This article appeared in a journal published by Elsevier. The attached copy is furnished to the author for internal non-commercial research and education use, including for instruction at the authors institution and sharing with colleagues.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

<http://www.elsevier.com/copyright>

Contents lists available at ScienceDirect

The International Journal of Biochemistry & Cell Biology

journal homepage: www.elsevier.com/locate/biocel

Review

Evolution and self-assembly of protocells

Ricard V. Solé^{a,b,*}^a Complex Systems Lab (ICREA-UPF), Parc de Recerca Biomedica de Barcelona, Dr Aiguader 88, 08003 Barcelona, Spain^b Santa Fe Institute, 1399 Hyde Park Road, Santa Fe NM 87501, USA

ARTICLE INFO

Article history:

Available online 17 October 2008

Keywords:

Cell replication
Evolution
Self-assembly
Protocells

ABSTRACT

Cells define the minimal building blocks of life. How cellular life emerged and evolved implies to cross the boundary between living and nonliving matter. Here we explore this problem by presenting several relevant components of the whole picture involving chemistry, physics and natural selection. Available evidence suggests that the basic logic of life can be understood and eventually translated into synthetic forms of cellular life. A simple, physically sound model of information-free protocell replication suggests that the basic logic of how to couple metabolism and container can be more relevant than the specific set of parameters used, thus indicating that the emergence of cells might have been easier than we would expect.

© 2008 Elsevier Ltd. All rights reserved.

Contents

1. Introduction	274
2. Self-reproducing machines	275
3. Self-assembly and physical constraints	276
4. Chemical constraints	277
5. Dynamical constraints	278
6. Computational modelling of protocell replication	278
6.1. Logic of the model	279
6.2. Physics and chemistry of protocell replication	279
6.3. Nanocell replication is robust	282
7. Discussion	282
References	283

The whole history of the world, as at present known, although of a length quite incomprehensible by us, will hereafter be recognised as a mere fragment of time, compared with the ages which have elapsed since the first creature, the progenitor of innumerable extinct and living descendants, was created Darwin (1859).

1. Introduction

Cells are the basic units of life. They have been often compared to machines, involving many components in interaction. These components, particularly proteins, define a set of functional structures

* Correspondence address: Complex Systems Lab (ICREA-UPF), Parc de Recerca Biomedica de Barcelona, Dr Aiguader 88, 08003 Barcelona, Spain.
E-mail address: ricard.sole@upf.edu.

able to cope with all cell requirements, including the gathering and processing of matter, energy and information (Harold, 2001). A cell is a machine made of macromolecules, able to survive in a given environment by properly exploiting available resources. However, living cells display a key feature not shared by artificial objects: they reproduce themselves.

Most descriptions of cellular behavior emphasize the role of DNA as the key molecule directing cell dynamics. This DNA-centered view is actually far from complete: cellular life cannot be described in terms of only DNA (or any other information-carrying molecule) nor as metabolism or as compartment (cell membrane) alone. Instead, it results from the coupling among these three components (Fig. 1a). A container is a prerequisite of biological organisation in order to confine reactions in a limited space, where interactions are more likely to occur (Deamer et al., 2002). It also provides a spatial location able to effectively facilitate division of labour

Fig. 1. Cellular life needs three key components (a) defining the three basic, cooperatively linked molecular machineries dealing with information, metabolism and container. The three of them are required and none can be considered in isolation. An early scenario of protocell evolution suggests that metabolism and container would be enough to achieve cell replication (b). In that case, information would not be necessary.

among reactions. Moreover, in modern cells, the membrane is an active component, channelling nutrients in (and waste products out) of the cell by means of specialised transport catalysts (Alberts et al., 2002). A metabolism provides the source of non-equilibrium and a mean of energy storage that is required in order to build and maintain cellular components. It is also required to allow cell growth to occur and eventually force splitting into two different (but similar) copies. From these basic ingredients, a protocell can be defined as a macromolecular system thermodynamically separated from the environment and able to replicate using available nutrient molecules and energy sources (Morowitz, 1992).

Understanding the early evolutionary steps towards cellular life has become one of the most important problems of biology (Gánti, 2003; Deamer et al., 2002; Jortner, 2006). How did cells emerge in the first place? Can we gain insight by looking at modern cells? Are the three components of modern cells needed in order to answer these questions? The last question is specially pertinent, since information might not be strictly necessary for a living cell to exist (Fig. 1b). Although information is essential to allow evolution to occur, early forms of cells might have lacked such component.

Answering most of these questions is a difficult task. In this context, although fossil remains of early cellular forms exist, they are likely to reflect already advanced forms with a complex membrane structure, metabolic pathways and large genomes (Jortner, 2006). Moreover, the logic of early life did not leave traces of itself. In order to achieve satisfactory answers to the previous questions, a theoretical perspective seems unavoidable, together with experimental approaches leading to the synthesis of artificial life.

The first forms of cellular life inhabited the thin frontier separating living from non-living matter (Pohorille and Deamer, 2002; Rasmussen et al., 2003; Szathmary et al., 2005; Solé et al., 2007; Rasmussen et al., 2008). Crossing the boundary between such two domains of organization required moving beyond pure chemistry and incorporating Darwinian evolution as a key ingredient. Without selection, any chemical soup of a huge number of molecular species would have failed to evolve towards order. Together with selection forces, some special properties of lipids might have been also crucial to the first steps towards living cells.

In this paper we review some fundamental issues regarding the molecular requirements allowing self-reproducing protocells to emerge. The key ingredients include the logic of self-replication, the physics of self-assembly of soft machines and how metabolism might have been able to couple with a container in such a way that reliable cell division would eventually occur. We pay special attention to a long forgotten component of this process, namely the embodied nature of protocells. These macromolecular assemblies are confined in a spatial domain and they need to change in

such a way that division effectively occurs. Most theoretical work in the field has concentrated in metabolism and information as the key players, with the container placed in a secondary position. However, ignoring how the container actually interacts with other components means ignoring a crucial element of cellular life. As discussed in (Ruiz-Mirazo and Moreno, 2004) the membrane must play a central role from very early stages.

2. Self-reproducing machines

Before we enter into the domain of wet biology and molecular interactions, let us first stop by looking at the problem of the logic of replication. Under such view, we can ignore (to some extent) the exact nature of the molecular players being involved. Instead, the focus is shifted towards the bare bones of the logic of self-reproduction and the minimal requirements for such process to take place (Freitas and Merkle, 2004).

The most influential work directed to answer the previous question was made by John von Neumann (Fig. 2a) who analysed this question under a purely abstract approximation (von Neumann, 1966). It is based on a computational picture of the logic of living systems (Nurse, 2008). The chief merit of such an approach (which might seem crude at first sight) is that it reduces the complexity of life replication to its bare bones: we only need to know what are the logic components to be present, no matter how they are made of and how they exactly interact. In von Neumann's picture, the self-replicating system was shown to be composed by a minimal set of closely related components, namely (see Fig. 2):

1. The constructor (A), able to build a physical, new system by using the available raw material in the surroundings.
2. The blueprint or instructions containing information on what has to be performed by the constructor.
3. A duplicator (B) which takes the instructions and duplicates them accurately.
4. The controller (C) required to guarantee that the whole process takes place in some well-defined sequence.

von Neumann's abstract ingredients can be easily mapped into the cellular machinery in terms of DNA, RNA, proteins and metabolism. We can identify the components of the automaton with those of living cells as follows: (a) the instruction set is the DNA molecule; (b) the duplicator is provided by the DNA polymerase and other components of the cell's replication machinery; (c) the constructor corresponds to the RNA polymerase and the translation machinery (allowing proteins to be formed) and (d) the controller is nothing but the regulation of transcription and translation.

Fig. 2. Von Neumann (left) found the general conditions for a self-reproducing system. In 1956 he found that this abstract problem had a general solution, schematically described in the diagram. A given “automaton” with the potential for replicating itself should include three basic elements. Although von Neumann’s result was obtained some years before the Watson–Crick discovery of the DNA and the emergence of molecular biology, this automaton contains precisely the elements of the machinery of life already observed in living cells.

Moreover, the automaton performs its operations in a spatial setting, since every component is constructed by following the internal instructions and using surrounding materials. In this context, hardware implementations of self-replicating robots have been successfully achieved (see Zykov et al., 2005). This work seems to indicate that, independent of the nature of the components being used to construct the copy, the logic of self-replication might be strongly constrained.

However, as mentioned at the introduction that there is an alternative path to self-replicating protocells which does not include information as a necessary requirement for a system to self-reproduce itself. In order to avoid this information requirement, we need to consider the presence of a spontaneous source of order provided by the natural tendency of some types of molecules to self-organize into macromolecular complexes defining a container.

3. Self-assembly and physical constraints

An important component of cellular life, namely the presence of a stable, closed compartment, is actually an easy part of the story (Mouritsen, 2005). The essential components of membranes, so-called amphiphiles, spontaneously self-organise in space within a water environment to form well-defined structures. Amphiphiles are polar molecules having a well-defined hydrophilic head group (showing attraction for water) and a tail group (usually a long chain) showing hydrophobicity. As a consequence of this conflicting relation towards water molecules (which are themselves polar too) amphiphiles can self-assemble into compact spatial structures showing a “phase separation” between water and the amphiphiles. This is illustrated in Fig. 3, where we display an example of the initial and final states of a computer simulation of vesicle formation. Here water molecules are not shown and the amphiphiles are represented as simple strings composed by a few connected beads (see for example Cook and Deserno, 2006 and references therein). Although molecules and molecular interactions are strongly simplified, the model is fully consistent with experimental results on self-assembly. This is a good example of self-organization (Nicolis and Prigogine, 1977; Luisi, 2006), where out of an initial mixture of molecules, order emerges.

Although the spontaneous self-assembly of molecules in space is a good starting point and might suggest that a key component

of cellular life is already in place, and important factor is clearly missing. The problem emerges from the fact that the process of self-assembly follows the standard energy minimization dynamics characteristic of any physical process (Jung et al., 2001; Bozic and Svetina, 2004; Mouritsen, 2005). Reaching the spherical vesicle state takes place through a spontaneous dynamical pathway and the resulting structure is typically very stable under many sources of perturbation. There is however an important relationship between amphiphile geometry and the type of macromolecular structures (Fig. 3). These molecules can be close to cylinders or instead have a strong asymmetry between head and tail. If they are like cylinders, they will pack forming parallel sheets. Instead, if the head is larger than the tail, a cone-like structure will allow packing of chains along a spherical object, where tails will be accommodated inside and heads will remain outside.

Specifically, the geometrical shape of a lipid can be most simply described by the so-called *packing parameter*, N_s (Evans and Wennerstrom, 1999; Bozic and Svetina, 2004). It captures the overall geometric shape in terms of length, area and volume, thus ignoring most molecular details. It is defined as:

$$N_s = \frac{v}{a_o L} \quad (1)$$

where v measures the volume of the hydrophobic portion of the molecule, L is the length of its hydrocarbon chain, and a_o the effective area of the hydrophilic head. As defined, N_s relates the volume of the hydrophobic portion to its length normed by the hydrophilic area. All the required parameters can be estimated and three examples of the basic shape associated to different packing parameter values are shown in Fig. 4. The packing parameter helps determining the curvature of the aggregate that will emerge:

1. for $N_s = 1$, the shape of the surfactant will be a cylinder and the typical aggregate will be a bilayer.
2. If $N_s = 1/3$, the surfactant has the shape of a cone and will form micelles in an aqueous solution.
3. For $1 < N_s < 1/3$, more sophisticated aggregates can be found including cylinders and even scaffolds of interconnected cylinders.
4. If $N_s > 1$, the aggregates will tend to form small closed vesicles.

Fig. 3. The shape of a lipid is essential in determining the class of macromolecular structure that can be formed. In this picture (adapted from Cook and Deserno, 2006) we show several examples of the equilibrium shapes obtained from computer simulations of aggregation of simple amphiphilic molecules. The basic geometry of these molecules, represented below, changes from a small head-large tail pattern to a large tail-small head one.

Since the key question is how cell replication can occur, we need to consider how likely is this to happen spontaneously by starting from a closed vesicle. If this vesicle as a radius R_0 , its volume would be

$$V(R_0) = \frac{4}{3}\pi R_0^3 \quad (2)$$

Now consider a potential process leading to vesicle splitting. If the resulting daughter vesicles are identical, their final volume V_1 would be such that $V_0 = 2V_1$. From this relation we can estimate the new radius of each cell:

$$R_1 = \sqrt[3]{\frac{1}{2}} R_0 \quad (3)$$

Using this radius, we can now calculate how energy has changed through the process. The energy associated to a membrane can be calculated on a first approximation by using the surface tension. Surface tension is the property of a surface $\Delta S = 4\pi(R_1^2 - R_0^2)$ that pervades its elastic behavior. The change in total surface, multiplied by the coefficient of surface tension γ will give us the energy change:

$$\Delta E = \gamma \Delta S = \gamma 4\pi(\sqrt[3]{2} - 1)R_0^2 > 0 \quad (4)$$

This positive value shows that the process would not take place spontaneously. Researchers have been aware of this problem and some models of early protocell formation consider the presence of external perturbations acting on lipid aggregates. Such physical forces would shake the prebiotic mixture in such a way that splitting would occur under environmental driving. Division would be imposed by external agitation (see Fernando and Rowe, 2007 and references therein).

Fig. 4. The shape of a given amphiphile can be captured by the packing parameter n_s (see text). This parameter allows determining the expected shape of the resulting self-assembled complex. Here three examples of these basic shapes are shown.

This simple argument tells us something important: although self-assembly is a highly energetically favoured mechanism and order can easily emerge from an initial set of randomly scattered molecules, the instability leading to replication is not expected to occur. As it happens with other self-organizing chemical systems, they have no control on the flow of energy and matter through them (Ruiz-Mirazo and Mavelli, 2007). Unless active mechanisms of membrane growth are in place in such a way that the membrane is forced to not just grow, but become unstable, no cell reproduction is expected to happen.

4. Chemical constraints

As noted by Harold Morowitz, in principle, the logic of replication could be separated from chemical constraints (Morowitz, 1992). Section II on self-replicating automata clearly makes a point in this direction. But the existence of physical constraints and the requirement of some geometrical properties allowing cell deformation to occur also place some limitations on the types of molecular components that can be relevant to achieve our goals.

Cellular life emerged somehow in a given chemical scenario where a high diversity of simple molecules was present (Fenchel, 2002). Moreover, an equilibrium state where reactions cannot generate further chemical complexity will forbid complex phenomena (including life) to take place. Out of equilibrium conditions might have been present in our planet in early stages of its geological history. Darwin himself considers this possibility in *Origins* (Darwin, 1859):

It is often said that all the conditions for the first production of a living organism are present, which could ever have been present. But if we could conceive in some warm little pond, with all sorts of ammonia and phosphoric salts, light, heat, electricity, etc., present, that a protein compound was chemically formed ready to undergo still more complex changes, at the present day such matter would be instantly devoured or absorbed, which would not have been the case before living creatures were formed.

Earth was a huge experimental arena for combining molecules and the work by Stanley Miller and others revealed the possibility of generating many simple building blocks (such as amino acids, nucleotides or sugars) starting from prebiotic initial conditions (Miller, 1953; Oró, 1961a; Monnard and Deamer, 2002). But the required elements for self-assembly of vesicles and micelles might have also arisen far from our planet. As early suggested by the Catalan astrochemist Joan Oró, comets might have been great contributors in supplying Earth with the requisite organic molecules for the process of chemical evolution (Oró, 1961b; Oró and Lazcano,

Fig. 5. Replicators and reproducers. At the chemical level, self-reproducing molecules can be obtained by means of an appropriate biochemical cycle. Here a given two-carbon molecule (C–C) can react with available monomers (C) leading to a four-carbon molecules C–C–C–C which then splits into two new C–C chains. This is known to occur in the so-called Formose reaction (see text).

1995). In this context, experimental work using UV photolysis applied to realistic interstellar ice analogs is able to generate complex molecules capable of self-assembly (Dworkin et al., 2001) thus confirming that complex assemblies can form spontaneously. The question is, of course, how these complexes can change, divide and even evolve.

There is also the problem of how the chemical reactions inside primitive compartments might have been able to generate useful chemicals. The best candidate is the presence of cycles as the one shown in Fig. 5. This corresponds to the so-called formose reaction (we do not indicate the exact molecular components, just the number of carbon atoms). In this reaction (Gánti, 2003; Szathmáry, 2006) we start from glycol aldehyde (C–C) which reacts with formaldehyde (C) leading to a three-carbon intermediate (glycerol aldehyde) producing a sugar (C–C–C). Finally, this sugar splits into two new glycol aldehydes and the cycle is completed. Although this is a nice (and observable) illustration of a simple, non-enzymatic catalytic cycle, other side reactions can also occur, draining molecular intermediates of the core cycle until it just stops (Shapiro, 1986). Actually, it is possible to reconstruct the graph of chemical possible reactions using a simplified quantum mechanical energy calculation (Benko et al., 2003a) which is shown to be a complex network with many potential chemical elements (Benko et al., 2003b). As discussed in (Szathmáry, 2006; Weber, 2007, 2008; see also Toxvaerd, 2005) this seems to be an important limitation to naked catalytic chemical systems.

5. Dynamical constraints

Before we engage in presenting a model of whole protocell replication, we need to consider a more general framework as described by simple molecular replicators (Maynard-Smith and Szathmáry, 1997). This is important for two reasons. First, replicator dynamics is a very general framework which allows modelling any class of system having reproduction and competition. That includes protocellular and cellular systems, although once again no embodiment is present. The second reason is that the type of dynamical behavior displayed will have an impact on the expected evolution of these systems.

Under a very coarse grained description, the replication process followed by a protocellular system P would be described by a

reaction:

where P , S and W indicate protocell, substrate monomers and waste, respectively. Although we emphasize here the requirement for a phase separation between protocell and the environment, it is useful to look at the problem in terms of population dynamics on a large scale approximation.

If we indicate by A the replicating molecule and by S the required precursor molecules used in the replication of A , the simplest scenario can be described as a simple reaction:

This process leads to an exponential growth of A (provided that S is constantly supplied. When two different molecules compete for a given substrate following this growth law, only the fastest replicator wins. This is somehow the skeleton of Darwinian dynamics: the survival of the most efficient (fittest) replicator. This type of reaction dynamics is thus what defines the “survival of the fittest” effect.

A different situation is described by a cooperative link between molecules, namely a reaction like:

when two copies of A would be needed (together with S monomers) in order to generate a new one. This mechanism leads to a so-called *hyperbolic growth*: a highly nonlinear population increase characterized by a finite time scale at which the number of individuals rapidly saturates. Hyperbolic growth also displays an interesting effect: the molecular species most common at the beginning is the winner, provided it replicates fast enough. This would lead to the “survival of the common” scenario. Growth rates would be relevant but not less than initial population sizes. This cooperative cycle is known as the *hypercycle* (Eigen and Schuster, 1977) and is characterized by the presence of two molecular species (not necessarily identical) enhancing their replication in a mutualistic fashion. The hypercycle represents one of the key innovations in the evolution of complexity (Schuster, 1999).

The relevant point to be made is that autocatalytic chemical reactions pervade Darwinian evolution and are required to achieve a successful propagation of biological order. All theoretical models considering the events pervading the emergence of these cooperative systems and their transition to cellular structures assume the importance of having containers able to isolate the cooperative cycles from the environment. Such containers would help protect the system from molecular parasites and allow a more efficient catalytic activity. But surprisingly, little has been done in explicitly modeling the emergence and dynamics of such protocellular structures. In this context, it is not clear how these containers will behave and their robustness to different sources of disorder. As will be shown below, a minimal model of protocell incorporating some class of cooperative growth shows that the path towards cellular life might be shorter than we expect.

6. Computational modelling of protocell replication

An important component of the protocell origins problem has to do with the physical embodiment associated with the replication cycle (Szostak et al., 2001). This is an obvious but largely forgotten point: in order to self-replicate, a cell needs to be out of equilibrium, grow in size, *change its shape* and eventually split. Several important results have been reported from experiments (Bachman et al., 1992; Walde et al., 1994; Oberholzer et al., 1995, 1999; Chen et al., 2004; Hanczyc et al., 2003; Hanczyc and Szostak, 2004; Monnard and Deamer, 2001; Oberholzer and Luisi, 2002; Chen and Szostak,

2004; Noireaux and Libchaber, 2004; Noireaux et al., 2005; Deamer, 2005; Mansy et al., 2008) On the other hand, standard models of cell replication have considered the problem of growth as the essential element (see for example Segre et al., 2001 and Mavelli and Ruiz-Mirazo, 2007 and references cited) so that the cycle gets reduced to a problem of increase in total number of molecules followed by a division. But the spatially embodied, physical process of cell division has not been explicitly addressed until recently (Fellermann, 2005; Solé et al., 2007; Macia and Solé, 2007a,b; Fellermann and Solé, 2006; Fellermann et al., 2007; Solé et al., 2008). However, as we just discussed, the behavior of micelles and vesicles is strongly dependent on the kind of molecule forming them, at least in terms of its geometrical features.

Protocells are embodied structures and as such their changes over time must be taken into account. Not every class of molecule – no matter how nicely it can contribute to some catalytic cycle – can form a compartment. What is more important: in order to be sure that cooperation among metabolism and membrane is successful in allowing a cell to replicate, they need to be both included in a spatially explicit framework. Unless such a full model is able to show that self-reproduction will occur, we cannot claim that a given model predicts such an event to occur. It could be the case that, once we include this essential component in the whole picture, getting a division cycle turns out to be a very difficult task. Or instead, it might be the case that such an event is not so difficult to be observed. In the following section we present an example of such spatially embodied protocell model revealing that, even under noisy conditions resulting from the nanoscale associated with our tiny protocells, a repeatable cell cycle can be obtained.

Any microscopic model of protocell replication including a physical description of molecular interactions will mean a hard computational problem. First, we need to make decisions on what to take and what to leave as (hopefully) not relevant. A full simulation system would require a molecular dynamics approach where every pair of interacting molecules is taken into account and the whole atomic complexity of every molecule is carefully considered. This is of course an ideal situation but even if such implementation were possible, exploring the large parameter space (see below) would be impossible. On the other hand, the mesoscopic behavior of micellar and vesicle assemblies has been shown to be fairly well replicated by using simplified models of amphiphiles. Even toy lattice models, where molecules are described as balls and sticks moving on a grid, can be very successful in capturing the truth of these macromolecular objects. Following this perspective, we can try to see if a physically sensible (even if oversimplified) model of protocell growth can undergo a replication cycle.

The goal of such a simple model would be to provide tentative answers to some key questions, such as:

1. Is there some complexity threshold that needs to be achieved in order to have replicating protocells?
2. Can random (thermal) fluctuations block or damage the process of replication, making difficult to obtain a reliable division cycle?
3. Is this process highly parameter dependent, and thus perhaps unlikely to happen or it is instead robust under wide parameter combinations?

Answering these questions would help providing powerful insight into the problem of the emergence of early cells and their potential for evolution. If only cells that are already chemical complex would be able to sustain protocell cycles, the emergence of cellular systems might have been an extremely rare phenomenon. The same would happen if small complexity allows to obtain a replicating protocell, but only under some fine-tuned parameter

combinations. Finally, since early forms of cell-like structures were likely to experience external fluctuations, the system should be able to resist the impact of such perturbations by forming robust assemblies and controlling their changes through the cycle.

6.1. Logic of the model

The basic logic of our model is roughly described in Fig. 6A–D, where we display the schematic “cell cycle” we wish to obtain. Here we consider a system involving water (not to be shown in the pictures), amphiphiles (H - T molecules) and precursors (T - T) which we assume to be hydrophobic and be formed by two tail beads. The model needs to include the basic, correct physics (see below) allowing amphiphiles to self-assemble into micelles. It will also consider a single chemical reaction happening close to the assembly (Fig. 6E) where precursors get transformed into amphiphiles while close to other amphiphiles.

We can see that there is a cooperative phenomenon here, where the size of the micelle increases by cooperatively transforming surrounding precursors into new micelle components. The hope here is that, with the appropriate molecular shapes and kinetic parameters, the micelle will grow until it roughly duplicates its size and then split into two similar daughter assemblies. Here the relevance of the amphiphile shape becomes clear: if the expected equilibrium shape of the small micelle is a sphere, amphiphiles should have a packing parameter $N_s \approx 1/3$. In order to obtain spherical micelles, the interaction parameters between different components was chosen in order to approach the $1/3$ value. But we also need the micelle to grow smoothly until the tails become separated and the elongated shape becomes too unstable (Fig. 6C) to persist. Luckily, the splitting event will generate two similarly sized daughter cells.

6.2. Physics and chemistry of protocell replication

Here we briefly describe a recent model of protocell self-assembly and replication introduced by Fellermann and Solé within the context of information-free systems (for details see Fellermann and Solé, 2006 and references therein). This was the first full model of protocell replication including all the relevant components under a realistic physico-chemical framework. The model involves very small sized micellar structures, thus introducing further sources of instability, since noise is expected to strongly influence self-assembly in both space and time. Since random fluctuations destroy information, a protocellular system able to experience a reliable cycle of growth and split in such changing environment

The model was implemented using a so-called dissipative particle dynamics (DPD) approach. This is a coarse grained simulation technique which has been successfully used to capture the dynamics of membranes, vesicles and micelles. In this computational view, the system is represented by a set of N particles. These are described by their class (water, head, tail), their mass m_i , position $\mathbf{r}_i$, and momentum $\mathbf{q}_i = m_i \mathbf{v}_i$. These particles (or “beads”) are not meant to represent individual atoms. Instead, they represent groups of atoms within a molecule (like several CH_2 groups within a hydrocarbon chain) or even a group of small molecules such as water.

The DPD method allows to efficiently compute the dynamical behavior of these systems. Each bead will move following Newton's Law of motion, i. e.:

$$\frac{d^2 \mathbf{r}_i}{dt^2} = \frac{1}{m} \mathbf{F}_i \quad (8)$$

where the force $\mathbf{F}_i$ acting on the i -th particle is actually a superposition of pairwise interactions, namely a sum over all other particles

Fig. 6. The basic model of the Fellermann-Solé model of nanocell replication. The goal of the model was to generate a stable cell cycle (A–D). Here small-sized micelles are formed by amphiphiles (here indicated as *H*–*T* connected pairs of balls). These amphiphiles have a hydrophilic head (*H*) and a hydrophobic tail (*T*). Precursor molecules are also shown as two connected, smaller open balls, both of them of hydrophobic character. Under the presence of amphiphiles, precursors are transformed into new amphiphiles (E) allowing the micelle to grow. When a critical size has been reached, the nanocell becomes unstable and splits into two aggregates thereby closing the cycle.

interacting with *i*. Three types of forces are considered:

$$F_i = \sum_{j=1}^N (F_{ij}^C + F_{ij}^D + F_{ij}^R) \quad (9)$$

The first component, F_{ij}^C is the standard (Newtonian) central force, to be written here as

$$F_{ij}^C = a_{ij} \left(1 - \frac{r_{ij}}{r_c} \right) \quad (10)$$

for $r < r_c$ and zero otherwise. This is actually a main aspect of DPD dynamics: the truncation of all pairwise interactions as given by some characteristic spatial scale r_c . Those particles at a distance larger than this value are thus considered as independent. In this way, computations become largely simplified. This central force can also be described as the negative gradient of a potential ϕ_{ij} , namely

$$F_{ij}^C = -\nabla \phi_{ij} \quad (11)$$

In the model described here, the conservative force is derived from a *soft-core potential* described by the function

$$\phi_{ij}(r) = \begin{cases} \frac{1}{2} a_{ij} r_c \left(1 - \frac{r}{r_c} \right)^2 & \text{if } r \leq r_c \\ 0 & \text{if } r > r_c \end{cases} \quad (12)$$

DPD considers further components to add to the force fields, including a so-called dissipative force introducing viscosity and a random force term, accounting for thermal effects. All these components are properly introduced in such a way they satisfy the so-called fluctuation–dissipation theorem, which establishes that a fluctuation in the system will die out as the system goes back to its equilibrium state. Such decay must be proportional to the fluctuation size (for a summary of these ideas and their implementation, see Fellermann, 2005).

In this model, as with most previous DPD studies, the conservative force is derived from a *soft-core potential* described by the

Fig. 7. Whole nanocell replication obtained from the DPD model (see text). Here water molecules and isolated precursors are not shown. In (A) an intermediate step is shown with two large $N \approx 30$ amphiphiles each. One of them (surrounded by a dashed square) keeps growing (the other one too, but we don't follow its growth) and in (B) and (C) we can see its larger size. As it grows beyond $N \approx 100$, it destabilizes and starts to deform (D). Eventually, two subsets get formed (E) and split (F).

function

$$\phi_{ij}(r) = \begin{cases} \frac{1}{2} a_{ij} r_c \left(1 - \frac{r}{r_c}\right)^2 & \text{if } r \leq r_c \\ 0 & \text{if } r > r_c \end{cases} \quad (13)$$

For the study of lipids and surfactants, covalent bonds between beads are commonly introduced as harmonic spring forces: on top of the above interactions, bonded beads interact according to the potential

$$\phi_{ij}^B(r) = \frac{br_b}{2} \left(1 - \frac{r}{r_b}\right)^2 \quad (14)$$

where b is the strength and r_b the optimal distance of covalent bonds. As usual, we use r_c , m , and $k_b T$ as units of space, mass, and energy, respectively. The time unit follows from Eq. (8) as $\tau = \sqrt{m/k_b T r_c}$.

To model the system under consideration, we define beads of type **W** (water), **H** (hydrophilic “heads”) and **T** (hydrophobic “tails”) of amphiphiles). Each pair interacts with some strength given by a symmetric matrix of weights:

	W	H	T
W	$25 k_b T$	$15 k_b T$	$80 k_b T$
H	$15 k_b T$	$35 k_b T$	$80 k_b T$
T	$80 k_b T$	$80 k_b T$	$15 k_b T$

For simplicity, all masses were fixed to unity. Precursor molecules are modeled as dimers of bonded **T** beads, surfactants as dimers of one **T** and one **H** bead. Here we used: $b = 125 k_b T$, $r_b = 0.5 r_c$ for all covalent bonds.

The objective behind this parameter set is to model surfactants that form spherical micelles. To achieve this, the effective head area must be large compared to the volume of the hydrophobic core (packing parameter $1/3$). This is expressed by $a_{TT} < a_{WW} < a_{HH}$. Furthermore, surfactant heads have a high affinity to water ($a_{HW} < a_{WW}$), which is usually due to charges in the hydrophilic groups of the molecules.

Finally (and that represented an important variation with respect to standard DPD) a simple chemistry was introduced by considering an amphiphile-catalyzed reaction transforming T - T

precursors into new amphiphiles (Fig. 6E) under the presence of nearby amphiphiles. The metabolic reaction under consideration takes the following form

This reaction is modeled by a stochastic process (?). Each precursor dimer can be spontaneously transformed into a surfactant molecule at a rate k_b but the process is enhanced by nearby surfactants whose catalytic influence decreases with the distance to the reactant up to a certain threshold ξ . Explicitly, the effect of several catalysts is modelled as follows: superposition:

$$k = k_b + \sum_{\langle mn \rangle} k_s \left(1 - \frac{r_c}{\xi} \right) \quad (16)$$

if $r < \xi$ and zero otherwise. The sum is performed over nearest neighbors (labelled as $\langle mn \rangle$). Here r_c is the distance of the catalyst and k_s the maximal catalytic rate per catalyst. In (Fellermann and Solé, 2006) k_{cat} was set to $1.0 \tau^{-1}$ and r_{cat} to $1r_c$. If a reaction occurs, the type of one random $\mathbf{T}$ bead is changed to $\mathbf{H}$, but positions and momenta are preserved.

6.3. Nanocell replication is robust

The previous model was successful in generating replicating protocells. In spite of the simplicity of the rules and the inevitable presence of fluctuations associated to the nanoscale being considered, it was not difficult to obtain reliable sets of parameters allowing cell replication to occur (Fig. 7). What is more important, the general replication cycle of micellar nanocells by metabolic turnover and division is very robust against changes in hydrophobicity and catalytic rates (Fellermann and Solé, 2006). It was shown that key statistical quantities such as the mean size are essentially dependent on the hydrophobicity of the surfactant (and precursor) as well as on the metabolic catalytic rate. Moreover, we also found that the rate of nanocell fission and surfactant dissociation depends on the size of the hydrophobic core of the nanocells, and is more likely to occur for small values in hydrophobicity and slow metabolic turnover. Daughter cells resulting from a fission event have been shown to vary significantly in size.

In summary, this model approach revealed that, under our basic mechanism (incorporation and turnover of precursor droplets followed by eventual aggregate division) a cell division cycle is achievable over a wide range of parameters. In fact, the most remarkable result of this work was that there is no parameter combination for which the general replication cycle has been rendered impossible: even at very low mean aggregation numbers and large dissociation rates, micelles grew and divided. This result strongly suggests that the basic set of rules and the logic of the process (more than the exact parameters) is the key for finding a self-replicating protocell. Such positive result indicates that very simple mechanisms of micelle-metabolism coupling in a primitive Earth scenario might have been able to trigger the proliferation of simple protocells. Since no information is included in this system, no further evolution is expected to occur. But this might have provided the initial, functional subsystem providing metabolism and embodiment for subsequent protocells of higher evolutionary complexity.

This picture fits the Oparin-Dyson view of life origins lacking information (such as RNA chains) which would be incorporated later. It is conceivable, that independently evolved information systems like RNA might have become incorporated into such functioning replicators. When the two formerly independent replication cycles of container and genome are orchestrated by coupling, such that each daughter cell of the dividing container is loaded with exactly one copy of the genomic information, one would obtain

Fig. 8. Classes of replicating systems within a complexity space including metabolism, membrane and information as the principal axes. The surface indicates the presence of a transition zone between cellular and non-cellular forms of organization. Below this surface, no replicating cells are present. This includes catalytic cycles, viruses or even stars, which have a complex spatial structure and a simple “metabolic” cycle involving thermodynamic reactions. The surface can be crossed by increasing chemical complexity.

a true self-reproducing protocell with the ability to metabolize, divide and evolve. Successful steps in modeling these type of protocells have been done (Fellermann et al., 2007).

7. Discussion

The winding road that leads to cellular life is marked by some special major events (Maynard-Smith and Szathmari, 1997; Schuster, 1999). The initial phases of such an evolutionary path, before information molecules and proteins took control over cell cycle, was necessarily influenced by the advantages and constraints derived from the physics of self-assembly and its interactions with primitive forms of metabolism. In Fig. 8 we roughly display a map of cell complexity using three axes: metabolic, spatial (membrane) and informational complexities. These axes are not representing any properly defined scale but can be thought as a basic guide for locating prototype classes of replicating systems. In this picture I use a surface separating several types of systems exhibiting some kind of replication cycle. These can include a container-free scenario or even a host-dependent environment as it happens with viruses. Once this twilight zone is crossed, cellular life can sustain itself. The surface displayed here does not derive from any theoretical argument but the basic message is clear: once some basic thresholds are reached, cells should be expected to emerge.

Self-replicating molecules might have been in place early on in our planet, as well as the first self-assembled lipid aggregates. But the transition to protocellular life has been always considered a much more difficult step. Such a conclusion seems reasonable if we consider all the constraints listed in our paper: several requirements seem needed before the right replicating machine is built. The model presented here is a good counterexample for these limitations. Although it just includes the bare bones of the physics and chemistry of a toy protocell model, it fairly well illustrates how robust is the coupling between self-assembled amphiphiles coupled to an external source of precursors and displaying a simple catalytic reaction. The robustness of the observed results supports

the view of cellular life as a likely event to happen provided that the basic molecular logic is in place.

Self-assembly is an essential component in the path towards cellular systems. The spontaneous generation of spatial order, allowing to easily define a container, is still at work at different scales of biological organization. Although modern cells have a sophisticated molecular machinery driving and controlling the whole replication process, the dynamics of many subcellular compartments take advantage of the physics of self-assembly. This is the case of vesicular transport systems which are being constantly formed, modified and degraded (Alberts et al., 2002). But it is also the mechanism used by viruses in order to encapsulate their genomes (Dimmock and Primrose, 1994). Self-assembly is used by viruses to generate viral particles and relies on a different geometrical principle, now associated to the mathematical constraints operating on a symmetrical object enclosing some empty space. In this case, a limited range of possibilities is at work, and not surprisingly, many viral capsids are easily recognized as icosahedrons formed by the regular packing of the same protein. Future work within the emergent field of synthetic biology (Solé et al., 2007) should throw light on how this events happened and perhaps recreate them again in the laboratory.

The author is indebted to the PACE/ICREA Complex Systems Lab dream team: Javier Macia, Andreea Munteanu, Sergi Valverde, Josep Sardanyes, and very specially to Harold Fellerman, with whom I did some of the work described here. I am also grateful to the European Center of Living Technology in Venice (particularly to Irene Poli) for hosting my group and allowing us to spend great days -and nights- of intense work. I would like to thank also Steen Rasmussen, Stuart Kauffman, Norman Packard, Gunter v. Kiedrowski and John McCaskill for so many interesting discussions on protolife. This work has been supported by EU PACE grant within the 6th Framework Program under contract FP6-0022035 (Programmable Artificial Cell Evolution), the James McDonnell Foundation and the Santa Fe Institute.

References

- Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. Molecular biology of the cell. New York: Garland; 2002.
- Bachman PA, Luisi PL, Lang J. Autocatalytic self-replicating micelles as models for prebiotic structures. *Nature* 1992;12(357):57–9.
- Benko G, Flamm C, Stadler P. A graph-based toy model of chemistry. *J Chem Inf Comput Sci* 2003a;43:1085–93.
- Benko G, Flamm C, Stadler P. Generic properties of chemical networks: artificial chemistry based on graph rewriting. *Lect Notes Comput Sci* 2003b;2801:10–9.
- Bozic B, Svetina S. A relationship between membrane properties forms the basis of a selectivity mechanism for vesicle self-reproduction. *Eur Biophys J* 2004;33:565–71.
- Chen IA, Szostak JW. A kinetic study of the growth of fatty acid vesicles. *Biophys J* 2004;87:988–98.
- Chen IA, Roberts RW, Szostak JW. The emergence of competition between model protocells. *Science* 2004;305:1474–6.
- Cook IR, Deserno M. Coupling between lipid shape and membrane curvature. *Biophys J* 2006;91:487–95.
- Darwin C. The origin of species by means of natural selection. London: John Murray; 1859.
- Deamer D. A giant step toward artificial life? *Trends Biotechnol* 2005;23:336–8.
- Deamer D, Dworkin J, Sandford S, Bernstein M, Allamandola L. The first cell membranes. *Astrobiology* 2002;2:371–81.
- Dimmock NJ, Primrose SB. Introduction to modern virology. Oxford: Blackwell Scientific; 1994.
- Dworkin JP, Deamer DW, Sandford SA, Allamandola LJ. Self-assembling amphiphilic molecules: synthesis in simulated interstellar/precometary ices. *Proc Natl Acad Sci USA* 2001;98:815–9.
- Eigen M, Schuster P. The hypercycle: a principle of natural self-organization. New York: Berlin Heidelberg, Springer; 1977.
- Evans DF, Wennerstrom H. The colloidal domain. New York: VCH Publishers; 1999.
- Fellermann H, Micelles as containers for protocells (master thesis). Germany: University of Osnabrück; 2005.
- Fellermann H, Solé RV. Minimal model of self-replicating nanocells: a physically embodied information-free scenario. *Philos Trans R Soc B* 2006;362:1803–11.

- Fellermann H, Rasmussen S, Ziock H, Solé RV. Life cycle of a minimal protocell: a dissipative particle dynamics study. *Artif Life* 2007;13:319–45.
- Fenchel T. Origin and early evolution of life. New York: Oxford U. Press; 2002.
- Fernando C, Rowe J. The origin of autonomous agents by natural selection. *Biosystems* 2007;91:355–73.
- Freitas RA, Merkle RC. Kinematic self-replicating machines. Landes bioscience. USA: Georgetown; 2004.
- Gánti T. The principles of life. Oxford: Oxford University Press; 2003.
- Hanczyc M, Szostak J. Replicating vesicles as models of primitive cell growth and division. *Curr Opin Chem Biol* 2004;8:660–4.
- Hanczyc M, Fujikawa SM, Szostak J. Experimental models of primitive cellular compartments: encapsulation, growth and division. *Science* 2003;302:618–22.
- Harold FM. The way of the cell. Oxford: Oxford U. Press; 2001.
- Jortner J. Conditions for the emergence of life. *Philos Trans R Soc B* 2006;361:1877–91.
- Jung HT, Coldren B, Zasadzinski JA, Iampietro DJ, Kaler EW. The origins of stability of spontaneous vesicles. *Proc Natl Acad Sci USA* 2001;98:1353–7.
- Luisi PL. The emergence of life: from chemical origins to synthetic biology. Cambridge UK: Cambridge U. Press; 2006.
- Macia J, Solé RV. Protocell self-reproduction in a spatially explicit metabolism-vesicle system. *J Theor Biol* 2007a;245:400–10.
- Macia J, Solé RV. Synthetic turing protocells: self-reproducing vesicles through symmetry-breaking instabilities. *Philos Trans R Soc B* 2007b:1821–9.
- Maynard-Smith J, Szathmáry E. The major transitions in evolution. Oxford: Oxford U. Press; 1997.
- Mansy SS, Schrum JP, Krishnamurthy M, Tobž S, Treco DA, Szostak JW. Template-directed synthesis of a genetic polymer in a model protocell. *Nature* 2008;454:122–5.
- Mavelli F, Ruiz-Mirazo K. A stochastic simulation model of minimal self-(re-)producing cellular systems. *Philos Trans R Soc B* 2007;362:1789–802.
- Miller S. A production of amino acids under possible primitive earth conditions. *Science* 1953;117:528–9.
- Monnard P, Deamer D. Nutrient uptake by protocells: a liposome model system. *Orig Life Evol Biosph* 2001;31:147–55.
- Monnard P, Deamer D. Membrane self-assembly processes: steps toward the first cellular life. *Anat Rec* 2002;268:196–207.
- Morowitz H. Beginnings of cellular life. New Haven: Yale U. Press; 1992.
- Mouritsen O. Life as a matter of fat: the emerging science of lipidomics. Berlin: Springer; 2005.
- Nicolis G, Prigogine I. Exploring complexity. New York: Freeman; 1977.
- Noireaux V, Libchaber A. A vesicle bioreactor as a step toward an artificial cell assembly. *Proc Natl Acad Sci USA* 2004;101:17669–74.
- Noireaux V, Bar-Ziv R, Godefroy J, Salman H, Libchaber A. Toward an artificial cell based on gene expression in vesicles. *Phys Biol* 2005;2:1–8.
- Nurse P. Life, logic and information. *Nature* 2008;454:424–6.
- Oberholzer T, Luisi PL. The use of liposomes for constructing cell models. *J Biol Phys* 2002;28:733–44.
- Oberholzer T, Wick R, Biebricher CK, Luisi PL. Enzymatic RNA replication in self-reproducing vesicles: an approach to a minimal cell. *Biochem Biophys Res Commun* 1995;207(250–257):250–7.
- Oberholzer T, Nierhaus KH, Luisi PL. Protein expression in liposomes. *Biochem Biophys Res Commun* 1999;261:238–41.
- Oró J. Mechanism of synthesis of adenine from hydrogen cyanide under possible primitive Earth conditions. *Nature* 1961a;191:1193–4.
- Oró J. Comets and the formation of biochemical components on the primitive earth. *Nature* 1961b;190:389–90.
- Oró J, Lazzano A. Comets and life in the universe. *Adv Space Res* 1995;13:81–90.
- Pohorille A, Deamer D. Artificial cells: prospects for biotechnology. *Trends Biotechnol* 2002;20:123–8.
- Rasmussen S, Chen L, Nilsson M, Abe S. Bridging nonliving and living matter. *Artif Life* 2003;9:269–316.
- Rasmussen S, Bedau MA, Chen L, Deamer D, Krakauer DC, Packard NH, et al., editors. Protocells: bridging nonliving and living matter. MIT Press; 2008.
- Ruiz-Mirazo K, Mavelli F. On the way towards basic autonomous agents: stochastic simulations of lipid-peptide cells. *Biosystems* 2007;91:374–87.
- Ruiz-Mirazo K, Moreno A. basic autonomy as a fundamental step in the origin of life. *Artif Life* 2004;10:235–59.
- Schuster P. How does complexity arise in evolution. *Complexity* 1999;2:22–30.
- Segre D, Ben-Eli D, Deamer D, lancet D. The lipid world. *Orig Life Evol Biosph* 2001;31:119–45.
- Shapiro R. A skeptic's guide to the creation of life on Earth. New York: Summit Books; 1986.
- Solé RV, Munteanu A, Rodríguez-Caso C, Mac'á J. Synthetic protocell biology: from reproduction to computation. *Philos Trans R Soc Lond* 2007.
- Solé RV, Macia J, Fellermann H, Munteanu A, Sardanyes J, Valverde S. Models of protocell replication. In: Rasmussen, et al., editors. Protocells: bridging living and non-living matter. Cambridge, Massachusetts: MIT Press; 2008.
- Szathmáry E. The origin of replicators and reproducers. *Phil Trans R Soc Lond B* 2006;361:1761–76.
- Szathmáry E, Santos M, Fernando C. Evolutionary potential and requirements for minimal protocells. *Top Curr Chem* 2005;259:167–211.
- Szostak JW, Bartel DP, Luisi PL. Synthesizing life. *Nature* 2001;409:387–90.
- Toxvaerd S. Homochirality in bio-organic systems and glyceraldehyde in the formose reaction. *J Biol Phys* 2005;31:599–605.

- von Neumann J. Theory of self-reproducing automata. Urbana, IL: Univ. of Illinois Press; 1966.
- Walde P, Wick R, Fresta M, Mangone A, Luisi PL. Autopoietic self-reproduction of fatty acid vesicles. *J Am Chem Soc* 1994;116:11649–54.
- Weber AL. The sugar model: autocatalytic activity of the triose-ammonia reaction. *Orig Life Evol Biosph* 2007;37:105–11.
- Weber AL. Sugar-driven prebiotic synthesis of 3,5(6)-dimethylpyrazin-2-one: a possible nucleobase of a primitive replication process. *Orig Life Evol Biosph* 2008;38:279–92.
- Zykov V, Mytilinaios E, Adams B, Lipson H. Self-replicating machines. *Nature* 2005;435:163–4.