Artificial Life press copy, March 26, 2003
LA-UR-02-7845

Bridging nonliving and living matter

Steen Rasmussen[1,2], Liaohai Chen[3], Martin Nilsson[1], & Shigeaki Abe[1,3]

[1]Self-organizing systems, EES-6, MS-T003
Los Alamos National Laboratory
Los Alamos NM 87545, U.S.A.

[2]Santa Fe Institute
1399 Hyde Park Rd
Santa Fe NM 87506, U.S.A.

[3]Bioscience Division
Argonne National Laboratory
Argonne IL 60439, U.S.A.

[steen@lanl.gov; nilsson@lanl.gov; lhchen@anl.gov; sabe@anl.gov]

Key words: *protocells, origins of life, nanotechnology, supermolecular chemistry, self-assembly, physical chemistry, photo-chemistry, multiscale simulation, evolution, artificial chemistry.*

Abstract

Assembling non-biological materials (geomaterials) into a proto-organism constitutes a bridge between nonliving and living matter. In this paper we present a simple step-by-step route to assemble a proto-organism. Many pictures have been proposed to describe this transition within the origins of life and artificial life communities and more recently alternative pictures are emerging from advances in nanoscience and biotechnology. The proposed proto-organism lends itself to both traditions and defines a new picture based on a simple idea: Given a set of required functionalities, minimize the physicochemical structures that support these functionalities, and make sure that all structures self-assemble and mutually enhance each other's existence. The result is the first, concrete rational design of a simple physicochemical system that integrates the key functionalities in a thermodynamically favorable manner as a lipid aggregate integrates proto-genes and a proto-metabolism. Under external pumping of free energy, the metabolic processes produce the required building blocks, and only specific gene sequences enhance the metabolic kinetics sufficiently for the whole system to survive.

We propose a concrete experimental implementation of the proto-organism with a discussion of our experimental results, together with relevant results produced by other experimental groups, and we specify what is still missing experimentally. Identifying the missing steps is just as important as providing the road map for the transition. We derive the kinetic and thermodynamic conditions of each of the proto-organism subsystems together with relevant theoretical and computational results about these subsystems. We present and discuss detailed 3-D simulations of the lipid aggregation processes. From the reaction kinetics we derive analytical aggregate size distributions, and derive key properties of the metabolic efficiency and stability. Thermodynamics and kinetics of the ligation directed self-replication of the proto-genes is

discussed, and we summarize the full life-cycle of the proto-organism by comparing size, replication time, and energy to biomass efficiency of contemporary unicells. Finally, we also compare our proto-organism picture with existing origins of life and protocell pictures.

By assembling one possible bridge between nonliving and living matter we hope to provide a brick in the ancient puzzle about who we are and from where we come.

1. Introduction

Like contemporary living cells, under appropriate laboratory conditions future proto-organisms will sustain themselves chemically, feed from the environment, self-reproduce, be capable of evolution, and be able to die under environmental stress. Unlike modern cells and their protocell ancestor, engineered proto-organisms have an artificial origin, environment and metabolism. They are not made or derived from existing cells; instead they are built "from scratch" to operate only in artificial environments. Although they may have some metabolic pathways common to those in natural cells, their designed metabolisms may exhibit completely different chemistries. Proto-organisms have not yet been constructed in the laboratory partly due to the physicochemical complexity of assembling such structures and partly because only very few focused efforts exist worldwide. However, recently several efforts have started.

There are two significantly different approaches to synthesize primitive life forms: A "bottom up" and a "top down" approach. The current contribution is based on a bottom up approach concerned with constructing a simple living system from nonliving organic and inorganic materials through self-assembly, with metabolic processes driven by an external supply of free energy. The top-down approach concerns itself with systematic simplification of very simple existing cells.

The presented work on assembling a proto-organism has grown out of a bottom up tradition, which has emerged over the past ten years at the intersection between the origins of life studies, nanoscience, new materials, supermolecular chemistry, and Artificial Life activities. Our work addresses several of the open questions formulated by the Artificial Life community at the 2000 conference in Portland, which can be found in Bedau et. al., 2000 [4]. The work mainly addresses question #1, *"Generate a molecular proto-organism in vitro"*, but it also touches

question #4, *"Simulate a unicellular organism over its entire lifecycle"*, and question #5, *"Explain how rules and symbols generated from physical dynamics in living systems"*.

1.1. Bottom up

Until recently it has been difficult to separate the ideas on how to bridge nonliving and living matter from the theories of the possible origins of life. Experimental bridging pathways that were believed to be likely have often been proposed as particular origins of life pictures and these pictures have often been tightly linked to specific molecules and/or processes. Pictures less dependent on the physicochemical details, usually lack experimental support, but may still be valuable from a theoretical point of view.

The Naked Gene approach to bridge nonliving and living matter has for more than a generation dominated the origins of life debate, which from the mid eighties, Cech 1986 [12], through the nineties proposed that the RNA was probably "the first living molecule". The RNA World [29,28] supporters had a very compelling argument, since RNA both have catalytic and information storage capabilities it could act simultaneously as DNA and protein does in contemporary life. Many different polymerization and replication approaches have been developed [24]. However, to develop a self-replicating RNA molecule turned out to be a very difficult task [39] and this problem has recently takes some of the enthusiasm out of the RNA World picture.

The Peptide World, stating that proteins probably are the first biomolecules, perhaps defines the oldest physicochemically based picture of the origins of life, as Oparin 1924 [68], pointed out that proteins can polymerize from amino acids in a prebiotic environment. The later

Miller and Urey experiments from 1959 [57], demonstrated a laboratory production of amino acids as well as many other key biological compounds under harsh prebiotic conditions. Today we know that most violent or extreme processes (high pressure, temperature, or irradiation) that involve simple elements produce traces of rather complex biomolecules. Examples of such processes include high explosives blasts [32] and meteorite content from cosmic chemistry [16]. One of the latest and most intriguing experimental developments within the Peptide World is the generation of a self-reproducing lipid system by Ghadiri and coworkers [45,82].

An alternative origins of life picture, the Lipid World, was originally developed by Luisi and coworkers [47], and by Deamer and Morowitz [60] 1988, who worked towards assembling proto-cells based on a self-reproducing lipid vesicle encapsulating a self-replicating RNA proto-gene. Self-reproducing lipid aggregates (micelles and lipososmes, single and bilayer lipid structures) have been developed successfully in several labs. It is also known that lipids are produces by cosmic chemistry and probably were readily available at the time of the origins of life. An interesting new twist on the Lipid World came recently with the proposed role of atmospheric aerosols in the origins of life [94]. For example, encapsulating of self-assembling structures, such as microtubules, within liposomes has also served as incomplete attempts to develop protocell models [35].

Recently important players from the RNA World community have joined this encapsulation approach as discussed in Szostack et al 2001 [89]. Their vision consists of a bottom up construction, but with rather high-level building blocks. They propose using an RNA-based chemistry, very similar to the information chemistries in contemporary natural cells. As they agree, the complexity of the catalysis needed to reproduce RNA (using an RNA replicase that is still a long way from being able to replicate its own coding RNA) as well as a metabolic encoding

of the building blocks (lipids and RNA) remain a barriers in their vision. In addition there is no natural physicochemical integration between the RNA and lipid container.

Another protocell vision is developed by Pohorille and Deamer, 2002 [70]. The starting point for their proposal is also the self-assembling dynamics of lipids to form lipid bilayer vesicles (liposomes), however they suggest to encapsulate modern cell organelles. They discuss different versions of artificial cells for different biotechnological purposes, with variations on RNA and DNA-based information chemistries, as well as metabolisms, noting, as do Szostak et. al., 2001 [89], that there are significant technical hurdles to making these complex chemistries work even when using existing cellular organelles. They conclude:

> "Many individual components needed for such structures have already been developed and many others are likely to be constructed in the near future. The main challenge now is to encapsulate them in a single cellular compartment and ensure that they will work in concert in a controlled manner."

This critique to some extent applies to all bottom up approaches.

The bottom up approach also has many theoretical and computational contributions. The notion of a hypercycle is a powerful theoretical picture proposed by Eigen, 1971 [20]. The hypercycle is a cooperative structure of self-replicating proto-genes. Also in 1971, Ganti [27] proposed a cooperative structure coupling a container, a metabolism, and a genetic system, which at least at a conceptual level is quite similar to our proposed proto-organism. This notion of cooperative structures being a key element for the bootstrapping of chemical systems to becoming biological systems has been the most important driver for the theory driven origin of life approaches. Such approaches span rather realistic model systems with autocatalytic sets of polymers as developed by Farmer et al, 1986 [23], Kauffman 1986 [42], and Bagley and Farmer, 1991 [3]. Lately, an interesting new idea has emerged based on autocatalytic lipid structures,

Segre et al, 2000 [83]. Wachterhauser [98] has proposed mineral surface based metabolic processes as the key bootstrap for the first self-reproducing lipid systems. Random graph generalizations of the cooperative feedback concept for proto-genetic systems were developed by Rasmussen, 1985 [75] and 1988 [76], and spatial hypercycle generalizations were developed by Hogeweg and coworkers [6]. Study of spatially extended systems showed that the hypercycle organization is persistent to parasites in such settings due to the formation of spiral waves. As a major simplification it was shown by McCaskill and coworkers [54,55] that hypercycles, or other reaction based cyclic cooperative feedback structures, are not necessary for the stabilization of distributed catalysis when one uses proximity in space. Recently, abstract self-reproducing, computational "protocells" has been developed by Ikegemi and coworkers [67]. These computational protocells have more realistic physical properties encoded than e.g. the intriguing self-reproducing loops by Langton, 1984 [44] and Sayama, 1998 [84], and for that reason they are more informative for the construction of physicochemical protocells.

In addition to the notion of cooperative structures in chemical networks, two other theoretical and computational traditions are important for our proto-organism work: (i) the detailed molecular dynamics (MD) simulation approach [100,51] together with (ii) the traditional lattice gas- [7], Lattice Boltzmann- [41] and the Ginzburg-Landau approaches [38]. Adding the intermediate time and length scale simulation technique, the MD lattice gas [52,66], we have the necessary components for a developing a predictive, 3-D, virtual proto-organism simulation. Rajagopalan, 2001 [74], has compiled a nice review of self-assembly simulations, but a comprehensive multi scale simulation is not yet available.

Our proto-organism design is both inspired by the above theoretical and computational ideas and by the experimental concepts for a protocell by Luisi and coworkers [49], but our model

also differs from this and other proposals in several aspects: (1) Our focal point is a minimalistic, thermodynamic coupling between the three functional structures: container, metabolism, and genes. We do not start with a self-replicating container or a self-replicating gene, which is then combined. (2) Instead of RNA we envision simpler molecules, such as peptide nucleic acids (PNA) [64,63], that may be much easier to couple with the lipid layer than RNA due to its hydrophobic backbone and which is also easier to synthesize. (3) Instead of sophisticated ribozymes we consider very simple short oligos that are capable of enzyme-less self-replication by means of a ligation mechanism [96,97]. (4) As in the other Protocell proposals we utilize the lipid to keep the cooperative structure together although the proto-genetic activity is not on the inside of a vesicle [1], but on the outside of a lipid aggregate. We can therefore work with simpler lipid structures such as micelles. (5) We make extensive use of the differences of the thermodynamic properties within the lipid phase compared to the water and the lipid/water interface and as a result we obtain a quite different chemistry.

1.2. Top down

The "Top-down" approach starts with contemporary cells with very small genomes. Experiments with a simple bacteria, *Mycoplasma Genetalium*, [26,36] indicate that cells can have much of their genome removed before the cell is no longer alive. Current estimates of minimal genome size based on this approach are ~300-350 genes. The minimal genome size obtained from these experiments may be considered an upper bound on genome size expected for simple proto-cells based on translated protein chemistry. Readjustment to deletions *via* retro-evolution of the cells and complementation has not been fully investigated. These minimal cells may still have artifacts in their genome that are needed to support physical and metabolic structures that have been

required along the evolutionary path of the original bacteria, but that may not be needed for proto-cells. Recently Venter and Smith, 2002 [30], have received financial support from the U.S. government (DOE) to create new more primitive life-forms using this top-down approach.

In parallel with this experimental top-down approach, a push for whole-cell simulations is under way within the Systems Biology community, see e.g. *Science* **299** (2002). These efforts are driven by a growing understanding of the details of the metabolic pathways and genetic networks. This tradition is concerned with representation and simulation of life as it is, as opposed to the Artificial Life community's focus on a synthesis and simulation of life as it could be. It is worth noting, however, that these whole-cell simulations could in a natural manner test and bench mark their approaches by simulating the much simpler proto-cells currently under development - including the proposed proto-organism in this paper.

1.3. Structure of the paper

This paper consists of several elements. Some parts constitute a traditional reporting of work done with associated results. Other parts consist of work that is not yet done, but work that is of key importance for the whole proto-organism project. If any one of these few missing links turn out to be insurmountable, our design has to be changed. However, due to the general interest in problem of generating life, we feel it is still of value to present this new picture. This mixture of tested and non-tested processes has however, made it challenging to write this paper. Secondly the paper has been challenging to write because it seeks to synthesize concepts and methods from very different traditions. We have attempted to address the latter problem by a sandwich construction of the paper with a general beginning and a general summary, interleaved with the necessary technical details.

Section 2 is a detailed conceptual presentation of the proto-organism. Section 3 discusses the experimental implementations, both what we know and what we don't know, and this section is rather technical. It requires some prior chemical knowledge and some familiarity with experimental techniques and procedures. Section 4 is also rather technical as it discusses the theoretical underpinnings of Section 3 and thus requites some prior knowledge of physical chemistry and computational physics. Unfortunately, some technical jargon is unavoidable to limit the page numbers of these two sections. Although this construction allows us to be precise and to the point as we present the different technical issues both in the experimental and in the theoretical and computational context, this presentation has the reader re-visit the key technical issues from three different perspectives: conceptually, experimentally, and theoretically. However, we are attempting to address a diverse and interdisciplinary community, so we apologize if the text seems too repetitive for the domain experts. Subsection 4.5 (the full life-cycle of proto-organism) is a summary of key elements from Sections 3 and 4 and it is again written in a more conceptual manner and so is the Discussion and the Conclusion (Sections 5 and 6).

2. Organizational and functional structure of a proto-organism

Which observable functional properties should the proto-organism have? An extensive discussion exists in the literature on the topic: "What is Life?" which we shall not enter here. For our purpose, it is sufficient to note that an ability (i) to evolve, (ii) to self-reproduce, (iii) to metabolize, (iv) to have adaptive response to environmental changes, and (v) the ability to die, are normally referred to as key properties of a living system. The proto-organism we are proposing has all of the above properties and these functionalities are generated as the different components of the proto-organism assemble and use free energy as it digests appropriate precursor molecules. To address the question about how

to implement these five functionalities in a minimalist manner, we may ask: *Which molecular aggregate interactions can carry or generate these functionalities?* – or using another language - *How can we assemble a dynamical hierarchy that ensures that these functionalities are generated?* An extensive discussion of the why we believe it is useful to operate with dynamical hierarchies is found in Rasmussen et al, 2001 [77].

We are now ready to present a thermodynamically downhill, step-by-step process, which combines into a single cooperative aggregate a proto-container, a proto-metabolism, and proto-genes defining a proto-organism. This physicochemical network derives energy from a coupled redo complex or photochemical reactions -- a simple form of metabolism -- and carries encoded information about the metabolic processes in a proto-gene, which together with the metabolic complexes are integrated in a lipid aggregate. Thus, this aggregate can self-replicate, use energy and nutrients available from its environment, undergo evolutionary change over time, and ultimately die. The high level organization of the proto-organism is shown in Figure 1.

Figure 1 here

A set of possible physicochemical implementations of the involved processes will now be presented together with a discussion of the thermodynamic and cooperative principles. A detailed discussion of the involved chemistry, thermodynamics, and kinetics follow in the next sections.

The simplest way to assemble a proto-organism seems to be through the combination of two physicochemical systems: a lipid-metabolic (redox or photo chemical) system and a lipid-templating system. In Figure 2 we discuss the structure and function of these systems and how they can be integrated. We seek to avoid initially getting lost in the physicochemical details so

that the main conceptual design can be communicated as clearly as possible, although we still want to include enough of the necessary details to clarify the main thermodynamics and kinetic issues. The top of Figure 2 (I) describes the self-assembly of the lipid container. The left-hand side (II) describes the steps involved in the integration of a simple metabolic system in an amphiphilic aggregate (micelle or vesicle bilayer). The right hand side (III) describes the integration of a templating polymer in an amphiphilic aggregate (micelle or vesicle bilayer). Finally, as these two systems are integrated (II+III) they define a simple proto-organism. Note that in this example with redox driven metabolic processes we assume that the lipid aggregate is a vesicle.

Figure 2 here

As shown in (1) in Figure 2 on the top (I), amphiphiles self-assemble into micelles and vesicles (or liposomes) that can host organic metabolic molecules. Since many organic redox or photoactive molecules are hydrophobic, thermodynamics will drive them to become integrated into the hydrophobic portion of the lipids. The energy source used by this metabolic system may either be chemical energy or light. As an example of the utilization of chemical energy, (2) in Figure 2 shows a vesicle that contains organic redox molecules, that becomes attached to a mineral surface containing appropriate metal atoms, due to charge differences between the head groups in the aggregate and the mineral interface.

Assuming that the hydrophobic redox molecules in the lipid phase and metal atoms in the mineral surface have appropriate redox potentials they can act as electron donors and receptors in an energy-rich exchange. As shown in (2) and (3), the redox molecules receive or donate electrons at the mineral surface, move to the inside of the lipid phase where the electrons can be exchanged.

This induces processes (4) that e.g. generate more amphiphilic (4 a) and/or organic redox molecules (4 b) and/or parts of templating polymers (4 c), from appropriate precursor molecules.

Vesicles self-reproduce (5) as more amphiphiles are produced, increasing the vesicle's surface area until the structure becomes unstable and surface tension breaks it into two new vesicles. The more interface area between vesicles and mineral surfaces, the more electron transfer reactions, which in turn will generate more amphiphiles and organic redox molecules. The more redox molecules the more redox molecules and the more lipid molecules can be produced. Thus, through two coupled autocatalytic reactions, cooperation can emerge between self-reproducing vesicles and a simple production of redox molecules.

The right hand side picks up where the left-hand side ended, with the self-reproduction of micelles (1') and/or vesicles (2'). Development of a proto-gene begins on the outer surface of a vesicle membrane (3'). The templating polymer is an oligomer with a hydrophobic backbone that sinks into the vesicle's surface (4')[1]. As an example, this polymer may be peptide nucleic acid (PNA) or related, because it has a simple backbone like that of a hydrophobic polypeptide rather than the more-complicated and charged RNA and DNA backbones composed of phosphates and sugars[2]. Since PNA has the same nucleobases as DNA or RNA, however, templating is possible. Note that without a hydrophobic backbone it would not be thermodynamically favorable for the templating polymer to sink into the lipid aggregate. For simplicity we use PNA throughout this paper as a representative templating polymer with a hydrophobic backbone. Please note that the attachment of e.g. RNA on a charged lipid interface would be possible, but the following reaction steps would be difficult.

[1] The idea to thermodynamically couple the template and the lipid aggregate in this manner was initially developed by Klaus Lackner and Steen Rasmussen in 1997.
[2] A detailed discussion of the hydrophobic PNA backbone properties is given in Sections 3 and 4.

The next step is to have two shorter, complementary polymer pieces, labeled "Ta" and "Tb," recognize the hydrogen-bond sites on the information template and form a double-stranded molecule (5'). Once this thermodynamic recognition has occurred, the templated molecule sinks further into the interior of the lipid aggregate (6') such that ligation can occur (at the asterisk) within the hydrophobic region (7'). Here a polymerization is more thermodynamically favorable[3] and the ligation could be driven by extraction of water (le Chatelier's principle). In addition the ligation process could be enhanced in several ways e.g. with a simple catalyst or in connection with a drying cycle of the lipid. Again, we want to emphasize that ligation is highly thermodynamically unfavorable in water and please also note that the charged backbones of RNA or DNA would not allow these molecules to sink into the hydrophobic phase.

The kinetics and thermodynamics of the association/dissociation processes for the two now complementary templating polymers within the vesicle must be in balance for dissociation to occur (8'). If the processes are balanced, both templating polymers will be able to perform template-directed self-replication and the process can be repeated (9'). The result of this process would be a template directed replication of the proto-gene polymer.

To obtain full cooperation between our vesicles, organic redox molecules, and proto-genes, the strands in the vesicle must as a minimum enhance one or more of the redox processes. Specific strands (e.g. conformation or chemical properties of nucleo bases) could e.g. enhance the rate of formation for either the amphiphilic or the organic redox molecules, at which point specific functionality would be "encoded" as a given sequence of monomers within the templating polymer. Such a feedback loop between the proto-genes, the redox processes and the lipid aggregate enables a Darwinian evolution: In the simplest situation assume we have two different proto-gene templates in two different aggregates. One gene that encodes a more efficient set of

[3] The ligation process is discussed in details in Sections 3 and 4.

redox reactions and one that has no positive influence on the redox reactions. Assuming everything else being equal a faster growth of the cooperative gene aggregate would result, and if there were competition for resources (available energy and building block molecules) the efficient gene aggregate would eventually dominate, as a simple example of proto-organism selection.

Although this self-reproducing molecular aggregate does not constitute a contemporary cell, we believe it is the first model of a concrete molecular system that results from thermodynamically downhill processes and that combines container, metabolism, and proto-genes in a cooperative manner. Figure 3 summarizes the causal structure of the full proto-organism.

Figure 3 here

In closing of the conceptual discussion, we review the proto-organism's key functional properties and how they are generated. This is a continuation of the discussion on dynamical hierarchies from [77] and it is not essential for the coming sections. It is clear that with a lipid aggregate we already have three natural levels of description with distinct observables: (1) the properties of the individual water molecules and monomers that make up the amphiphilic polymers, (2) the properties of the amphiphilic polymers, e.g. elasticity, and for the amphiphilic aggregates we can observe (3) an inside, which is hydrophobic, and an outside, which is mostly hydrophilic, and where both parts have distinctly different chemical (catalytic) properties. Also the aggregate has a loading ability of a variety of different molecules, which we later return to in details. As we add molecular objects to the lipid aggregate the fundamental level of description (mainly the length scale) does not change, but the observed properties of the composed structure may change dramatically. We see this happen when a proto-metabolic process (redox or photo drive) is implemented, which enables

the structure to grow and reproduce: As more lipid molecules are produced the aggregate grows and eventually divides, which defines self-reproduction. We may define the order of (emergent) functionality that is generated simply by paying attention to which of the substructures that are necessary for the observed functionality. Then we may enumerate (order) each of the functionalities successively as they are generated (by the assembly of more subsystems), under the condition that each earlier functionality is a prerequisite of the next following functionality. In this context, e.g. the proto-container (3^{rd} order functionality) is needed before a redox driven proto-metabolism (4^{th} order functionality, but still 3^{rd} level structure) can be established, and further the proto-genes (5^{th} order functionality, but still 3^{rd} level structure) require the proto-metabolism to exist. However, our proposed photo-driven metabolism requires a particular electron relay chain to function properly, and this electron relay is implemented within the templating polymers. Thus, the proto-metabolism and the proto-genetic system may both be characterized as 4^{th} (or 5^{th}) order functionalities in this context.

3. Proposed physicochemical implementation of the proto-organism

3.1. Assembly, stability and self-reproduction of the lipid proto-container

It is important to note that the lipid aggregates in our approach are utilized in a somewhat different manner compared to most other model studies of the origins of life or artificial cell approaches. Our lipid aggregates have three key functionalities, where the first is rather unique for our approach.

(i) As a proto-container it holds together the other two key aggregates, the proto-metabolism and the proto-genes. However, it does so either at the exterior surface of the aggregate

or deep within the lipid phase. The aggregate does not contain these other aggregates e.g. in the interior (water) volume of a vesicle, as they do in the typical Lipid World picture.

(ii) Also due to the container property, close spatial proximity results in very high local concentrations, which is key for many of the processes.

(iii) Finally, the interior lipid phase as well as the water/lipid interface has very different physicochemical properties compared to that bulk water phase and as such these three different physicochemical environments can be used to enhance certain reactions, in sequence. Both the lipid phase and the lipid/water interface act as catalysts.

Simple self-reproduction of micelles and vesicles from fatty acid molecules are demonstrated in the literature, in particular by Luisi and coworkers [47,48,49,50,98]. It is also shown that long chain fatty acids are able to form bilayer structures in water depending on the pH value and salt concentration of the solution, which has been investigated by Deamer and coworkers [17,18,19,1,58]. For an example, vesicles are formed when the solution pH is equal to the pKa of the acid in the bilayer. By generating (introducing) more fatty acid molecules to a lipid aggregate, the micelles or vesicles will reach their maximum loading conditions (size) and eventually they will split into two smaller assemblies via budding. For our purpose it is important to correlate the fatty acid structures with their ability for micellation, vesicle formation, and their stability properties.

Based on extensive studies of self-assemble of chromophore derivatized fatty acids Whitten at. al., 1998 [103] as well as work done by Deamer and coworkers [17,18,19,1,58], who have aggressively addressed the fundamental questions associated with simple fatty acid bilayer structures over the past years, we can summarize the following important facts about the self-assembling of simple fatty acids.

No vesicles are observed under any conditions using carboxylic acids with chain lengths shorter than 8 carbons, but all fatty acids from 8 to 12 carbons in length produced obvious vesicles. Shorter hydrocarbon chains form micellar structures. The minimum concentration of monocarboxylic acid necessary to form vesicles in aqueous solution is a function of chain length. Concentration for vesicle formation ranges from 130 mM for octanoic acid to 10 mM for dodecanoic acid. Carboxylic acids with chain lengths of 8 or more carbons forms vesicles in the pH range near the pKa of their terminal carboxyl group. Else they form micelles. It is found that the optimum pH for each chain length also varies slightly with the concentration of the acid in solution. Longer chain lengths required slightly higher pH ranges for stability. The addition of alcohols of the same chain length as the monocarboxylic acid forming the bilayer membrane dramatically increases the pH range of stability and decreased the concentration of fatty acid necessary for stable vesicles to form. For example, nonanoic acid requires 20 mM concentration of acid in the presence of 2 mM nonanol to form stable vesicles, which are stable at any pH above 6.5.

Vesicle membranes composed of nonanoic acid are relatively impermeable to ionic solutes such as KCl, with the half-time of osmotic gradient decay of approximately 40 minutes. The vesicles are more permeable to smaller polar solutes like glycerol, which has a half-time of 3 minutes. It is established that carboxylic acids having chain lengths or 8 or more carbons are able to self-assemble into stable vesicles within the constraints of certain concentrations and pH ranges. The addition of small amounts of fatty alcohol greatly expands the pH range of stability and decreases the concentration of acid necessary for vesicle formation and stability. Depending on the chemical structures of fatty acid and chromophore molecule, the lipid phase

of carboxylic acid vesicles are able to store up hydrophobic in the aqueous solution up to millimolar range.

Two approaches are demonstrated in the literature to form fatty acid molecules inside bilayer structures. The first approach involved a buffered water solution of containing preformed vesicles, overlaid with a small amount of insoluble carboxylic anhydride [48]. Under those conditions, a significant increase of the hydrolysis rate of the anhydride can be observed with respect to a reference system, which does not contain vesicles in the water phase. The second approach involved irradiating a dispersion of the photocleavable water-insoluble precursor didecyl-2-methoxy-5-nitrophenyl phosphate to form didecyl phosphate surfactant in the vesicles [50]. The phosphate-phenolate bond was selectively hydrolyzed and was released in the medium. Light microscopy was used to assess the presence of vesicles, which were generally in the range of 1-10 μm.

In summary: the experimental conditions for lipid self-assembly, stability, and self-reproduction are well known, although the specific loading capabilities of the aggregates for the proposed precursor molecules are not well known.

3.2. Lipid proto-metabolic system (photo-driven)

We now present a simple *lipid proto-metabolic implementation* as summarized in Figure 4. In this system a proto-gene, a PNA (peptide nucleic acid) [65] assembly within an amphiphilic aggregate (proto-container) encodes the metabolic production of the lipids as well as the precursor genes.

As we already discussed in Figure 2, initially amphiphilic monomer e.g. carboxyl acids and alcohols assemble into amphiphilic aggregates. A PNA strand, a proto-gene, will attach to the membrane due to its hydrophobic nature. Additional hydrophobic groups can be attached to the PNA backbone to ensure an appropriate partition in the membrane as necessary[4]. For simplicity we use PNA as a representative of such a class of templating polymers without insisting on a particular backbone chemistry or nucleic acid alphabet. It is due to this favorable thermodynamic properties of PNA and not RNA is utilized in this system.

We propose a direct autocatalytic feedback between the PNA proto-genes and the production of both lipid molecules and more PNA precursor molecules. This feedback can be implemented by using a modified PNA as a photo-catalyst due to it's charge transfer capabilities, see Figure 4. Charge transfer in DNA (same bases as in PNA) is well established experimentally and theoretical models have been developed. We return to a detailed discussion of PNA charge transfer in Section 4.3.

In this proposed system, we use the photolysis of various phenacyl esters ($PhCOCH_2$-OCOR, where R denotes a long carbon chain) to produce surfactant molecules. It has been well established that phenacyl esters can undergo photo-induced C-O bond scission to form acetophenone ($PhCOCH_3$) and the corresponding carboxylic acid (RCO_2H, surfactant) in the presence of appropriate electron donor molecules [22]. However, due to the fast back electron transfer, the quantum efficiency is very low (~0.01 – 0.05). A sensitizer coupled with an electron relay system can be introduced to block back electron transfer process, and thus increase the quantum yield of surfactant production, which has been successfully used in other systems. This provides an opportunity to achieve the surfactant reproduction controlled by the

[4] Synthesis of PNA monomers with such backbones has been demonstrated by Peter Nielsen and coworkers and will be discussed in next Section (3.3).

proto-genes. We may incorporate a sensitizer molecule to the PNA backbone and use an appropriate sequence of nucleobases as the electron relay system. Accordingly, we will anchor sensitizer (S) directly on the PNA backbone and use adenine (A) and guanine (G) as the effective electron relay unit.

We predict the following dynamics in this proposed system. Upon the excitation of S by irradiation, the excited state of S can be quenched by the phenacyl derivative to generate a contact ion pair of S cation radical and a phenacyl anion radical. Due to the close geometry of the PNA strand, the S cation radical can be efficiently scavenged by adenine A to yield an adenine cation radical and the recovery of the sensitizer. The A cation radical can further react with adjacent G to yield a G cation radical, which is eventually quench by electron scavenger e.g. HS^-. Since the formed phenacyl anion radical and the G anion radical are far away from each other, back-electron transfer will be avoided and the formed phenacyl anion radicals will have enough time to undergo the fragmentation reaction to produce the corresponding carboxylic surfactant molecules. Without the presence of the nucleobase sequence as the electron relay unit, the back-electron transfer between the senstizer cation radical and phenacyl anion radical will dominate the reaction and the yield of production of surfactant will be substantially low. Therefore, surfactants can be replicated only when a certain base sequence is present. The reaction can be easily followed by NMR (nuclear magnetic resonance) and FTIR (Fourier transform infrared) of the formation of carboxylic surfactant, as well as by a transient spectroscopic study of monitoring the formation of A and G cation radicals.

Figure 4 here

We may apply the same chemistry to demonstrate the production of PNA repeat units (Figure 4) by converting amino ester to amino acid monomer. Under these conditions, precursor PNA dimer (or short oligomers) will be effectively converted to functional PNA dimer (or oligomer), which consequently can be polymerized (the use of PNA dimer instead of monomer will diminish the cyclozation reaction of PNA monomer).

To summarize the proposed photo-metaboloc system: in the presence of adenine, guanine and sensitizer modified PNA (an electron relay system), photo-induced electron transfer reaction occurs between the sensitizer and the phenacyl ester, consequently, the sensitizer cation radical can receive an electron from adenine forming an adenin cation radical. This adenin cation is then reacted with adjacent guanine to yield a guanine cation radical, which is eventually scavenged by HS^-. The amphiphilic molecules can only be synthesized when the particular AG base sequence is present. As a result of the increased production of amphiphilc (lipid) molecules the aggregate grow, become unstable, and eventually divide. By using the same photo-fragmentation process a precursor PNA (dimer or oligomer) forms a functional PNA oligomer, which can hybridize with a complementary PNA template, which all together defines the catalytic properties. This autocatalytic feedback system establishes our first goal of coupling the proto-container, the proto- metabolism, and the proto-genes. However, note that this system does not yet constitute a full proto-organism as defined in the last section. There is no template directed replication of the proto-genes, which is introduced in Section 3.3.

Our preliminary experiments support the above approach. As an experimental model system, we use an amino-pinacol derivative as the electron donor molecule and a phenacyl ester as the electron acceptor molecule. As depicted in Figure 5 (a), the excited state of the pinacol can be quenched by the phenacyl derivative (with quenching constant of 10^2 M^{-1}) to generate a

contact ion-pair of pinacol cation radical and a phenacyl anion radical. However, due to the fast back-electron transfer reaction rate, 98% of the contact ion pairs return back to the starting materials. A small fraction of contact ion pairs undergo the charge separation, which eventually led to the formation of ketone molecules for the pinacol cation radical and carboxylic acid product for the phenacyl ester anion radical. The formation of ketone molecules can be easily monitored by their fluorescence (Figure 5 (b)). From the fluorescence spectra, we can deduce that the product quantum yield is only about 2% due to the dominant back-electron transfer processes. The quantum yield of the ketone product can be increased significantly when the same reaction is carried out in the presence of a high concentration (5 M) of a second electron acceptor, e.g. carbon tetrachloride (CCl_4). As shown in Figure 5 (c), CCl_4 can efficiently intercept the phenacyl ester anion radical inside the contact ion pair to generate CCl_4 anion radical. Thus 95% percent of contact ion pairs underwent charge separation and led to the high quantum yield of product formation, evidentially demonstrated in the fluorescence spectra. This proves our point that an electron relay system can dramatically enhance the photo fragmentation yield, although we would need to introduce an electron donor and not an electron acceptor as here. Also, since it is a homogenous reaction, a high concentration of CCl_4 is essential to react with the contact ion pairs; while in the heterogeneous case with proper molecular alignment such as attached to the PNA backbone, a low concentration of sensitizer should achieve the same effects. However, this has not yet been proven experimentally for this particular system.

Figure 5 here

3.3. Lipid proto-gene subsystem

To implement a *template directed ligation of PNA in a lipid aggregate* with a subsequent replication process the above systems can be expanded as proposed. Initially, note that it is highly thermodynamically unfavorable to attempt a ligation in water. Assume as in Figure 2 that we have lipid aggregates with PNA strands (templates) attached. If this system is supplied with two complementary pieces of PNA, a hybridization reaction between the original template and two complementary oligomers will occur, as a consequence of thermodynamics. As this new, three component complex sits in the hydrophobic environment, a ligation (polymerization) is possible, driven by the expulsion of water from the lipid layer (le Chatelier's principle). The possibility of polycondensation of amino acids in the lipid phase is already experimentally demonstrated by Luisi and coworkers [5]. The ligation reaction may be further be enhanced either by a cyclic drying of the lipid aggregates or by catalysis. Simple catalysts may also speed up the kinetics, including a general base catalysis using pyridine or a Lewis acid catalysis using a transition metal[5]. The next step in the template directed replication of PNA in the lipid aggregate requires a reasonable equilibrium to be established between the hybridized (double stranded) and single stranded PNA molecules. This is necessary to enable further templating reactions. Experimentally this can be established by variations in temperature or pH, but employment of modified PNA backbones that decreases hybridization energies (such as the β-alanine type backbone [37]) is also an option. The result of this proposed process is a

[5] Originally proposed by Shelley Copley.

multiplication (replication) of the proto-genes. However, this lipid catalyzed PNA replication process has not yet been experimentally demonstrated.

Preliminary experimental results that support the above approach are available from Nielsen and coworkers. PNA-PNA and PNA-DNA melting curves in water are already characterized extensively [79,90]. PNA directed chemical synthesis of another sequence complementary PNA strand is demonstrated in the aqueous phase using a C10 template onto which G-dimers were oligomerized via carbodimide activation [11]. Furthermore, PNA monomers of thymine in which the backbone glycine is replaced with phenylalanine, isoleucine or valine are described in [33,73] and it is demonstrated that these monomers can be incorporated into PNA oligomers without major loss in hybridization potency. Based upon the structure of aminoethylglycine PNA, we expect that it will only partition slightly into the organic phase, which indicates that it might be necessary with an addition of a hydrophobic group to the backbone. How much extra backbone hydrophobicity is necessary is an experimental question that could be mitigated with detailed MD and MD lattice gas simulations.

Template directed ligation has to be experimentally demonstrated in water with energized dimmers, but it has not yet been shown in a lipid phase. However if polymerization of PNA is achieved in lipid layers, one can turn to the more difficult problem of replication of PNA molecules. This process requires that a strand of PNA serve as a template for the assembly of a second complementary strand of PNA. A complex between a template strand of PNA and two monomers might form in either the aqueous phase or the lipid phase. The condensation reaction would occur within the lipid phase, driven by the expulsion of water from the hydrophobic environment within the lipid layer. True replication requires that, once a double-stranded PNA decamer is formed by template-directed synthesis, the decamer can

dissociate and both strands then serve as templates for another round of replication. The choice of PNA (with respect to backbone and base composition) will eventually have to be determined based upon theoretical calculations and the experiments described above, which will help choose molecules that form metastable complexes, that can separate to provide templates for subsequent rounds. Identification of appropriate conditions for metastable PNA complexes probably defines the most uncertain experimental task.

Initially, in order to facilitate the ligation, it is possible to study the reaction using PNA pentamers with activated (through an active ester function (NHS/pentafluorophenyl)) carboxyl group. By this approach one can "bypass" the uphill thermodynamics of the ligation and in more detail study the issue of multiple turnover. Again, analogous experiments in aqueous solution have already been performed and published [11].

These proposed experiments will provide insight into a problem that has vexed pre-biotic chemists for decades. It long has been recognized that the condensation reactions required to form polymers such as nucleic acids and proteins are thermodynamically unfavorable without strong activating groups, but likely activating groups and conditions for driving the endothermic formation of activated monomers have not been identified. The special environment provided by a lipid layer could certainly promote these types of reactions.

3.4. Proto-lipid, -metabolic, and –genetic integration

In order for PNA to qualify as a true proto-gene, it has to include "genomic" information that "encodes" the production of parts of the proto-organism. Further, the genomic information has to be inheritable such that the newly generated daughter proto-organism has a similar genome to

the parent proto-organism. Here, we propose a proto-organism where the replication is controlled by a catalytic unit having a particular base sequence embedded in the PNA and go through the predicted dynamics of the full assembly. Since only certain sequence of nucleobase pairs can catalyze the production of the amphiphilic molecules, the replication of the lipid aggregate will be determined by the PNA. Given our ability to control the surfactant reproduction through certain nucleobase pairs, we should be capable of using PNA as a genomic code to guide the replication of surfactant assemblies. As proposed in Figure 6, we may start with a surfactant self-assembly (a supported bilayer, a micelle, or a vesicle) containing a PNA oligomer (hexamer) with a TCTCTC base sequence and sensitizers S attached to the C monomers. The synthesis of sensitizer attached PNA monomer can be achieved by following the reaction scheme developed by Nielsen and coworkers. (A simpler implementation scheme, which does not involve the integrated PNA/sensitizer, is discussed in Figure 7). This PNA oligomer not only serves as the genetic information source, but also functions as a template for the PNA replication.

Now all the components for the production of surfactant and PNA dimer, as well as the PNA polymerization catalyst are present in the lipid phase. However, since no A/S/G complex is yet available, the surfactant replication process is inhibited by the back electron transfer from the sensitizer cation radical and phenacyl anion radical as mentioned earlier. The replication process will eventually start with the formation of functional PNA dimers (or oligomers), which are able to polymerize, as outlined in the previous section. While the formed PNA dimers (or oligomers) are floating around in the lipid phase, the longer PNA template will attract the complementary PNA dimers (AG, AG, AG) to form a duplex. Consequently, these dimers will align themselves according to the sequence of bases on the PNA template. Since efficient

A/S/G complexes are available now, they will speed up the co-sensitization process for the photofragmentation reaction. Consequently, surfactant replication will be initialized. The production of surfactant will increase as the alignment of A, S, G along the template PNA is increased, since the sensitizers (S, A and G) are not consumed in this case. Meanwhile, template-directed ligation and synthesis of PNA (perhaps in the presence of the catalyst) will also start and a complementary PNA oligomer AGAGAG will be formed. Under certain conditions, the double stranded PNA is subject to dissociation into two single strands[6], while at the same time the continued production of surfactants from the hydrophobic precursers embedded in the aggregate eventually causes the aggregate to become unstable and split into two aggregates. The newly formed surfactant assemblies embedded with PNA templates are subject to further replication to form even more assemblies as more PNA- and lipid precursors are supplied. At the final stage, the population of the proto-organism assemblies containing the original sequence of PNA template will predominate.

Integrating this template directed replication process with the above template encoded lipid and template production processes, a cooperative (autocatalytic), feedback has now been established between the three key elements in the proto-organism: *(i) proto-container, (ii) proto-metabolism, and (iii) proto-genes,* as we originally set out to do.

This autocatalytic feedback also forms the basis for a Darwinian evolution of the system. E.g., starting with two different proto-genes on different lipid aggregates, one template with and one without AG components will result in an "evolutionary" takeover by the aggregates with AG rich string, because it mitigates the production of (proto-container) lipids. Due to the predicted parabolic replication kinetics (Section 4.4), these aggregates could also carry ("parasitic") coexisting genes that can replicate, but cannot function as an electron relay.

[6] The PNA disassociation process (balance) is extensively discussed in Section 4.4.

Although this proposed self-reproducing molecular aggregate, or proto-organism, does not constitute a contemporary cell, to our knowledge, it is the first model of a concrete molecular system that results from thermodynamically downhill processes and that combines container, metabolism, and genes into a cooperative structure. However, to obtain a fully integrated and operational experimental system is a very difficult task, even if each of the subsystems is working in isolation. We cannot exclude that as yet unknown emergent properties make it impossible. We don't think this will occur, but it a question that eventually has to be experimentally addressed. The full causal structure of this proto-organism is discussed in Figure 8. For completeness we discuss an alternative physicochemical implementation of the proto-organism processes base on a redox metabolism, although in less detail.

Figure 7 here

Figure 8 here

3.5. Alternative implementations (redox-driven)

In Figure 9 the production of both surfactant and functional PNA molecules are driven by chemical energy sources. Head groups commonly present in amphiphilic molecules such as surfactants include carboxylic or sulfonic acids are usually synthesized by using relatively harsh oxidation conditions such as light. Alternatively, amine groups, which can also act as a hydrophilic head groups, are often synthesized through a mild reduction process. For example,

organic nitro compounds can easily be reduced to amines. Many reducing agents can reduce nitro compounds, the most common ones being zinc, tin or iron cations, as well as catalyzed hydrogenation. One such nitro reduction reaction is the so-called Zinin reduction [72]. This reaction uses sulfides or polysulfides, which actually are plausible prebiotic reducing agents since the presence of HS^- or H_2S is extensive in volcanic geothermal environments. This process may occur through the formation of single chain lysophosphatidyl ethanolamine (lysoPE), a derivative of a natural phospholipid molecule. It has been reported that single chain phosphates assemble into closed bilayer vesicles, which can be easily observed by the optical or electron microscopy [34]. We may use phosphatidyl nitroethanol as the precursor reactant and utilize chemical energy from HS^- to drive the reduction process, as shown in Figure 9. Since the HS^- is located in the aqueous solution and the nitro-phosphates will reside in the membrane (organic) phase, we may incorporate appropriate electron transfer compounds such as dinitrophenolindophenol or phenazine methosulfate in the membrane to facilitate the redox reaction. Standard chromatographic methods may be used to monitor the formation of the single chain phosphatidyl ethanolamine, and products can be verified by electrospray mass spectrometry, NMR and FTIR. We also expect that as the synthesis occurs, the newly formed lysoPE, will incorporate into the existing bilayer vesicles to produce a growth process.

In order to demonstrate the production of PNA repeat units, a nitro derivative of PNA dimer may be used as the precursor for the synthesis of PNA dimer[7]. Due to the hydrophibicity of the PNA precursor, it will reside in the vesicle membrane (organic) phase as originally discussed in figure 2, while the HS^- ion is located in the aqueous solution. Therefore, an electron relay system may be employed to ensure the electronic flow between the energy source

[7] Again, the reason to use PNA dimer instead of monomer is again to diminish the cyclozation reaction for the monomer.

and the precursor. Under these conditions, the nitro-group will be effectively reduced to the amine group and thus the PNA dimer will be formed.

By blending all the above components in the lipid phase, the proposed system is capable of self-generating more surfactants and PNA dimers, both of which are subject to assemble to more organized structures (vesicle) or to polymerize to more PNA oligomer. It is expected that a thermodynamic coupling of these two processes still exists. Since the total free energy is decreasing as the surfactants assemble and the dimers at the same time polymerize, this slight energetic drive will ensure that a weak coupling of two replication processes exists, although this coupling is much weaker that the metabolic coupling discussed in the previous sections.

So far this scheme does not account for the production of more redox molecules (such as dinitrophenolindophenol or phenazine methosulfate), so the proto-organism has to be on "life-support" in the sense that these redox molecules have to be part of the available resources. The same scheme can be used for the proto-gene replication, but it is not clear how the encoded base sequence in a direct manner can influence (catalyze) the redox kinetics. It is, however, conceivable that the kinetics will depend on the specific properties of the PNA stand's properties, e.g. chemical or folding, but without direct experiments we do not dare to make any specific predictions.

Figure 9 here

In the previous sections we have been focusing on a light driven proto-metabolism, due to its simplicity in the laboratory context, but it may not have been the energy form utilized at the origins of life. Evidence elsewhere seems to indicate that the surface of the early Earth is too harsh

an environment due to the late accretion bombardments. The protected pore space in the subsurface (in the planetary crust) may very well have provided the richest and most appropriate chemistry in terms of water, hydrocarbons, minerals and interfaces on the young Earth. In addition phylogenetic studies based on RNA sequences seem to indicate that the redox driven metabolism is earlier than the photo-driven metabolism. This all points to a subsurface redox driven origins of life [15].

4. Thermodynamic and kinetic aspects of the proto-organism

The key thermodynamic and kinetic issues associated with the proto-organism can be divided into five main problems: (i) the assembly of the lipid aggregates, (ii) the association (loading) of the proto-genes and the metabolic precursor molecules in the lipid aggregates, (iii) the thermodynamics and kinetics of the metabolic processes, (iv) the thermodynamics and kinetics of the template directed ligation and replication processes, and (v) the thermodynamic and kinetic of the full system replication.

Under standard conditions (constant pressure and temperature) the change in Gibbs free energy ΔG for a given physicochemical process is defined as

$$\Delta G = \Delta U + P\,\Delta V + \Sigma_i\,\mu_i \Delta n_i - T\,\Delta S = \Delta H - T\,\Delta S \qquad (1)$$

where ΔU essentially defines the overall change of internal (molecule to molecule) potential energy, ΔS defines the internal entropy change, and we assume that we can ignore volume changes ΔV. P is the pressure and T is the absolute temperature. Since the system does involve chemical

reactions through the metabolic processes (breakage and formation of covalent bonds), additional terms $\Sigma_i \mu_i \Delta n_i$ are needed to account for the changes in ΔG due to the changes in the presence of chemical species. Δn_i defines the concentration change in species i and μ_i defines the chemical potential corresponding to species i. ΔH, the enthalpy, is just a short hand for the first three terms. The thermodynamic potential ΔG is discussed in connection with the key proto-organism processes in Figures 10, 14 and 15. Although most of the proposed subprocesses involved in the proto-organism dynamics are well know from earlier experiments in different contexts, only very little is known about an integration of these reactions which is necessary to form the full life-cycle of the proto-organism.

Recall that thermodynamics does not involve any time scales. The thermodynamic properties are determined by the equilibrium conditions for the system including the equilibrium concentrations for the involved physicochemical species. The system time scales, as defined by the kinetics, are generated by the details of the assembly and reaction mechanisms. The time scales are influenced by catalysis, but catalysis does not change the thermodynamic conditions. There is a clear connection between the thermodynamic conditions and the kinetic reaction constants as reaction constants are defined by the ratio between the thermodynamic equilibrium concentrations.

4.1. Thermodynamics and kinetics of the lipid aggregation processes

Self-assembly of lipids in water is an extensively studied area, which includes experimental and theoretical work by our team. Self-assembly fundamentally involves microscopic processes of the form

$$A(n) + A(m) \leftrightarrow A(n+m) \tag{2}$$

where *A(n)* is a lipid aggregate of size *n* (consists of *n* lipid molecules). One special case of (2) is important: $n = 1$, $m >> 1$, as it defines the assembly and disassembly of individual lipid molecules into a micelle or vesicle. Equation (2) also accounts for the fusion of smaller aggregates and the break down (division) of larger aggregates. Breakdown of larger components into multiple components (more than two) can be described through successive iteration of (2).

In this subsection we examine three interrelated theoretical approaches to address issues for the molecular self-assembly of aggregates: (i) thermodynamics, (ii) kinetics and (iii) detailed 3-D simulations, where the mathematical complexity tends to grow in each successive approach.

Lipid self-assembly has two thermodynamic drivers. One driver, perhaps the easiest to understand directly, is a decrease in ΔU as potential bond energy through self-assembly. Through the self-assembly the hydrophobic tails of the amphiphilic molecules are pushed together and out of the way of the water molecules such that more of the stronger water-water bonds can form. This yields lower internal potential energy because the water-water hydrogen bonds (~20 kJ/mol/bond) are stronger than the water-hydrophobic dipole-induced dipole bonds (~2 kJ/mol/bond). Thus, lipid self-assembly has a tendency to minimize the interface between the water molecules and the hydrophobic molecules.

In addition lipid assembly has an entropy driver, which is actually stronger in absolute terms than the energy driver. This driver is due to an overall increase in the degrees of freedom for all molecules as the total water-lipid interface areas decrease. The water-lipid interface generates a layer, or network, of relatively stable water-water bonds. This happens because the

water next to the hydrophobic lipid interface only has less "choices" to build strong water-water hydrogen bonds. Thus, an entropy gain ΔS is generated in lipid self-assembly. The entropic part of the process is often referred to as the Hydrophobic Effect. Together the entropy gain and the decrease in internal energy yield a negative (Gibbs) free energy, $\Delta G < 0$, recall equation (1).

In the following we discuss a variety of approaches to uncover the nature of lipid self-assembly processes and the specific properties of the generated aggregates, e.g. size distribution, stability, and the aggregate dependence on the physicochemical parameters. A deeper understanding of the proto-container dynamics (as well as the dynamics of the other sub-systems) is necessary prior to the assembly of the more complex, full proto-organism.

The free energy change in the micellation process can be determined from the equilibrium between the concentration A of n free monomers and the concentration A_n of the n-aggregate (micelle of size n) concentration, recall equation (2), using the fundamental relation

$$\Delta G = RT \ln K \tag{3}$$

where R is the gas constant and the equilibrium constant is defined as

$$K = A_n / (A)^n. \tag{4}$$

Substituting (1) into (3) and rearranging the terms we get the well known relation

$$\ln K = (- \Delta H / R) (1/T) + (\Delta S / R). \tag{5}$$

Thus, the function of $ln\ K$ as a function of $1/T$ yields ΔH and ΔS from the line's slope and the intercept, respectively. Experimentally this relation can be constructed by measuring K at different temperatures.

Lipid phase diagrams are usually very complicated. Figure 10 (a) discusses general lipid self-assembly, and micellation is perhaps the simplest relevant example of such a process in the context of the proposed proto-organism. As a rule of thumb, the micellar structure may be derived from three simple parameters defining the lipid molecules: (i) the hydrocarbon chain length l_c, (ii) the head group size Δ, and (iii) the head group area a_0, as it can be defined on the surface of a micellar aggregate [81]. Obviously, we can define the total micellar surface area A and the (internal) hydrophobic micellar volume V_c as

$$A = n\, a_0 = 4\, \pi\, (l_c + \Delta)^2, \text{ and } V_c = {}^4/_3\, \pi\, l_c^3, \tag{6}$$

respectively, where n is the aggregation number (micellar size). Empirical studies show that we may approximate the resulting aggregates by

$$V_c/(l_c\, a_0) = \begin{cases} 0 - 0.33, & \text{generates spherical micelles in water} \\ 0.33 - 0.5, & \text{generates rods in water} \\ 0.5 - 1.0, & \text{generates lamellar structures in water} \\ > 1.0, & \text{reverse micelles in non-polar medium,} \end{cases} \tag{7}$$

which also depends on the lipid concentration. These relations may be useful as we may assume that the simple proto-container operates between spherical micelles and rod shaped

structures as we move between pure lipid aggregates and "loaded" aggregates containing larger amounts of precursor lipids, see Figure 10 (c). However, equation (7) is only a simplified version of the story.

Lipid self-assembly has a rich dynamics and the complex higher order structures of the assembled lipids depend on the properties of the specific lipid molecules. The assembly structure depends on the number and length of the hydro carbon chain(s) (*m* molecules of the type $-CH_2-$), the charge and structure of head group (W), as well as the properties of the solvent, which is normally water, together with e.g. pH, salt concentration, and ion mixture. A typical lipid molecule is about (*m*+2) 3 Angstrom in length ~ 2-4 nanometer, where m is the number of "mid tail" hydrophobic monomers. Up to a certain critical lipid concentration (the critical micellation concentration or CMC), no micelles can form, but at concentrations above the CMC, the micelles will start to assemble. CMC strongly depends on the lipid and the solvent. Many micelles start forming in water at concentrations between 10^{-2} and 10^{-4} M (mole/liter) with the higher CMC for nonionic amphiphiles as carboxyl acids [81]. The micellation time, the time it takes an initially "randomly" distributed solution of lipid molecules to assemble into micelles, is about 10^{-6} seconds.

The change of free energy contribution from each lipid molecule assembling into a micelle may also be factored out into the different components (monomers) of the lipid molecule:

$$\Delta G_{mic} = \Delta G_{mic}(\text{W-}) + m\ \Delta G_{mic}(\text{-CH}_2\text{-}) + \Delta G_{mic}(\text{-CH}_3), \qquad (8)$$

where ΔG_{mci}(W-) indicates the change of free energy due to the hydrophobic head group of the lipid, m ΔG_{mic}(-CH_2-) the change due to the m hydrophobic mid tail hydrocarbon groups, and ΔG_{mic}(-CH_3) the change due to the hydrophobic end tail hydrocarbon group [81]. The free energy gain associated with a typical (simple) lipid is about -30 kJ/mol (~ - 300 meV/lipid molecules) which means that each -CH_2- monomer contributes with about -2 kJ/mol (~ - 20 meV/hydrocarbon monomer). The head group's contribution is dismal. For comparison the thermal energies at 300 K (~25 C) are about 25 meV. Typical aggregate sizes *n* for non-ionic micelles range from about 25 to several hundred lipid molecules.

Figure 10 here

To experimentally determine the micellation size *n*, we can define *M* as the concentration of monomers in *n*-aggregates at equilibrium, and we can define M_0 as the total concentration of monomers. Then $M_0 - M$ defines the concentration of free monomers and equation (4) can be rewritten to

$$ln\, M = n\, ln\, (M_0 - M) + ln\, K\,. \qquad (9)$$

It turns out that both M_0 and *M* can be determined directly by means of experimental observables [Mittal et al., 1984] as long as the micellation causes a change in the absorbance spectrum, which it usually does,

$$M_0 = E_{mice}\, K\, L \qquad \text{and} \qquad M = R_a\, (I_1 - R_m\, I_2) / (R_a - R_m), \qquad (10)$$

where E_{mice} is the extinction coefficient at the absorption maximum for pure micelle; K is again the equilibrium constant, L is the length of the optical path; R_a is the ratio of micelle absorbance at λ_1 to that at λ_2, which can be obtained from the pure micelle spectrum; R_m is the ratio of the monomer absorbance at the same two wavelengths λ_1 and λ_2, which can be obtained from the pure monomer spectrum, and I_1 and I_2 are the individual absorbencies of micelle at λ_1 and λ_2 at any given concentration M_o. The slope from the line defined by *ln(M)* as a function of $ln(M_0\text{-}M)$ will give the average micellation size based on the equation (9).

The details of the aggregation size distribution is not fully understood and strongly depends on the properties of the individual lipid molecules and the physicochemical conditions, which translates into how the individual lipid molecules join and separate in solution, how lipid molecules join and leave existing micelles, how micelles becomes unstable and split, as well as on how smaller micelles join to form larger micelles, recall equation (2). As a simple example we may define the detailed lipid aggregate kinetics to consist of only three processes ignoring the assembly of larger aggregates

$$\begin{aligned} A_1 + A_n &\xrightarrow{k} A_{n+1}, \\ A_m &\xrightarrow{s} A_n + A_{m-n}, \\ A_n &\xrightarrow{d} nA_1, \end{aligned} \tag{11}$$

describing how lipids join and leave micelles, how larger micelles divide into two smaller micelles and how micelles break apart, and where the parameters *k, s*, and *d* are assumed known functions of the aggregate size. Under the assumption that we can define the size of a lipid aggregate as a

continuous variable *x*, it can then be shown [95] that the dynamics of the probability *P*(*x,t*) of finding aggregates of size *x* at time *t* is determined by

$$\partial_t P(x,t) = -\partial_x (k(x)P(x,t)) - s(x)P(x,t) - e(t)d(x)P(x,t) + \int_x^\infty T_x(y)s(y)P(y,t)dy\,, \qquad (12)$$

where ∂_t and ∂_x are the time and aggregate size derivatives, where the term $-\partial_x (k(x)P(x,t))$ defines the single lipid association process, the term $-s(x)P(x,t)$ defines the aggregate division/splitting process as an aggregate of size x is split into two new aggregates, and the term $\int_x^\infty T_x(y)s(y)P(y,t)dy$ defines the addition of new aggregates of size *x* from the division processes of aggregates of size larger than *x*. $T_x(y)$ defines the probability of generating an aggregate of size *x* (and one of size *y* - *x*) from an aggregate of size *y*. Finally, the term $-e(t)d(x)P(x,t)$ defines the outflow (or normalization) from the aggregates from the steady state chemostat. This term ensures that *P(x,t)* is normalized to be a probability distribution.

Analytical steady state solutions for (12) can e.g. be obtained with *d(x)* =1 and *k* a constant, at which point the equilibrium aggregate size distribution is defined by

$$P(x) = \frac{2}{\sqrt{k}} x\beta(x,k)e^{-\beta(x,k)x}\,, \beta(x,k) = \frac{1}{\sqrt{k}} + \frac{x}{2k}\,. \qquad (13)$$

For details please see [95]. Figure 11 discusses the micellar size distribution from (13).

Figure 11 here

The most comprehensive theoretical understanding of the detailed lipid self-assembly dynamics probably comes for a study of the processes in 3-D bottom up simulations.

Depending on the relevant level of details (time and length scales) for the questions we ask, we may use: (i) the molecular dynamics (MD) method for the small length and time scale processes, (ii) the MD lattice gas method for the intermediate length and time scales, and (iii) the lattice Boltzmann or Ginzberg-Landau continuous methods for the larger length and time scales. In Section 5 a more detailed discussion of such a multilevel coupling of 3-D simulation can be found. Perhaps the most appropriate simulation method for directly addressing molecular self-assembly processes is the MD lattice gas method, although it does not have enough predictive power as a stand-alone method.

The MD lattice gas simulation technique [52,53,66,77] addresses molecular interactions at nanometer to micrometer length scales and over time scales up to milliseconds on workstations and seconds on supercomputers. This intermediate range suits them to model molecular self-organization and self-assembly processes involving ions, monomers, complex polymers, polymer aggregates (supermolecular structures), complex charged surfaces, and chemical reactions. The MD lattice gas method enables direct simulations of amphiphilic polymers self-assembling into micelles in aqueous environments (where the water molecules are simulated explicitly). The computational technique is a combination of a traditional molecular dynamics and a lattice gas cellular automaton. Molecules (not atoms) are spatially confined to a three-dimensional lattice (i.e., both translational and rotational degrees of freedom are discrete), whereas the momentum space is continuous (particle velocity distributions are Maxwellian).

The translational movements are probabilistic and proportional to the velocities of the molecules. The rotational dynamics is an annealing process, which attempts to minimize the local potential energy. The collision operator locally preserves momentum, angular momentum

and kinetic energy. The particle interactions are represented by continuous fields, which are defined at discrete points between the molecular lattice points. During translational and rotational movement of the molecules kinetic and potential energy is exchanged, but the total energy is conserved. The main advantage of using a discrete lattice is that the discreetness of the spatial dimensions numerically stabilizes the simulations and therefore enables simulations time- and length-scales unreachable by traditional molecular dynamics techniques.

Using a continuous momentum space together with non-trivial molecular interactions gives a model rich enough to be thermodynamically interesting for molecular self-assembly processes. Figure 12 depicts a preliminary MD lattice gas study of PNA partition and loading in micelles, which we discuss in Section 4.2.

Figure 12 here

4.2. Loading (solubilization) of the lipid aggregates

A thermodynamic driving as describe above also occurs when hydrophobic polymers assemble in water, see Figure 10 (b). In this case less organized hydrophobic aggregates or "blobs" are formed, and depending on the overall hydrophobicity of the polymers a clear phase separation may occur as e.g. seen for mineral oil in water. The oil-water picture may be a useful analogy for the lipid precursor assembly in water.

The thermodynamics of the loading of the lipid aggregates with precursors and templates is discussed in Figure 10 (b) and (c). As discussed in Figure 10 (b) the precursor lipids, which are hydrophobic polymers, will most likely act similarly to mineral oil in water and phase separate by forming large non-structured hydrophobic clusters in the water. Adding amphiphilic lipid

molecules, which are produced from the fragmentation of these precursor molecules, tends to break up the larger hydrophobic clusters into smaller aggregates and thereby form more interface area [58]. This occurs because the lipids act as surfactants placing themselves with their hydrophobic tails in the precursor blobs while sticking their hydrophilic head groups into the water phase. Contrary, adding a small amount of all hydrophobic precursor lipids to a solution of micelles in water tends to "load" the micelles with the precursor molecules making the micelles bigger by "swelling". Since the micelles now have a larger hydrophobic volume they may be more rod-like when loaded, recall equation (7).

Thus, as more surfactants (lipids) are being generated within the loaded micelles, as discussed in Figure 10 (c), the larger and more hydrophobic blobs will have a tendency to break up into smaller aggregates. In the end, if all precursor lipid is transformed into lipid, only micelles are left, and they will be of a particular size typical for the give surfactant and the physicochemical conditions. The important lesson to remember in the proto-organism context is that the precursor "loaded" micelles have a tendency to be larger of size than the structures formed by pure lipids. Thus there is a thermodynamic driving to "divide" the aggregates into smaller units once the metabolic processes have run their course and transformed all the (mineral oil-like) precursor lipid into functional surfactant molecules. This aggregate division process of course has to be coordinated with the PNA replication process.

Also the precursor PNA molecules will attach to the lipid aggregates, how strongly depend on the detailed composition of the PNA backbone. As a first approximation we assume that the base components of the PNA do not have a preference for the lipid phase (not hydrophobic; they are identical to their DNA counterparts), such that we can assume that the main hydrophobic part of the PNA is within the backbone. Then we may approximate the free energy advantage for

joining the lipid phase by the adding the ΔG component contributions from the backbone. For each PNA monomer the backbone polymer piece is composed of a set of molecular groups (monomers) [91], where the brackets indicate the repeat of the structure

$$\ldots NH\text{-}]\text{-}COH\text{-}CH_2\text{-}N(Base)\text{-}CH_2\text{-}CH_2\text{-}NH\text{-}[\text{-}COH\text{-}\ldots \qquad (14)$$

The three CH_2 monomers are the only clearly hydrophobic groups of the backbone so as a first approximation we may assume that the free energy of lipid attachment is less than 3 $\Delta G_{attach}(CH_2)$, which is ~ –6 kJ/mol, recall the discussion of equation (8), which is a weak free energy advantage. The free energy advantage of lipid attachment can be made much stronger if necessary e.g. by the attachment of strongly hydrophobic peptide groups as phenylalanine, valine, or leucine, recall the discussion in Section 3.3. Also a direct attachment of the photo-sensitizer to the PNA backbone, recall the discussion in figure 6, will surely anchor the PNA backbone in the lipid phase. Preliminary MD lattice gas simulation studies of the PNA lipid aggregate (micelle) interactions are discussed in Figure 12.

In summary, it should be noted that we at present do not know the detailed overall dynamics of the complex aggregates as a function of the lipids, the loading of precursor molecules, and the PNA. It is clear that self-assembly will occur due to the thermodynamic driving as described above, and we may also assume that a loaded micelle will become less stable as the precursor lipid within the aggregates are metabolized into functional surfactants (lipids). However, either direct experiments or detailed multilevel 3-D simulation investigations have to be performed before we can make predictions about the dynamics of these complex aggregate systems.

4.3. Thermodynamics and kinetics of the proto-metabolism

Photo-fragmentation processes are found both in non-biotic and biotic systems and technological applications exist Chen and Whitten and Chen et al [13,14]. In Section 3 we discussed several implementations of photo-fragmentation processes that produce lipids and/or oligomers for the templating process, where the oligomers also act as photo catalysts. The overall energetic scheme of such a process is depicted in Figure 14, which defines a photo-fragmentation of precursor lipid and precursor oligo molecules into functional lipids and oligomers. Since the fragmentation process involves a breakage of a covalent ester bond the photon energy must be larger than this covalent bond energy. Typical free energy differences between the ground state and the excited state are 400 kJ/mol and typical free energy differences between the excited state and end products are 300 kJ/mol [62]. The actual energy levels as well as the kinetic reaction constants depend on the physicochemical details.

Photo-induced fragmentation reactions have been studied by our team as well as by other groups [13,14,71]. The key issue for a successful and high yield photo fragmentation reaction is to bring the reactant into a state where the bond to be broken is significantly weakened. Thermodynamic properties and the quantum efficiencies of these reactions depend upon the molecular structure of the reactants in addition to the reaction conditions. For the fragmentation reaction of phenacyl esters ($PhCOCH_2$-OCOR) as mentioned in Section 3, which can undergo photo-induced C-O bond scission in the presence of appropriate electron donor molecules, it is suggested that the conformation of the anion radical intermediates may play an important role in determining the rate of the bond fragmentation, and consequently the quantum

efficiency of the reaction. The reactivity of anion radical can be predicted based on the thermochemical cycles shown in Figure 13.

Figure 13 here

Since the free energy change is independent of pathway, the free energy change for the anion radical cleavage (ΔG_C) can be estimated from equation (15) where ΔG_{BDE} represents the bond dissociation free energy of the neutral molecule, ΔG_R the free energy required for one electron reduction of the radical fragment, and ΔG_M the free energy of one electron reduction of the neutral molecular.

$$\Delta G_C = \Delta G_{BDE} + \Delta G_R - \Delta G_M. \qquad (15)$$

The facile fragmentation (ΔG_C) of the photo generated anion radical upon one electron reduction (ΔG_M) can then understood from the thermochemical cycle depicted in Figure 13: In general, the electron affinity (or reduction potential) of a neutral compound of starting material is always higher than that of the corresponding fragment radical ($\Delta G_R < \Delta G_M$). Therefore there is frequently a sizable difference in redox potentials, or electron affinity, between the neural starting material and the fragments produced by homogeneous cleavage of the C-O bond (ΔG_{BDE}) on the one hand, and the starting material and the cleavage of the C-O bond in the anion radical (ΔG_C) on the other hand, which leads to the prediction of greatly reduced C-O bond energies in the anion radical compared to the neutral molecule. As a consequence, the

dissociation energies of specific bonds will be drastically lowered in the anion radicals compared to the corresponding neutral compounds.

To be more quantitative, the difference in free energy of reduction (ΔG_R - ΔG_M) can be shown[8] to be proportional to (with some factor F) the difference in reduction potential of the radical and the neutral molecule ($E^{ER}{}_{½}$ ($A^{\cdot}$) and $E^{RE}{}_{½}$ (AB)). Thus, the free energy change for anion cleavage, equation (15), can be rewritten as equation (16), where F defines the before mentioned proportionality factor. To evaluate the process by experimentally accessible observables such as enthalpy (ΔH), equation (16) can be transformed into equation (17) using equation (1).

$$\Delta G_C = \Delta G_{BDE} - F\,[E^{ER}{}_{½}\,(A^{\cdot}) - E^{RE}{}_{½}\,(AB)] \qquad (16)$$

$$\Delta H_C = \Delta H_{BDE} - F\,[E^{ER}{}_{½}\,(A\,) - E^{RE}{}_{½}\,(AB)] - T\,[\Delta S(AB) - \Delta S(AB^{-\cdot})]. \qquad (17)$$

The last term in equation (17) represents the difference in entropy change for the cleavage of the neutral molecule and the corresponding radical anion. If it is assumed that for these large delocalized species, the solvation of the radical anion is similar to the solvation of the carboanion fragment formed upon cleavage, and further if it is assumed that the solvation entropies of the neutral molecule and the radical fragment are also similar to each other (or are relatively small compared to the entropies of the charged species), this entropy term in equation (17) will be very small, i.e., $\Delta S(AB) = \Delta S(AB^{-\cdot})$ and thus the bond enthalpy for the bond cleavage can then be estimated from equation (18).

[8] Values for $E^{RE}{}_{½}$ are tabulated and can be found in standard tables [62].

$$\Delta H_C = \Delta H_{BDE} - F\,[E^{RE}_{\frac{1}{2}}(A^{\cdot}) - E^{RE}_{\frac{1}{2}}(AB)]. \qquad (18)$$

Relating the bond dissociation enthalpy to the rate of cleavage of the radical anion require the further assumption that there is no additional activation energy for this process, i.e., there is no activation energy for the reverse reaction of the radical fragment with the carboanion. While this may be true for the reaction in the gas phase, in solution there may be some activation energy associated with solvent reorganization, particularly in polar solvents. Nevertheless, equation (18) should be used as the first approximation to predict reactivity of the anion radical. For example, the redox potential for the reduction of phenacyl ester $PhCOCH_2$-OCOR is estimated to be -0.7V vs SCE (standard calomel electrodes) and the redox potential for the reduction of radical $PhCOCH_2^{\cdot}$ is estimated to be great than 1.3 V vs. SCE. Based on equation (18), the bond cleavage energy for phenacyl ester radical anion will be much less compared to the bond cleavage energy for phenacyl ester, which should be manifested by rapid fragmentation reactions in photo-generated radical ion.

Figure 14 here

As a complement to the above thermodynamic calculations detailed quantum chemical calculations (*ab initio*) could also be applied e.g. in terms of deriving the bond energies for the different excitation stages of the photo driven processes and to provide key parameters in the higher-level simulations [2,85,25,92,93]. Quantum-chemical calculations are currently feasible for finite molecular systems up to hundreds of atoms in size, although they are excessively expensive computationally. Such calculations could serve as a basic tool for understanding main

physical and chemical processes in small molecular systems and feed calculated input to the coarser models as well as to the experiments and the synthesis.

Quite a bit is known about the experimental and theoretical properties of charge transfer in the DNA double helix under different conditions [10,69,106,40,43]. Based on this body of work, we can make well-informed guesses about the charge transfer properties of PNA[9]. We expect that charge transfer is faster in PNA than in DNA in particular if the PNA duplex is in a lipid phase. The stronger the hybridization energies are, the colder it is, and the fewer ions and water molecules we have near the PNA, the faster the charge transfer is expected to be. Charge transfer rates around 10^{12} s^{-1} are to be expected.

The key kinetic question for the proto-metabolic system is to develop a set of processes with a good quantum yield as discussed above. This means that the balance between the back reaction (see Figure 14) and the electron charge transfer reaction is in favor of the charge transfer reaction. The typical reaction constant for a photo activation process is ~10^{-15} sec and the typical back reaction constant is ~10^{-10} sec. The charge transfer reaction constant thus needs to be somewhat smaller ~10^{-11} sec, which we also expect it to be from the above discussion. The rate-determining factor for the production of building materials is given by the efficiency of the charge transfer chain. Thus, the presence of an electron relay system determines whether the yield of new lipid and PNA is high enough to counter the unavoidable degradation and delusion processes.

Since there is no feedback with amplification (autocatalysis) neither in the photo-segmentation process nor in the general redox process, the overall kinetics for the proto-metabolism is exponential damped as a function of the driving (the energetic input). To demonstrate this important quality of the kinetics, and not to focus solely on photo-driven systems,

[9] Personal communication with Alexander Burin, George Kalosakas, and Kim Rasmussen.

we can for example assume a general water-membrane-water redox scheme for a vesicle as discussed in Figure 2. That gives us the following well-known simplified redox scheme

$$\begin{aligned} A + C &\xrightarrow{k_{AB}} B + D, \\ D + F &\xrightarrow{k_{CF}} C + E, \end{aligned} \tag{19}$$

where the input to the system is energy rich molecules A and precursor molecules F. The waste products are B. The hydrophobic redox compound in the membrane exists in two oxidation stages C and D, which it is being cycled between. A acts as the external energy input to C (in the membrane), which is then transformed into D (also in the membrane), which in turn can modify some other compound F (precursor) into E (proto-organism building blocks). The solution to the kinetic equations derived from (19) yields (under symmetric initial conditions assumptions and for $k_{AC} = k_{BF} = 1$)

$$\begin{aligned} A(t) &= F(t) = a_0 \, exp(- c_0 \, t) \\ B(t) &= E(t) = b_0 + a_0 \, [1 + exp(- c_0 \, t)] \\ C(t) &= D(t) = c_0, \end{aligned} \tag{20}$$

where a_0, b_0, and c_0 defines the initial concentrations for A, B, and C. As expected, it essentially is a smooth exponential decay towards thermodynamic equilibrium unless A is continuously added and B is continuously removed. As for the photo-driven reactions, the key redox kinetic question is to ensure high enough reaction rates k_{AC} and k_{BF}, such that the system produces enough building materials to sustain itself and grow.

In summary, the rate limiting steps for the proto-organism stemming from the metabolic processes will either be: (a) too low yield in the production of functional oligomers, and/or (b) too low yield in the production of lipids molecules. In particular if the proto-metabolic reactions are of photo-chemical nature, it is important to note that: (c) The reactions involved in the photo-kinetics are many, many orders of magnitude faster than the kinetics involved with the lipid aggregate (assembly, growth, and division) as well as the template directed ligation and replication reactions (see the next subsection). (d) The proto-metabolic reactions will asymptotically reach a steady state production of the building blocks for the proto-organism, if the system is continuously feed with the energy rich electrons or photons and the appropriate precursor molecules.

4.4 Thermodynamics and kinetics of the proto-genes

The key thermodynamic questions for the proto-gene replication process are: (i) How tightly is the PNA backbone associated (free energy) with the lipid aggregate? We already discussed this question in section 4.2. (ii) How tightly associated are the PNA-PNA complexes (fee energy) in water at the water/lipid interface, and in the lipid phase? (iii) What is the activation barrier (free energy) for the template directed amide bond formation in the membrane? Without a template directed ligation (polymerization) no replication. (iv) Under which conditions are the PNA-PNA-lipid-water system in a "balanced state" with both hybridized and on-hybridized PNA? Or, under which external conditions can we alter this balance towards disassociation without destroying the aggregate complex? Disassociated PNA-PNA duplexes are important for a continued template reactions, which requires single stranded PNA.

Melting curves for PNA duplexes indicate a high melting temperature of around 70C^{o} even for PNA decamers (a higher melting temperature than that for the corresponding RNA or DNA decamers) [79]. To destabilize the PNA duplex it may be problematic just to raise the temperature, since we can easily loose the integrity of the lipid aggregate at such high temperatures, although some surfactant aggregates have their most stable regime at around 60 C [80]. Experimentally it is shown [91] that the free energy of base paring (hydrogen bond formation in the PNA duplex) is about ΔG_{base}~ -8 kJ/mole per base pair or PNA monomer for PNA in water, which corresponds to a melting temperature of about 70C at low salt concentration (The melting temperature decreases some with increasing salt concentration, but not very much). Thus, the PNA-PNA association is stronger than the estimated ΔG_{lipid} PNA-lipid association, which is less than -6 kJ/mol per PNA monomer backbone segment. However, this balance may be significantly changed in the other direction by modifications of the PNA backbone, recall the discussion in section 4.2. Also, the attachment of hydrophobic groups to the PNA backbone directly influence (decrease)[10] the value of ΔG_{base}. As an example, let us assume that we attach a small hydrophobic side chain to the PNA backbone (-CH-$(CH_3)_2$). As a first approximation we may assume that ΔG_{base} stays the same (~- 8 kJ/mol), but this side chain will approximately double the hydrophobisity of the backbone to about ~ - 12 kJ/mol per PNA monomer. Shifting this balance may also create a "hydrophobic torque" from each of the side chains that weaken the base pairing and thereby generates a metastable PNA complex.

The stability of PNA-PNA duplexes in organic solvents and in liposomal (miceller) solutions is not yet known. PNA-PNA duplex formation in aqueous solution is enthalpically driven with a large entropy loss [80]. Such duplexes are stabilized by a combination of

[10] Peter Nielsen, private communication.

hydrophobic, base stacking interactions and hydrogen bonding between the base pairs. We believe that the hydrophobic contribution dominate in water as the nucleobases to a large extent exchange the hydrogen bonding to water molecules with those to the complementary nucleobase. In contrast it would be predicted[11] that the situation should be opposite in organic solvents, where the formation of hydrogen bonds should provide a significant energetic stabilization whereas base stacking should contribute less energy. Preliminary data in water-organic solvent mixtures indicates a decreased PNA-PNA duplex stability in non-aqueous solvents[10]. Thus the thermodynamic of these systems will eventually have to be determined by thermal denaturation (melting) experiments. Furthermore, it is expected that PNA duplexes will be relatively more soluble in lipids than the single stranded PNAs due to shielding of the hydrophilic, hydrogen binding faces of the nucleobases.

These are all beautiful question to address with detailed MD simulation combined with experiments.

At present we do not know the exact activation energy associated with a lipid assisted amide bond formation in the template directed PNA ligation, but lipid assisted amide bond formation from free peptides is experimentally demonstrated by Luigi and coworkers [5] with reasonable kinetic yields (products in 24 hours). These experiments also demonstrated a certain differences in the polymerization yields depending on the detailed peptide chains. It is not unreasonable to expect a higher yield in a template directed amide bond formation, since the free energy barrier associated with the amide bond formation in the juxtaposed situation of a template directed ligation should be lower than it is for amide bond formation between free polypeptides. In addition please recall Section 3.3 that simple catalysts or changes in the external conditions could

[11] Peter Nielsen, private communication.

also enhance the ligation kinetics. Experiments will eventually be able to settle the PNA ligation question.

Assuming the above thermodynamic issues have been settled, a single-stranded PNA template *C* is located at the lipid-water interface, exposing the hydrophilic part of the bases into the water, while the hydrophobic backbone is sunk into the lipid layer. The system is fed with oligos *A* and *B* from the aqueous phase (or from oligos already attached to the lipid aggregate) so initially only partial double strands (*AB* or *AC*) are obtained by specific hydrogen bond formation. Once a full double strand *ABC* is formed the surface of the whole surface of the complex is hydrophobic so

Figure 15 here

it will tend to sink further into the lipid phase. Here the ligation step can occur, which requires a peptide bond formation. The ligation results in two full plus and minus templates *CC'*. A small fraction of *CC'* will diffuse back to the lipid-water interface where it may disassociate into its two strands, perhaps as the system goes through some cycle in external salt concentration or temperature. The balance between the hybridized *CC'* complex and the free *C* and *C'* strands depends on the hybridization energies (which depends on the PNA backbone properties), the temperature, the pH and the salt concentration of the aqueous phase and several other factors. The important ingredient in our reasoning is that both *C* and *C'* remain anchored at the surface of the lipid aggregate such that the proto-genes are permanently associated with the lipid aggregate. The kinetics of the replication is composed of two templating processes, followed, by a template directed ligation step, with a subsequent disassociation, which can be expressed as

$$
\begin{aligned}
&A + C \longleftrightarrow AC, B + C \longleftrightarrow BC, \\
&AC + B \longleftrightarrow ABC \longleftrightarrow BC + A, \\
&ABC \longrightarrow CC' \longleftrightarrow 2C
\end{aligned}
\tag{21}
$$

where all reactions are reversible except for the ligation process. A detailed discussion of the dynamics of this process can be found in [78], where we use the methods developed by [104,88], and [8]. Perhaps the most interesting finding is that the kinetics supports a co-existence of multiple proto-genes due to an overall parabolic growth rate for the template directed replication process

$$
c(t) \sim t^2, \tag{22}
$$

where c is the total concentration of template in all its association form equation (21). The parabolic growth dynamics is generated from the interaction between a strict replication process and a product inhibation process. Note that this growth law is ruled out by the original quasispecies model developed by Eigen,1971 [20]. The quasispecies model does not allow for coexistence beyond the mutant cloud of the master sequence in homogenous solution. The possibility of fragmented genomes in principle opens up for a much richer set of scenarios for further evolution. However, the template directed gene replication at the lipid interface, of course, does not bypass the limits on the amount of heritable information posed by the error threshold [20]. However, to address the overall consequences of this growth law, a more complete reaction kinetic study is needed, which integrates the details of the lipid aggregates growth and reproduction kinetics as well as the metabolic kinetics.

To summarize, two key experimental questions are still open with respect to proto-gene self-replication: (i) Although Luigi and coworkers have demonstrated peptide oligomer polymerization catalyzed by a lipid aggregate [5], the same process has not yet been demonstrated for PNA-like molecules. The thermodynamic problem associated with this process is discussed in Figure 14 (b), where the free energy barrier of activation is depicted. (ii) Another open question is how to accomplish the necessary dehybridization of the double stranded complex *CC*', to allow a new templating process to occur. It is not yet clear under which conditions an appropriate force hierarchy can be established to form metastable PNA duplexes. The rate limiting steps in this lipid assisted PNA replication scheme are still to be determined experimentally, but we strongly suspect that both the ligation- and the dehybridization processes define the two most problematic (and slowest) reaction steps.

4.5. Thermodynamics and kinetics of the full system replication:

Life-cycle of the proto-organism

To complete the full life-cycle of the proto-organism, defining the generation of one new copy of itself, requires (i) the production of enough lipid to bud off a new aggregate coordinated with (ii) the production of (at least two) new oligomers that are ligated to form (one or more) minus templates, and finally (iii) a partition of the templates which each aggregate has at least one template.

In the following we need to know that proportions of the proto-organism. The weight of the proto-organism is given by the approximately 100 lipid molecules, the PNA 10-mer, and one or more photo-sensitizer molecules, which comes to about 100×(200g/mol) for the lipids

~12 carbon molecules + 10×(192 g/mol per PNA monomer) + say 5×(250 g/mol) for the photo-sensitizer molecules (twice that much if each PNA monomer has a sensitizer molecule attached), which is approximately 23,000 g/mol. Thus, a single proto-organism weighs 23×10^{3} (g/mol)/6×10^{23} (mol^{-1}) ~ 4×10^{-20} g, which means that it is ~ 10^{7} times smaller than the smallest known unicellular organism (at ~ 0.1×10^{-12} g). With a density close to water (1 g/ml) the proto-organism volume is about 4×10^{-20} ml, with a diameter of about 4.5 nm (10^{-9} m). In summary: the proto-organism is unch smaller and much simpler than modern unicells.

Although we at this point do not know how to experimentally orchestrate the proto-organism's life-cycle, it should be clear from the calculations in the previous subsections (4.1-4.4) that multiple conditions should enable such an orchestration. Our picture of the full proto-organism life-cycle is discussed in Figure 16.

The experimental conditions could e.g. be controlled such that the photo-driven PNA oligomer (3- to 5-mer) generation of PNA precursors occurs first, followed by the PNA template (a 6- to10-mer) directed ligation. It is assumed that two or more PNA 6- to10-mer strands can stay attached to the same aggregate. Now having generated new (negative) proto-gene templates, precursor lipids could be released into the solution, which would result in a loading of precursors into the aggregates. The lipid aggregate swell up as the mineral oil like precursor lipids are incorporated. As the proto-driven generation of new lipid molecules creates more molecules that function as surfactants, the larger aggregate becomes unstable and splits into two new stable aggregates. If several PNA templates (balance between the single stranded PNA and the duplexes necessary) are attached to the original aggregate a certain (random) partition between the two lipid aggregates will occur. If we assume that the rate limited step is the template directed ligation and the subsequent dehybridization process, we have reasons to

believe that the full life-cycle could occur in 24 hours or less, just based on the kinetics of the amide bond formation process reported by Luisi and coworkers [5].

The kinetic relation between the PNA sequence and the photo-metabolitic processes determines the evolutionary properties of the proto-organism. Recall that the metabolic kinetics is enhanced by a certain PNA sequence (GAGA…), which is the basis for an error prone, coupled aggregate/proto-gene selection, defining the elements of a Darwinian evolutionary process, and which is one of only a few known universal principles of living systems. However, recently theory has been developed to explain the universal nature of allometric scaling laws relating mass to other contemporary biological observables by West et al., 1997 and 2002 [101,102]. Modern biology satisfy remarkably simple empirical scaling relation ships e.g. between the mass of an organism and the power of its metabolism. The proto-organism, due to its minimalist design, defines a possible origo for this relationship.

Figure 16 here.

Using the proto-organism's mass and estimated metabolic rate, we obtain an effect of about 5×10^{-22} W for the 4×10^{-20} g system, which is about two orders of magnitude below the expected metabolic power limit, if the observed scaling relation ship was extrapolated down to the proto-organism size scale. There also seems to be an universal energy efficiency for the life-cycle of contemporary unicells, which is independent of cell size ($\sim 0.1\ 10^{-12}$g up to $\sim 10^{-6}$ g), cell organization (prokaryotic vs. eukaryotic), and cell metabolism (arobic vs. anaerobic), and it comes out to about [105]

$$\sim 500 \text{ J/g} \qquad \text{compared to} \quad \sim 1{,}200 \text{ J/g} \tag{26}$$

based on the life-cycle of the proposed proto-organism. 1,200 J/g "biomass", come from about 100 new lipid molecules and about two new PNA oligomers, which are able to polymerize (ligate) into a new template, within its life-cycle. Each new lipid molecule costs about 300 kJ/mol to produce and it weighs about 2000 g/mol, which yields about 1,500 J/g, which will be adjusted down a little if we also account for the generation of the new template with associated photo-sensitizer, thus giving ~ 1,200 J/g. Thus, the proto-organism has about half the energy to biomass efficiency compared to a contemporary unicell.

Two significant conditions, with opposite effect, have to be emphasized when making a comparison between a contemporary cell and the proposed protocell: (i) The proto-organism is on "life-support" compared to the environment most contemporary cells live in. Our proto-organism is provided with precursor molecules, which are quite similar in makeup to its molecular building blocks. This condition, however, should yield a *higher* proto-organism energy-biomass efficiency compared to the contemporary cell. (ii) Contemporary cells have evolved into highly efficient synthetic production units. One hand they are able to utilize resources of a much lower grade than their building blocks, and on the other hand, which is key to address the energy efficiency issue, their metabolic processes operate much closer to the reversible regime compared to the proto-organism. A modern aerobic cell is e.g. able to produce up to 36 ATP molecules for each ingested sugar molecule[12]. These 36 ATP molecules can drive the synthesis of multiple high-grade biomolecules [46].

From these two significant conditions, we may define two different useful "distances" or metrics from the above energy efficiency discussion. The first metric may be defined as the

(free) energetic distance between low-grade inorganics as CO_2 and NH_4, which the autolitothrophes[13] can utilize, and the high-grade organics as e.g. the provided precursor lipids (phenacyl ester) for the proto-organism. This distance defines the degree to which a novel life-form is on "life-support" compared to an autolithotroph.

The second metric may be defined as the (free) energy distance to an ideal reversible metabolic conversion of the energy input all the way through the free energy utilized to biosynthesis. Modern life has evolved enzymes, which are used to lower the activation energies both in the ATP production and in the ATP driven synthesis, thereby lowering the required free energy for each of the metabolic steps. This metric defines the overall free energy efficiency of the metabolism. Using the first, life-support metric, it's clear that our proposed proto-organism, does not address the question of how an autolitothrophic life-form can emerge, which e.g. is the focus of Morowitz, 1992 [61]. However, the first life-forms may not have been autolithotrophes after all, as the primitive Earth might have had a much richer composition of complex organics compared to what we find today. If that is true, it seems reasonable to assume that the first life-forms would have utilized the simplest possible metabolic processes and not at the onset developed more sophisticated and energy efficient metabolisms. Later life may have depleted this initial "easy food", which would have forced more efficient metabolisms to evolve in response to this environmental change. We know that life has dramatically changed so many other environmental conditions over the history of the Earth (e.g. the high oxygen content of the atmosphere is biogenic), so why not also the initial global composition of free organic molecules?

---

[12] A sugar molecule should also be considered as a high-grade molecule in this context.
[13] Microorganisms that can exist without any organic input, neither as building blocks nor as energy sources. The organisms are currently the true primary producers in our biosphere.

The universal unit of contemporary life is a collection of physicochemical processes in a negative-entropic environment, which defines a contemporary cell. The ability of a cell to generate negative entropy consists of its ability to utilize free energy and building blocks from the outer entropic environment and transform them into biomass. In this sense life is a natural free energy accumulating system. In the same manner the proto-organism defines a cluster of physicochemical processes in a negative-entropic environment (the lipid-PNA-sensitizer assembly).

Very importantly, it should be pointed out that in the above energetic calculations only the supplied external free energy used by the metabolic processes is included. The free energy associated with the self-assembly processes as they form the new aggregate, is not accounted for. For the proto-organism the estimated free energy generated from self-assembly of the new aggregate is approximately ~ 100×30kJ/mol lipid + ~ 150 kJ/mol PNA (with photo-sensitizer) which yields about ~ 3,150 kJ/mole. This has to be compared to about ~ 100×300 kJ/mol +600 kJ/mol ~ 30,600 kJ for the metabolic production of the new building blocks. Thus, the self-assembly processes accounts for only about 10% of the total free energy (negentropy) in one proto-organism life-cycle although no life as we know it would be possible without self-assembly. Also in modern life it is the self-assembled organelles that keep all functionalities together and makes life possible.

5. Discussion

In this section we summarize important issues already touched in the previous technical sections as well as address important issues about potential future implication of our work. We provide a simple "systemic" solution to the question, "How to generate a molecular proto-organism by rational design?", as we demonstrate how to integrate a container, a metabolism, and genes in a thermodynamically downhill manner. We provide experimental evidence for a subset of the processes, and we provide circumfencial- and theoretical evidence for the remaining processes in terms of thermodynamic and kinetic arguments. However, experimental evidence for the integrated processes (the full proto-organism) does not yet exist.

What is known and what is not known: Since it is just as important to emphasize what is still unknown as what is known about the proposed bridge between nonliving and living matter, we review the steps in Figure 16 and summarize the uncertainties. Processes (1-2) are well understood although the exact partitioning of PNA into the aggregates as a function of backbone details is not yet known. Processes (2-3) are well understood although the exact loading capabilities of the different lipid aggregates are not yet known. Processes (3-4) still have open questions, as it has not been demonstrated that PNA can act as a photo-catalyst for both the lipid- and the PNA precursors in praxis. Although every process is based on known chemistry there is no guarantee that the composed system will work as the known chemistry predicts - even if all the sub-systems work in isolation. For processes (4-5) it is known that the production of more surfactants will destabilize the aggregates, but the appropriate conditions for template directed ligation of PNA in lipids with a subsequent metastable duplex state have not yet been demonstrated experimentally. Most of this paper is devoted to an exploration of the underpinning issues associated with the processes (3-4) and (4-5). Finally, processes (5-6) follow from the generated lipid aggregate instability although

the partition of PNA templates (and photo-sensitizes) in the two daughter aggregates could in principle become an issue.

Origins of Life: It is hopefully clear that we do not claim to have addressed the historical question of the origins of life, although we do address the question of how to bootstrap a self-reproducing molecular machine, a "proto-organism" from simple non-biological molecular species. We are seeking an example of the simplest and most fundamental organizational and structural bauplan of a proto-organism, not necessarily the earliest proto-organism.

We are focusing on a light driven proto-metabolism, which may not have been the energy form utilized at the origins of life. Evidence elsewhere seems to indicate that the surface of the early Earth is too harsh an environment due to late accretion bombardments. The protected subsurface (the planetary crust) may very well have provided a richest chemistry in terms of water, hydrocarbons, minerals and interfaces on the young Earth. In addition phylogenetic studies based on RNA sequences seem to indicate that the redox driven metabolism is earlier than the photo-driven metabolism. This all point to a subsurface redox driven origins of life [15].

Artificial Life: It is argued in the previous sections, that the proposed proto-organism mainly addresses the open Artificial Life question #1,*"Generate a molecular proto-organism in vitro"*, but it also addresses important aspects of questions #4, *"Simulate a unicellular organism over it's entire life-cycle"*, and #5, *"Explain how rules and symbols are generated from physical dynamics in living systems"* [4]. Questions 4 and 5 are discussed in succession:

Question #4: It is clear that a stepwise approach is necessary to obtain a predictive simulation of the full life-cycle of a modern unicell. It is also clear that a full virtual proto-organism could be constructed using one of several different simulation methods or it could be constructed by means of a 3-D multilevel simulation coupling different levels of description. We

believe that the proposed proto-organism defines a useful stepping-stone for the work leading up to a modern virtual unicell. The proposed proto-organism together with other protocell proposals could act as drivers and test-beds for part of the effort in Systems Biology that ultimately aims at developing whole cell simulations, see e.g. [86]. Also, the virtual proto-organism constitutes a well-defined problem, which we hope that part of the Artificial Life community could engage in as it lives at the boundary between nonliving and living matter as well as between life as it could be and life as we know it.

Traditionally, theoretical and computational studies of 3-D molecular self assembly systems have been performed by microscopic lattice models, Ginzburg-Landau theories, membrane theories, simple lattice gas models, and lately also by molecular dynamics (MD) simulations. While many of these models have been successful in explaining qualitative experimental results (mainly producing topologically corrects phase diagrams), only MD includes structural details of the interacting molecules; water, lipids and other molecules. Many aspects of the dynamics of amphiphilic self-assembly depend strongly on the details in the structure of the molecules and their interactions. Water molecules form hydrogen networks, the length of the lipid polymer determines the stability of assembled structures, the strength and geometric properties of the interaction between amphiphiles, and water can directly affect the size distribution of micelles, etc. However, MD is still too computationally expensive for large-scale (long time) simulations of self-assembly processes. The discrete spatial dimensions in the molecular dynamics (MD) lattice gas technique, numerically stabilize the time evolution and therefore enables simulations time- and length-scales unreachable by traditional molecular dynamics techniques. At the same time, using a continuous momentum space together with non-trivial molecule interactions gives a model rich enough to be thermodynamically interesting. Therefore it seems reasonable to couple (i) the

molecular dynamics (MD) method [51,90,100] for the small length and time scale processes, (ii) the MD lattice gas method [53,66,77] for the intermediate length and time scales, and (iii) the lattice gas and lattice Boltzmann [7,21,41] or Ginzberg-Landau [38,87] continuous methods for the larger length and time scales simulating small proto-organism populations over many life-cycles. Although the MD lattice gas as a stand-alone technique can address 3-D molecular self-assembly processes it should be noted that this method alone is not reliable enough to make precise experimental predictions. Only in concert with the more detailed MD simulations *and* experimental results from related situations as a reference can we expect to make true predictions of experimental systems at a complexity level as the proposed proto-organism.

Question #5 in the proto-organism context, *"Explain how rules and symbols are generated from physical dynamics in living systems"*: The "emergence of a biological system from chemistry" comes from the observed functionalities of the proto-organism, one of these functionalities being the presence of (proto-) genes which can induce higher kinetic efficiency of the organism's metabolism. The nucleobases "encode" the metabolic efficiency in a very direct (physical) manner as they act as an electron relay chain. Nevertheless we content that it is a demonstration example right at the transition between physical dynamics and symbolic encoding dynamics, because of the nature of the relation between the functionality of the gene sequence and the way it is transmitted (encoded) to the next generation.

Continuing this way of thinking, we may be tempted to conjecture that "chemistry becomes biology" at the point where an organizational closure is established between physicochemical processes that naturally occur and interactions that "tweaks" or enhance their mutual yield. By this we may ask whether the initial biological organization just is a slight – later as life evolves it of course becomes a significant - control ("tuning") on top of physicochemical processes that occur in

any event? Our insight from molecular self-assembly processes and the impact of the gene sequences on the metabolic reaction kinetics suggests this view.

Living Technology[14] applications: Development of proto-cells will be an accomplishment that derives value not from a singular application, but from proto cells' ability to open the door to a huge array of applications. Threshold technological advances such as this include the silicon revolution (development of the transistor), mapping of the human genome, and quantum computing if we are successful. Pohorille and Deamer, 2002 [70] recently make the case for the technological potential of protocells:

> "Artificial cells designed for specific applications offer unprecedented opportunities for biotechnology because they allow us to combine properties of biological systems such as nanoscale efficiency, self-organization and adaptability for therapeutic and diagnostic applications … it will become possible to construct communities of artificial cells that can self-organize to perform different tasks and even evolve in response to changes in the environment."

Protocells have not been developed in the laboratory yet because the necessary cell chemistry and architecture has previously been regarded as too complex to yield to engineering applications. We believe the time is now ripe for their development for the following reasons: (i) A novel and significantly simpler design has now been developed and presented in this paper, (ii) Most of the necessary chemical mechanisms have been independently implemented *in vitro* and calculations indicate that a full system is feasible, (iii) Combinatorial chemistry and automated laboratory techniques enable more efficient search of possible operation conditions for protocells, (iv) Advances in the theory and simulation of self-assembly molecular self-organization have reached the point that they can propel the technology to realize the protocell

concept. (v) Last but not least, vast economic markets will become accessible as the protocell technology spans many application areas including, chemical, medical, environmental, structural, and computational technology. Protocells will be a key component of vast array of different application of Living Technology all based on autonomy, robustness, self-repair, intelligence, and self-replication. It also seems necessary to resolve the fundamental problem of assembling a living system before key macroscopic (robotic) applications become available [9].

Ethical concerns: Generating life *de novo* will create public reactions. The reactions will probably be along two lines: (i) Environmental concerns that the life-producing technology could "get out of control", and (ii) Religious and moral concerns, based on the beliefs that humankind must restrain from certain endeavors on grounds that they are fundamentally amoral.

Ethical issues related to creation of cells other than natural cells have a long history, but were taken up in the press recently as fallout from announcement of the research on minimal cell research of the Institute for Genomic Research [26,36]. Their press release and links to subsequent news stories may be found on their website[15]. The resulting minimal cell was dubbed "Frankencell" in the non-technical press. Similar concerns about nanostructures capable of proliferating in natural environments have been formulated in the Nanotechnology Community [56]. However, genetically coded proliferation itself is not a risk in itself (*cf.* its current ubiquitous use in molecular biology labs in the form of PCR).

Regarding the first class of objections ((i) above), the special control of proto-cell design and complementation, implicit in the approach proposed here, enables the designers to build in various forms of protection, e.g., dependence of the metabolism and the genetics on substances

[14] Concepts and perspectives about the emerging Living Technology is developed by Mark Bedau, John McCaskill, Norman Packard, and Steen Rasmussen, June 2001.

that do not occur naturally and must be provided to allow the cells' continued existence and proliferation.

Regarding the second class of objections ((ii) above), reaction in the press directly after the Frankencell ethical panel report had no explicit religious overtones, but subsequent references to Frankencell on web discussions had extensive negative religious overtones. The ethical issues seem small with respect to those presented by stem-cell and cloning research. In fact, proto-cells could provide an ethically cleaner route to providing subsets of functionality found in these research areas.

Finally we should also mention that the development of new carbon chemistry based life-forms at least theoretically could have biothreat application. Long term it should be possible to engineer or evolve e.g. unknown human pathogens [30]. As with all technologies the good and the evil always come in pairs. We agree with the conclusions of the ethical panel reviewing Venter's earlier Frankencell enterprise, which stated that: If the ultimate goal was to benefit humankind and if all appropriate safeguards were followed, then the project could be regarded as ethical.

[15] www.tigr.com

6. Conclusion

A proposal for a simple step-by-step, thermodynamic downhill assembly of a proto-organism is proposed. The resulting aggregate is capable of feeding form the environment, grow and replicate, die under environmental stress, as well as undergo Darwinian evolution. In short, the proto-organism is alive at least with respect to the most common definition of the term. This is perhaps the first concrete proto-organism picture where all molecules and processes are explicitly given and it is certainly much simpler compared to other recent proposals, Luisi et al 1994 [49], Szostack et al 2001 [89], Pohorille and Deamer, 2002 [70]. Physicochemical implementations of the proto-organism are presented and the associated experimental results and problems are discussed. Using micellar structures as a lipid proto-container with a photo-driven proto-metabolism it is shown that it is possible to produce lipids as well as PNA oligomers from appropriate precursor molecules and it is shown how the metabolic kinetics depends on the encoded proto-gene sequences.

The thermodynamic and kinetic properties of the detailed subsystems are derived or simulated and important conclusions can be drawn about the appropriate balance between key processes. Reaction kinetic analysis of the metabolic and the template directed replications is presented. Also reaction kinetic and detailed 3-D simulations of the lipid self-assembly are presented.

A full life-cycle analysis of the proto-organism, partly based on experimental results and partly derived from calculations, shows that it is operating at about half the overall energy efficiency (~1,2 kJ/g biomass) compared to contemporary unicells. The estimated generation time of the proto-organism is of the order of 24 hours, and the proto-organism can be as small as 5 nm in diameter and weigh about 10^7 times less than the smallest modern unicell (diameter ~ 0.5 μm). It should be noted that the free energies associated with the molecular self-assembly processes

only accounts for about 10% of the necessary metabolic energies. However, it is this 10% that enables life to exist as this modest free energy enables the formation of the functional basis structures.

Finally, it should be reiterated that the minimalistic thermodynamic coupling between the container, the metabolism, and the genes could also be realized using a redox-based metabolism. It might even be possible to use RNA or some other charged templater instead of PNA for the proto-genes by using oppositely charged lipid aggregates. We believe that the proposed proto-organism picture is quite general although we are aware that multiple alternative transitions between nonliving and living matter are possible. Despite several experimental pieces in the proto-organism puzzle are still unresolved, it is our hope that this detailed account of a particular bridge from nonliving to living matter can help focus more attention to this fascinating problem.

Acknowledgement

Development of the presented level of detail for the proposed proto-organism has taken several years and it has involved help from many people. A special thanks to Klaus Lackner, Luigi Luisi, Peter Nielsen, Shelly Copley, David Deamer, David Whitten, Peter and Barbel Stadler, Daniel Yamins, Stirling Colgate, and Kolbjorn Tunstrom, who have all contributed either with important ideas or with technical details to this work. Over the last year Mark Bedau, John McCaskill, and Norman Packard have also provided important insight and perspectives to the proto-organism effort, in particular w.r.t. future protocell applications and the ethical context as

well as its place in the broader context of Living Technology. In addition we want to thank Charles Apel, Bernd Mayer, Gottfried Kohler, William Newman, Donatella Pasqualinie, Andrew Pohorille, and Lau Sennels for valued discussions and support. This work is supported in part by an LDRD grant from the *Center for Space Science and Exploration (CSSE)* and the LDRD grant entitled *Computational proto-organisms*, both at Los Alamos National Laboratory.

References

1. Apel C., Deamer D., and Mautner M. (2002). "Self-assembled vesicles of monocarboxylic acids and acloholes: Conditions for stability and for the encapsulation of biopolymers", *Biochemical Biophysics Acta,* **1559** 1-9.

2. Alexander A.J. and Zare R.N. (2002). "Molecular tennis – Flat smashed and wicked cuts", *Accounts of Chemical Research*, **33**, no.4, 199-205.

3. Bagley, J.R. and Farmer, J.D. (1991). "Spontaneous emergence of a metabolism", in C. Langton, C. Taylor, J.D. Farmer, and S. Rasmussen (Eds.), *Artificial Life II* (p 93-140). Redwood City, CA, Addison-Wesley.

4. Bedau, M., McCaskill S.J., Packard, N.H., Rasmussen, S., Adami, C., Green, D.G., Ikegami T., Kaneko, K., and Ray, T.S. (2000). "Open problems in artificial life", *Artificial Life,* **6**, 363-376.

5. Blocher, M., Liu, D., Walde, P., and Luisi, P.L. (1999). "Liposome assisted selective polycondensation of α-amino acids and peptides", *Macromolecules,* **32**, 7332-7334.

6. Boerlijst, M.and Hogeweg, P. (1991). "Self-structuring and selection: Spiral waves as a substrate for prebiotic evolution", in C. Langton, C. Taylor, J.D. Farmer, and S. Rasmussen (Eds.) *Artificial Life II*, SFI Vol. X (p 255-276), Redwood CA, Addison-Wesley.

7. Boghosian, B., Coveney, P., and Love P. (2000). “A three dimensional lattice gas model for amphiphilic fluid dynamics”, *Proceedings of Royal Society of London*, **456A**, 1431-1454.

8. Borghans, J.A.M., de Boer, R.J., and Segel L.A. (1996). “Extending the quasi-steady state approximation by changing variables”, *Bull. Mathematical Biology*, **58**, 43-63.

9. Brooks, R. (2001), “The relationship between matter and life”, Nature **409**, p. 409-411.

10. Bruinsma R., Gruner G., D’Orsogna M.R., and Rudnick J. (2002), *Physical Review Letters*, **85**, 4393-4396.

11. Böhler, C. , Nielsen, P.E., Orgel, L.E. (1995). “ Template switching between PNA and RNA oligonucleotides”, *Nature,* **376:** 578-581.

12. Ceck, T.R. (1986).“A model for RNA-catalized replication of RNA”, *Proceedings of the National Academy of Science USA,* **83**, 4360-4363.

13. Chen, L., D. Whitten, D. (1995). **“**Photoinduced electron transfer double fragmentation: An oxygen-mediated radical chain process in the cofragmentation of aminopinacol donors with organic halides”**,** *Journal of the American Chemical Society,* **117,** 6398-6399.

14. Chen, L., Lucia, L., and Whitten, D. (1998). “Cooperative electron transfer fragmentation reactions: Amplification of a photoreaction through a tandem fragmentation f acceptor and donor pinacols”, *Journal of the American Chemical Society,* **120**, 439-440.

15. Colgate, S., Rasmussen, S., Solem, and Lackner, K. (2002) “An entropy and astrophysically based strategy for understanding a possible universal origin of life”, *Advances in Complex Systems*, in press.

16. Deamer, D. W. and Pashley, R. M. (1989). "Amphiphilic compomenets of the murchison carbonaceous chondrite: Surface properties and membrane formation*." Origins of Life and Evolution of the Biosphere,* **19**, 21-38.

17. Deamer, D.W. (1992). “Polycyclic aromatic hydrocarbons: Primitive pigment systems in the prebiotic environment”, *Adv. Space. Res*., **12**, 183-189.

18. Deamer, D.W. (1997) The first living systems: A bioenergetic perspective. *Microbiol. Mol. Biol. Rev.* **61**, 239 - 262.

19. Deamer, D.W. (1998). “Membrane compartments in prebiotic evolution”, in A. Brack (Ed.) *The Molecular Origins of Life,* Chap. 8 (p 189-205) Cambridge.

20. Eigen, M. (1971). “Self-organization of matter and the evolution of macromolecules”, *Naturwissenshaften*, **58**, 465-523.

21. Evans D.F. and Ninham, B.W. (1986). "Molecular forces in the self-organization of amphiphiles", *Journal of Physical Chemistry*, **90**, 226-234.

22. Falvey, D. and Banerjee, A. (1997). "Protecting Groups That Can Be Removed through Photochemical Electron Transfer: Mechanistic and Product Studies on Photosensitized Release of Carboxylates from Phenacyl Esters", *Journal of Organic Chemistry*, **62**, 6245-6251.

23. Farmer, J.D., Kauffman S., and Packard, N.H. (1986). "Autocatalytic replication of polymers", *Physica D*, **22**, 50-67.

24. Ferris, J. P., Hill, Aubrey R. Jr., Liu, R. and Orgel, L., E. (1996). "Synthesis of long prebiotic oligomers on mineral surfaces", *Nature* **381**, 59-61.

25. Foresman J.B., Headgordon M., Pople J.A., and Frisch M.J. (1992). "Toward a systematic Molecular-orbital Theory for Exicite-states", *Journal of Physical Chemistry*. **96,** 135-149.

26. Fraser, C.M., Gocayne, J.D; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.; Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.; Small, K.V.; Sandusky, M.; Furhmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.M.; Tomb, J.-F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.; Lucier, T.S.; Scott

Peterson N.; Smith, H.O.; Hutchison III, C.A.; and Venter, J.C., "The Minimal Gene Component of Mycoplasma Genetalium", *Science* **270**, 397-403

27. Ganti, T. (1997). "Biogenesis itself", *Journal of Theoretical Biology*, **187**, 583-593.

28. Gestland, R., Cech, T., and Atkins, J. (1999). *The RNA World*, Cold Spring Harbor.

29. Gilbert, W. (1986). "The RNA World", *Nature* **319** (1986) 618.

30. Gillis, J. (2002). "Scientist planning to make new form of life", *Washington Post*, Nov. 21, 2002, A01.

31. Goetz, R., Gompper, G., and Lipowsky, R. (1999). "Mobility and elasticity of self-assembled membranes", *Physics Review Letters,* **82**, No. 1, 221-224.

32. Greiner N.R., Rogers Y., and Spall D. (1990) "Chemistry of detonation soot: More diamonds and volatiles", *Los Alamos Report LA-11837-MS*. See also: Badziag P., Verwoerd W.S., Ellis W.P., and Greiner N.R. (1990). "Nanometer-sized diamonds are more stable than graphite", *Nature* **343**, 244.

33. Haaima, G., Lohse, A., Buchardt, O., and Nielsen, P.E. (1996)."Peptide nucleic acids (PNAs) containing thymine monomers derived from chiral amino acids: hybridization and

solubility properties of D-lysine PNA." *Angewente Chemie, International Edition English*, **35,** 1939-1941.

34. Hargreaves, W.R. and Deamer, D. (1978). "Liposomes from ionic, single-chain amphiphiles", *Biochemistry* **17**, 3759-3768.

35. Hotani, H., Lahoz-Beltra, R., Combs, B., Hameroff, S., and Rasmussen, S. (1992). "Liposomes, microtubules, and artificial cells", *Nanobiology*, **1**, 1, 61-74.

36. Hutchison, C.A., Peterson, S., Gill, S., Cline, R., White, O., Fraser, C., Smith, H., and Venter, C. (1999). "Global transposon mutagenesis and a minimal mycoplasma genome", *Science* **286** 2165-2169.

37. Hyrup, B., Egholm, M., Nielsen, P.E., Wittung, P., Nordén, B., and Buchardt, O. (1994). "Structure-Activity Studies of the Binding of Modified Peptide Nucleic Acids (PNAs) to DNA"**.** *J American Chemical Society*, **116,** 7964-7970.

38. Jiang, Y., Lookman, T., and Saxena, A. (2000). "Phase separation and shape deformation of two-phase membranes", *Physical Review E,* **61,** R57.

39. Jonston, W., Unrau, P., Lawrence, M., Glasner, M., and Bartel, D. (2001). "RNA-catalyzed RNA polymerization: Accurate and general RNA-template primer extention", *Science* **292**, 1319-1325.

40. Kalosakas, G., Rasmussen, K.O., Bishop, A.R. (2003) "Charge trapping in DNA due to intrinsic vibrational hot spots", *Journal of Chemical Physics*, **118**, no. 8, p. 3731-3737

41. Kang, Q., Zhang, D., Shiyi Chen, and He, X. (2002). "Lattice Boltzmann simulation of chemical dissolution in porous media", *Physical Review E,* **65**, 036318.

42. Kauffman, S. (1986). "Autocatalytic sets of proteins", *Journal of Theoretical Biology*, 119, 1-24.

43. Komineas, S., Kalosakas, G., Bishop, A.R. (2002). "Effects of intrinsic base pair fluctuations on charge transfer in DNA", *Physical Review E*, **65**, no. 6, pt.1, p. 061905-1905.

44. Langton, C.G. (1984). "Self-reproduction in cellular automata", *Physica D*, **10**, 135-144.

45. Lee, D., Granja, J., Martinez, J., Severin, K., and Ghadiri, M. (1996). "A self-replicating peptide", *Nature,* **382**, 525-528.

46. Lengeler, J.W., Drews, G., and Schlegel, H.G. (1999). *Biology of the prokaryotes*, Blacwell Science Pub.

47. Luisi, P.L., Giomini, M., Pileni, M., and Robinson, B. (1988). "Reverse micelles as hosts for proteins and small molecules", *Biochimica et Biophysica Acta,* **947**, 209-246.

48. Luisi, P.L., Bachmann, P.A., and Lang, J. (1992), *Nature*, *357*, 57.

49. Luisi, P.L., Walde, P., and Oberholzer, T. (1994). “Enzymaitic synthesis in self-reproducing vesicles: An approach to the construction of a minimal cell”, *Ber. Bunsenges. Phys Chem.,* **98**, 1160-1165.

50. Luisi, P.L., Veronese, A., and Berclaz, N. (1998). “Photoinduced Formation of Bilayer Vesicles”, *Journal of Physical Chemistry B*, **102** (37); 7078-7080.

51. Maillet J.-B., Lachet, V., and Coveney, P. (1999) “Large scale molecular dynamics simulation of self-assembly processes in short and long chain cationic surfactants”, *Physical Chemistry Chemical Physics*, **1**, 5277-5290.

52. Mayer, B., Koehler, G., and Rasmussen, S. (1997). “Simulation and dynamics of entropy driven, molecular self-assembly processes”, *Physical Review E*, **55**, 1-11.

53. Mayer, B. and Rasmussen, S. (2000). “Dynamics and simulation of self-reproducing micelles”, *International Journal of Modern Physics C*, **11**, 809-826.

54. McCaskill, J.S. (1993). "Inhomogene molekulare Evolution" in Jahresbericht 1992/1993 Institut fur Molekulare Biotechnologie IMB, Jena, Germany.

55. McCaskill, J.S. (1997). “Spatially resolved *in vitro* molecular ecology”, *Biophysical Chemistry,* **66,** 145-158.

56. Merkle, R. (1992). “The Risks of Nanotechnology”, in B. Crandall & J. Lewis (Eds.), *Nanotechnology -Research and Perspectives* (p. 287-294). Cambridge, MA: MIT Press.

57. Miller, S. and Urey, H. (1959). “Organic compound synthesis on the primitive Earth”, *Science* **130**, 245-251.

58. Mittal, K.L. and Lindman, B. (Eds.) (1984). “Surfactants in Solution”, Plenum Press, New York.

59. Monnard, M., Apel, C., Kanavarioti, A., and Deamer, D.W. (2002). “Influence of Ionic Inorganic Solutes on Self-Assembly and Polymerization Processes Related to Early Forms of Life: Implications for a Prebiotic Aqueous Medium”, *Astrobiology,* **2**, 139-152.

60. Morowitz, H., Deamer, D., and Heinz, B. (1988). “The chemical logic of a minimal protocell*”, Origins of Life and Evolution of the Biosphere,* **18**, 281-287.

61. Morowitz, H.J. (1992). *The Origins of Cellular Life: Metabolism Recapitulates Biogenesis.* New Haven: Yale University Press

62. Murov, S.L., Carmichael, I., and Hug, G.L. (1993). *Handbook of Photochemistry*, Marcel Dekker, New York.

63. Nelson, K., Levy, M., and Miller, S. (2000), "Peptide nucleic acid rather than RNA may have been the first genetic molecule", *Proceedings of the National Academy of Science USA,* **97**, 3868-3871.

64. Nielsen, P.E. (1993). "Peptide nucleic acid (PNA): A model structure for the primordial genetic material?", *Origins of Life and Evolution of the Biosphere,* **23**, 323-327.

65. Nielsen, P.E., Engholm, M., Berg, H., and Buchart, O. (1991) "Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide", *Science* **254,** 1497-1500.

66. Nilsson, M., Rasmussen, S., Mayer, B., and Whitten, D.G. (2002) "Constructive molecular dynamics lattice gases: 3-D molecular self-assembly", in D. Griffeath and C. Moore (Eds.) *New constructions in cellular automata* (p. 157-184), Oxford U. Press.

67. Ono, N., and Ikegami, T. (1999). "Model of self-replicating cell capable of self-maintenance" in Floreano,D. et al. (Eds.) *Advances in Artificial Life* (p.399-406), Springer-Verlag, Berlin.

68. Oparin, A.I. (1924). "The origin of life on the Earth" (1st ed. 1924, Pabochii) 3rd ed. 1957, Oliver and Boyd.

69. Peyrard M. and Bishop A.R. (1989). "Statistical mechanics of nonlinear model for DNA denaturation", *Physical Review Letters*, **62**, 23, 2755-2758.

70. Pohorille A. and Deamer D. (2002). “Artificial Cells: Prospects for biotechnology,” *Trends in Biotechnology* **20**, 123.

71. Popielarz, R. and Arnold, D.R. (1990). “Radical ions in photochemistry. Part 24. Carbon-carbon bond cleavage of radical cations in solution: theory and application”, *Journal of American Chemical Society*, **112**, 3068-3082.

72. Porter, R. (1973). “Zinin reduction”, *Organic Reactions*, **20**, 455-481.

73. Püschl, A., Sforza, S., Haaima, G., Dahl, O., and Nielsen, P.E. (1998). “Peptide nucleic acids (PNAs) with a functional backbone.” *Tetrahedron Letters,* **39:** 4707-4710.

74. Rajagopalan, R. (2001). “Simulation of self-assembly systems”, *Current Opinion in Colloid and Interface Science*, **6**, 357-365.

75. Rasmussen, S. (1995). “Aspects of instabilities and self-organizing processes”, Ph.D Thesis, Department of Physics, Technical University of Denmark.

76. Rasmussen, S. (1988). “Toward a quantitative theory of the origin of life”, in C. Langton (Ed.) *Artificial Life*, Redwood City, Addison-Wesley, 1988.

77. Rasmussen, S., Baas, N.A., Mayer, B., Nilsson, M., and Olesen, M.W. (2001). "Ansatz for dynamical hierarchies", *Artificial Life,* **7**, 329-353.

78. Rasmussen, S., Chen, L., Stadler, B., and Stadler, P. (2002). "Proto-organism kinetics: Evolutionary dynamics of lipid aggregates with genes and metabolism", *Origins of life and evolution of the biosphere* (in press).

79. Ratilainen, T., Holmen, A., Tuite, E., Haaima, G., Christensen, L., Nielsen, P.E., and Norden, B. (1998). "Hybridization of peptide nucleic acid", *Biochemistry*, **37**, 12331-12342.

80. Ratilainen, T., Holmén, A., Tuite, E., Nielsen, P.E., and Nordén, B. (2000). "Thermodynamics of Sequence-Specific Binding of PNA to DNA." *Biochemistry*, **39:** 7781-7791.

81. Rosen, M.H.J. (1988). *Surfactants and interfacial phenomena*, Wiley Interscience. Pub.

82. Saghateliaqn, Y. Yokobyashi, K. Soltani, and M.R. Ghadiri, "A chiroselective peptide replicator", *Nature* **409** (2001) 797-401.

83. Sagre, D., Ben-Eli, D., and Lancet, D. (2000). "Composational genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies", *Proceedings of the National Academy of Sciences, USA,* **97**, 4112-4117.

84. Sayama, H. (1998). “Introduction of structural dissolution into Langton’s self-replicating loops”, in Adami et. al., (Eds.), *Artificial Life VI*, MIT Press, 1998, 114-122.

85. Schnell M., Muhlhauser M., and Peyerimhoff S.D (1992). “*Ab initio* MRD-CI study of excited states of chloromethanol ClCH2OH and photofragmentation along C-O and C-Cl cleavage”, *Chemical Physics Letters*, **344,** 519-526.

86. Schwehm, M. (2002). “Parallel stochastic simulation of whole-cell models”, Univ. of Tubingen, Internal Report (8 pages).

87. Smit, B., Hilberts, P.A.J., Esselink, K., Rupert, L.A., and van Os, N.M. (1991). “Simulation of oil/water/surfactant systems”, *Journal of Physical Chemistry*, **95**, No. 16, 6361-6368.

88. Stadler, B., and Stadler P. (2002) “Molecular replicator dynamics”, Santa Fe Institute preprint series 02-09-049.

89. Szostak, J., Bartell, D., and Luisi, P.L. (2001). “Synthesizing life” *Nature* **409**, 387-390.

90. Tieleman, T., van der Spoel, D., and. Brendsen, H.J.C. (2000). “Molecular dynamics study of dodecyclphosphocloride micelles at three different aggregate sizes”, *Journal of Physical Chemistry B*, **104**, No. 27, 6380-6388.

91. Tomac, S., Sarkar, M., Ratilainen, T., Wittung, P., Nielsen, P.E., Norden, B., and Graslund, A. (1996). "Ionic effects on the stability and conformation of peptide nucleic acid complexes", *Journal of American Chemical Society,* **118**, 5544-5552.

92. Tretiak S. and Mukamel S. (2002). "Density matrix analysis and simulation of electronic excitations in conjugated and aggregated molecules", *Chemical Reviews*, **102**, 9, 3171-3212.

93. Tretiak S., Saxna A., Martin R.L., Bishop A.R. (2002). "Conformational dynamics of photoexicited conjugated molecules", *Physical Review Letters*, **89**, 9, 097402-7402.

94. Tuck, A. (2002). "The role of atmospheric aerosols in the origins of life", *Surveys In Geophysics*, **23**, 379-409.

95. Tunstrom, K. (2002). "Mathematical toy models for micelle self-assembly", MS Thesis, Department of Matematics, Trondheim NTNU and Los Alamos National Laboratory, April 8, 2002.

96. von Kiedrowski, G. (1986). "A self-replicating hexadeoxynucleotide", *Angwante Chemie International Edition English,* **25**, 932-935.

97. von Kiedrowski, G. (1993). "Minimal replicator theory I: Parabolic versus exponential growth", in *Bioorganic Chemistry Frontiers*, Vol. 3, 115-146, Berlin.

98. Wachterhauser, G. (1997). "The origin of life and its methodological challenges", *Journal of Theoretical Biology*, **187**, 483-494.

99. Walde, P., Wick, R., Fresta, M., Mangone, A., and Luisi, P.L. (1994). "Autopoietic self-reproduction of fatty acid vesicles", *Journal of American Chemical Society,* **116**, 11649-11654.

100. Watanabe, K., Ferrario, M., and Klein, M.L. (1988). "Molecular dynamics study of a sodium octanoate micelle in aqueous solution", *Journal of Physical Chemistry,* **9**, 819-821.

101. West, G.B., Brown, J.H., and Enquist, B.J. (1997). "A general model for the origin of allometric scaling laws in biology", *Science* **276**, 122-126.

102. West, G.B., Woodruff, W.H., and Brown, J.H. (2002). "Allometric scaling of metabolic rate from molecules and mitochondria to cells and mammals", *Proceedings of the National Academy of Science USA,* **99**, Suppl. 1: 2473-2478.

103. Whitten, D., Chen, L., Geiger, H., Perlstein, J., and Song, X. (1998). "Self-assembly of aromatic-functionalized amphiphiles: The role and consequences of aromatic-aromatic

noncovalent interactions in building supramolecular aggregates and novel assemblies", *Journal of Physical Chemistry B*, **102**, 10098-10111.

104. Wills, P., Kauffman, S., Stadler, B., and Stadler, P. (1998). "Selection dynamics in autocatalytic systems: Template replicating through binary ligation", *Bulletin of Mathematical. Biology,* **60**, 1073-1-98.

105. Woodruff, W.(2002) private communication. This proto-organism work ties in with the very active area on universal scaling relations in biosystems, as it defines the origo in the charts used. For a comprehensive discussion of this active interdisciplinary research areas see e.g. "Scaling relationships in biology" LDRD-DR project at Los Alamos National Laboratory, 2002-2005 (PI: D. Breshears, EES Division).

106. Yoo K.H., Ha D.H., Lee J.O., Park J.W., Kim J., Kim J.J., Lee H.Y., Kawai T., and Choi H.Y. (2001). "Electrical conduction through Poly(dA)-Poly(dT) and Poly(dG)-Poly(dC) DNA Molecules", *Physical Review Letters*, **87**, no. 19, 198102.

Figures

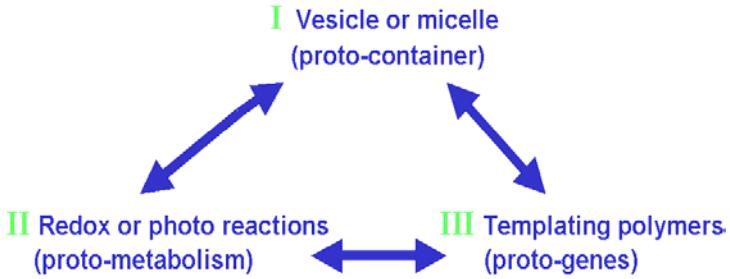

Figure 1: *The functional organization of the involved aggregates in the simple proto-organism. Multiple physicochemical implementations of this structure are possible, but they all pose difficult thermodynamic and kinetic challenges.*

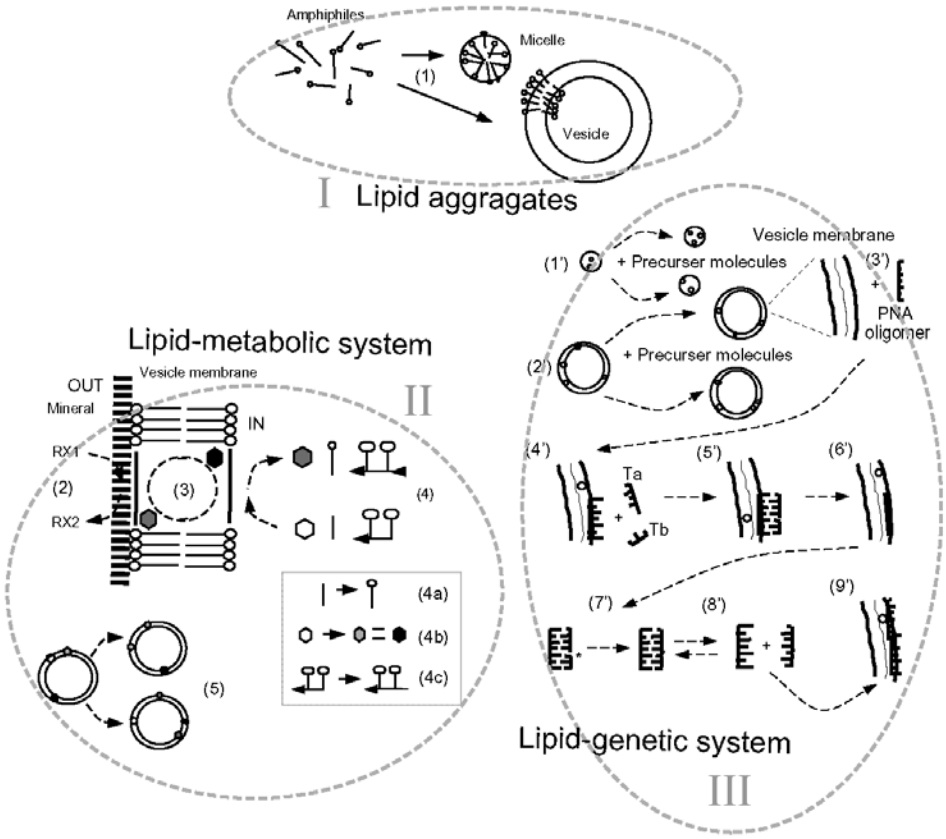

Figure 2. *Two experimental systems II and III need to be combined to form a simple proto-organism. The basis for II and III is the self-assembly of lipid aggregates driven by the hydrophobic effect as shown in I. The left hand side II defines the redox (or photo) reactions within the lipid aggregate and the right hand side III defines the templating reactions within the lipid aggregate. Either (chemical) redox or photo-energy is driving the proto-metabolic processes for subsystem II. The proto-genetic system, subsystem III, is based on a delicate balance between the hydrophobic effect, hydrogen bond formation and –dissociation together with a ligation reaction. See text for details.*

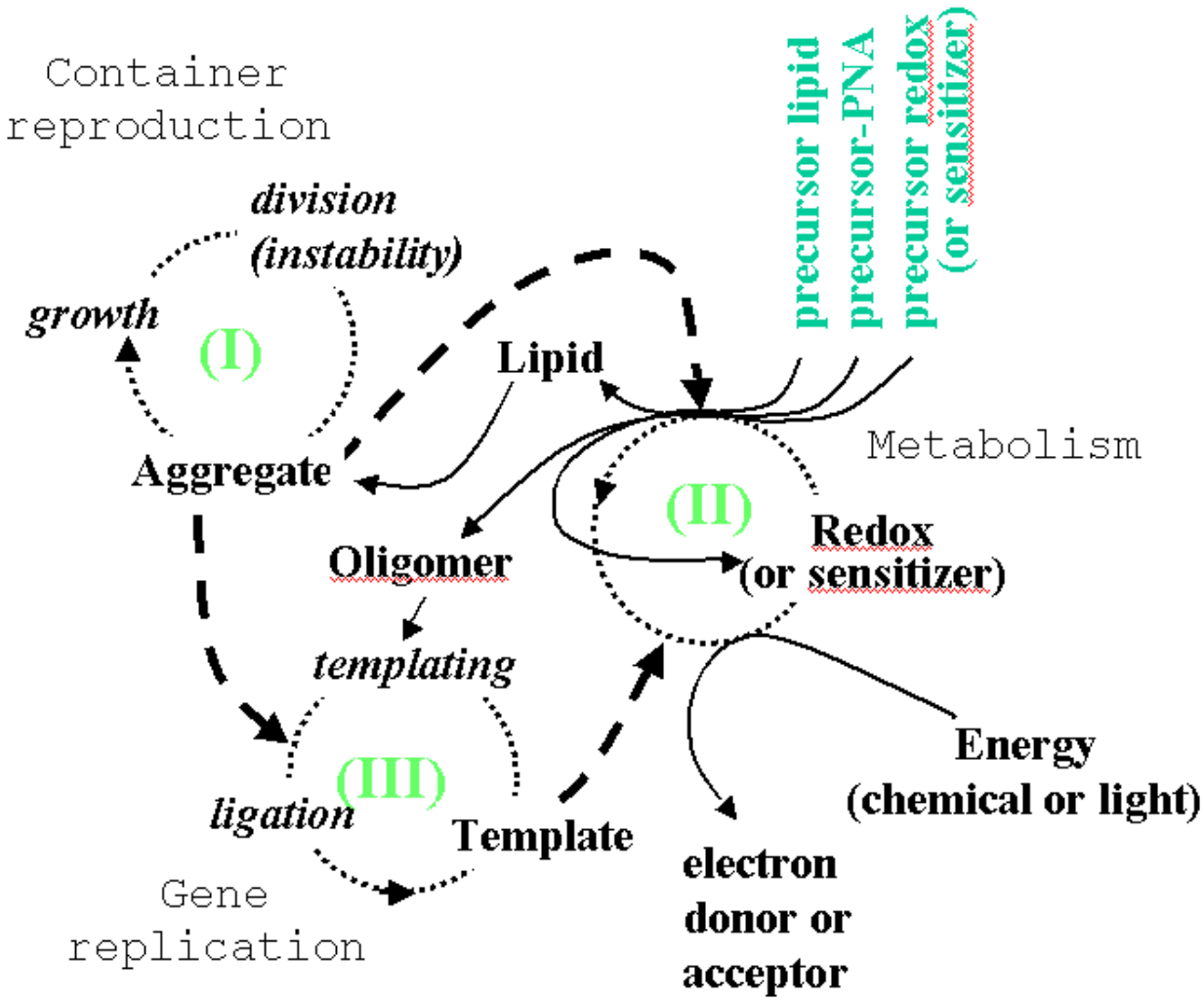

Figure 3. *Causal structure of the full proto-organism discussed in Figure 2. Note the three aggregates with associated "production cycles": container reproduction (I), metabolic cycle (II), and gene-replication (III). The black (full) arrows indicate production and the heawy (dotted) arrows indicate catalysis. The (gray) precursor molecules constitute the "food resources" and the whole system is fueled either by inorganic chemical- or light energy.*

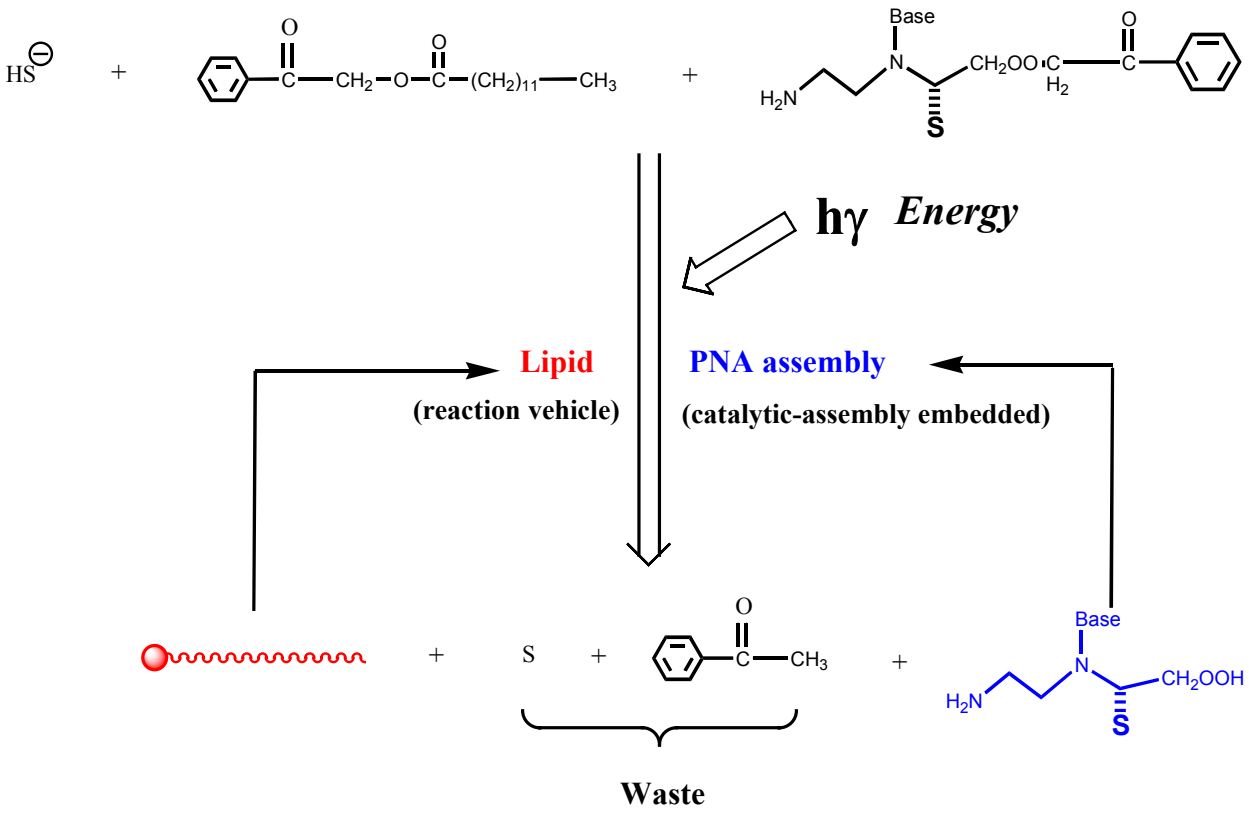

Figure 4. *Simple physicochemical implementation of coupling between proto-genes, proto-container, and proto-metabolism. The metabolic chemistry in this system is based on two autocatalytic photo-segmentation reactions where an ester is converted into a carboxylic acid derivative, one for the simple lipid and one for the proto-gene monomers (or oligomers). The PNA with attached sensitizer acts as a photocatalyt for both segmentation processes once the PNA is polymerized. Note that S both symbolizes free sulfur (end product after donation of an electron from sulfite) and the photo-sensitizer molecule attached to the PNA backbone.*

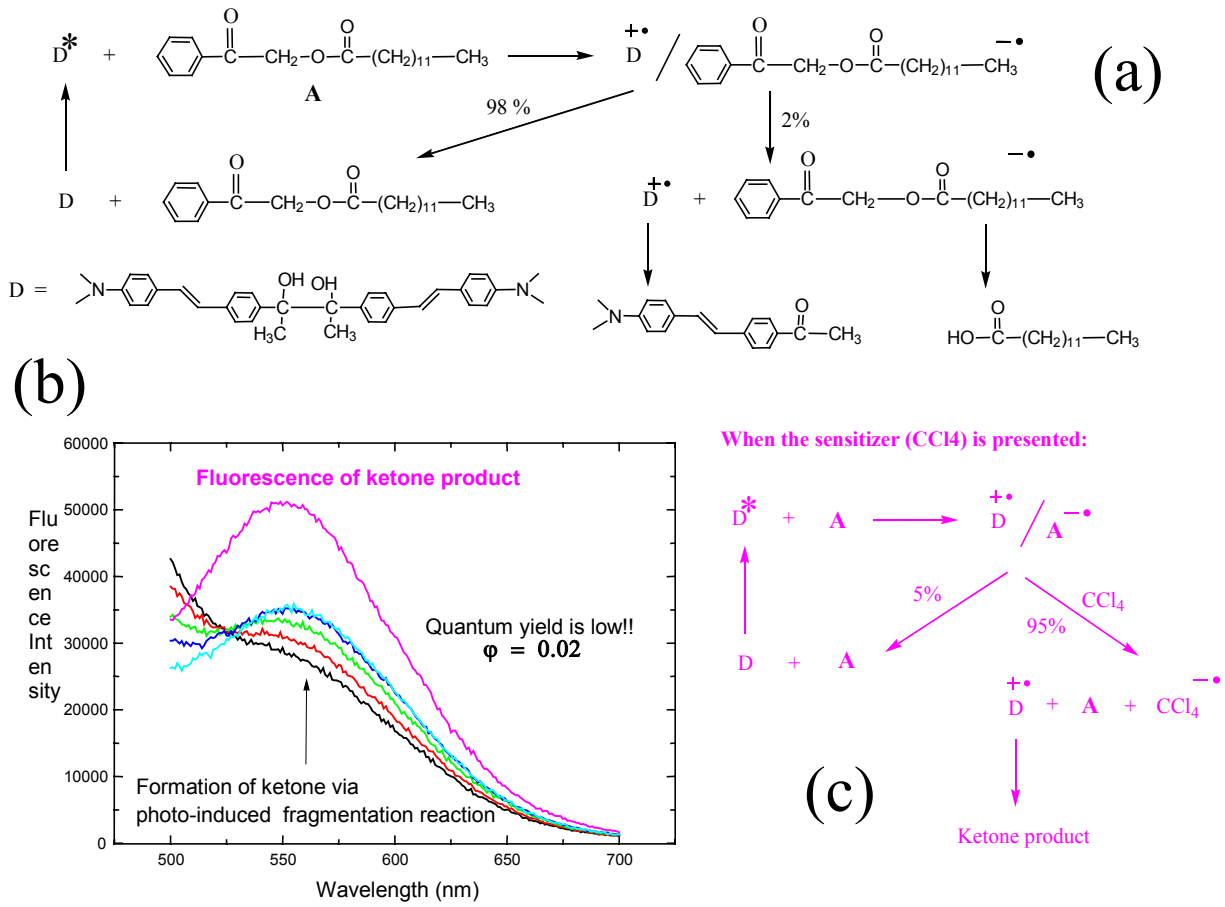

Figure 5. *Photo-induced fragmentation reaction of an amino-pinacol derivative as the electron donor molecule and a phenacyl ester as the electron acceptor. (a) The excited state of the pinacol can be quenched by the phenacyl derivative to generate a contact ion-pair of pinacol cation radical and a phenacyl anion radical. (b) From the fluorescence spectra we can deduce that the product quantum yield is only about 2%. Due to the dominant back-electron transfer processes (recall (a)). (c) CCl_4 can efficiently intercept the phenacyl ester anion radical inside the contact ion pair to generate a CCl_4 anion radical. Thus about 95% of contact ion pairs will now undergo charge separation and lead to a high quantum yield of product formation. See text for details.*

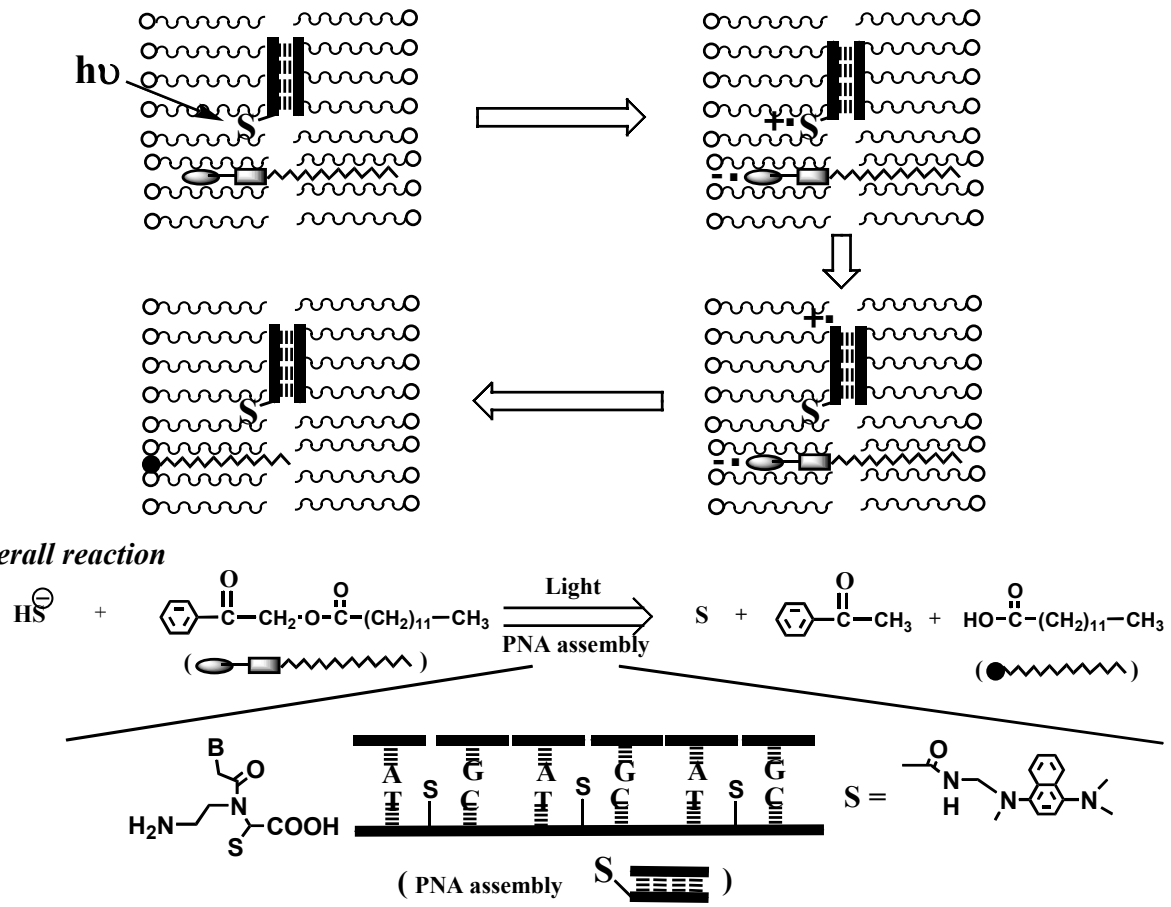

Figure 6. *Scheme for proto-organism replication. Top left tableau depicts the excitation of the sensitizer due to an energy rich photon. The energized sensitizer causes a charge separation between the sensitizer and the lipid precursor (penacyl), in the upper right tableau. Middle right tableau depicts the neutralization of the sensitizer by the delivery of an electron from the PNA electron relay chain. The precursor lipid molecule is still energized and will eventually break (fragment) into a lipid (carboxyl acid) and a phenyl group, which is depicted in the middle left tableau. The breakage occurs at the ester bond. The overall reaction is summarized in the lower part of the figure where the structure of the sensitizer enriched PNA is also indicated. Note that S is sulfur in the overall reaction, while S indicates the sensitizer molecule attached to the PNA backbone in the other tableaus. See text for details.*

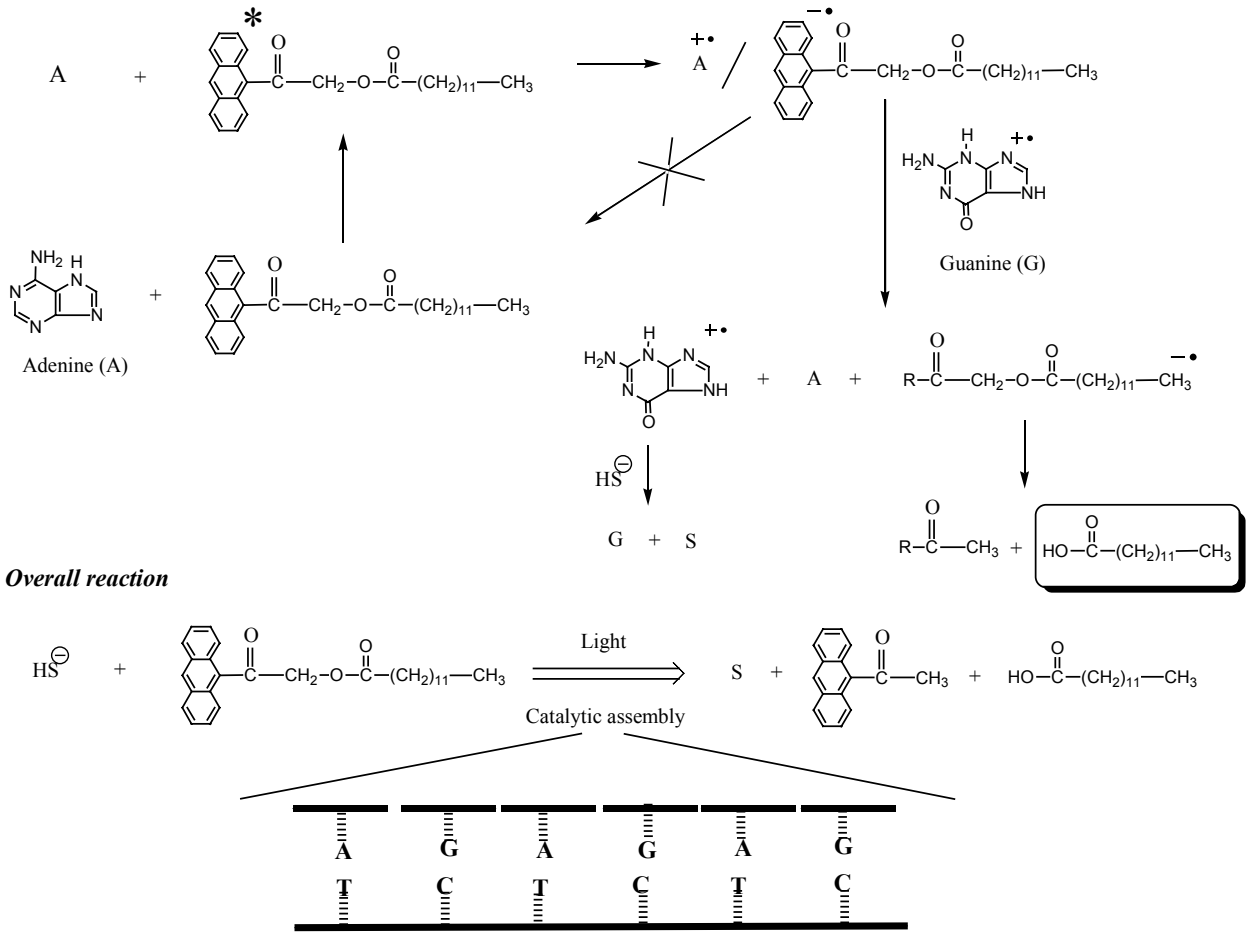

Figure 7. *Alternative physicochemical implementation of the photo-driven segmentation reactions without a direct attachment of the sensitizer to the PNA backbone. This approach could be used as an alternative to the PNA monomers with the sensitizer attached. Overall scheme is essentially the same as discussed in Figure 8, although sensitizer molecules have to be provided as part of the resources.*

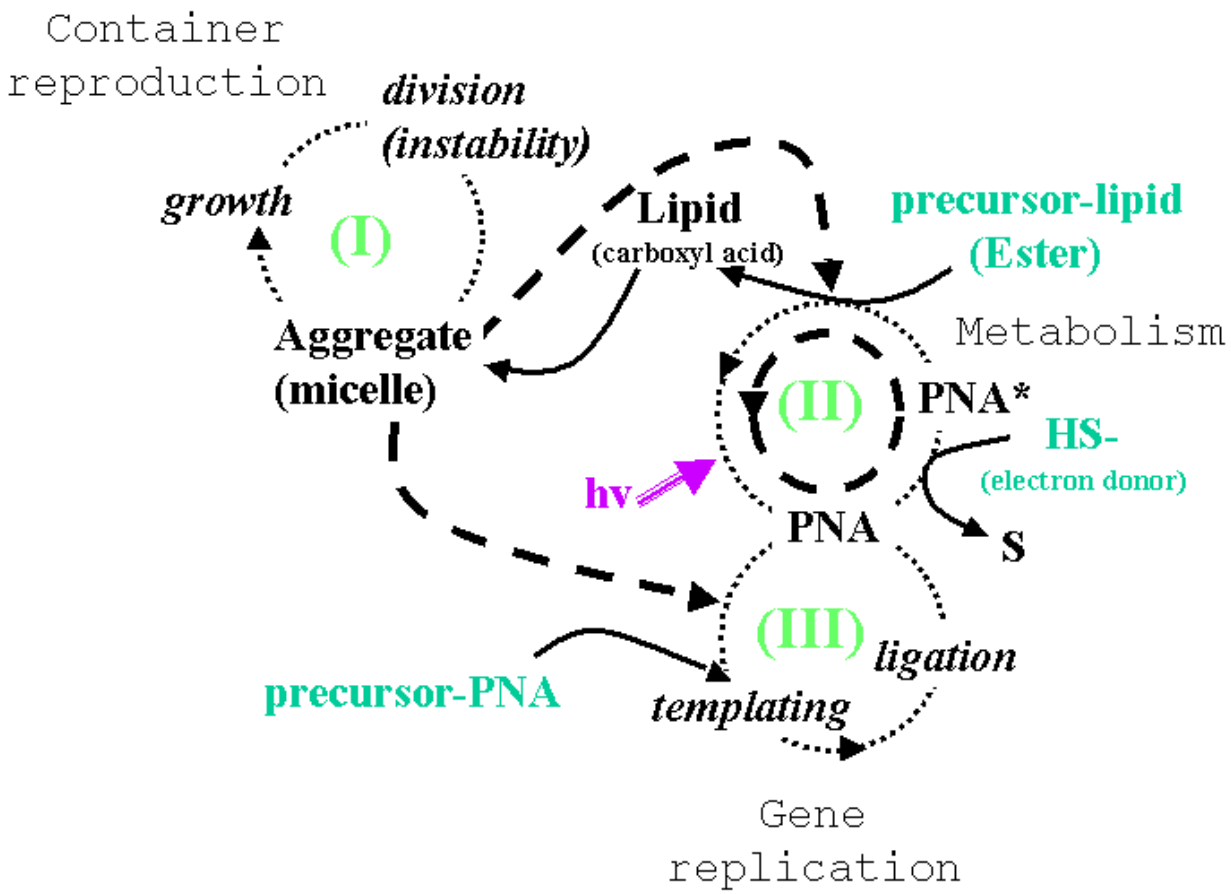

Figure 8. *Causal structure of the simple physicochemical implementation of the full proto-organism discussed in Figure 2 and in the text. Compare with Figure 3. Again the "resources" for the structure is indicated with gray: precursor-lipid (a hydrophobic ester), precursor (oligomer) -PNA, and an electron donor (sulfite), as well as photo energy in terms of light (light gray). Also note the three main physicochemical cycles: the aggregate growth and division cycle (I) feed by the production of lipids; the metabolic (PNA) cycle (II) feed by light (note how the PNA exists in a neutral and an energized form PNA*); and the PNA replication cycle (III) feed by short (pre-template) PNA strands. Each of these cycles are both kept together and catalyzed by the lipid aggregate (heavy dashed arrows), but in this system the proto-metabolic- and the proto-genetic molecules are identical. Thus, the gene-replication cycle (III) and the metabolic cycle (II) are linked more tightly compared to Figure 3.*

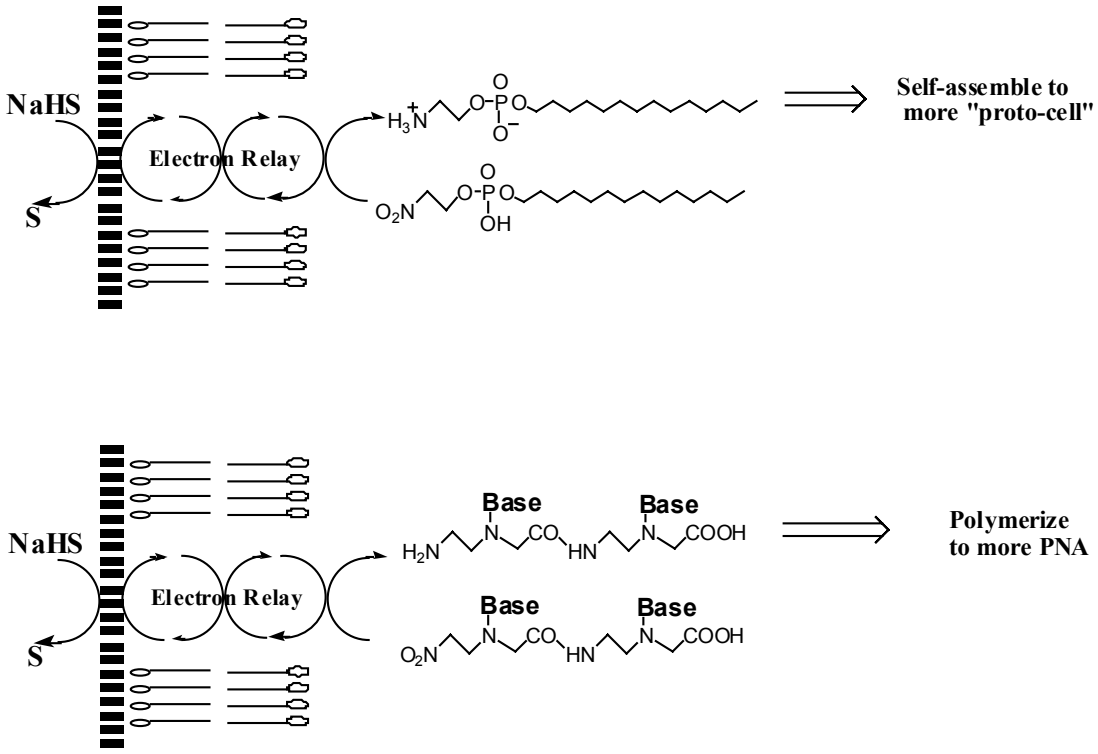

Figure 9. *Reductive synthesis of surfactant and PNA dimer in a vesicle system. See text for details.*

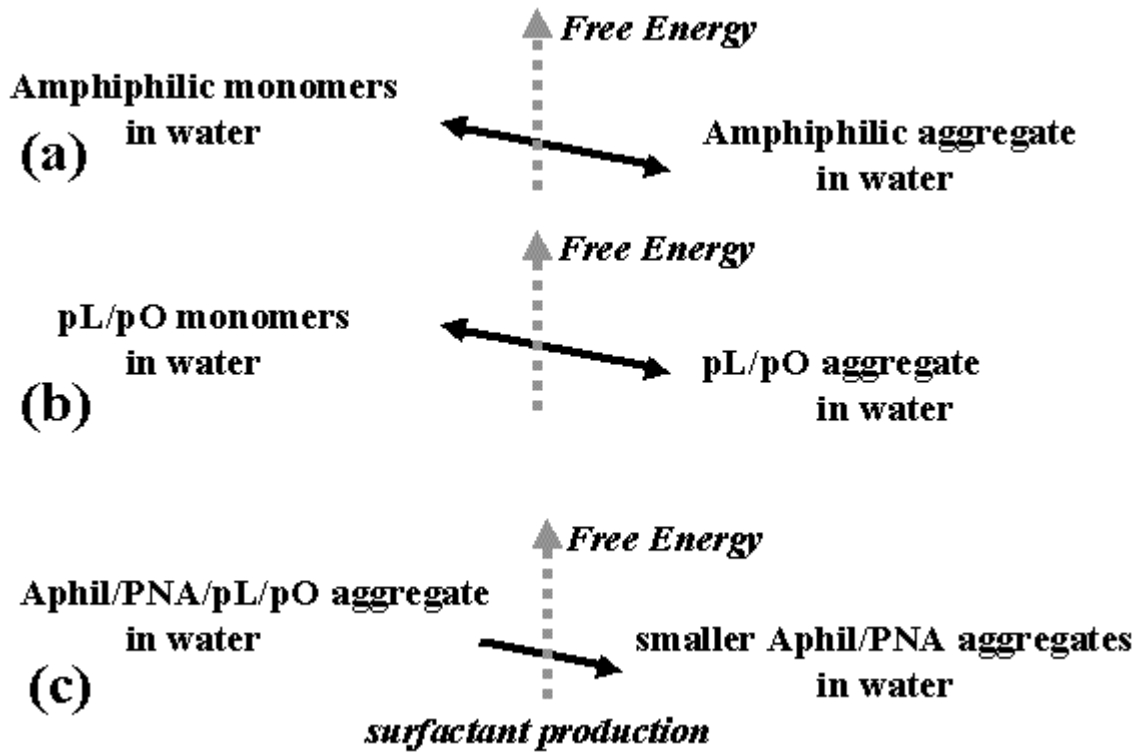

Figure 10. *(a) The thermodynamics of the lipid self-assembly, (b) assembly (phase separation) of the hydrophobic precursor lipids (and a few PNA precursors) in water, (c) the association (loading) of many precursor lipids and one or a few proto-genes and precursor oligomers into the lipid aggregates, which a following lipid production. The vertical axis indicates the relative free energy level of each component. (a) and (b) all have reversible reactions and so is the loading of the precursor molecules into the aggregates, whereas the metabolic production of lipids and functional PNA from the loaded presursor molecules (c) is irreversible. Note that the relative free energy levels both depend on the detailed chemical species and their relative concentrations on each side of the reaction arrow. See text for details.*

P(micellar size)

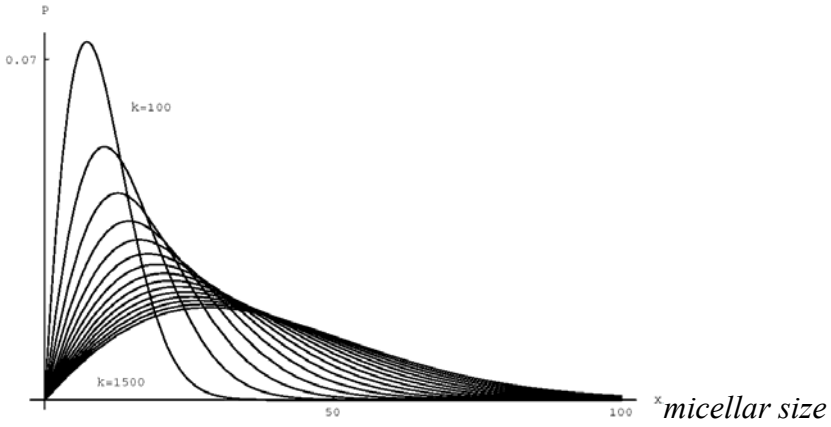

Figure 11. *Equilibrium micellar size distribution from equation (17) for values of k ranging from 150 to 1,500. Note how the choice of parameters yields small micelles with maximal probability of finding micelles of size* $x_{\max} = \sqrt{k}(\sqrt{3}-1)$, $(\partial_x P(x) = 0)$. *See text and [Tunstroem, 2002].*

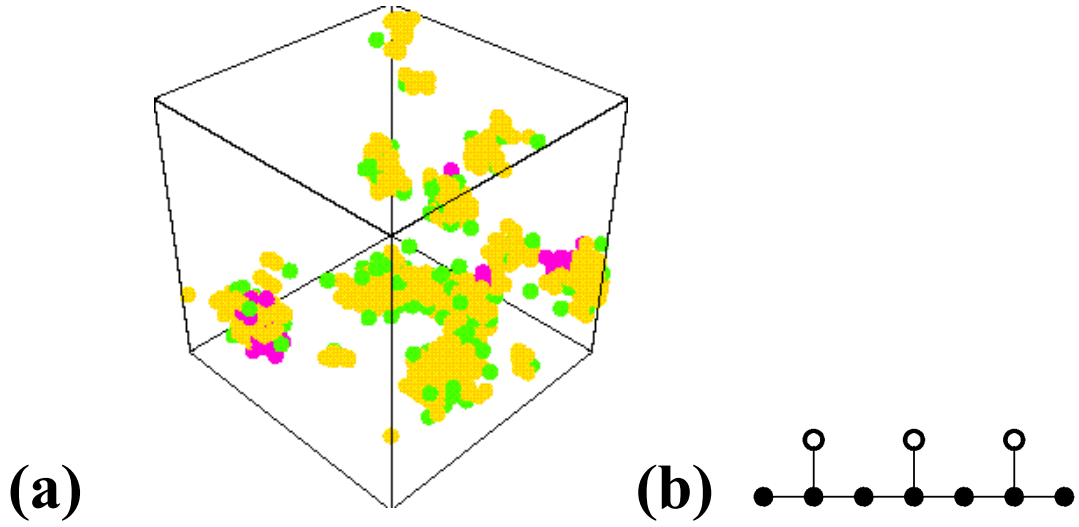

Figure 12. *Miceller self-assembly and interactions with PNA-like polymers. Water is not shown. The yellow monomers indicate the hydrophobic tail parts of the lipids and the green monomers indicate the hydrophilic head groups of the lipids. The red polymers are short simplified PNA trimer strings, see (b), which in the simulation are composed by a linear hydrophobic backbone with a side group on every other backbone element. Each PNA polymer has a hydrophobic backbone, which results in their affinity for the lipid aggregates. No free PNA molecules are found. Each lattice cube is about 9 nm on each side and the typical simulation time for the shown situations are about a millisecond. Note that different lipid molecules probably should be used such that larger micelles naturally are produced. Larger lipid structures are necessary because they are better able to host longer PNA strings and a loading of precursor molecules.*

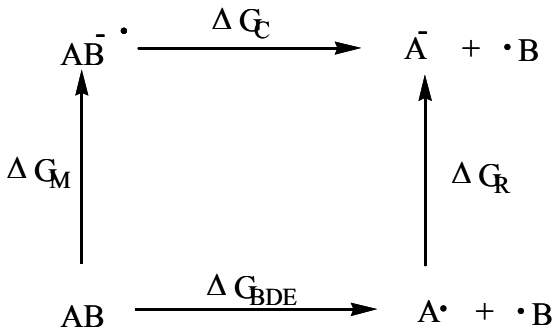

Figure 13. *General scheme for the thermochemical cycles (or pathways) of the photo-fragmentation reaction. Starting in the lower left corner two different pathways are in principle possible for the products in the upper right corner: Direct breakage of the C-O bond in the neutral molecule AB to form two radicals A· and B·, followed by the one electron reduction of the lipid to form the desired products A·⁻ and B·, or alternatively, a one electron reduction of AB to form the radical AB·⁻ followed by the cleavage of the C-O bond in the charged radical. The upper pathway with an initial electron reduction of AB significantly weakens the C-O bond such that the fragmentation step happens more readily (and predictably) resulting in higher yields. Since both pathways result in the same products the overall free energy difference for the two pathways is obviously the same. See text for details.*

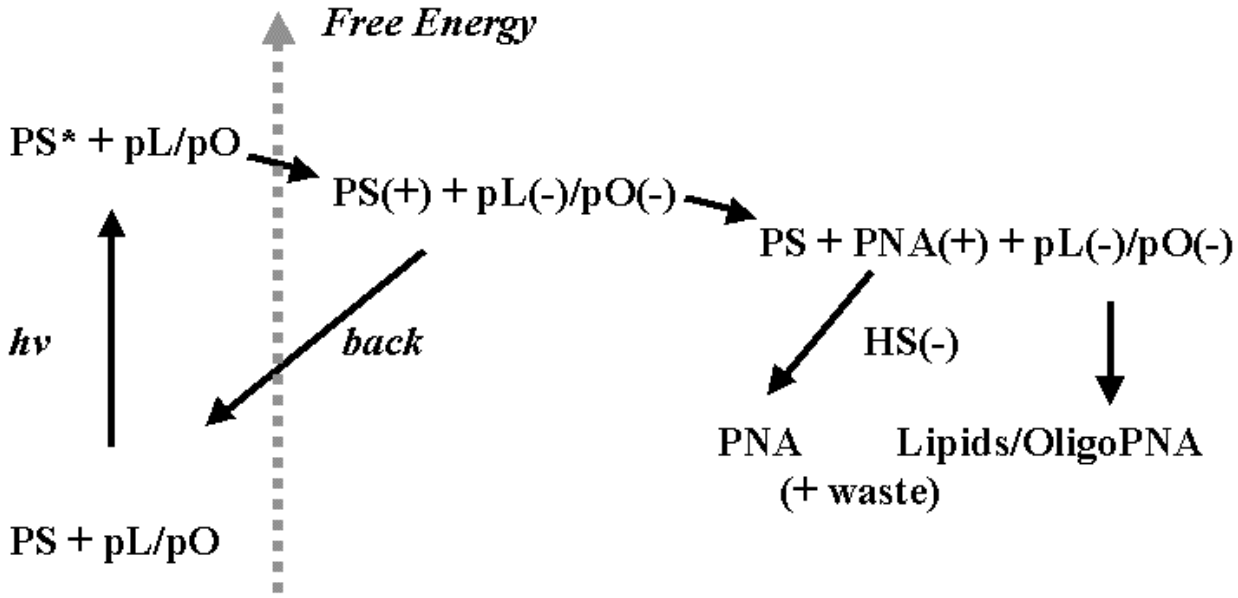

Figure 14. *The thermodynamics of the proposed photo-driven proto-metabolism: The vertical axis indicates the relative free energy of the components. Only the overall reactions are depicted not each detailed reaction step. From the bottom left: S, pL, pO, and PNA defines a sensitizer, precursor lipid, precursor (PNA) oligomer, and PNA respectively. Approximate values for the reaction constants are indicated for each main reaction. As light energy (hv) is pumped into the system the sensitizer is initially energized. This energy causes a charge separation between the sensitizer and precursor molecules (either pL or pO). This charge separation is very short lived (back reaction), unless a nearby electron relay system is present. Adenine within PNA is the immediate electron donor for the now activated precursor molecule, and adenine immediately thereafter gets an electron from guanine, which eventually will harvest an electron from a final electron donor as sulfite (HS^-). The resulting energized and negatively charged precursor molecule then fragments at the ester bond site after which a functional lipid - or PNA oligomer - is created together with "waste". The overall reaction is an example of a photo-fragmentation reaction.*

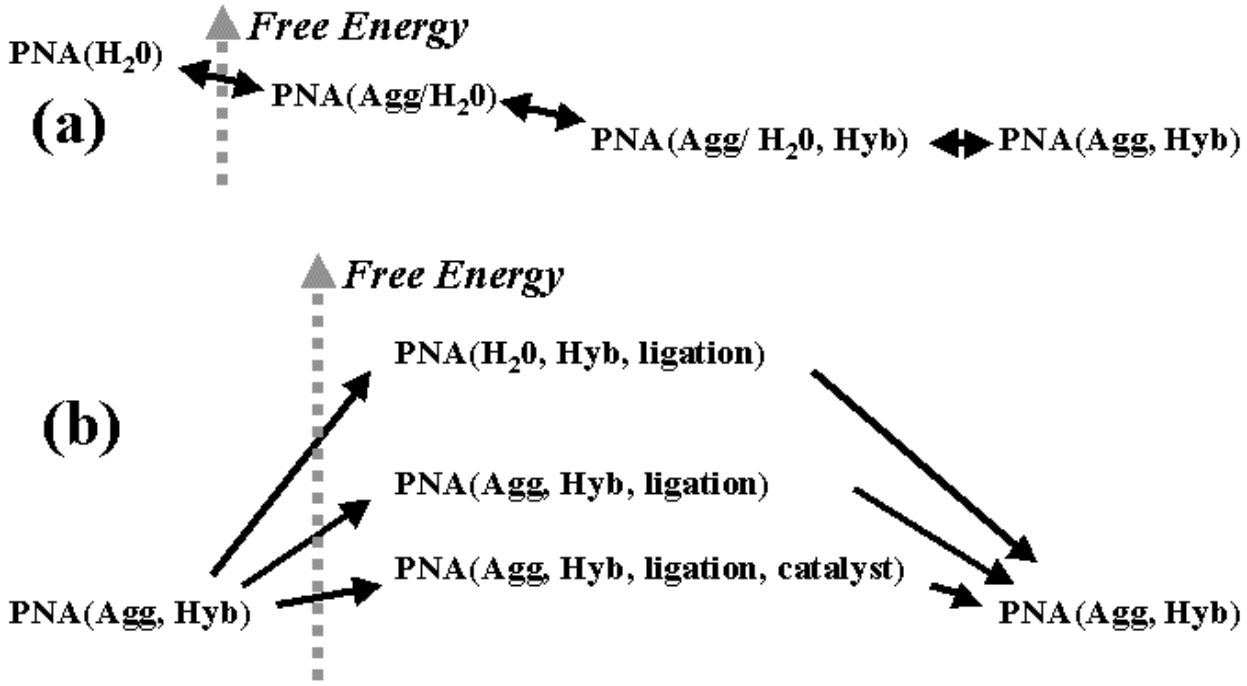

Figure 15. *Thermodynamics of the lipid catalyzed PNA replication scheme. (a) PNA in water will attach to the lipid aggregate due to its hydrophobic backbone. If complementary strings meet they will hybridize to form duplexes. Neutral diffusion of the duplex between the lipid aggregate surface and the interior of the aggregate can occur because the exterior of the duplex is hydrophobic. Since each step is thermodynamically down hill (or neutral) from left to right it poses a problem to dehybirdize the duplex after a ligation has occured, which is necessary for successive templatin processe . A balance has to exist between the effective duplex hybridization energy and the energy gain from the exclusion of the hydrophobic backbone from the water phase. (b) Once the ABC duplex is formed, a condensention reaction has to occur between A and B to form C'. The activation energy in water is very high, whereas the activation energy is significantly lowered in the lipid aggregate. Perhaps the introduction of simple organic catalysts could lower the amide bond activation barrier even further.*

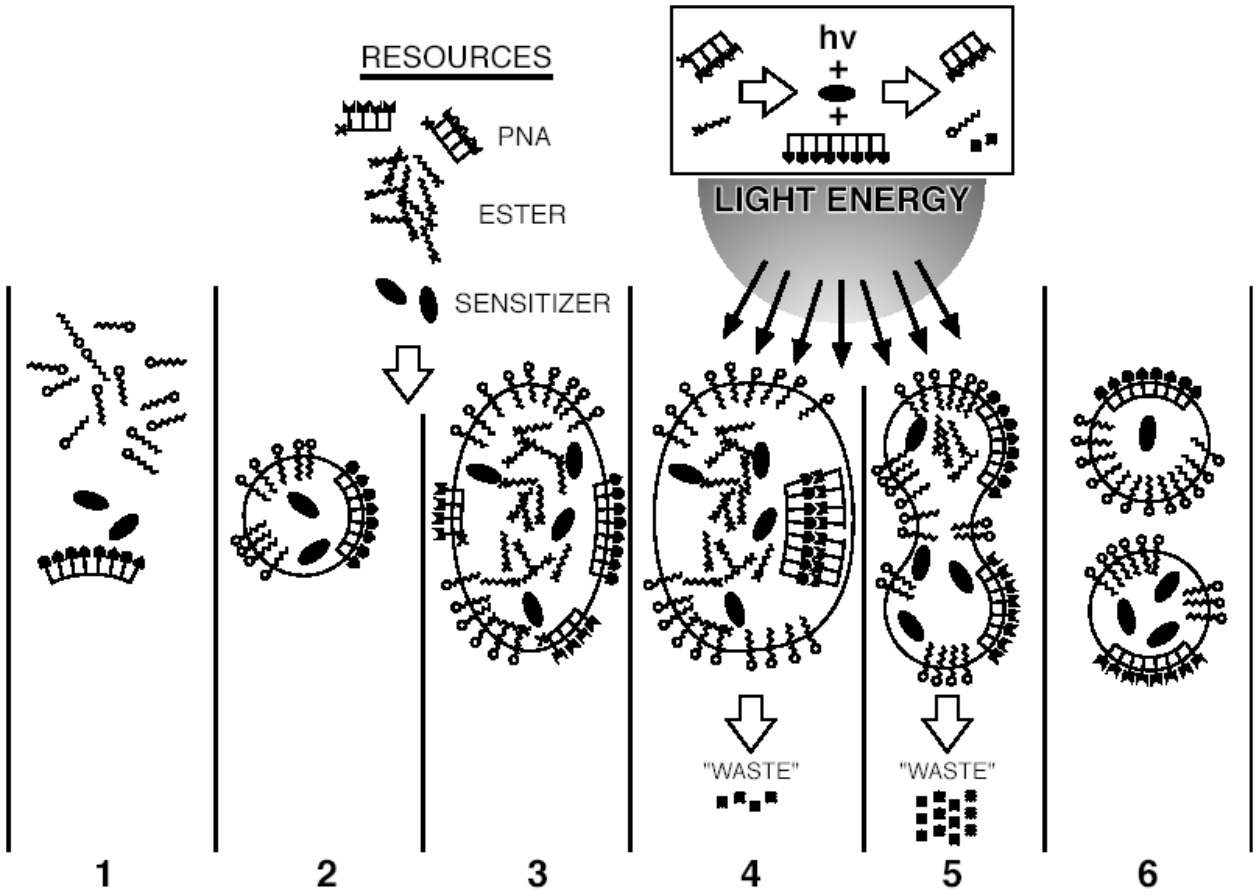

Figure 16. *Full life-cycle of the proposed proto-organism. (1)-(2) Self-assembly of the proto-organism components, the lipids, the PNA, and the sensitizer molecules (Sections 3.1, 3.2, and 4.1). (3) Feeding the proto-organism with PNA- and lipid precursors and more sentisizer molecules. The proto-organism swells up in particular as it is loaded with the many precursor lipids (Sections 3.1, 3.2, and 4.2). (4) As light is being provided the phenyl group is first fragmented from the precursor PNA forming PNA oligomers that can ligate, once they are aligned by the template (Section 3.1, 3.2, 3.3, and 4.3). (5) As the precursor lipids are turned into lipids (surfactants) the large aggregate becomes unstable and stars breaking up. A thermodynamic balance between hybridized and non-hybridized (double- and single stranded) PNA has to be established to allow a reasonable partition of the proto-genes between the two aggregates (Sections 3.3 and 4.4). (6)The life-cycle is complete as the proto-organism has*

generated a copy of itself. The overall rate limiting steps are believed to be the PNA template directed ligation process and the balanced lipid-PNA and PNA hybridization kinetics. See text for more details.