabolism, they could increase the capabilities of the protocell. Eventually, this process could lead to the emergence (or utilization) of nucleic acids and their coupling with peptides into a genomic system.

To examine the evolutionary potential of a non-genomic system, we have developed a simple, computationally tractable model that is still capable of capturing the essential biochemical features of the real system. In this model, the specific identities of amino acids are not considered, and the catalytic efficiencies of all peptides of the same length are normally distributed. As peptides grow longer, their mean efficiency also grows, but non-monotonically: only when peptides reach sufficient lengths to adopt ordered structures do the average efficiencies increase markedly. Peptides act as both catalysts and reaction substrates. In the simplest implementation of the model, only two catalyzed reactions were considered: the formation (polymerization) and destruction (hydrolysis) of peptide bonds. The properties of the products of these reactions are

related to the properti[es] [o]f the reactants, and so the model is called t[he] [I]nherited Efficiencies Model.

Simulations of this [mo]del were carried out using a Monte Carlo m[e]thod [20]. For many choices of model para[me]ters, the overall catalytic efficiency of the [pro]tocell was observed to increase. Two prop[er]ties strongly affected the ability of the prot[ocell] to evolve: the balance between the proba[bi]lities that a peptide is an efficient protease and [an] efficient ligase, and the strength of the pre[fe]rence toward the hydrolysis of unstructured [pe]ptides. It was found that when efficient liga[se]s were not difficult to form (Fig. 3, left panel) [e]r when there was a strong preference for the [hy]drolysis of unstructured peptides (Fig. 3, [ri]ght panel), any long, highly efficient peptides [fo]rmed would not be destroyed before they c[ou]ld accelerate the rate at which new, and pos[si]bly efficient, peptides were formed. Under th[ese] conditions, the ability of the protocell to i[mp]rove its catalytic capabilities was high. Th[ese] results demonstrate the possibility of a n[ew] mechanism of early protocellular evolution [tha]t does not require the presence of a genom[e] [n]or any form of sequence complementari[ty] the exact replication of protein sequen[ce]s. In fact, the sloppy replication of protein se[qu]ences is an advantage in the earliest phase of [evo]lution because it allows for the rapid exp[lo]ration of the space of proteins and the discov[ery] of new functions. It is the preservation of th[e] functions and their inter-relationships whi[ch] must be maintained during this early stage [of] evolution, not the identity of the actors [in] forming those functions.

7 Acknowledge[me]nts

This work has been su[pp]orted by the NASA Astrobiology Institute a[nd] the NASA Exobiology Program.

References

[1] Morowitz, H.J., B. He[ad] D.W. Deamer: The chemical logic of a mi[nim]um protocell. Origins of Life Evol. Biosphere [21], 281–287, 1988.

[2] von Keidrowski, G.: [A] self-replicating hexadeoxynucleotide. Ang[ew.] Chem. Int. Ed., 25, 932–935, 1986.

[3] Unrau, P.J., D.P. Bar[tel] RNA-catalysed nucleotide synthesis. Natu[re] 395, 260–263, 1998.

[4] Lee, D.H., G.R. Granja, J.A. Martinez, K. Severin, M.R. Ghadiri: A self-replicating peptide. Nature, 382, 525–528, 1996.

[5] Yao, S., I. Ghosh, R. Zutshi, J. Chmielewski: Selective amplification by auto- and cross-catalysis in a replicating peptide system. Nature, 396, 447–450, 1998.

[6] Walde P, M. Fresta, A. Mangone, P.L. Luisi: Autopoietic self-reproduction of fatty acid vesicles. J. Amer. Chem. Soc., 116, 11649–11654, 1994.

[7] Wintern, E.A., J. Rebek: Autocatalysis and the generation of self-replicating systems. Acta Chim. Scand., 50, 469–485, 1996.

[8] Mautner, M., R. Leonard, D.W. Deamer: Meteorite organics in planetary environments: hydrothermal release, surface activity, and microbial utilization. Planet. Space Sci., 43, 139–147, 1995.

[9] Deamer, D.W.: The first living systems: A bioenergetic perspective. Microbiol. Mol. Bio. Rev., 61, 239–261, 1997.

[10] Gogarten, J.P., L. Taiz: Evolution of proton pumping ATPases — Rooting the tree of life. Photosynth. Res., 33, 137–146, 1992.

[11] Pitard, B., P. Richard, M. Dunach, J.L. Rigaud: ATP synthesis by the F_0F_1 ATP synthase from thermophilic *Bacillus PS3* reconstituted into liposomes with bacteriorhodopsin 2: Relationships between proton motive force and ATP synthesis. Eur. J. Biochem, 235, 779–788, 1996.

[12] Pinto, L.H., L.J. Holsinger, R.A. Lamb: Influenza virus M_2 Protein has ion channel activity. Cell, 69, 517–528, 1992.

[13] Sansom, M.S.P., I.D. Kerr, H.S. Son: The influenza A virus M_2 ion channel: A molecular modeling and simulation study. Virology, 233, 163–173, 1997.

[14] Kraulis, P.J.: Molscript: A program to produce both detailed and schematic plots of protein structures. J. App. Crys., 24, 946–950, 1991.

[15] Pinto, L.H., G.R. Dieckmann, C.S. Gandhi, C.G. Papworth, J. Braman, M.A. Shaughnessy, J.D. Lear, R.A. Lamb, W.F. deGrado: A functionally defined model for the M_2 proton channel of Influenza A virus suggests a mechanism for its ion selectivity. Proc. Natl. Acad. Sci. USA, 94, 11301–11306, 1997.

[16] Schweighofer, K.J., A. Pohorille: Computer simulation of ion channel gating: The M_2 channel of Influenza A virus in a lipid bilayer. Biophys. J., 78, 150–163, 2000.

[17] Deamer, D.W.: Polycyclic aromatic hydrocarbons — Primitive pigment systems in the prebiotic environment. Adv. Space Res., 12, 183–189, 1992.

[18] Roberts, R.W., J.W. Szostak: RNA-peptide fusions for the in vitro selection of peptides and proteins. Proc. Natl. Acad. Sci. USA, 94, 12297–12302, 1997.

[19] Konz, D., M.A. Marahiel: How do peptide synthetases generate structural diversity? Chemistry and Biology, 6, R39–R48, 1999.

[20] New, M.H., A. Pohorille: An inherited efficiencies model of non-genomic evolution. Simulation Practice and Theory, 8, 99–208, 2000.