Life Cycle of a Minimal Protocell—A Dissipative Particle Dynamics Study

Harold Fellermann
Self-Organizing Systems
EES-6 MS-D462
Los Alamos National Laboratory
Los Alamos NM 87545
and
ICREA-Complex Systems Lab
Universitat Pompeu Fabra (GRIB)
Dr Aiguader 80
08003 Barcelona, Spain
harold.fellermann@upf.edu

Steen Rasmussen
Self-Organizing Systems
EES-6 MS-D462
Los Alamos National Laboratory
Los Alamos NM 87545
and
Santa Fe Institute
1399 Hyde Park Road
Santa Fe NM 87501
steen@lanl.gov

Hans-Joachim Ziock
Self-Organizing Systems
EES-6 MS-D462
Los Alamos National Laboratory
Los Alamos NM 87545
ziock@lanl.gov

Ricard V. Solé
ICREA-Complex Systems Lab
Universitat Pompeu Fabra (GRIB)
Dr Aiguader 80
08003 Barcelona, Spain
and
Santa Fe Institute
1399 Hyde Park Road
Santa Fe NM 87501
ricard.sole@upf.edu

Abstract Cross-reactions and other systematic difficulties generated by the coupling of functional chemical subsystems pose the largest challenge for assembling a viable protocell in the laboratory. Our current work seeks to identify and clarify such key issues as we represent and analyze in simulation a full implementation of a minimal protocell. Using a 3D dissipative particle dynamics simulation method, we are able to address the coupled diffusion, self-assembly, and chemical reaction processes required to model a full life cycle of a protocell composed of coupled genetic, metabolic, and container subsystems. Utilizing this minimal structural and functional representation of the constituent molecules, their interactions, and their reactions, we identify and explore the nature of the many linked processes for the full protocellular system. Obviously the simplicity of this simulation method combined with the inherent system complexity prevents us from expecting quantitative simulation predictions from these investigations. However, we report important findings on systemic processes, some previously predicted and some newly discovered, as we couple the protocellular self-assembly processes and chemical reactions.

1   Introduction

The twilight zone that separates nonliving matter from life involves the assembly of and cooperation among different subcomponents, which we can identify as metabolism, information, and compartment. None of these ingredients are living, and none of them can be ignored when looking at life as a whole. When assembled appropriately in a functional manner, their systemic properties constitute minimal life.

Understanding the transition from nonliving to living matter requires one to consider far simpler cells than those of modern life. Cells as we know them in our current biosphere are highly complex. Even the simplest, parasitic cellular forms involve hundreds of genes, complex molecular machineries for energy harvesting and utilization, and intricate membrane structures [2]. Such modern organisms are presumably far away from the initial simple forms of cellular life that inhabited our planet a long time ago, whose primitive early cousins we are now attempting to assemble in the laboratory [26].

Several complementary designs of protocells have been proposed that differ in the actual coupling between their various internal components [12, 16, 19, 27, 28]. One particularly important problem here, beyond the specific physical and chemical difficulties associated with the assembly of these protocells, is the problem of modeling the coupling of the possible kinetic and structural scenarios that lead to a full cell cycle. None of the current proposed designs has yet been formulated in a full mathematical model that in a 3D simulation is able to generate the possible outcomes of a successful coupling between the three prime components: the genes, the metabolism, and the container. We believe that a physically well-grounded modeling approach can provide critical insight into what can be expected from a coupled set of structures and reactions, how the nanoscale stochasticity can jeopardize appropriate molecular interactions, and even what are the effects of molecular information carriers in helping accurate replication to occur. In this article we present such a minimal 3D model, which in connection with ongoing experimental efforts is aimed at assembling and understanding a new class of nanoscale-sized protocells: the so-called *Los Alamos bug*.

In the Los Alamos bug, the container is built of amphiphilic surfactants. Due to their interaction with water, the surfactants spontaneously self-assemble into micelles with the hydrophobic ends of the surfactant molecules in the interior of the micelles and their hydrophilic ends in contact with the surrounding water. The interactions between the micelle and the other components of the Los Alamos bug, namely the photosensitizer, the genome, and the container precursors, allow the micelles to host these other components.

The genomic biopolymer (possibly decorated with hydrophobic anchors) is also an amphiphile, and due to the specific nature of its interactions with water and the micelle, it will tend to reside at the surface of the micelle (see Figure 1b). The sensitizer is a hydrophobic molecule and will therefore reside in the interior of the micelle. Once self-assembled, the protocell aggregate is fed with precursor molecules for the surfactants (oily esters), sensitizers, and genomic precursor oligomers. As surfactant precursors are hydrophobic, they will agglomerate inside the proto-organism and form a hydrophobic core (Figure 1c). Light energy is used by the metabolism to transform precursors into new building blocks (surfactants and oligomers) of the protocell. The genomic oligomers that are complementary with particular stretches of the template strand will hybridize with it (Figure 1d). The fully hybridized template-oligomers complex, which now only has hydrophobic elements exposed, will move into the interior of the container, where polymerization of the oligomers occurs, followed at some later time by a random dissociation of the fully polymerized double-stranded genome into two single-stranded templates that move back to the surface. This process could also be enhanced by a temperature cycle around the gene duplex melting point that is gentle enough to preserve the integrity of the aggregate.

As surfactant precursors are digested, the core volume of the protocell decreases while at the same time new surfactants are produced. The resulting change in the surface-to-volume ratio causes the micelle to become unstable (Figure 1e), until it finally splits into two daughter cells (Figure 1f). Assuming that components of the growing parent micelle are appropriately distributed upon division, the two daughter cells will be replicates of the original organism, thus completing the protocell cycle.
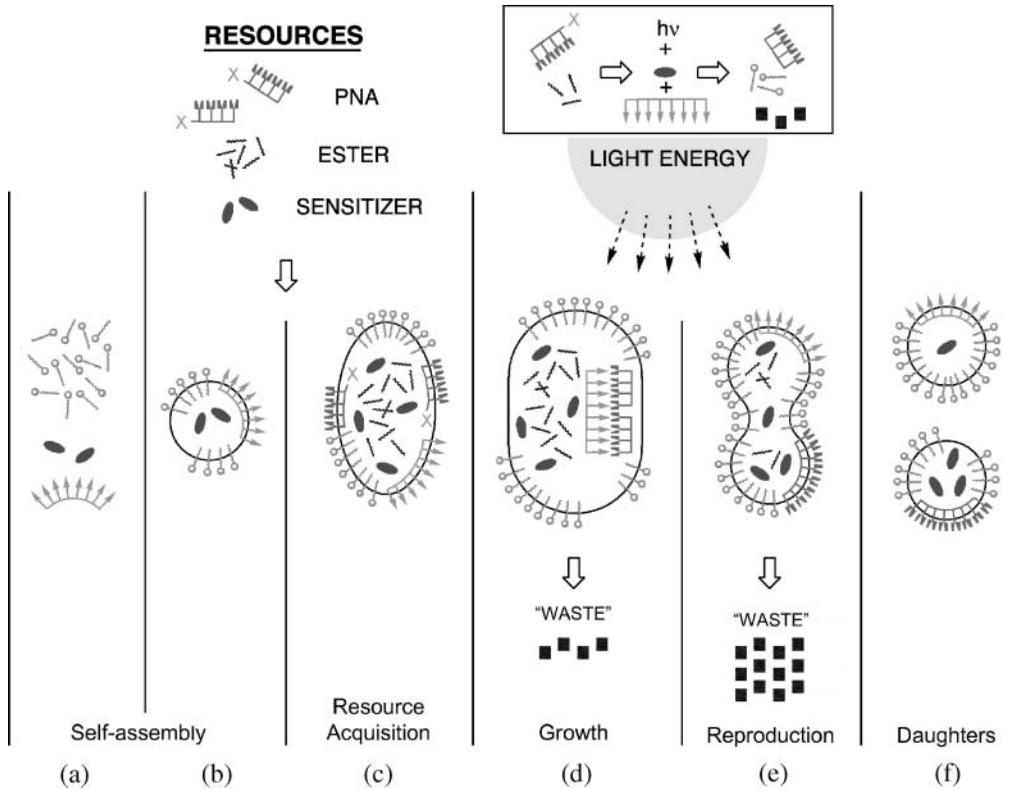
Figure 1. Schematic of the life cycle of the Los Alamos bug. (a) The system consists of surfactants, sensitizers, and a biopolymer that acts as a genome. (b) The surfactants spontaneously self-assemble into a micellar container within which the sensitizer resides while the biopolymer sticks at the surface of the container — this forms a complete protocell. (c) Resources (genomic oligomers, sensitizers, and surfactant precursors in the form of esters) are added to the system and get incorporated into the container. (d) The existing information carrier acts as a template for supplied oligomers to hybridize and effectively replicate the genome. (e) Light energy is used to convert the surfactant precursor and the oligomer precursors into actual surfactant, oligomers, and waste. The container grows as new surfactants are produced. (f) Once the container reaches a critical size, it becomes unstable and divides into two daughter cells. This completes the life cycle of the protocell.

In the above setup, the container, genome, and metabolism are coupled in various ways. Obviously, both the replication of the container and replication of the genome depend on a functioning metabolism, as the latter provides building blocks for aggregate growth and reproduction. In addition to that, the container also has a catalytic influence on the replication of both the metabolic elements and the genome: The micellar structure provides a compartment, which brings precursors, sensitizers, and nucleic acids into close vicinity, thereby increasing local concentrations and thus metabolic turnover. Furthermore, the micellar interface catalyzes the hybridization of the informational polymer with its complementary oligomer. Once the hybridized complex enters the water-poor or -free interior of a micelle, the thermodynamics should change sufficiently to allow a dehydration reaction to occur whereby the oligomers become polymerized. Alternatively, either the water-lipid interface could itself act as a ligation catalyst or the addition of simple amphiphilic catalysts could facilitate the gene polymerization process. Last, but not least, the nucleic acid catalyzes the metabolism, which otherwise is extremely slow. A summary of the subsystem coupling is shown in Figure 2.

2 The Model

Dissipative particle dynamics (DPD) is a mesoscale simulation method introduced by Hoogerbrugge and Koelman in 1992. The method has been improved as a result of theoretical support, revision, and
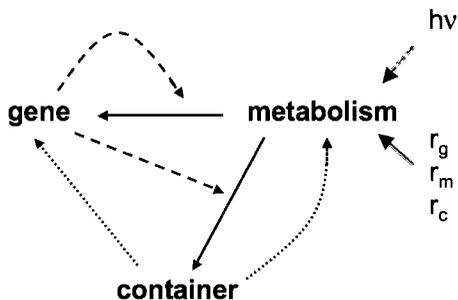
Figure 2. Functional coupling between container, metabolism, and genome. Note how the gene catalyzes (dashed arrows) the metabolic production (solid arrows) of both gene and container building blocks. The container ensures high local concentrations (proximity) and facilitates thermodynamic reaction conditions (dotted arrows) of both the metabolic molecules and the amphiphilic replicator polymers. The free energy is provided by light ($h\nu$), and the provided resources are precursor lipids r_c, precursor gene oligomers r_g, and sensitizers r_m.

expanded capabilities [8, 15, 20], and has been applied to a number of biological systems such as membranes [14, 31], vesicles [33, 34], and micelles [13, 35]. Also chemical reactions have been in-corporated into the DPD method [5, 6]. In the context of protocells, DPD has recently been applied to study a self-replicating micellar system [10]. The DPD formalism used in this work is the revised version from Groot and Warren [15] that has become the de facto standard of DPD.

In general, DPD is a coarse-grained particle method in which a single particle does not represent an individual atom, but instead a group of atoms in a large molecule or several small molecules like water. This grouping allows one to average over fast degrees of freedom [11] and treat them as noise and friction. While this is the case for many particle-based methods (coarse-grained molecular dynamics, Brownian dynamics, etc.), the unique feature of DPD is a thermostat that conserves local momentum, and therefore the hydrodynamics of the system.

2.1 Dissipative Particle Dynamics

A DPD simulation consists of a set of N particles located in three-dimensional continuous space with Euclidean metrics. Each particle i has a position $\mathbf{r}_i$, mass m_i, and momentum $\mathbf{q}_i$, from which one can derive its velocity $\mathbf{v}_i = \mathbf{q}_i/m_i$. Its motion is determined by a force field $\mathbf{F}_i$ through Newton's second law of motion:

$$\frac{d^2\mathbf{r}_i}{dt^2}(t) = \frac{1}{m_i}\mathbf{F}_i(\mathbf{r}_i(t)). \tag{1}$$

The force acting on particle i can be decomposed into pairwise interactions, which are each the sum of three different components—a conservative, a dissipative, and a random one:

$$\mathbf{F}_i = \sum_{j \neq i}\mathbf{F}_{ij} = \sum_{j \neq i}\left(\mathbf{F}_{ij}^C + \mathbf{F}_{ij}^D + \mathbf{F}_{ij}^R\right), \tag{2}$$

where $\mathbf{F}^C$, $\mathbf{F}^D$, and $\mathbf{F}^R$ are defined by

$$\mathbf{F}_{ij}^C = -\nabla\phi_{ij}, \tag{3}$$

$$\mathbf{F}_{ij}^D = -\eta\omega^D(r_{ij})(\mathbf{n}_{ij}\cdot\mathbf{v}_{ij})\mathbf{n}_{ij}, \tag{4}$$

$$\mathbf{F}_{ij}^{R} = \sigma\omega^{R}(r_{ij})\xi_{ij}\mathbf{n}_{ij}. \tag{5}$$

For each particle pair (i, j), $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$ is the relative position, $r_{ij} = |\mathbf{r}_{ij}|$ the center-to-center distance, and $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$ the relative velocity. We denote with $\mathbf{n}_{ij} = \mathbf{r}_{ij}/r_{ij}$ the (unit) direction between the two particles. A detailed discussion of the different forces $\mathbf{F}_{ij}^{X}$ now follows.

The conservative force $\mathbf{F}_{ij}^{C}$ is expressed in the usual way as the negative gradient of a potential $\phi_{ij} = V_{ij} = V(r_{ij})$. In most DPD simulations, a pure repulsive soft core potential of the form

$$V_{ij}(r) = \begin{cases} \frac{a_{ij}}{2}(r - r_c)^2 & \text{if } r < r_c, \\ \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

is used for all particle interactions. a_{ij} and r_c are constants that define the strength and range of the particle interaction. The magnitude of the resulting force decreases linearly from $|\mathbf{F}_{ij}^{C}(0)| = a_{ij}$ to $|\mathbf{F}_{ij}^{C}(r_c)| = 0$. The a_{ij}'s depend on the type of interacting particles—and are therefore the appropriate place to parameterize the model. In addition, different particle pairs could be given different values of r_c if one wanted to effectively give particles different radii. However, in the current work we choose $r_c = 1$ for all bead interactions, which is the standard in almost all DPD simulations.

For the study of information polymers and amphiphiles, individual DPD beads can be covalently bonded. A bond between bead i and bead j is formalized by an additional harmonic potential

$$V_{ij}^{s}(r) = \begin{cases} \frac{b}{2}(r - r_b)^2 & \text{if } (i, j) \text{ are bonded}, \\ \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

with bond strength b and range r_b. In addition to that, we introduce a bending potential to stiffen longer polymer strands: In a chain $i - j - k$ of interconnected polymer beads, the angle θ_j formed by the two bonds of the central bead j induces an additional harmonic potential

$$V_{ijk}^{\theta}(\theta_j) = 1/2\, c_{ijk}(\theta_j - \theta_{\text{eq}})^2, \tag{8}$$

where θ_{eq} is the equilibrium angle and c_{ijk} denotes the strength of the bending potential.

The dissipative force $\mathbf{F}_{ij}^{D}$ is a function of the relative velocity of the two particles. It models the viscous damping of the fluid. The friction coefficient η in Equation 4 scales the strength of this force, and ω^{D} is a distance weighting function not determined by the general formalism.

The random force $\mathbf{F}_{ij}^{R}$ accounts for thermal effects. It is scaled by a strength parameter σ and a second weighting function ω^{R}. In Equation 5, ξ_{ij} is a Gaussian distributed random variable with $\langle \xi_{ij}(t) \rangle = 0$, $\langle \xi_{ij}(t)\xi_{kl}(t') \rangle = (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk})\delta(t - t')$, and $\xi_{ij} = \xi_{ji}$.

In order to reproduce the right thermodynamic behavior, the DPD formalism must satisfy the fluctuation-dissipation theorem. As a consequence, the equilibrium state will obey Maxwell-Boltzmann statistics and therefore allow the derivation of thermodynamic properties. As shown by Español and Warren [8], DPD satisfies the fluctuation dissipation theorem if and only if the weighting functions ω^{D} and ω^{R} obey the relation

$$\omega^{D} = (\omega^{R})^2. \tag{9}$$

In agreement with the DPD standard, we set

$$\omega^D(r) = \left[\omega^R(r)\right]^2 = \left[2\left(1 - \frac{r}{r_c}\right)\right]^2.$$ (10)

If Equation 9 is fulfilled, $\mathbf{F}_{ij}^D + \mathbf{F}_{ij}^R$ acts like a thermostat to regulate the temperature of the system, and the equilibrium temperature k_bT is given by

$$k_bT = \frac{\sigma^2}{2\eta},$$ (11)

where k_b denotes the Boltzmann constant. In molecular dynamics simulations, a variety of thermostats have been explored, but only the DPD thermostat is guaranteed to conserve linear and angular momenta of the particles and thus flow properties of the fluid (because all involved forces are central: $\mathbf{F}_{ij} = -\mathbf{F}_{ji}$). It is therefore the only thermostat that allows the study of transport processes [30]. However, it has to be pointed out that mass transport in the original DPD method is too fast compared to transport of momentum. Namely, the Schmidt number (i.e., the ratio of the kinematic viscosity to the mass diffusivity) of a DPD fluid is 1000 times lower than the one of actual water [15]. Thus, it is unclear how well diffusive processes are actually captured by the model.

In agreement with the DPD standard, we use r_c and k_bT as our units of length and energy. All particles have unit mass $m_i = 1$. From Equation 1 we can derive the unit of time as $\tau = r_c\sqrt{m/k_bT}$. We will give an estimate of the order of magnitude of the physical length in Section 3.

2.2 Incorporation of Chemical Reactions

A rigorous way to extend the DPD formalism to allow for chemical reactions would need to couple the free energy change of a reaction to the kinetic and potential energy reservoirs in the model in such a way that reactions result in local heating or cooling. This approach is highly nontrivial, and further method development in this direction is certainly needed. However, it is also not clear that this is needed in order to achieve the qualitative results that we are seeking. Thus, we employ a simple stochastic process on the basis of reaction rates that is executed between every two DPD updates. This algorithm was used in earlier Brownian dynamics simulations [25]. Our choice can be legitimated by the observation that chemical reactions are relatively rare events in our simulations and thus their impacts are rapidly equilibrated by the DPD thermostat.

Chemical reactions in our system occur between two reactants and fall into two different classes:

- transformation: $\mathbf{U} \rightarrow \mathbf{V}$;

- polymerization (of different bases or oligomers): $\mathbf{X} + \mathbf{Y} \rightarrow \mathbf{XY}$.

Each reaction has a given rate k_s for spontaneous occurrence.

The spontaneous reaction rate can be enhanced by the presence of nearby catalysts. The catalytic effect decreases linearly with increasing distance to the reactant up to a cutoff distance r_{cat}, beyond which it is zero. For simplicity, the effect of several catalysts is modeled as a superposition. Thus, the overall reaction rate is given as

$$k = k_s + \sum_{\mathbf{C}} f_{cat}(r_{\mathbf{C}})$$ (12)

with

$$f_{\mathrm{cat}} = \begin{cases} k_{\mathrm{cat}}\left(1 - \frac{r_{\mathbf{C}}}{r_{\mathrm{cat}}}\right) & \text{if } r_{\mathbf{C}} < r_{\mathrm{cat}}, \\ \\ 0 & \text{else.} \end{cases} \tag{13}$$

In these equations, the sum runs over all catalyst beads, with $r_{\mathbf{C}}$ denoting the distance to the first reactant, r_{cat} the maximal catalytic range, and k_{cat} the catalytic rate. Polymerization has the further restriction that the distance between the reactants must be less than a maximal reaction range R. To deduce probabilities from the reaction rates, we used an agent-based-like algorithm that is given in the Appendix.

If a reaction occurs, we change the particle types of the reactants from $\mathbf{X}$ to $\mathbf{Y}$ and/or establish or remove a bond between the reactants, depending on the type of reaction. Particle positions and momenta are conserved.

We also introduced particle exchange into the model to mimic the supply of chemicals into the system, which drive it out of its equilibrium. Our model allows us to define regions in which particles of a certain class can be exchanged with a given probability to reestablish a nonequilibrium state so that the system can continue to evolve. Note that total particle number is kept constant. Likewise, in chemical reactions we conserve positions and momenta when exchanging particles.

2.3 Components of the Minimal Protocell Model

We model the protocell with the following components: water, surfactant precursor, surfactant, sensitizer, information templates, and information oligomers and their precursors. Water ($\mathbf{W}$) and sensitizer ($\mathbf{Z}$) are single DPD particles. Surfactants are modeled as amphiphilic dimers: one hydrophilic head ($\mathbf{H}$) and one hydrophobic tail particle ($\mathbf{T}$) connected by a covalent bond. Precursor surfactants are dimers of two hydrophobic particles ($\mathbf{T}-\mathbf{T}$). Interaction parameters (as multiples of $k_b T$) for the water and amphiphiles have been taken from [13] (where surfactants are modeled as dimers as well):

a_{ij}	$\mathbf{W}$	$\mathbf{H}$	$\mathbf{T}$
$\mathbf{W}$	25	15	80
$\mathbf{H}$	15	35	80
$\mathbf{T}$	80	80	15

The bond parameters are $b = 150 k_b T$ and $r_b = 0.5 r_c$. These parameter values were originally used to analyze polymer-surfactant interactions. Later, the phase diagram for varying surfactant concentrations was analyzed [35].

In order to keep the number of different parameters as low as possible, we express further interactions with the same parameters as the ones above: Sensitizer beads are hydrophobic. Thus, their interaction parameters are equal to those for surfactant tails: $a_{\mathbf{Z}j} = a_{\mathbf{T}j}$.

2.3.1 Genes

The gene is modeled as a strand of covalently bound monomers ($\mathbf{A}$ and $\mathbf{B}$) with hydrophobic anchors ($\mathbf{T}$) attached to it. We assume the gene is similar to a peptide nucleic acid (PNA) decorated with lipophilic side chains to the backbone. The reason why we are utilizing PNA and not DNA or

RNA is that we want to have a non-charged backbone for the gene molecule to enhance its lipophilic properties. For details, see [28]. We note that the use of PNA decorated with lipophilic side chains in conjunction with an amphiphilic surface layer will cause the genetic material to have a behavior that is quite different from that of DNA or RNA in water. In particular, it is not at all clear that the two complementary macromolecules locally will lie in a common plane when hybridized with each other. Thus we investigated a number of possible different orientations.

By numbering the monomers within each strand, we introduce an orientation of the molecule that mimics the orientation of the actual peptide bond given by its *C*- and *N-termini*. This allows us to define the following vectors for each gene monomer bead: $\mathbf{u}_i$ is a unit vector pointing from the previous monomer towards the current one. For the first monomer in the strand, $\mathbf{u}_i = 0$. Likewise, $\mathbf{v}_i$ is a unit vector pointing towards the next monomer in the strand (or 0 for the last monomer). $\mathbf{z}_i$ is a unit vector pointing from the actual monomer towards its anchor bead. To obtain the association of PNA to the micellar surface, the molecule is modeled as interconnected amphiphiles. For the hydrophobic anchors, we use the same bead type $\mathbf{T}$ as used for the surfactants and precursors, while nucleotide beads share the interaction parameters of the hydrophiles: $a_{\mathbf{A}j} = a_{\mathbf{B}j} = a_{\mathbf{H}j}$.

We need to introduce additional interactions that describe the affinity of complementary gene monomers. Due to the rather complex combination of hydrogen bond formation, cooperative stacking, and π stacking between real gene monomers, we cannot expect the complementary monomer bead forces to be as simple as the bead-bead interactions introduced earlier. We now implement and test several alternative representations of such base affinities, as discussed below.

a. *Undirected attraction:* The obvious extension of $\mathbf{F}_{ij}^C$ to include attractive interactions is a combination of attractive and repulsive components. Thus, in the first representation, we replace $\mathbf{F}_{\mathbf{AB}}^C(\mathbf{r})$ by the stepwise linear function

$$\mathbf{F}_{\mathbf{AB}}^{C_1}(\mathbf{r}) = \mathbf{F}_{\mathbf{AB}}^C(\mathbf{r}) + \begin{cases} a_2(r_{c_2} - r)\mathbf{n} & \text{if } r < r_{c_2}, \\ \\ 0 & \text{else} \end{cases} \tag{14}$$

with $r_{c_2} > r_c$ and $a_2 < 0$. Different attraction strengths a_2 will be used and compared in later computer simulations (Section 3.4.1). To compensate strong attractions for small values of r, we will vary the repulsion strength $a_1 = a_{\mathbf{AB}}$ accordingly. Note that another generalization of $\mathbf{F}_{\mathbf{AB}}^{C_1}$ compared to $\mathbf{F}_{\mathbf{AB}}^C$ is the change in the interaction range, which, in addition to the standard r_c dependence, now also depends on the actual pair $(\mathbf{A}, \mathbf{B})$ through r_{c_2}.

b. *Directed perpendicular attraction:* In the real gene system, hybridization is partly due to the formation of H bonds between the complementary nucleotides. H bonds share features with covalent bonds, which are better characterized by directed than by undirected interactions. Hence, in the second representation, we introduce directed attractions parallel to the $\mathbf{A}-\mathbf{T}$ and $\mathbf{B}-\mathbf{T}$ axes. Here, we replace $\mathbf{F}_{\mathbf{AB}}^C$ by

$$\mathbf{F}_{\mathbf{AB}}^{C_2}(\mathbf{r}) = \mathbf{F}_{\mathbf{AB}}^C(\mathbf{r}) + \begin{cases} a_2(r_{c_2} - r)(\mathbf{z} \cdot \mathbf{r})\mathbf{n} & \text{if } r < r_{c_2}, \\ \\ 0 & \text{else} \end{cases} \tag{15}$$

with the above definitions for $\mathbf{r}$, $\mathbf{z}$, and $\mathbf{n}$. Again, different attraction coefficients a_2 will be compared in the later simulations. The value $a_1 = a_{\mathbf{AB}}$, on the other hand, can be held fixed, because the

attraction vanishes when r approaches 0. We set $a_1 = 35k_bT = a_{AA} = a_{BB}$. We call this interaction *perpendicular*, because the strongest attraction will be perpendicular to the surface of the micelle, once the PNA strand is attached to the surface of the micelle (Figure 3a).

c. *Directed tangential attraction:* The third representation is similar to the second, except that attraction is now perpendicular to the backbone and to the **AT** (or **BT**) axis. The force is attractive towards one side of the PNA and repulsive towards the other—hence, this is the only implementation that catches the directionality of the molecule:

$$\mathbf{F}_{AB}^{C_3}(\mathbf{r}) = \mathbf{F}_{AB}^{C}(\mathbf{r}) + \begin{cases} a_2(r_{c_2} - r)\left(\frac{(u+\mathbf{v})\times\mathbf{z}}{|(u+\mathbf{v})\times\mathbf{z}|} \cdot \mathbf{r}\right)\mathbf{n} & \text{if } r < r_{c_2}, \\ 0 & \text{else.} \end{cases} \tag{16}$$

This force is expected to be strongest tangential to the surface of the micelle. As in the last case, we will vary a_2, but keep a_1 fixed at a value of $35k_bT$ (Figure 3b).

Covalent bonds within PNA strands have a bond strength of $b = 150k_bT$ with an ideal bond length $r_b = 0.5r_c$ for bonds between nucleotides and anchors, and $r_b = 0.75r_c$ for bonds between the nucleotides themselves. In addition, we introduce stiffness (Equation 8) within the PNA strand: Angles of interconnected strands prefer to be stretched out ($\theta_0 = 180°$, $c_{ijk} = 10k_bT$). With the stiffness we model folding restrictions of the peptide bond, as well as π-π electron stacking of nearby nucleotides. This affects only the PNA chain, not the bonded hydrophobic anchors, as they do not experience any bending potential. Table 1 summarizes the chosen set of parameters.

2.3.2 Reactions

For the above-listed components we introduce the following chemical reactions.

First, we define a reaction that transforms the oil-like precursor surfactants into actual surfactants. In the real chemical implementation of the protocell, the precursors are fatty acid esters. The ester bond of the precursor surfactant breaks, thereby producing a fatty acid—the surfactant—and

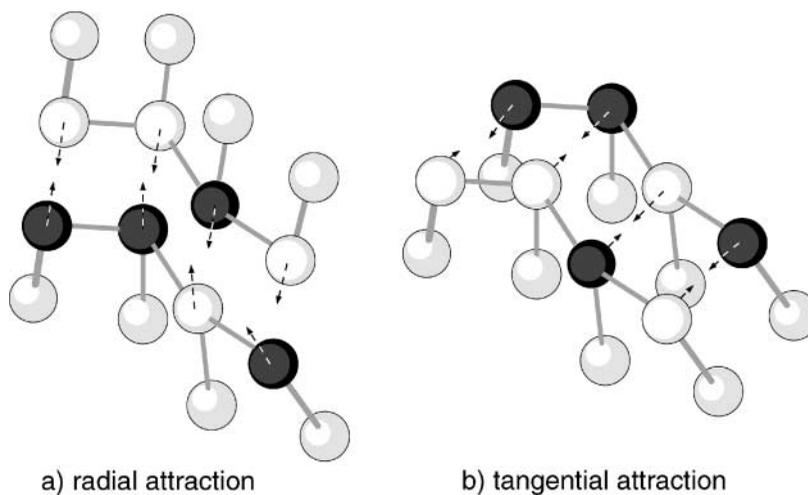a) radial attraction   b) tangential attraction

Figure 3. Hybridization complexes for (a) perpendicular and (b) tangential attraction between complementary bases. Bases are shown as black and white beads, hydrophobic anchors in light gray. Arrows denote the direction of strongest attraction.

Table I. Interaction strength a_{ij} (as multiples of k_bT) for all bead types defined in the model. The force (*) between complementary nucleotides **A** and **B** has attractive parts and cannot be expressed by a single interaction parameter a_{AB}. Three different force fields have been considered for such interactions. See the text for details.

a_{ij}	**W**	**H**	**T**	**A**	**B**	**Z**
W	25	15	80	15	15	80
H	15	35	80	35	35	80
T	80	80	15	80	80	15
A	15	35	80	35	(*)	80
B	15	35	80	(*)	35	80
Z	80	80	15	80	80	15

some small aromatic molecule, which is considered waste. Disregarding the production of the waste, we model this reaction by the scheme

$$\mathbf{TT} + \mathbf{Z} \rightarrow \mathbf{HT} + \mathbf{Z}, \tag{17}$$

which reflects that both parts of the ester are hydrophobic, while the resulting surfactant is an amphiphile. For simplicity, the spontaneous reaction rate is set to $0\tau^{-1}$. The sensitizer acts as a catalyst with a catalytic radius of $1.0r_c$. In our simulation, the catalytic rate of the sensitizer can be turned on ($k_{cat} = 1.0\tau^{-1}$) and off ($0\tau^{-1}$) interactively by a switch. This mimics the photoactivity of the sensitizer.

Second, we introduce reactions to form covalent bonds between the terminal monomers of pairs of oligomers:

$$\mathbf{A} + \mathbf{B} \rightarrow \mathbf{AB},$$

$$\mathbf{A} + \mathbf{A} \rightarrow \mathbf{AA}, \tag{18}$$

$$\mathbf{B} + \mathbf{B} \rightarrow \mathbf{BB}.$$

These syntheses are only applied to the terminal monomers in the PNA strands and involve no catalysts. The maximal range is $0.75r_c$; the maximal reaction rate is $k_{max} = 0.1\tau^{-1}$. The actual reaction rate between monomers i and j further depends on the orientation of the ligating strands: We set

$$k_{ij} = \frac{1}{2} k_{max} \left(\frac{\mathbf{u}_i + \mathbf{v}_i}{2} \cdot \frac{\mathbf{u}_j + \mathbf{v}_j}{2} + 1 \right). \tag{19}$$

This formulation also prevents covalent bonds between complementary strands (which are anti-parallel, and thus have an effective k close to zero).

3   Results

We use the model discussed above to study various aspects of the life cycle of the Los Alamos bug as depicted in Figure 1. In particular, our simulations address the spontaneous self-assembly of proto-

cells (Figure 1a, b), the incorporation of resources (b, c), the metabolic growth of the protocell (d, e), template reproduction, and finally fission into two daughter cells (e, f). We will further analyze some of the catalytic coupling processes explained in the introduction.

All simulations are performed in three-dimensional space with periodic boundaries. We set σ to 3 and η to 4.5, which leads to an equilibrium temperature of $1k_bT$. A total bead density $\rho = 3.0r_c^{-3}$ is used for all simulations. System size and number of iterations are noted for each individual simulation run. We integrate Equation 1 numerically with the DPD variant of the leapfrog Verlet integrator discussed in [15], with $\lambda = 0.5$ and a numerical step width of $\Delta t = 0.04\tau$.

3.1 Self-assembly of Micelles

We initialize a cubic box of size $(12.5r_c)^3$ randomly with 2.9 water beads and 0.05 surfactant dimers per unit volume, or 5664 water beads and 98 dimers in the box. Simulations are performed for $0\tau < t < 1000\tau$ with the interaction parameters summarized in Table 1 and the model parameters given in the introduction to this section. We observe the formation of spherical micelles with aggregation numbers up to about 20, with a peak around 12. This is shown in Figure 4, where once the system had reached an equilibrium state, we followed its behavior. For each time step we recorded the number of aggregates of a particular aggregation number and hence the total number of surfactants in the aggregates of that size. The average of this result over the number of time steps was than histogrammed. We also observe a continuous exchange of surfactants with the bulk phase. As a result of these associations and dissociations, we find a number of free monomers and submicellar aggregates in the bulk phase. These observations qualitatively fit theoretical and experimental results (see, e.g., [9]).

Although we do not intend to model specific chemicals, we can roughly estimate the order of magnitude for the physical length scale of our simulation, using a procedure proposed by Groot and Rabone [14]. Our calculation is based on sodium alkane sulfates, as these are well-studied surfactants with properties similar to the fatty acids used in the real chemical implementation. Table 2 lists the critical micelle concentration (CMC), that is, the minimal concentration at which micelles spontaneously form. The table also gives the mean aggregation number and the volume of these molecules.
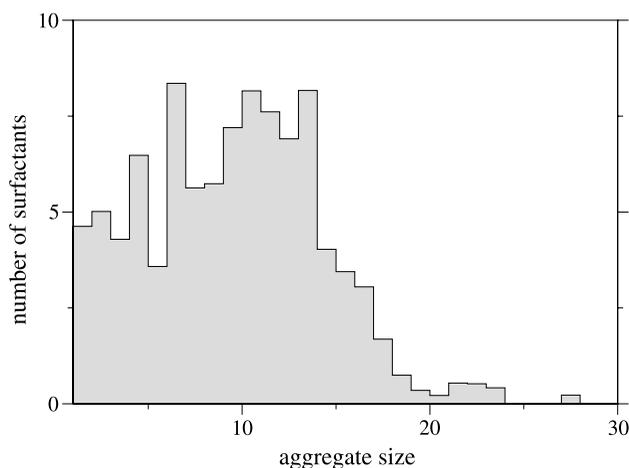

Figure 4. Micellar size distribution for a system containing 2.9 water beads and 0.05 surfactant dimers per unit cube. To obtain the aggregate size histograms from a system state, every two surfactants whose **T** beads are separated by less than r_c are considered to belong to the same aggregate. 20,000 states of an equilibrated system ($200\tau < t < 1000\tau$) are averaged in the distribution shown.

Table 2. Data for sodium alkane sulfate surfactants with varying tail length. For each surfactant, CMC and mean aggregation number are listed from [3]. The molecular volume is estimated from the number n of carbon atoms using the formula $V = 27(n + 1)$ Å³ [9] plus a constant 88.51 Å³ for the sulfate group [whose value is derived from the molecular mass (98.08 g/mol) and density (1.84 g/cm³) of sulfuric acid]. The coarse-graining parameter N_m, the physical length scale r_c, and the total surfactant concentration are the interpretations of model parameters in the case that the model dimer represents the respective surfactant. Finally, the fraction of a micellized surfactant is the prediction of the closed association model for the surfactant and the calculated concentration [9].

Surfactant	CMC (mol/l)	Aggregation number	Surfactant vol. (Å³)	N_m	r_c (Å)	Surfactant conc. (mol/l)	Predicted micellization ratio
$NaC_6H_{13}SO_4$	0.42	17 ± 6	278	4.625	7.467	0.201	1×10^{-5}
$NaC_7H_{15}SO_4$	0.22	22 ± 10	305	5.075	7.701	0.183	2.5×10^{-3}
$NaC_8H_{17}SO_4$	0.13	27	332	5.525	7.923	0.168	0.2
$NaC_9H_{19}SO_4$	6.0×10^{-2}	33	359	5.975	8.132	0.156	0.6
$NaC_{11}H_{23}SO_4$	1.6×10^{-2}	52	413	6.875	8.521	0.135	0.935
$NaC_{12}H_{25}SO_4$	8.2×10^{-3}	64 ± 13	440	7.325	8.703	0.127	1
$NaC_{14}H_{29}SO_4$	2.1×10^{-3}	80 ± 16.5	494	8.225	9.046	0.113	1

Under the simplifying assumption that all DPD beads have equal effective volume, we can derive the molecular volume of a single DPD bead, and—knowing the molecular volume of water $(V_{H_2O} = 30$ Å³$)$—we get the so-called *coarse-graining parameter*

$$N_m = \frac{\frac{1}{2} V_{surf}}{V_{H_2O}}, \tag{20}$$

which tells us how many water molecules are represented by a single DPD bead. The average number of DPD water beads per unit cube is ρ, each one of them representing N_m molecules. Therefore, the physical length scale r_c resolves to

$$r_c \equiv \left(\rho N_m V_{H_2O} \right)^{1/3}. \tag{21}$$

We will work with solutions that are quite dilute and hence dominated by water. Noting that a liter of water has $1000/18 = 55.56$ moles of water in it, while a volume of r_c^3 has ρN_m molecules of water in it, we find that a concentration of 1 particle/r_c^3 yields a unit of concentration as

$$1 r_c^{-3} \equiv 55.56 \text{ mol}/\rho N_m. \tag{22}$$

With these estimations, we find that the lipid concentration in the above simulation represents between 0.11 and 0.20 mol/l. It is somewhat arguable how to estimate the concentration of free lipids in the bulk phase, because our simulations do not yield a sharp distinction between free lipids—that is, submicellar aggregates—and proper micelles. Assuming that the most reasonable choice for such a distinction is the first minimum in the micellar size distribution at aggregates of size 5 or less, from Figure 4 we get an average of 22.9 free surfactants in the bulk phase out of 98 lipids in

the total volume, that is, 76.6% of the surfactant is micellized, and the free lipid concentration lies between 0.03 and 0.05 mol/l. Knowing the physical surfactant concentration, we can compare this finding with the prediction of the closed association model [9]. According to this model, surfactants are either in bulk phase (S) or in micelles of aggregation number N (S_N). With the pseudo-chemical reaction $NS \rightleftharpoons S_N$ and the condition that $d[S]/d[S]_{total}|_{CMC} = dN[S_N]/d[S]_{total}|_{CMC} = 0.5$, one can calculate the fraction of micellized surfactant for any total surfactant concentration $[S]_{total} = [S] + N[S_N]$. The corresponding ratio $N[S_N]/[S]_{total}$ is also given in Table 2.

We find that our model best matches the aggregation numbers of short chain surfactants ($NaC_6H_{13}SO_4$), while our micellization ratios more closely match the predictions for the somewhat longer chains ($NaC_9H_{19}SO_4$). Although our model representation of surfactants as dimers is rather simplistic, we find a reasonable match (at least in the order of magnitude) between experiment, simulation, and theory. It should be noticed that the micellization parameters for fatty acids, which are the container surfactants of choice in the Los Alamos bug, are qualitatively similar to the listed sodium alkane sulfate surfactant parameters, which are the best-studied surfactants in the scientific community. Given the easy availability of relevant parameters for alkane sulfate surfactant parameters and the level of coarse graining in our DPD model, we can safely use these experimental data to calibrate our simulation. It is conceivable that closer matches might be found by changing interaction parameters or the representation of surfactants. We have however decided to stick to the standard parameter set in order to get comparable results to earlier DPD simulations [10, 13, 35].

Next, we analyzed a ternary mixture of water, surfactant, and oil. In the system described above, we exchanged an additional 0.1 water bead per unit volume by 0.05 hydrophobic oil dimers ($\mathbf{T}$–$\mathbf{T}$), which represent the lipid precursors of the Los Alamos bug. Starting from a random initial condition, the system forms loaded micelles: The precursors aggregate into a core in the interior of the individual micelles because of their high degree of hydrophobicity. This core is coated by surfactants, which shield it from water. We observe a stabilizing effect from the hydrophobic core: The rate of monomer dissociation from the aggregates decreases by a factor of 4 to 5. Dissociation of oil dimers does not happen during the simulations. Over the simulated time span ($0\tau < t < 1000\tau$), these loaded micelles constantly fused to form bigger aggregates. At $t = 250\tau$, the system is composed of five micelles with aggregation numbers 12, 13, 16, 24, and 32, where the aggregation number just counts the surfactants in an aggregate and not any of the precursors or other components. At $t = 500\tau$ we find four micelles (with sizes 16, 24, 25, 32) and finally, for $t = 1000\tau$, the system consisted of only two micelles with aggregation numbers 43 and 53. It remains unclear whether this was the equilibrium solution, or whether the two micelles would finally fuse to form a single aggregate. It is known that any given mixture of surfactants and oil in water results in some equilibrium aggregate structure, some such structures being useful and some less useful as protocellular container substrates; see, for example, the recent summary discussion in [23].

In general, the addition of hydrophobic precursors allows aggregates to grow far beyond their micellar aggregation number, while at the same time, monomer dissociations from the assembly fall by a factor of 4 or more. This is consistent with simulation results from earlier studies of a similar surfactant-precursor-water system [10]. However, a more systematic DPD investigation is necessary to address the relation of dynamics, stability, and size distribution in this context.

3.2 Self-assembly of the Protocell

In this subsection, we study the self-assembly of protocells. We initialize a cubic box of size $(7.5r_c)^3$ with 1212 water particles, 21 surfactant dimers, four sensitizer particles, and one PNA strand that is four nucleotides in length. All other simulation parameters are as before. Using these numbers, we achieve the same overall particle density and the same surfactant concentration as in the previous section.

Starting from an arbitrary initial condition, we observe the spontaneous formation of a protocell, that is, a micelle that is loaded with sensitizer, that has PNA attached to its surface, and whose

nucleotides are exposed to the aqueous phase (see Figure 5). Aggregation happens within a remarkably short period: After only 10 time units, we already find complete protocells. The lipid aggregation number of this micelle is around 14, with few associations and dissociations of monomers. The slight increase in aggregation number along with a decrease of monomer dissociations is most probably due to the stabilizing effect of the additional sensitizers.

3.3   Replication of the Container

The dynamics of a surfactant-precursor-water system similar to the one under consideration has been studied in detail in [10]. Considering precursor and surfactant kinetics, the previously analyzed system differs from the one discussed here in that (i) the catalytic role of sensitizers is performed by the surfactants themselves, and (ii) the metabolic turnover is not regulated by turning the light on and off, but instead only follows chemical mass kinetics. Using simulations based on classical lattice gas methods, Coveney et al. [7] in 1996 reproduced the micellar self-replication experiments of Bachmann et al. [4]. In 1998 and 2000 Mayer and Rasmussen developed an extended lattice polymer approach [21, 22] for explicitly including polymers and chemical reactions similar to the current DPD approach, and they were also able to reproduce the experimental findings by Luisi's group [4]. The purpose of this section is to show that the reported dynamics also hold for the metabolic reaction scheme of the Los Alamos bug.

A system of size $(10r_c)^3$ is initialized with a micelle consisting of 15 surfactants and loaded with four sensitizer beads in its interior. Model parameters are given in the beginning of this section. In a single spherical region of radius $2r_c$ located away from the micelle, pairs of water particles are replaced by surfactant dimer precursors with an overall exchange rate of $\approx 2.5 \times 10^{-3}$ precursors per time unit.

Because of their hydrophobic nature, the precursor molecules tend to agglomerate into oil-like droplets. The diffusion of such droplets becomes progressively slower the bigger they are. This initiates a positive feedback: The bigger the droplets, the more slowly they diffuse out of the exchange region. The more slowly they diffuse, the more likely they are to accumulate additional precursors before they diffuse out of the exchange volume. By varying the volume of the exchange region and/or the rate of exchange, one can set the mean size of the precursor droplets that are formed. Due to the positive feedback, the effect will not be linear with either the exchange region size or the exchange rate.

Since we do not want the non-continuous exchange events to disturb the system's dynamics too much, we restrict particle exchange to a region of $2.0r_c$ (3% of the total system volume). By varying the exchange rate used to introduce precursors, we find that 5.0×10^{-5} is close to the optimum for which droplets of precursor molecules are provided at a reasonable rate, yet are still small enough to diffuse at a reasonable speed. With these values, the precursor droplets consisted of 5 dimers on average. Once in the vicinity of a micelle, the droplets are immediately absorbed.

(a)   (b)   (c)
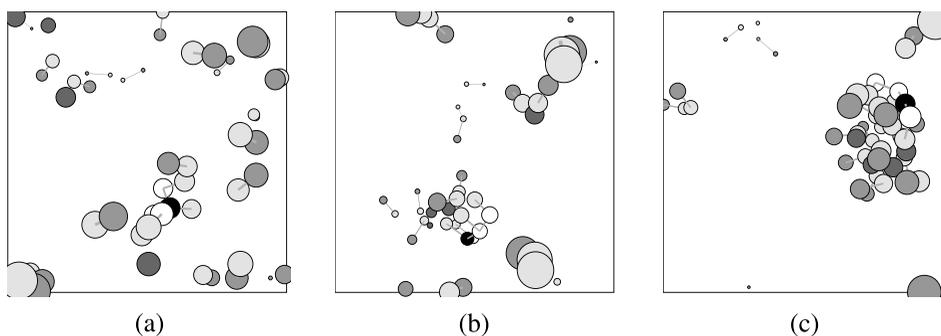
Figure 5. Self-assembly of the protocell from a random initial condition. The diagrams show the state of the system at times (a) $t = 0\tau$, (b) $t = 4\tau$, and (c) $t = 10\tau$. Surfactants are shown in medium gray (head bead) and light gray (tail), the sensitizers in dark gray, the PNA backbone in light gray, and the PNA monomers in black and white.

When the micelle absorbs 15 precursor molecules into its interior, we stop supplying additional precursors and trigger the catalytic activity of the sensitizer by turning on the light. During the metabolic turnover, the micelle grows in amphiphile number, while losing few, if any, amphiphiles due to the stabilizing effect of the remaining precursors as was discussed previously. It responds to the changing surfactant-to-precursor ratio by changing its shape from spherical to rodlike. The elongation continues until nearly all the precursors are metabolized. At some moment, the elongated aggregate becomes unstable and divides into two daughter cells (see Figure 6). With the parameters used, overall precursor turnover and fission takes place in approximately 20 time units (i.e., 500 time steps).

We compared the above findings with simulations of an unregulated system, where the precursor supply and catalytic rate are not triggered, but instead held constant over the whole simulated time span. The objective behind this simulation was to find whether the system might feature inherent self-regulation: As the precursor forms droplets in the bulk phase, their incorporation into the micelle occurs in spurts rather than continuously. If the introduction rate of precursors into the system is locally fast enough to allow larger droplets to form (especially due to the positive feedback effect), a larger number of precursors can simultaneously enter the protocell. Then if the metabolic turnover rate is sufficiently fast, the turnover of the large number of precursors might be sufficient to trigger container division rather than having a slow but continual loss of newly formed amphiphiles.

To investigate this possibility, we performed simulation runs for a system of size $(10r_c)^3$ initialized with a micelle of 15 surfactants and four sensitizer beads. Other model parameters are the same as given in the beginning of this section. Precursors were supplied by the same mechanism and rate as before. We observed the incorporation of droplets between 3 and 9 precursor dimers in size. As the transformation of precursors happened significantly faster than the precursor supply, nearly each droplet was transformed separately. When only a few precursors were absorbed at once (i.e., in a small droplet), the micelle responded by rejecting several surfactants into the bulk phase. Such loose surfactants then formed submicellar aggregates or attached to precursor droplets when present. However, when the incorporated droplet was big enough, the outcome of the metabolic turnover was a proper cell division. A micelle that consisted of 15 surfactants and four sensitizers, for example, split in two after the absorption and turnover of 8 precursors. The fission products were two micelles, one with 14 surfactants and three sensitizers and the other with 9 surfactants and one sensitizer.

This result suggests that the explicit regulation of the metabolic turnover by light bursts might not be necessary to obtain the replication cycle of the container, as a similar regulation can be obtained by a careful regulation of the provided precursor droplet sizes. Light control might, however, still serve as a convenient mechanism to synchronize container and genome replication if they occur on separate time scales.
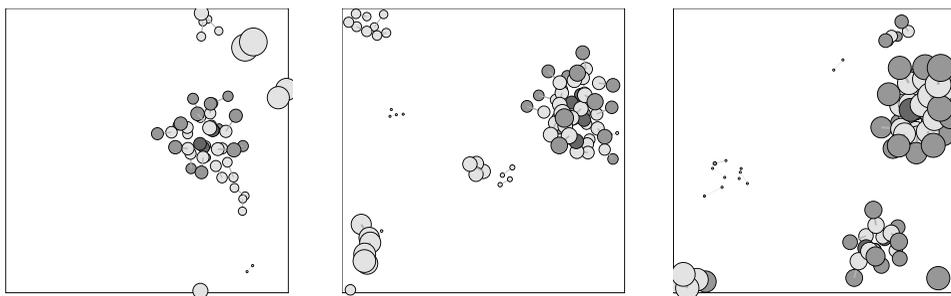

Figure 6. Replication dynamics of the container. Precursors are fed into the system far from the micelle at the (periodically reflected) edge of the system space. They form droplets in the aqueous phase, which are absorbed by the protocells as a whole. Protocells grow by incorporation of precursors. After a critical amount of precursor is transformed into surfactant, the assembly loses its stability and splits into two daughter cells (right frame).

3.4   Replication of the Genome

In our experience, the most difficult component of the protocell to model with DPD methods is the genome and its behavior. Furthermore, the DPD hybridization process seems more ill defined than the ligation process, which is why our discussion of the replication of the genome is divided in two consecutive steps: hybridization and ligation. Recall that hybridization denotes the alignment of short PNA oligomer sections along the template PNA strand and "hydrogen" bonding to it, while ligation—or polymerization—is the reaction that turns aligned oligomers into an actual (complementary) copy of the template.

3.4.1   Hybridization

Replication of the genome essentially depends on the stability of the hybridized complex: It can only occur if the double strands are stable for a time long enough for all the needed oligomers to diffuse to and align with the template. It should be noted that if more than two oligomers are involved, the joining of additional oligomers and their polymerization can occur sequentially, so the unpolymerized templates need not all be simultaneously attached. As will be shown further below, once some polymerization has occurred, that section will be more stable in hybridized form. We studied the stability of the hybridization with the following simulation: A system of size $(5.5r_c)^3$ was initialized with an oil layer that is meant to mimic a two-phase system (single beads of type **T** are confined to lie below a plane above which the water is located). The overall particle density is $\rho = 3r_c^{-3}$, as in the earlier experiments, in order to make the hybridization process as simple as possible. As we shall see later, aggregate surfactant dimers tend to tangle with the gene anchors, which both slows down the hybridization process and makes it less accurate. A four-monomer long PNA template was placed at the oil-water interface with its anchors pointing down toward the oil and its bases pointing up toward the aqueous phase. A pair of 2-nucleotide-long complementary oligomers was placed at a distance of $0.5r_c$ from this strand at a location and orientation for proper hybridization. The location and orientation were varied to match the different hybridization cases studied. In the case of directed perpendicular attraction, this meant that all the beads of the complementary PNA molecules are outside the interface plane, with their hydrophobic anchors pointing away from the hybridization site. In contrast, in the case of tangential attraction, both the template and the oligomers span the interface region, as shown in Figure 7.

In the system modeled, we only had two different types of monomers (**A**, **B**), with **A** and **B** being complementary to each other, but not self-complementary. All possible tetramer templates were used (e.g., **AAAA**, **AAAB**, **ABAB**, **ABBA**), excluding symmetric configurations (e.g., **BBBB**, **BBBA**, **BABA**).
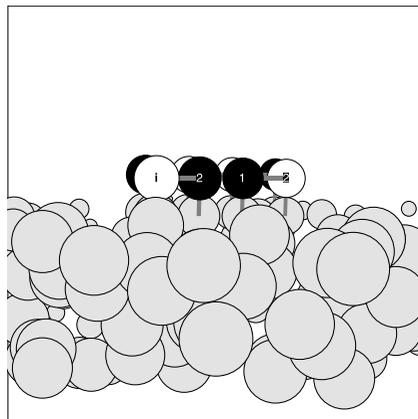

Figure 7. Initial setup of the hybridization simulations. The system is initialized with an oil layer that mimics the oil-water interface of a two-phase system. A tetramer template and two complementary dimers are placed at the interface so that they form a hybridization complex. The association time of such hybridization complexes is measured for different PNA implementations and attraction forces.

For each template only the proper complementary dimer oligomers were used. The different tetramer configurations can differentially hinder the ability of the complementary bases to slide along the template.

During the simulations, the distances between all four complementary base pairs were measured at every time step. When one of these distances exceeded $1.5r_c$ (the maximal interaction range for complementary bases), the PNA strands were considered to be dehybridized. The time it took for the double strands to dehybridize—that is, the association time of the hybridized complex—serves as a measure of the stability of that state. After a maximum of $t = 100\tau$, simulations were truncated and the hybridization was considered to be stable.

For the three different representations of PNA hybridization (see Section 2.3.1, cases a, b, and c), we performed simulations for all possible combinations of four bases excluding symmetrical combinations. Strengths for attractive forces were set with respect to the repulsive force parameter a_{AB} so that complementary bases attracted each other but did not overlap by more than $0.6r_c$. The association times were measured using 10 to 20 runs for each combination. Results are shown in Figure 8.

a. *Undirected attraction:* In the case of undirected attraction, we found mean association times between 2.12τ for $a_1 = 50k_bT$, $a_2 = -10k_bT$, and 7.76τ for $a_1 = 65k_bT$, $a_2 = -20k_bT$. For strong attractions, association times tended to increase with the number of equal (preferably nearby) nucleotides in the template (**AAAA** is the most and **ABBA** is the least stable sequence). However, these differences were rather small.

b. *Directed perpendicular attraction:* For directed perpendicular attraction, the mean association times ranged from 0.45τ for $a_2 = -10k_bT$ to 0.98τ for $a_2 = -30k_bT$ ($a_1 = a_{AB} = 35k_bT$ for all cases) without any significant variation for different sequences. For most simulation runs, it took only a few time steps for the initial complex to dehybridize. The reason for the poor hybridization of the PNA for the perpendicular attraction is quite obvious: Due to the amphiphilic character of PNA, the strands will arrange so that nucleotides point towards water and the anchors towards oil. Thus, the attraction is directed perpendicular to the oil-water interface and into the aqueous phase, where the oligomers do not want to reside. Because of the dot product in Equation 15, the attraction between two PNA molecules on the interface is marginal and the association time is essentially a matter of diffusion.

c. *Directed tangential attraction:* In contrast to the other tested situations, in the case of directed tangential attraction, one can see significant differences in the association time of the initial hybridized complexes, provided the attraction is strong enough: For gene sequences with pairs of equal bases at terminal positions (e.g., **AAAA** and **AABB**), hybridization is usually less stable than for sequences without equal bases at terminal positions (**ABBA** and **ABAB**). The association time of sequences with only one such dimer lies between the values for the above two situations. Examination of the simulations reveal the cause of this trend: A continuous group of two or more equal monomers, one of which is a terminal position of the template, allows the attached dimer to slide along the template strand without a strong penalty in potential energy, and eventually protrude beyond the end of the template. In this misaligned configuration, the dimer can easily distort from the parallel alignment, thereby reducing the overall attraction to the template, until it finally disassociates from the complex. Distinct bases at terminal positions, on the contrary, prevent this sliding along and then off of the template, thereby significantly stabilizing the hybridized state.

For the more promising PNA implementations—undirected and tangential attraction—we further measured the mean distance between complementary bases (hybridization distance) and the distance between those bases in the oligomers that are supposed to polymerize (ligation distance). We performed these measurements using the sequence **AAAA** for the undirected and **ABBA** for the tangential attractions (interaction parameters are given in the caption of Figure 9). Simulations are performed for $0\tau \leq t \leq 1000\tau$. The resulting time series are shown in Figure 9.

In the case of tangential hybridization one finds two alternating domains in the hybridization distance time series: (i) When oligomers are aligned to the template, the mean hybridization distance is
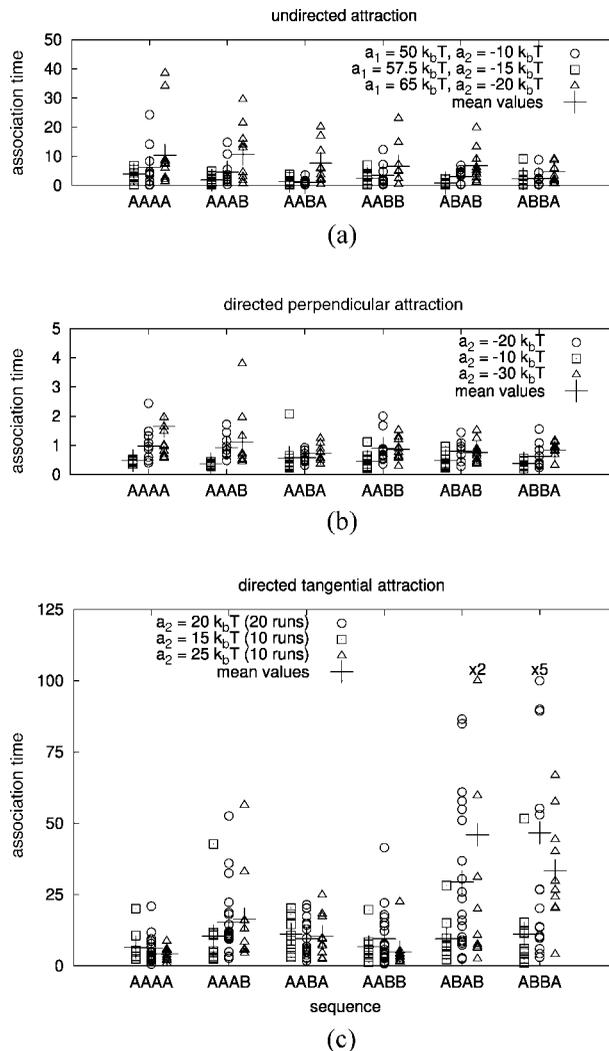
Figure 8. The association time (i.e., the time until the initially hybridized complex becomes dehybridized) for different PNA template sequences of length 4 using (a) undirected, (b) perpendicular, and (c) tangential attraction. For each implementation, three different attraction strengths are compared, as given in the key for each part. a_1 denotes the coefficient of the repulsive part, and a_2 the coefficient of the attractive part, of the interaction force. In the case of directed attraction (b and c), a_1 was set to $35k_bT$ independent of the respective value of a_2. In c, the plotted averages are lower bounds for the actual averages, as simulations were truncated at $t = 100\tau$. If runs were truncated, the multipliers above that run designate how often this was done.

around $1.04r_c$ with only small fluctuations and an average ligation distance of $1.01r_c$ (e.g., $430\tau \leq t \leq 450\tau$ and $700\tau \leq t \leq 780\tau$ in Figure 9). In between such periods, (ii) oligomers dissociate from the template and diffuse over the interface, which is indicated by the large variance in hybridization distance.

Undirected attraction, in contrast, yields hybridization distances around $1.07r_c$ with significant continual fluctuations and a mean ligation distance of $1.158r_c$. One cannot observe the locking of the hybridized state that is apparent for the tangential attraction: Although the oligomers preferably stay in the vicinity of the template, they are not forced into any particular orientation. Investigation of simulation states reveals that oligomers align along different sites of the template or even cross the template strand. Thus, although it appears from a quick look at Figure 9 that the undirected attraction performs better on average, it is only during the locked-in period that the desired reactions
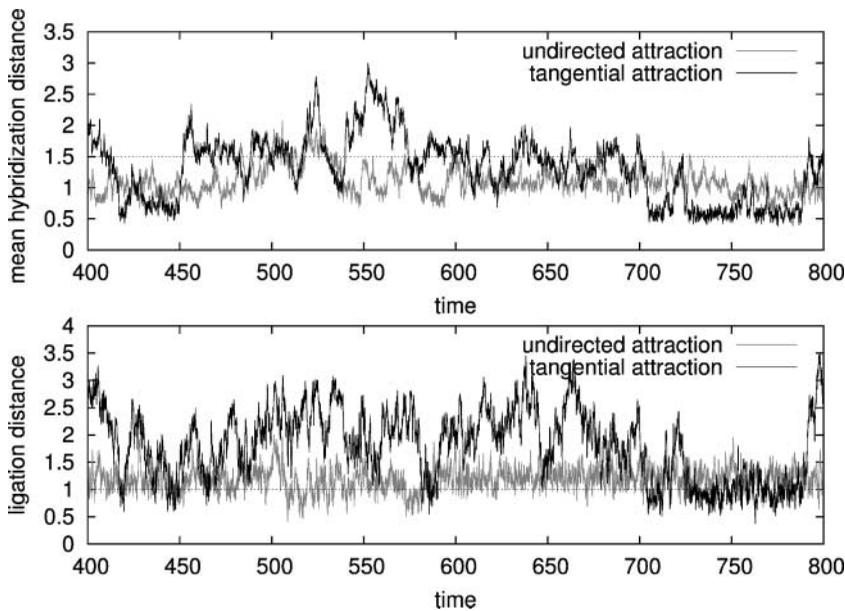
Figure 9. Mean hybridization (upper panel) and ligation distance (lower panel) for the PNA templates (and corresponding oligomers) **AAAA** using undirected attraction with $a_1 = 65k_bT$, $a_2 = -20k_bT$ (gray), and for **ABBA** using tangential attraction with $a_1 = 35k_bT$, $a_2 = 40k_bT$ (black). By hybridization distance, we mean the average distance between complementary nucleotides; by ligation distance, we mean the minimal distance between two terminal nucleotides that are supposed to polymerize. The maximal values of the various distances are limited by the small size of the box.

occur. We can therefore conclude that only the implementation of PNA using tangential attraction is able to generate a proper hybridization and base recognition approximation.

It is assumed that the PNA replication is catalyzed by the oil-water or surfactant-water interface. This is because: (i) lipophilic PNA concentrates at the oil-water interface and thus obtains a much higher local concentration there than in water; (ii) the interface contains a lower water concentration than the bulk phase; (iii) the interface might act directly as a catalyst for amide bond formation; and (iv) the PNA is more spread out (linear) when attached to the interface. To test the geometric part of this hypothesis, we also performed simulations of hybridization in pure water. We randomly initialized a box of size $(5.5r_c)^3$ with water, PNA template (ABBA), and complementary oligomers, using directed tangential forces (the overall bead density was $\rho = 3r_c^{-3}$). Simulations were performed for $0\tau \leq t \leq 1000\tau$. Hybridization and ligation distances are plotted in Figure 10.

The mean hybridization distance in this scenario is $1.41r_c$ (which is close to the maximum radius r_{c_2} at which attraction of complementary nucleotides still exists), with a standard deviation of $0.34r_c$.
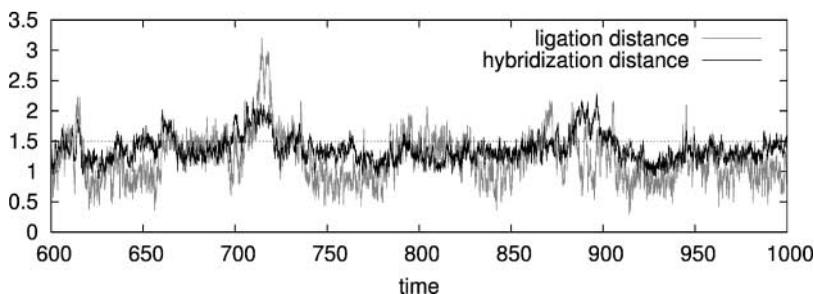

Figure 10. Hybridization and ligation distances of PNA template and complementary oligomers in water. For PNA, tangential directed attraction with $a_2 = 40k_bT$ has been used. The nucleotide sequence is **ABBA**.

Moreover, there is no clear separation between hybridized and dehybridized states. In contrast to the scenario for the oil-water interface, the oligomers never completely dissociate from the template. However, the oligomers are not properly hybridized either. Instead, the template and complementary strands attract each other mainly due to the hydrophobic interactions between the tail beads of these strands rather than due to forces between their bases. Inspection of the simulated states shows that oligomers are seldom aligned parallel to the template. The overall structure has more resemblance to that of a micelle with geometries defined by the amphiphilic properties of the molecules, than to a double strand defined by base affinities. The ligation distance has an average value of $1.12r_c$ with a standard deviation of $0.39r_c$. Unfortunately, this is smaller than in the previous simulations. This might result in ligation rates higher than those on the surface. However, if we decide to make the ligation probability depend on the angle between PNA backbones, the effective ligation rate is smaller than at the oil-water interface.

Last but not least, it is notable that we cannot achieve reliable hybridization without a stiffness potential in the PNA chain. In the absence of such stiffness, complementary bases within one strand tend to bind to each other and form sharp hairpin loops, even for very short strands. This effectively hinders any proper hybridization except for a very few sequences that do not offer any possibility for loop formation (such as **AAAA**).

3.4.2 Ligation

To study the polymerization reaction, a tetramer template strand and two complementary dimers are placed randomly on the surface of a loaded micelle (20 surfactants, 20 precursors) within a system of size $(10r_c)^3$ and total density $\rho = 3.0r_c^{-3}$. As **ABBA** was identified as forming the most stable hybridization complex, we restrict polymerization experiments to this particular sequence, using the PNA representation with tangential directed attraction (see Figure 11).

Of the performed simulations, eight out of ten generated proper template directed ligation, while the remaining two reactions occured spontaneously in the absence of the template strand and constitute the expected background reaction [24]. In our simulations, one of the two spontaneous ligation results was a correct complementary copy of the template strand, while the other was not. Note that in our simulation, polymerization has not been explicitly restricted to happen only between C- and N-terminals, which means that both ends can be concatenated with any other end. When ligation is template directed, six out of eight runs led to correct complementary sequences, while the other two resulted in mispairings of the form **BABA**. In summary, we find that correct replication is about 50% more reliable when directed by the template. If one prohibits the ligation of equal terminal beads (C−C and N−N), the reliability of replication is expected to further increase.

The simulations reveal that it can take a surprisingly long time for the oligomers to form a ligated hybridized complex with the template. Ligation occurs after 90τ in the fastest and after 674τ in the
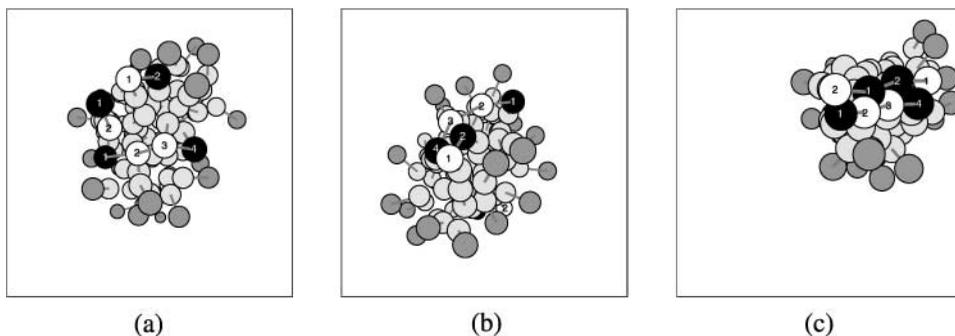

Figure 11. The three steps of template directed replication: (a) Template (**ABBA**) and oligomers (**BA** and **AB**) diffuse over the surface of the micelle, (b) oligomers form a hybridization complex with the template strand, and (c) oligomers polymerize to yield a complementary copy of the template.

slowest run. The average time is estimated as 223.2τ. The huge variance is due to the random walk of template and oligomers over the surface of the micelle. Compared to the oil-water interface of the previous section, oligomer motion is further slowed down by the head particles of the amphiphiles as well as the dimer structure of the aggregate building blocks.

It is worth mentioning that, as expected, the hybridized complex is significantly more stable after the ligation has occurred than before. None of the hybridized complexes that formed in the above simulations showed any sign of dissociation within 750 time units after ligation took place.

3.5 Full Protocell Division

The last elemental step in the life cycle of the Los Alamos bug is the fission of the grown cell into two daughter cells as shown in Figure 12. In addition to what was discussed in Section 3.3, here we studied the fission of the whole protocell after the replication of its genome, that is, a micelle loaded with some lipid precursors, sensitizers, and two complementary PNA templates. The objective is to illuminate how templates and sensitizers are distributed among the daughter cells. Although not addressed by simulations in earlier sections, here the influence of the number of sensitizers is also investigated.

Proper division into two daughter cells requires the melting of the double-stranded PNA resulting from genome replication, which may be achieved by a temperature cycle. In the DPD formalism, temperature translates into the interaction parameters a_{ij}. To study the impact of a temperature cycle on the whole system, one would need to exchange the interaction parameters between all DPD beads. For simplicity in these initial investigations, and in the absence of a rigorous calibration of our model, we chose to invoke melting by simply turning off the attractive hybridization interactions between the PNA bases.

We performed simulations of a system of size $(10r_c)^3$ with an initial protocell consisting of 20 surfactants, 20 precursors, four to eight sensitizers, and two PNA template strands randomly located on its surface. Otherwise, the standard parameters listed in the beginning of this section were used. Snapshots of the system are shown in Figure 12. In all cases, metabolic turnover initiated the division of the aggregate at times of between 50τ and 100τ after the start of the simulation. Fission times were found to be longer than in the former experiments. This was because the aggregate consisted of more particles and because the template strands stabilized the rodlike aggregate that precedes protocell division. It was observed that PNA strands were preferably located along the elongated part of the aggregate, rather than at the caps. We believe that due to the stiffness parameter (Equation 8) of the PNA strands, the aggregate tends to elongate in a direction that is parallel to the PNA's long axis.

Using only four sensitizers, the distribution of sensitizers and PNA among the daughter cells was rather diverse: In one out of ten runs, all sensitizers and templates remained in one of the fission products, while the other consisted of only 11 surfactants. In seven of the runs the partition was nearly even: Both sensitizers and templates were equally distributed between the two daughter

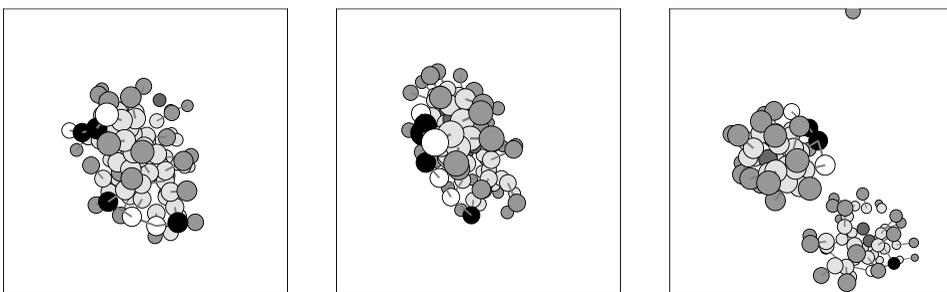Figure 12. The division of the whole protocell completes the life cycle of the Los Alamos bug. A mature protocell is loaded with precursor molecules, sensitizers, and two complementary PNA strands. During the metabolic turnover of precursors, the aggregate elongates and divides. Both PNA strands and sensitizer molecules tend to distribute evenly among the daughter cells when only a few sensitizers are present.

cells, which differed in aggregation number by at most 3 surfactants. Last but not least, we also observed two runs where the other components were distributed equally, but one of the daughter cells contained both template strands. We note that although it was not observed, it might be possible for a template to connect two otherwise divided aggregates by attaching to both their surfaces.

One might expect the equipartition of sensitizers is more likely when their number is increased. Our simulation results, however, showed quite the opposite: Protocells loaded with eight sensitizers instead of four almost always responded by rejecting an average of 11 to 12 surfactants. By doing so, the protocell was able to maintain a stable spherical shape even with an aggregate number of 27 surfactants. This is due to the collective stabilizing effect of the strongly hydrophobic core of sensitizers within the aggregate. The more sensitizers are added, the more they will tend to stick together. The more they stick together, the less likely they will partition into different daughter cells. Thus they are better able to stabilize the amphiphilic dimers in the aggregate. For an initial protocell that holds six sensitizers, proper division can still be observed, but the results are less reliable than in the case of four sensitizers. For six sensitizers, equipartition of sensitizers was only achieved in one out of five simulations. The other runs led to empty micelles or a situation where one of the daughter micelles had only one sensitizer bead. Equipartition of PNA could not be achieved for the cases with either six or eight sensitizer beads.

4   Discussion

Because of the inherent simplifications of the aggregated DPD simulation technique and the inherent complexity of our protocell system, accurate predictions of neither the detailed kinetic nor the thermodynamic properties could be expected. However, insights into generic issues and likely system behavior could be obtained by the illumination of the systemic properties of the proposed protocell design. In particular, we were able to see how the global behavior emerges from the simple and well-defined properties of the underlying molecular ingredients. Interpolation between several simulation methods combined with experimental data is necessary to obtain predictive understanding of this protocellular system. Investigations based on quantum mechanics, molecular dynamics, and reaction kinetics, combined with these and other DPD studies, hopefully can address the quantitative prediction issues in a more complete manner [26].

We found that the micellar kinetics that underlie the container replication are highly affected by hydrophobic molecules present in the solution. In the design of the Los Alamos bug, these hydrophobic molecules can be the metabolic precursors and sensitizers. As these molecules are incorporated into the protocell, they form a core that stabilizes the aggregates. Such loaded micelles have a larger aggregation number than micelles in a pure surfactant-water system, and the surfactant exchange with the bulk phase is strongly decreased. The simulations thus suggest that a three-component (ternary) surfactant-oil-water system is more suitable for yielding a suitable container than a two-component system based on surfactant and water only.

We also observed that protocells grow in spurts rather than continuously, even with a continuous supply of resource molecules. This is because the oil-like precursor molecules form droplets before they are absorbed by the aggregates. Furthermore, due to slower diffusion of larger objects, once the droplets start to form, volumewise they will tend to grow ever more rapidly the larger they become prior to being absorbed. The spurtlike support of resources might be sufficient to initiate the division process of the aggregate if these droplets have the appropriate size. If so, the system would be self-regulated and no further triggering of the metabolism (as with an external light switch) would be necessary. Whether or not this self-regulation enables a reliable replication of the whole organism also depends on a number of other factors, such as the rate of precursor supply compared to the replication rate of the genome. Further simulation investigations will be necessary to identify whether the metabolic self-regulation is sufficient when the precursor supply rate is not carefully balanced.

Our representations of the biopolymer that stores genomic information can be considered to be the crudest feature of the model. None of the implementations relate in detail to the actual

physicochemical traits of the real PNA molecule. The behavior of the PNA molecule with hydrophobic side chains in our protocell is also found to be quite different from that seen for DNA or RNA in water. Unlike DNA, where hybridized base pairs are radially opposite, in our PNA the hybridized bases are more likely to line up side by side in our attempts to model them. Furthermore, we have not been able to achieve an appropriate modeling of the balance between the hydrogen bond formation and the π stacking between the bases, in large part due to the hydrophobic and amphiphilic elements involved. More work and new ideas are needed here. However, we believe that the most fundamental properties of the biopolymer used—a PNA strand decorated with hydrophobic anchors that is able to hybridize with another PNA strand via H-bonds—is captured, at least in a qualitative manner. Against the background of this caveat, two findings are of particular interest: The simulations reveal that even our simple template representations are sufficient to affect the stability of the hybridization complex. In other words, it is observed how a molecular fitness function emerges from very few assumptions about the underlying molecular implementation. Furthermore, this fitness function is not a simple superposition of the individual monomer properties, but rather depends on the *sequence* of nucleotides in the genome. This finding is consistent with experimental studies on non-enzyme template-directed replication of RNA [1, 17].

It should be noted that an aggregate catalyzed gene replication could be realized in a variety of ways. In this implementation, we have assumed that the aggregate-water interface catalyzes the oligomer ligation process. Alternatively, one could imagine that the hybridized PNA complex sinks into the interior of the aggregate as a consequence of H-bond saturation. Here ligation, which is a dehydration reaction, might be favored due to the low local water content. None of our genome implementations can capture this behavior. In order to achieve it, one would probably need to alter the hydrophilicity of the surfactant head groups, depending on their hybridization state. We did not explore this route any further, since it has not yet been clarified how the hybridization complex behaves in the experimental system.

Equipartition of the components among the daughter cells after the division was achieved only when a few hydrophobic sensitizers are present in the protocell. Above a minimal number of sensitizers, equipartition becomes less probable as the number of sensitizers is further increased. This counterintuitive finding is connected to the fact that sensitizers, like precursors, form a hydrophobic core in the interior of the micelle, thereby increasing the allowed size of stable aggregates, in addition to stabilizing them overall. Since the stability of the core itself increases with its size, once large enough, it becomes nearly impossible for the core and therefore the protocell as a whole to divide. Instead, the instability caused by the excess surfactants is addressed by rejecting excess individual surfactants one at a time. The results suggest that the volume of the sensitizer molecules most likely will affect the fission dynamics when a certain threshold is reached.

Many open questions about systemic issues are still left unanswered by these initial investigations. The main open issues include:

1. What is the effect of heating the whole system in order to dehybridize the gene templates? Obviously, the lipid aggregate has to be more heat tolerant than the gene duplex.

2. What is the effect of defining the gene duplex as the photocatalyst, as in the originally proposed protocell design [28]? In our simulations, the sensitizer has been assumed to do the photofragmentation without any genetic catalysis. Also, what is the effect of having the sensitizer as a separate molecule (as reported here) versus covalently linking it to the gene, for example, as one of the lipophilic anchors?

3. What is the effect on the overall protocell replication if both the gene precursors (oligomers) and the lipid precursors are supplied to the solution and have to diffuse to the protocell? In such a case, will we see the coordinated gene and container growth based on reaction kinetics predicted by Rocheleau et al. [29]? As gene replication is necessary before container division for two viable daughters, can that be ensured in other ways than through a sequential resource supply?

4. What new issues arise when the protocell goes through more than one generation of its life cycle, for example, due to complementary resource sequence supplies?

Subsequent work in this area must also relate the DPD simulation implementation in this publication and its dynamics with corresponding molecular dynamics simulations [32] and reaction kinetics studies [18] as well as experimental findings as they arise.

5 Conclusion

The overall replication dynamics that constitute the life cycle of the Los Alamos bug was implemented using DPD simulations. In particular, we investigated the dynamics of container, metabolic complex, and genome subsystems, as well as the mutual interaction between these individual components. Component diffusion, self-assembly, precursor incorporation, metabolic turnover, template-directed replication of the gene, and finally the protocellular division were studied in various simulations. The main systemic findings are:

1. Metabolic growth orchestration can be coordinated by a switchable light source and/or by a continuous light source together with regulation of the size and frequency of the oily precursor package injection; this was not anticipated.

2. As anticipated, there is a trade-off between the lipophilic strength of the genetic backbone that makes it stick to the aggregate and its ability to hybridize easily with a complementary string.

3. As anticipated, for PNA with hydrophobic side chains, three-dimensional structure formation that can potentially inhibit appropriate hybridization is more likely in water than at an oil-water or lipid-water interface, although this is in part also dependent on the type of hybridization attraction.

4. Gene replication is easier at the surface of a micelle with a substantial oil core than for a micelle with a little or no oil core. The larger the oil core is, the easier the gene replication becomes, due to the aggregate stability and the ability to have a linear template.

5. As anticipated, the stability of two full complementary gene strings is much higher than that of a gene template and two complementary unligated gene pieces.

6. We observe that the template directed replication rate is dependent on the monomer component sequence and not only on the monomer component composition.

7. Partition of lipids, sensitizers, and gene between daughter cells strongly depends on the size of the oil core. The smaller the oil core is, the more balanced the partition becomes, which was not anticipated.

These systemic findings are now being considered in the experimental designs being pursued as part of the Protocell Assembly (PAs) and Programmable Artificial Cell Evolution (PACE) collaborations, and their validity will eventually be addressed as the experiments are executed.

Acknowledgments

References

1. Acevedo, O. L., & Orgel, L. E. (1987). Non-enzymatic transcription of an oligodeoxynucleotide 14 residues long. *Journal of Molecular Biology, 197*(2), 187–193.

2. Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Watson, P. (2002). *Molecular biology of the cell*. New York: Garland.

3. Aniansson, E. A. G., Wall, S. N., Almgren, M., Hoffmann, H., Kielmann, L., Ulbricht, W. J., Zana, R., Lang, J., & Tondre, C. (1976). Theory of the kinetics of micellar equilibria and quantitative interpretation of chemical relaxation studies of micellar solutions of ionic surfactants. *Journal of Physical Chemistry, 80*, 905.

4. Bachmann, P., Luisi, P., & Lang, J. (1992). Autocatalytic self-replicating micelles as models for prebiotic structures. *Nature, 357*, 57–59.

5. Bedau, M., Buchanan, A., Gozzala, G., Hanczyc, M., Maeke, T., McCaskill, J., Poli, I., & Packard, N. (2006). Evolutionary design of a DDPD model of ligation. In *Proceedings of the 7th International Conference on Artificial Evolution EA'05* (pp. 201–212). Berlin: Springer.

6. Buchanan, A., Gazzola, G., & Bedau, M. A. (2006). *Systems self-assembly: Multidisciplinary snapshots* (chapter "Evolutionary design of a model of self-assembling chemical structures"). London: Elsevier.

7. Coveney, P. V., Emerton, A. N., & Boghosian, B. M. (1996). Simulation of self-reproducing micelles using a lattice-gas automaton. *Journal of the American Chemical Society, 118*, 10719–10724.

8. Español, P., & Warren, P. (1995). Statistical mechanics of dissipative particle dynamics. *Europhysics Letters, 30*, 191–196.

9. Evans, D., & Wennerström, H. (1999). *The colloidal domain: Where physics, chemistry, biology, and technology meet*. New York: Wiley-VCH.

10. Fellermann, H., & Solé, R. (2006). Minimal model of self-replicating nanocells: A physically embodied, information-free scenario. *Philosophical Transactions of the Royal Society London B,* DOI: 10.1098/rstb.2007.2072, in press.

11. Forrest, B. M., & Suter, U. W. (1995). Accelerated equilibration of polymer melts by time-coarse-graining. *Journal of Chemical Physics, 102*(18), 7256–7266.

12. Ganti, T. (2003). *The principles of life*. Oxford, UK: Oxford University Press.

13. Groot, R. D. (2000). Mesoscopic simulation of polymer-surfactant aggregation. *Langmuir, 16*, 7493–7502.

14. Groot, R. D., & Rabone, K. L. (2001). Mesoscopic simulation of cell membrane damage, morphology change and rupture by nonionic surfactants. *Biophysical Journal, 81*, 725–736.

15. Groot, R. D., & Warren, P. B. (1997). Dissipative particle dynamics: Bridging the gap between atomistic and mesoscale simulation. *Journal of Chemical Physics, 107*(11), 4423–4435.

16. Hanczyc, M., & Szostak, J. (2004). Replicating vesicles as models of primitive cell growth and division. *Current Opinion in Chemical Biology, 8*, 660–664.

17. Joyce, G. F. (1984). Non-enzyme template-directed synthesis of RNA copolymers. *Origins of Life and Evolution of the Biosphere, 14*, 613–620.

18. Knutson, C., Benkö, G., Rocheleau, T., Mouffouk, F., Maselko, J., Shreve, A., Chen, L., & Rasmussen, S. (2006). Metabolic photo-fragmentation kinetics for a minimal protocell. *Artificial Life,* DOI: SFI-WP 06-12-047, in press.

19. Luisi, P., Walde, P., & Oberholzer, T. (1994). Enzymatic RNA synthesis in self-reproducing vesicles: An approach to the construction of a minimal synthetic cell. *Bunsengesellschaft für Physikalische Chemie. Berichte, 98*, 1160–1165.

20. Marsh, C. (1998). *Theoretical aspects of dissipative particle dynamics*. Ph.D. thesis, University of Oxford, Oxford, UK.

21. Mayer, B., & Rasmussen, S. (1998). Self-reproduction of dynamical hierarchies in chemical systems. In *Proceedings of the Sixth International Conference on Artificial Life* (pp. 123–129). Cambridge, MA: MIT Press.

22. Mayer, B., & Rasmussen, S. (2000). Dynamics and simulation of micellar self-reproduction. *International Journal of Modern Physics C, 11*, 809–826.

23. McCaskill, J., Packard, N., Rasmussen, S., & Bedau, M. (2006). Evolutionary self-organization in complex fluids. *Philosophical Transactions of the Royal Society London B,* DOI: SFI-WP 06-12-049, in press.

24. Nielson, P. (2007). *Protocells: Bridging nonliving and living matter* [chapter "Peptide nucleic acid (PNA) as prebiotic and abiotic genetic material"]. Cambridge, MA: MIT Press.

25. Ono, N. (2001). *Artificial chemistry: Computational studies on the emergence of self-reproducing units.* Ph.D. thesis, University of Tokyo, Tokyo.

26. Los Alamos National Laboratory, Protocell Assembly (PAs), http://protocells.lanl.gov; European Commission, Programmable Artificial Cell Evolution (PACE), http://www.istpace.org/.

27. Pohorille, A., & Deamer, D. (2002). Artificial cells: Prospects for biotechnology. *Trends in Biotechnology, 20*, 123–128.

28. Rasmussen, S., Chen, L., Nilsson, M., & Abe, S. (2003). Bridging nonliving and living matter. *Artificial Life, 9*, 269–316.

29. Rocheleau, T., Rasmussen, S., Nielson, P. E., Jacobi, M. N., & Ziock, H. (2006). Emergence of protocellular growth laws. *Philosophical Transactions of the Royal Society B,* DOI: 10.1098/rstb.2007.2076, in press.

30. Trofimov, S. (2003). *Thermodynamic consistency in dissipative particle dynamics.* Ph.D. thesis, Technische Universiteit Eindhoven, Eindhoven, The Netherlands.

31. Venturoli, M., & Smit, B. (1999). Simulating self-assembly of model membranes. *PhysChemComm, 10*, 45–49.

32. Weronski, P., Jiang, Y., & Rasmussen, S. (2007). Molecular dynamics study of small PNA molecule in lipid-water system. *Journal of Biophysics, 92*, 3081–3091.

33. Yamamoto, S., & Hyodo, S. (2003). Budding and fission dynamics of two-component vesicles. *Journal of Chemical Physics, 118*(17), 7937–7943.

34. Yamamoto, S., Maruyama, Y., & Hyodo, S. (2002). Dissipative particle dynamics study of spontaneous vesicle formation. *Journal of Chemical Physics, 116*(13), 5842–5849.

35. Yuan, S., Cai, Z., & Xu, G. (2002). Dynamic simulation of aggregation in surfactant solution. *Acta Chimica Sinica, 60*(2), 241–245.

Appendix: Algorithm for Chemical Reactions

Between every two DPD time steps, the following algorithm is applied to perform chemical reactions: For every reaction scheme, we successively check all possible pairs of reactants A, B, and compare their effective reaction rate k with a number taken from a suitably normalized pseudorandom number generator. If the reaction rate is smaller than this value, we perform the reaction and go on to the next pair of possible reactants. However, A and B will not be considered again in this step. The exact algorithm—notated in the Python programming language—reads as follows:

```
shuffle(reaction_list)

for reaction in reaction_list :

  for A in space.particles(reaction.educt_A) :

    if reaction.is_synthesis :

      # if reaction is a synthesis, possible

      # reaction partners are particles

      # of type educt_B in the vicinity of A.

      partners = A.neighbors(

        reaction.educt_B,reaction.R

      )
```

```
else :
    # otherwise, possible reaction partners
    # are particles of type educt_B bonded to A.
    partners = A.bonded(reaction.educt_B)

for B in partners :
    # compute effective reaction rate
    k = reaction.k
    for C in A.neighbors(
        reaction.catalyst,reaction.r_cat
    ) :

        k += reaction.k_cat *
            (1−(A.pos−C.pos).length()/reaction.r_cat)

        if random() < dt * k :
            # perform reaction
            react(A,B,reaction)
            # and leave loop over partners
            continue
```