Physically Embedded Minimal Self-Replicating Systems – Studies by Simulation

Dissertation
zur Erlangung des Grades
eines Doktors der Naturwissenschaften (Dr. rer. nat.)

Harold Fellermann
Diplom-Systemwissenschaftler

Fachbereich Mathematik/Informatik
Universität Osnabrück

August 2009

To my family and friends.

Summary

Self-replication is a fundamental property of all living organisms, yet has only been accomplished to limited extend in manmade systems. This thesis is part of the ongoing research endeavor to bridge the two sides of this gap. In particular, we present simulation results of a minimal life-like, artificial, molecular aggregate (i.e. protocell) that has been proposed by Steen Rasussen and coworkers and is currently pursued both experimentally and computationally in interdisciplinary international research projects.

We develop a space-time continuous physically motivated simulation framework based on the method of dissipative particle dynamics (DPD) which we incrementally extend (most notably by chemical reactions) to cope with the needs of our model. The applicability of the method over the entire length scale of interest is reintroduced, by rejecting a concern that DPD introduces a freezing artifact for any model above the atomistic scale. This is achieved by deriving an alternative scaling procedure for interaction parameters in the model.

We perform system-level simulations of the design which attempt to account for theoretical, and experimental knowledge, as well as results from other computational models. This allows us to address key issues of the replicating subsystems – container, genome, and metabolism – both individually and in mutual coupling. We analyze each step in the life-cycle of the molecular aggregate, and a final integrated simulation of the entire life-cycle is prepared.

Our simulations confirm most assumptions of the theoretical designs, but also exhibit unanticipated system-level dynamics. These findings are used to revise the original design of the Los Alamos minimal protocell over the course of the analysis.

The results support the hypothesis that self-replication and probably other life-like features can be achieved in systems of formerly unanticipated simplicity – if these systems exploit physicochemical principles that are immanent to their physical scale.

Contents

Chapter 1

Introduction

> "Anyone who looks at living organisms knows perfectly well that they can produce other organisms like themselves... Furthermore, its equally evident that what goes on is actually one degree better than self-reproduction, for organisms appear to have gotten more elaborate in the course of time." – von Neumann [1]

1.1 Theoretical models of self-replication

1.1.1 Replication versus self-replication

On first thought, replication seems to be easily graspable conceptually. We say that something has replicated if it has produced a sufficiently similar copy of itself. To be more than a mere model, the replica has to resemble the original at least in both structure and function. Concerning physical objects, the required degree of similarity can also extend to their material properties.

We have to distinguish self-replication from other forms of replication. For example, we would not speak of a sheet of paper as being self-replicating, merely because we can make a copy of it. In general, a structure is not able to self-replicate when the replicating act is only due to its environment. A first indicator that a system is indeed self-replicating is found in its environment: if everything that is embedded in the environment will be replicated, the act of replication has to be attributed to the environment rather than the system. For this reason, mathematical and formal structures that are generated by iterated function systems or recursive grammars (such as Lindenmayer systems [2]) cannot be considered to be self-replicating objects, because these environments replicate any given input. As a minimal condition, the set of self-replicating systems has to be a proper subset of the set of all systems.

This first requirement is necessary but not sufficient, as we could easily equip a copying device that was hitherto able to copy all objects with a sensor that recognizes and excludes an arbitrary subset of objects (such as modern photo copiers that recognize and prevent the replication of banknotes). Anything the enhanced copier does still replicate would conform to a definition of self-replication that is solely based on the proper subset condition.

This motivates the demand that a self-replicating system somehow actively participates in the process of replication, rather than being passively copied [3]. As it turns

out, however, the distinction between active and passive replication is not easily expressed in a rigorous concept, since one can always argue that any action is caused by natural laws (or by the external application of rules in a formal calculus) which puts all objects in the passive position of being mere recipients of an action. As a brief historical overview of self-replicating objects will elucidate, there is no clear distinction but a smooth transition between replication and self-replication.

We briefly highlight some of the accomplishments in the history of artificial self-replicating systems. The presentation is not meant as a review but only supports background for the argument we will make in section 1.1.5. Comprehensive reviews of the subject can be found in Refs. [4, 5, 6].

1.1.2 Von Neumann's universal constructor

The mathematical study of self-replicating systems originated in the late 1940s, when Hungarian-American mathematician John von Neumann became interested in the behavior of complex automata and machines [see e.g. 6]. Interested in the evolution of complexity in living systems, von Neumann tackled the problem whether an automaton would be able to produce an at least equally complex automaton. In particular, von Neumann searched for automata that are able to replicate their own structure. Until his death in 1957, von Neumann had outlined five concepts for self-replicating machines that all derive from a common sufficiency proof.

Von Neumann concluded that the following properties suffice for a machine X to allow it to self-replicate:

1. X must be able to construct, i.e. it must be able to manipulate the material, information, and energy it consists of.

2. X must be constructionally universal, i.e. it must be possible to program it by a sequence of instructions to construct any machine within a system S of machines.

3. X must be a member of the system S.

As Barry McMullin points out [7], the notion of universal construction is context dependent with respect to the system S which is implicitly defined by the constructive operations of X, similar to how set operations define an algebra. This implies considerable freedom in the design of a universal constructor which is only partly constrained by condition 3.

Starting from these prerequisites, von Neumann designed a machine that consists of four components: First, a universal constructor A which can build any machine Y in S when given an appropriate instruction tape; second, a copying device B that can replicate an arbitrary instruction tape; third, a controller C to control the operations of both the constructor and copier; and forth, an instruction tape $\Phi(Y)$ that tells the universal constructor how to build machine Y. If Φ encodes instructions on how to build the machine $(A+B+C)$, the setup allows for self-replication: first, the controller initiates the constructor A to build a replica $(A' + B' + C')$; second, the controller commands the copier B to replicate the instruction tape $\Phi(A + B + C)$. Finally, the controller activates the newly formed copy so that it starts its own replication cycle.

Note that in this setup, the instruction tape appears both as a program (to control the universal constructor) and as data (when copied to another tape). This ambiguity resolves an infinite regress that would otherwise appear if the instructions where not

copied but would instead directly instruct the constructor to build the instruction tape de novo ("build a tape with instructions to build a tape with instructions to ..."). In parenthesis, we mention that this ambiguity parallels the double role of DNA in natural organisms, which serves both as code for the production of proteins and as data when these very proteins copy the genome during DNA replication. Also, the resolution of this infinite regress has become standard in modern general purpose computer architectures which employ a unique memory space for both data and program instructions.

The original framework von Neumann had in mind for the above proof was what he later referred to as the kinematic machine [6]: a hypothetical physical device located in an environment of building blocks (such as a stockroom of spare parts) in which the machine could move to access and assemble parts. As a mathematician however, von Neumann abandoned the physical scenario in favor of the more rigorous framework of lattice models following a suggestion of his colleague Stanislav Ulam. The class of lattice models that von Neumann designed as the environment for the self-replicating machine has now become widely popular and known as cellular automata [8].

Cellular automata (CA) consist of a grid of locally linked finite state machines (cells) where each machine is in one of a finite set of states and changes its state according to its current state and the current input from connected cells. This update is assumed to happen synchronous in all cells and in discreet time steps. A CA is uniquely defined by its state set, topology (i.e. linking of cells and boundary conditions), and set of transition rules. In the CA model, a machine corresponds to a pattern (i.e. configuration of cell states). The performance of the machine is given by the time evolution of the state space which is completely determined by the definition of the CA.

In von Neumann's CA (a two-dimensional square lattice with a four-cell orthogonal neighborhood), condition 1 is trivially met by observing that CA transition rules change cell states in response to the current pattern. In other words, the pattern that constitutes a machine in a CA is able to manipulate its constituting elements. In order to show that conditions 2 and 3 can be met by a CA, von Neumann laid out an existence proof by explicitly designing a particular CA that can replicate a pattern.

Von Neumann's cellular automaton consists of 29 states, with one of them representing a 'ground' state or substrate that only changes into another state upon actions from neighboring non-ground states [8]. The other states enable the machine to directionally transmit and process information thereby modifying substrate cells. Equipped with this framework, von Neumann designed basic elements such as logic gates, delay stations, and their wiring. The universe of machines that can be built out of these elements (the system S in the above outline) holds a memory tape, read/write elements, a "construction arm" which are finally combined to form the entire self-replicating machine $(A + B + C)$ [9]. Due to its complexity, the automaton was first implemented on a digital computer by 1995 by Nobili and Pesavento [10] but required a memory tape that was too long to allow for self-replication within reasonable time. In 2008, Nobili [11] incorporated run-length compression into the reading device to shorten the memory tape which allowed for the first actual in silico replication of von Neumann's machine (see Fig. 1.1).

The CA model is certainly the furthest developed framework of self-reproducing automata that von Neumann envisioned. But it was only one of five general concepts. The idea of a mechanical automaton in the form of the kinematic machine has already been mentioned. Other than these two, von Neumann also envisioned a variant of his CA that employs the McCulloch-Pitts model of artificial neurons [12] (referred to as the neuron-type machine by A. Burks [8]); a differential equation based continuous variant of the neuron-type machine, and finally a stochastic version thereof. For details, the

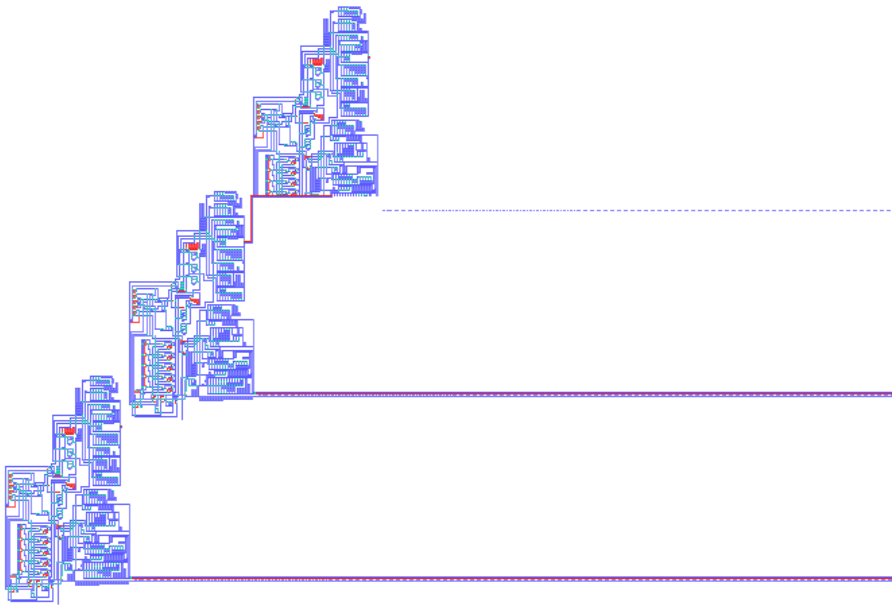


Figure 1.1: Implementation of von Neumann's self-replicating machine by Nobili [11]. The pictues shows the replicating machine in the middle of its second replication cycle. Each machine consists of a universal constructor (upper part), a tape reader (lower right part), a control unit (lower left part), a memory tape (extending from each automaton to the right), and (while in operation) a construction arm (red link connecting second and third automaton). To simplify the layout, the cellular automaton employs 32 rather than von Neumann's original 29 states and employs run-length compression to reduce the length of the instruction tape. (Public domain: `http://en.wikipedia.org/wiki/Image:Nobili_Pesavento_2reps.png`, version of August 1., 2008.)

reader is referred to Refs. [6, 13].

At the same time, the CA based automaton has given raise to numerous variations that link the capabilities of von Neumann's and alike machines to self-inspection [14] (thereby removing the need for explicit instructions on a tape), self-repair [15], and evolution [16] while the original design has been greatly simplified by Codd [17], reducing it to 8 states instead of 29.

1.1.3 Artificial Life models of self-replication

While von Neumann presented a mathematical proof, that self-replication is achievable in a mechanical universe, the work does not seek for a particularly simple solution. Clearly, enabling a machine to construct any machine, instead of only its own unique structure, adds to the complexity of von Neumann's approach: while a universal constructor is sufficient for self-replication, it is certainly not a necessity. This observation lead to a search for minimal structures in CAs, that would be able to accomplish self-replication (and nothing else). Since "minimality" is ambiguous and can refer to the number of states, cells, or iterations needed for replication, there is no single minimal solution, but a set of candidates to be considered.

In 1984, Langton [18] published his work on a "self-replicating loop": an eight state cellular automaton (instead of von Neumann's 29) that consists of mere 86 non-quiescent cells (rather than about 150,000 cells in von Neumann's design implemented in Ref. [9]) and can self-replicate in 151 time steps (as opposed to 63 billion in Ref. [9]).

Basically, Langton's CA consists of a looped rectangular tape that is sheathed in "walls" with a construction arm extending it at one corner (see Fig. 1.2). Cell states that code for instructions are constantly transmitted among the cells in the circular memory tape. They are duplicated at the junction of the storage loop and the construction arm. Instructions are transmitted without affecting the geometry of the device, as long as there is a free tape cell other than the one they have been transmitted from in the last time step. This enables the device to store and transmit information without actually being altered. When instructions do reach a tape cell, however, from which no other tape cell is reachable– i.e. at the end of the construction arm – these instructions cause the device to extend its instruction tape by one cell, either orthogonal or perpendicular, depending on the respective instruction. By this mechanism, the construction arm can be programmed to construct a second loop adjacent to it. As instructions are automatically duplicated at the junction, nothing is needed to equip the newly constructed loop with its own set of instructions. The replication cycle ends when the first instruction of the copied code reaches the end of the parent's construction arm and initiates it to separate. A total of 207 CA transition rules are needed to define this behavior.

Langton's contribution started a race toward ever more simple self-replicating looplike CA structures of which Byl's loop [19] with 6 states and 12 cells and Reggia et al.'s loop [20] with 8 states but only 5 cells are the current record holders. Still, these implementations require a set of about one hundred transition rules. Furthermore, looplike self-replicating devices have been modified to allow for robustness [21] and evolution [22, 23]. Also, Mange et al. [24] linked self-replicating loops back to universal construction: in their approach, the instruction tape can be programmed to perform arbitrary modifications of the cell space after replication of the loop.

In general, the Artificial Life movement of which Langton was the premier spokesman [4] demonstrates the relative ease with which life-like properties can be obtained from simple mathematical models. It does so by adopting the computational paradigm of von Neumann's distributed, locally interacting, and decentralized CAs in order to

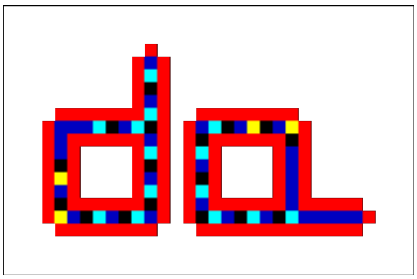


Figure 1.2: Langton's loop after replication. Two replicating loops are shown, each consisting of a circular instruction tape and a construction arm (dark blue cells) sheathed by walls (red cells) with instructions to extend the arm orthogonal (cyan cells) or perpendicular (two yellow cells). Courtesy of Thomas Schoch taken from `http://commons.wikimedia.org/wiki/Image:Langton_Loops_after_replication.gif`, version of April 16, 2007.

emphasize the emergence of global behavior and properties from simple, distributed, locally interacting components. However, where von Neumann populates his cellular world with devices that are nowadays recognized as conventional engineered compounds (such as logic gates, random access memory, etc.), Artificial Life research seeks for behavior that is an immediate consequence of the underlying computational model. As a result, Langton notes that "the essential machinery of living organisms is quite a bit different from the machinery of our own invention, and we would be quite mistaken to attempt to force our preconceived notion of abstract machines onto the machinery of life." [4, page 21].

At the same time, the Artificial Life movement set its focus on the logic of emergent behavior rather than its physical foundation. In the words of Langton: "Since we know that it is possible to abstract the logical form of a machine from its physical hardware, it is natural to ask whether it is possible to abstract the logical form of an organism from its biochemical behavior. The field of Artificial Life is devoted to the investigation of this question." [4, page 21].

1.1.4 Mechanical self-replicating devices

Mechanical implementations of self-replicating systems stem from the original conception of von Neumann's physically embedded kinematic machine. Notable early solutions are the work of Penrose and Penrose in 1957 [25], who manufactured mechanically connectable building blocks that could replicate a specific local connection configuration (the "seed") when the container in which they reside is shaken. Similarly, Jacobson in 1958 [26] designed a railroad system, in which a specific sequence of self-propelled toy train coaches replicates by detecting passing coaches and directing their coupling. Freitas and Merkle [6] note "that a great deal of functionality essential for replication [in Jacobson's system] resides in the environment". This concern corroborates the hypothesis of a the smooth transition from replicators to self-replicators that has been postulated in section 1.1.1.

The pioneering work of Penrose and Jacobson lead to the engineering of self-replicating robots which has succeeded in recent years [27, 28, 29, 30]. These designs basically imitate the conceptions of Penrose and Jacobson, by allowing a modular

robot to detect and assemble spare modules to eventually combine them into a copy of its own structure. While being remarkable in achieving actual mechanical replicators, these solutions share the engineer's approach of von Neumann, in that they employ conventional building blocks in order to mimic life-like behavior.

Contemporary approaches in robotics that directly apply to the self-replicating systems presented in this thesis are the research areas of physical embodiment and morphological computation, i.e. the ability of passive mechanical elements to perform complicated computational tasks [31, 32, 33, 34]. A paradigmatic example of morphological computation is the passive dynamic walker by McGeer [35], a bi-ped robot that is able to stably walk down a slope without employing any active element such as motors or sensors. The "robot" gains this ability from a carefully balanced mass-spring system that transforms gravitational force into directed motion. Pfeiffer notes: "We showed that by exploiting morphology, materials, and system-environment interaction, hard tasks such as rapid locomotion or grasping can be achieved in a 'cheap' manner." [33]. Paul goes beyond that statement by noting that "the fact that such computation can be so simply achieved suggests that it is not simply a rare phenomena, but possibly a pervasive characteristic of physical structures" [31].

1.1.5 Physically embedded self-replication

The present work is an attempt to apply the capabilities of morphological computation to the area of self-replicating structures. The central hypothesis of our work is that natural self-replicating systems are not only embedded in a physical environment, but that their material, structure, and function is intimately shaped by the natural laws that govern their physical environment. We argue that self-replicating structures could grow to abundance in our world, because they employ mechanisms and components that are favored by the underlying laws of nature. In other words: existing natural self-replicators have been selected for being simple. Therefore, we argue that self-replication, and likely other features hitherto unique to life are deeply immersed in the physical laws of our universe.

We support this hypothesis by a computational model which is based on a simplified representation of the physical laws relevant to natural self-replicating systems. Basically, our framework relies on two physically motivated processes for motion and reactive turn-over of building blocks. We then design structures that directly exploit the modeled physicochemical laws of the framework in order to self-replicate. The simplest of such structures consists of only three types of components with few parameterized interaction laws (see chapter 4). Due to the conceptual difference of our model compared to CA models, it is unclear how the presented self-replicator compares in terms of complexity and minimality to the systems described in section 1.1.3. However, we can demonstrate that self-replicating structures in our framework do not require a pre-designed seed, but spontaneously self-assemble from constituting parts, and that our mechanism of self-replication is dynamically stable against random fluctuations and robust (i.e. structurally stable) over a wide range of system parameters – all features that are difficult to achieve in CA based systems.

Choosing a conceptual framework grounded in physical laws has further advantages: (i) the relative closeness to physical theories allows to connect to the vast amount of knowledge from the areas of statistical physics, physical chemistry, and biophysics, which allows e.g. for calibration of the method. (ii) As a consequence, concepts behind the presented replicating structures become testable both by computation and experiments. (iii) The framework can not only be employed to study self-replicating objects

as abstract mathematical structures, but rather allows for modeling real self-replicating systems or conceptions thereof. A major part of this work is therefore dedicated to the modeling capabilities of the employed framework (chapters 2 and 3) as well as the actual model of a proposed molecular system that is envisioned to exhibit self-replication and other life-like properties (chapters 5 and 6).

1.2 Self-replicating chemical structures

1.2.1 Emergence of self-replicating structures - the origin of life

The emergence of self-replicating systems in our world is directly connected to the origin of life on Earth or in the universe. As it is our hypothesis that natural self-replicating systems are simple because they directly exploit physical laws for their functioning, it seems mandatory to briefly review what is known about simple self-replicating physico-chemical systems and their possible interplay in the origin of life.

Throwing light on the origin of life is particularly difficult due to the lack of knowledge about the early Earth's atmosphere, the multitude of possible physico-chemical environments and the absence of any fossil records of life's very origin. Even the most simple living organisms abundant today as well as the oldest available fossils display a complexity that renders their spontaneous appearance unconceivable. Nevertheless, the organizational structure of living systems elucidates its origins by rendering some scenarios more likely than others.

At the simplest level of description, living organisms consist of a container, or body, that separates the organism from its environment, a metabolism, and a carrier of inheritable information. In simple procaryotic cells, the body is a bilayer lipid membrane composed of phospholipids, glycolipids, steroids (such as cholesterol) and embedded membrane proteins [36] (Lipid membranes are present in all organisms known today, but higher developed organisms possess additional container components such as cell walls or a multicellular body plan [37]). Furthermore, living organisms possess a metabolism that harvests energy from the environment in order to produce their constituent components. Needless to say, the very first organisms must have been autotrophic and able to drive their metabolism with either chemical (redox) or light energy. Finally, all present living organisms possess inheritable information, i.e. a genome, in the form of DNA. Upon replication, an organism must duplicate its container, its genetic information, as well as the constituents of its metabolism.

As has been stated in section 1.1.2, DNA plays a double role in nowadays organisms as it encodes information for protein synthesis on the one hand, while being operand of a working protein machinery on the other hand. This poses an apparent "chicken and egg paradox" for both DNA and proteins must exist in order for the other to function: DNA directed protein synthesis is only possible in a world with DNA, whereas DNA requires a working proteome to function. To resolve this paradox, Woese [38] and Gilbert [39] have formulated the hypothesis of an RNA world, in which RNA as a single constituent would act both as information carrier and as enzyme (sometimes called ribozyme) that catalyzes its own or the replication of other RNA molecules or its constituents [40]. The widely accepted RNA world hypothesis is supported by experimental findings of ribozymes [41, 42, 43], as well as the non-encoding but catalyzing role of rRNA in nowadays organisms' protein synthesis – indeed, the catalytic centers of ribosomes, the pivotal point of protein synthesis, have been identified as being made of rRNA rather than proteins [44, 45, 37]. Finally, scenarios have been conceived by

which an RNA world might have given birth to the familiar DNA/protein world [39].

With respect to the discussion in sections 1.1.2 and 1.1.3, self-replicating biopolymers in the RNA world constitute a simple scenario in which self-replication and construction are tied together in that construction *is* replication. Only with the advent of a DNA/protein world, evolution discovered universal construction in the sense of von Neumann, in which a multipurpose constructor (the ribosome and its associated biomolecular machinery) can be programmed (via encoding DNA/mRNA strings) to construct arbitrary machines (proteins). As the biopolymer in the RNA world does not encode building instructions, the term "genome" is sometimes rejected. Nevertheless, RNA is capable to store inheritable information also in the absence of an encoding, as will be shown in section 1.2.3.

We will now discuss the replicative abilities of prebiotic lipid aggregates and biopolymers, as they might have contributed to the container and inheritable information of the very first organisms.

1.2.2 Self-replicating lipid aggregates

Whereas the membranes of contemporary cells exhibit a complex composition of lipids and proteins which is adapted to the environment and function of the cell [37], origin of life scenarios commonly envision the first lipid aggregates to be of a much simpler composition [46]. For a review on the availability and abiotic synthesis of lipids in prebiotic scenarios, see Ref. [47]. In general, it is believed that single-chain surfactants such as the fatty acids and alcohols that contribute the amphiphilic property of contemporary lipids are likely candidates of prebiotic and early biotic membranes [47]. In particular, fatty acid lipids and amphiphilic alcohols have been shown to form vesicles and micelles by spontaneous self-assembly in aqueous solution [48].

The phase diagram of these aggregates is subject to a variety of molecular and systemic parameters: whether an amphiphile solution forms micelles, vesicles, or other (less prominent or less defined) structures, is influenced by the length and possible branching of the hydrocarbon chain, the characteristics of its head group, its pK value, pH, temperature, and other systemic parameters. Ternary mixtures of oil, surfactant and water exhibit an even richer phase behavior. Notably, fatty acid surfactants can stabilize otherwise unstable oil water emulsions, giving raise to surfactant coated oil droplets known as micro-emulsion compartments. Unfortunately, the subject of soft condensed matter systems is too broad to allow for a concise overview. The reader is referred to Ref. [49]. The following overview will be restricted to unilamellar vesicles, micelles and micro-emulsion compartments as conceivable containers of prebiotic organisms, although other lipid phases might have also played a role in prebiotic evolution [50].

Induced budding and fission of vesicles

Spontaneous division of vesicles has not been reported in the literature. Division of vesicles requires bending of the bilayer membrane to form a bud small enough for lipids of the adjacent bilayer sheets to rearrange. For phospholipid (and likely also fatty acid) membranes, the bending energy of the membrane imposes an energy barrier that is unlikely to be overcome by thermal motion [37]. For this reason, the division of contemporary cells is orchestrated by the complex machinery of the cytoskeleton which itself is subject to the proteomics of the cell cycle [37]. Needless to say, if protoorganisms were based on a vesicular embodiment, they lacked this advanced machinery and must have relied on other (possibly external) means of division.

Division of vesicles can be induced externally. Experimentally, this is most easily achieved by extrusion, a standard procedure of pressing amphiphile solutions through a porous filter [51]. As Hanczyc and Szostak demonstrated [52], extrusion can force giant unilamellar vesicles to devide without significant rapture of the membrane that would lead to leakage of encapsulated material (in their case a dye). In a separate step, vesicles of the next generation are grown by supplying additional lipids to the solution, until they reach the original size.

Apart form extrusion, externally induced budding off of small vesicles from a giant "mother" vesicle has been employed as a mechanism for vesicular division. Budding can be enforced for example by (i) osmotically changing the surface to volume ratio of the vesicle, (ii) by selectively increasing the area of the outer bilayer leaflet through a temperature difference between the internal and external solution, or (iii) through an asymmetry in the density of inner and outer membrane leaflets [52, and references therein]. Budding can further be supported by the boundaries of domain forming lipid compositions [53]. Theoretical studies have suggested additional means of induced vesicle division based on adhesive nano-particles [54] or osmotic pressure [55, 56].

Replication of micelles, reverse micelles, and oil droplets

Whereas spontaneous replication scenarios of vesicles are not known, autonomous division processes have been reported for micelles, and reverse micelles [57, 58, 59]. Bachmann et al. [57] first reported self-replication of reverse micelles which are water droplets stabilized in organic solvent by an amphiphile layer. If the reverse micelle hosts a catalyst for a metabolic reaction, an autonomous growth and division cycle can be achieved by the following means: A hydrophobic ester (in their case octanoic acid octyl ester) is provided to the organic solvent of the system and serves as nutrient for the single metabolic reaction of cleaving the ester bond by hydrolysis. Reaction products are fatty acids (octanoic acids) and alcohols (octanol) – which are essentially the surfactants of the ternary system. Ester cleavage is enhanced by a hydrophilic catalyst (here LiOH) that will reside in the aqueous interior of the reverse micelle. The setup guarantees that the metabolic turnover of nutrients occurs at the micellar interface. Ref. [59] reports that ester cleavage is also found in the absence of a catalyst as the metabolic reaction can instead be mediated by milieu effects in the micellar structures themselves. Ref. [58] presents a similar recipe for aqueous micelles in which the geometry of the aqueous and lipid phase are reversed.

The supposed replication process in these system is best described for the original setup of Ref. [57]: surfactants that are newly produced by ester hydrolysis arrange at the lipid water interface of the reverse micelles as a result of their amphiphilic properties. Under the course of the reaction, this leads to a change in the surface-to-volume ratio of these aggregates, as the water is entrapped in the interior of the reverse micelles. It is supposed that this induces an elongation of the structure up to a point where thermal fluctuation suffices to divide the aggregate in two. A different interpretation of the results, however, suggests that the surface-to-volume change of the aggregates may be prevented by the rapid exchange of surfactants with the bulk phase [60] where excess amphiphiles would form micellar structures de novo. The exact pathway of micellar self-replication has not been identified in their publications [57, 58, 59] and is likely subject to system parameters. Self-replication of surfactant coated oil-droplets is computationally studied in chapter 4 of the this work.

1.2.3 Self-replication of biopolymers

Biopolymers, such as deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), are the premier self-replicating molecules in living organisms. Their chemical structure – linear heterogeneous polymers of pairwise complementary nucleotides that are able to align to each other by specific and energetically weak Watson-Crick binding [36] – immediately suggests the replication mechanism that is found in all contemporary cells [37]. Nevertheless, even in the most simple procaryotes, biological DNA replication is controlled by enzymes in almost every single step [37]. A significantly less complex, non-ezymatic replication mechanism must have preceded contemporary replication, which might have partially employed external factors [50].

Basically, the replication of biopolymers requires the processes of hybridization, ligation, and melting. Hybridization is the alignment of complementary nucleotides or oligomers along a template strand. Ligation is the forming of covalent bonds between the aligned nucleotides. Finally, melting separates the double strand into the original and its complementary copy.

As the ligation reaction is energetically uphill, monomers or oligomers need to be activated in order to perform ligation. Nonenzymatic template-directed RNA polymerization of elementary activated nucleotides (monomers) shows little yield in aqueous solution but can be enhanced at surfaces (e.g. clay) or through up-concentration in water-ice [50]. Template-directed replication from shorter activated oligomers, on the other hand, has be shown to produce high yields for both RNA and DNA [61, 62, and references therein]. The shortest DNA strand that has been experimentally replicated in the absence of enzymes is a hexamer with complementary trimers [63].

1.3 Approaches toward artificial cells

The last section has discussed the replication of prebiotic molecular systems as they are thought to have predated contemporary life, namely replicating lipid aggregates and biopolymers. Trying to tie these subsystems to each other and to a metabolism has lead to the field of protocell research [64, 65]. Protocells, or artificial cells, are molecular aggregates that are able to grow, replicate, and possibly provide means to grow in complexity, such that they might eventually undergo Darwinian evolution. In contrast to the top-down approach of minimal genome research [66], protocell research takes a bottom-up approach by combining inanimate matter to systems that can be identified as being alive. They are not confined to employ only biological components, and they are not primarily intended to be plausible in early life scenarios. Two proposed protocellular designs will now be presented. For a recent comprehensive overview of the field, see Ref. [65, 67].

1.3.1 The chemoton

The chemoton (short for chemical automaton) is the design proposal of a minimal artificial life-like system that was introduced by Gánti in the 1970s and has been continuously redefined since then [68, and references therein] .

In its current conception [69], the chemoton consists of the three subsystems container, metabolism, and genome. The genetic information carrier (a biopolymer system) is conceived to reside in a lipid container (vesicle) where it replicates by consuming monomer material produced via an autocatalytic metabolism [68]. The subsys-

tems are stoichiometrically coupled: in particular, the metabolic production of genetic, metabolic, and container building blocks is controlled by a differential feedback mechanism that maintains an orchestrated growth of the entire system [70].

In total, the chemoton model employs 5 metabolic reagents that constitute the autocatalytic cycle, one metabolic precursors for membrane molecules and one for monomers of the genetic systems. With additional nutrients and waste components, the chemoton adds up to 12 components which interact through 10 chemical reactions [70].

It has to be emphasized, however, that the biopolymer in the chemoton does not carry any information that would affect the behavior of the system apart from its need of being reproduced in order for full replication to occur. Therefore, the chemoton does not feature a true information component.

1.3.2 The Los Alamos minimal protocell

The protocellular design conceived by Rasmussen et al. [71] is currently subject to combined experimental and computational studies. The model presented in chapters 5 and 6 of this work contributes to the computational portfolio of the Los Alamos protocell assembly project. A recent report of the overall project status can be found in Refs. [72, 67].

In adherence with the simplicity paradigm of the Artificial Life movement, the underlying design principle of this protocell is to minimize the number and complexity of the physicochemical structures that a molecular aggregate requires in order to self-assemble, grow, self-replicate, and exhibit limited evolutionary potential [71, 72]. To achieve this, the Los Alamos minimal protocell features several major simplifications (a detailed presentation of the system will be given in chapter 5):

1. The metabolic and genetic complexes operate at the external interface of a lipid aggregate, rather than in the interior volume of a vesicle. Both fatty acid vesicles and oil droplets are considered as proto-container.

2. The metabolism is reduced to a single reaction mechanism, namely a light driven ester hydrolysis similar to the reaction described in section 1.2.2.

3. The information carrier (biopolymer) directly influences the rate of the metabolic turnover reaction rather than encoding for the production of catalysts (such as enzymes).

4. Catalytic coupling between container, metabolism, and information carrier attempts to harvest robustness from self-balancing system dynamics, rather than relying on a tight and more fault-prone stoichiometric coupling.

As this design tries to exploit ordering mechanisms which emerge from fundamental physicochemical processes and properties, we argue that the Los Alamos minimal protocell is a paradigmatic example of the physically embedded self-replication outlined in section 1.1.5.

As has been shown in section 1.2.1, a non-encoding biopolymer does not provide programmability of the system. This allows for self-replication as in section 1.1.3, but not for universal construction as in section 1.1.2. As a consequence, selection is expected to generate adaptation of the system toward better performing configurations (i.e. inherited nucleotide sequences), but the evolutionary potential of the system might turn out to be limited [73].

The reader is invited to understand this thesis as two-sided: on the one hand, the work presents an abstract self-replicating structure that resides in a mathematical universe along the lines of von Neumann's and Langton's cellular automata. Its most notable difference of being space-time continuous arguably allows both for the mathematical framework as well as the replicating structure itself to be of unpreceded simplicity. On the other hand, the work presents a computational model of an envisioned physicochemical system. It has to be emphasized, however, that the inherent simplifications of the employed modeling framework do not allow one to blindly draw quantitative predictions from simulation outcomes. Rather than prediction, the model has been used within the protocell assembly project to support or reject design decisions, illuminate system-level dynamics, and similar quantitative considerations. Only the comparison with laboratory results and simulation results of other computation methods (such as mass reaction kinetics and Molecular Dynamics simulations) will allow for reliable interpretation of the presented results. Where available, these results are included in the discussion.

Outline of the remainder of this work

The remainder of this work is organized as follows: chapter 2 introduces to the modeling techniques of Brownian dynamics and dissipative particle dynamics (DPD), its physical foundations along with recent extensions to cope with chemical reactions, the latter being used in the upcoming simulations. The chapter presents implementation details, gives a comprehensive overview of recent applications in the field of Artificial Life, and presents available software tools to perform said simulations.

Chapter 3 is dedicated to method development and rejects a claim that DPD is inapplicable in the mesoscopic range, i.e. on the length scale of micro- to millimeters, on which most lipid dynamics of our interest take place. To achieve this, scaling relations are derived for all simulation parameters which respect physical and geometrical constraints of the coarse-graining procedure, while simultaneously exploiting gauge freedoms in the choice of units. The derived scaling relations render simulation parameters scale free and it is shown that a single DPD simulation represents an entire family of physical systems that cover the entire mesoscopic regime. We note that the applicability of DPD to the entire scale is helpful but not crucial for the validity of the later simulations, as those take place on a length scale on which the modelling technique has not been questioned.

Chapter 4 presents a simple model of physically embodied, self-replicating protocells that couples a minimal autocatalytic metabolism to a simple container. The setup is shown to be able to self-replicate, and the effect of critical model parameters on the ability and performance of replication is analyzed.

In chapter 5 the minimal self-replicating system of chapter 4 is extended by an inheritable information molecule and the metabolic reaction is modified in order for the system to closely model the Los Alamos approach towards artificial protocells. The model is used to identify and analyze systemic issues over the entire life cycle of the protocell as well as the coupling of its subsystems in an integrated simulation framework. Results are compared to experimental data and findings of more detailed computational studies.

Chapter 6 collects presently unpublished simulation results on specific issues of the protocellular toy model that are not addressed in the previous chapters. Specifically, the chapter presents means to vary temperature in DPD, and analyses the effect

of temperature cycles on the melting behavior of the container and genome. The original conceptual design of the protocell is revised to prevent the occurrence of otherwise infertile offspring. Namely, this is done by choosing double stranded information carriers over single stranded one. Preliminary results indicate that this design change may cause product inhibition which may ultimately prevent successful self-replication of the system.

Parts of this thesis have been published individually as peer reviewed contributions:

- Chapter 2: H. Fellermann, Spatially resolved artificial chemistry, In: A. Adamatzky and M. Komosinski (eds.), *Artificial Life Models in Software* 2nd edition, Springer, 2009
- Chapter 3: R. Füchslin, H. Fellermann, A. Eriksson, and H.-J. Ziock, Coarse-graining and scaling in dissipative particle dynamics, *J. Phys. Chem.* 130(21), 2009
- Chapter 4: H. Fellermann, and R. Solé, Minimal model of self-replicating nanocells: A physically embodied, information-free scenario, *Philos. Trans. R. Soc.* Ser. B 362(1486):1803-1811, 2007
- Chapter 5: H. Fellermann, S. Rasmussen, H.-J. Ziock, and R. Solé, Life-cycle of a minimal protocell - a dissipative particle dynamics (DPD) study, *Artif. Life* 13(4):319-345, 2007

Acknowledgments

I want to thank my supervisor, Steen Rasmussen, for his goal-oriented guidance, and the constant support of advice and optimism that was necessary to accomplish the presented work; Hans-Joachim Ziock for uncountable comments, and careful proof-reading of the publications; and Ricard V. Solé, Anders Eriksson and Rudolf Füchslin for useful discussion during development of the simulation framework. The participants of the European Community funded PACE project and he Los Alamos sponsored Protocell Assembly project are acknowledged for providing a dynamic and stimulating interdisciplinary research atmosphere. Finally, I want to thank my mother, Helga Fellermann, for without her support this work would not have been accomplished.

Chapter 2

Spatially Resolved Artificial Chemistry

Reprinted with kind permission of Springer Science and Business Media from H. Fellermann, Spatially resolved artificial chemistry, In: A. Adamatzky and M. Komosinski (eds.), Artificial Life Models in Software 2nd Edition, Springer, 2009.

2.1 Introduction

Although spatial structures can play a crucial role in chemical systems and can drastically alter the outcome of reactions, the traditional framework of artificial chemistry is a well-stirred tank reactor with no spatial representation in mind. Advanced method development in physical chemistry has made a class of models accessible to the realms of artificial chemistry, that represent reacting molecules in a coarse-grained fashion in continuous space. This chapter introduces the mathematical models of Brownian dynamics (BD) and dissipative particle dynamics (DPD) for molecular motion and reaction. It reviews calibration procedures, outlines the computational algorithms and summarizes examplary applications. Four different platforms for BD and DPD simulations are presented which differ in their focus, features, and complexity.

The traditional mindset of artificial chemistry (AC) is a well-stirred tank reactor with possible inflow and outflow of substrates (see Dittrich et al. [74] for a review of AC). In this framework, spatial heterogeneities within a chemical solution are usually not taken into account: the focus lies on the mere presence, absence or concentration of chemicals rather than their spatial organization in the reaction vessel.

It is well known, however, that spatial structures can play a crucial role in chemical systems and can drastically alter the outcome of reactions. Such structures can be either imposed from the outside or they can be the result of the chemical reactions themselves. An example for imposed heterogeneities are upheld concentration gradients along a reaction vessel. An example for self-organized structures are self-assembled lipid aggregates which play a prominent role in molecular biology and its primordial origins and are thus of primary interest for the field of artificial life (AL) and AC.

The surplus of complexity that results from the presence of lipid structures is manifold: closed impermeable or semipermeable membranes (liposomes) compartmentalize the reaction space thereby allowing for a variety of chemical regimes at once; proteins embedded in such membranes allow for specific transport of substances (under the use

of energy even against a concentration gradient); lipid structures can increase the effective concentration of hydrophobic chemicals and alter rate constants of reactions therein (milieu effect).

Notable AL and AC models that explicitly incorporate space are the pioneering works by Varela and Maturana [75], the lattice molecular automaton (LMA) by Mayer and Rasmussen [76, 77], and a model of self-reproducing abstract cells by Ono and Ikegami [78]. The common ground of all these models is that space is treated by means of cellular automata: molecules or parts of molecules are modeled as point particles that occupy sites on a 2D or 3D lattice. Motion of these particles and reactions between them are defined by transition rules and a (possibly stochastic) updating algorithm is used to follow a system state through time.

Advanced method development in computational physical chemistry paired with the dramatic increase in computational power made a new class of models accessible to the realm of AC. While being similar to the above mentioned cellular automata in that space is modeled explicitly, these models operate off-lattice, i.e. particle positions are not confined to sites on a grid. Instead of a fixed set of transition rules, particle motion is determined by means of differential equations. Among these new methods, Brownian dynamics (BD) and dissipative particle dynamics (DPD) are the most prominent ones.

Off-lattice simulation techniques have several advantages compared to cellular automata: i) not being constrained to fixed lattice sites and having translational invariant interactions results in a significantly smaller set of required transition rules, ii) continuous particle coordinates allow to closely connect to mechanistic physical theories, iii) unconstrained motion of the particles avoids some of the artifacts found in lattice models (see e.g. [79]). However, the price one has to pay for these advantages is generally a higher computational effort both in implementation and runtime.

The remainder of the chapter is organized as follows: Sect. 2.2 presents the theoretical concepts of BD and DPD. First, the general principle behind coarse-grained off-lattice simulation methods are outlined in Sect. 2.2.1. This overview is followed by detailed discussions of individual aspects of the physics (Sect. 2.2.2 through 2.2.4) and implementation (Sect. 2.2.5). The theoretical section is concluded by a summary of recent applications in Sect. 2.2.6. Sect. 2.3 is dedicated to the presentation of software to perform off-lattice simulations with the possible capacity for chemical reactions. The software packages – ESPresSo (Sect. 2.3.1), Spartacus (Sect. 2.3.2), Smoldyn (Sect. 2.3.3) and LAMMPS (Sect. 2.3.4) – are summarized in one section each that list respective features and shows examplary simulation setups where appropriate.

2.2 Concepts

2.2.1 Basic principles of coarse-grained, off-lattice simulation techniques

BD and DPD are instances of coarse-grained modeling techniques in which the spatial structure of molecules is represented explicitly, though not in full atomistic detail. Instead, groups of atoms within a molecule are lumped together into point particles, usually called *beads*. These beads are then connected by elastic springs to form the whole molecule. Small molecules such as water, are even considered to be lumped together into a single bead by groups of 3 to 5 molecules. The number of solvent molecules per bead is referred to as the *coarse-graining* parameter. While it is possible to relate coarse-grained representations to physical molecules [80], qualitative studies

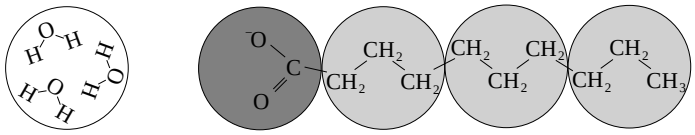


Figure 2.1: Example coarse-grained representation of decanoic acid (a fatty acid surfactant) and water. The molecules are modeled by three types of beads: one representing the hydrophobic tail of the surfactant (3 hydrocarbon groups), one representing the hydrophilic carboxyle group, and one representing water molecules of approximately the same molecular volume.

often content themselves with simply specifying functional groups like hydrophobic or charged parts of a molecule without a particular reactant in mind.

Concerning the treatment of solvent molecules, one can either explicitly represent them by beads or implicitly account for their effect on the interactions of other beads. If one is not interested in the dynamics of the solvent (e.g. hydrodynamic modes of the system), its implicit treatment can save significant computational effort, since most of the calculations are typically spent on solvent-solvent interactions.

An off-lattice simulation consists of a set of N beads in a two- or three-dimensional box. In explicit simulations, the space is considered to be densely filled, whereas in implicit simulations, the bead density will depend on the concentration of solved molecules. Since only a finite volume is simulated, boundary conditions need to be defined to determine the outcome of collisions with the wall. Most common are periodic boundary conditions, but closed, reflecting boundaries or a combination of the two can also be found in the literature. Each bead has a position, velocity, and a type (for an example configuration, see Fig. 2.2). The type of the bead determines its mass and its interactions with other beads. The motion of the beads follows *Newton's Second Law of Motion* according to which a particle accelerates proportional to the force that acts on it and inversely proportional to its mass. Written as an ordinary differential equation (ODE) in bead position and velocity, the law reads:

$$\begin{aligned} \dot{\mathbf{r}}_i(t) &= \mathbf{v}_i(t) \\ \dot{\mathbf{v}}_i(t) &= \frac{1}{m_i}\mathbf{F}_i(t), \end{aligned} \tag{2.1}$$

where $\mathbf{r}_i$ is the position, $\mathbf{v}_i$ the velocity, and m_i the mass of bead i, dots denote time derivatives (instantaneous changes in time). $\mathbf{F}_i$, the force that acts on bead i, collects all bead interactions.

If all atoms were simulated individually (as is done in fully atomistic Molecular Dynamics (MD)), the force could be expressed as the negative gradient of some potential ϕ_i such that Eq. 2.1 constantly transforms kinetic into potential energy and vice versa. But when individual atoms are lumped together into coarse-grained beads, energy is also exchanged with the internal energy of these beads. Instead of explicitly accounting for the internal energy of each bead, coarse-grained methods employ a *Langevin* formalism to express the energy exchange with internal degrees of freedom: the force $\mathbf{F}_i$ is expressed by three additive components, a conservative one ($\mathbf{F}_i^{\mathrm{C}}$), a dissipative one ($\mathbf{F}_i^{\mathrm{D}}$), and a random one ($\mathbf{F}_i^{\mathrm{R}}$):

$$\mathbf{F}_i = \mathbf{F}_i^{\mathrm{C}} + \mathbf{F}_i^{\mathrm{D}} + \mathbf{F}_i^{R}. \tag{2.2}$$

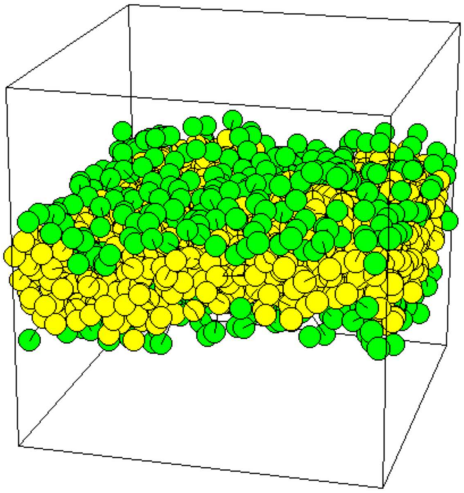


Figure 2.2: Example configuration of a coarse-grained simulation: the image shows a bilayer membrane composed of **HTT** trimers (green-yellow surfactants) in water (not shown).

The dissipative force models friction by which kinetic energy is dissipated into successively less coherent motion, thereby removing energy from the explicitly represented mesoscale into the assumed underlying microscale (internal energy of the beads). This energy flow is counteracted by a random force, that models the effect of microscopic thermal noise on the mesoscale (Brownian motion). As the name suggests, $\mathbf{F}_i^{\mathrm{R}}$ involves random variables, turning Eq. 2.1 into a stochastic differential equation, such that the change in system energy can be captured by an Uhlenbeck-Ornstein process. Combined, the dissipative and random forces act as a thermostat to regulate the temperature, i.e. the kinetic energy of the explicitly modeled mesoscale.

Several proposals have been made to combine off-lattice simulations with chemical reaction kinetics. To enable this, a set of chemical reaction rules is added to the system description that defines allowed transitions between bead types. To account for these reactions, the numerical integrator is intertwined with a stochastic process that applies bead transformations to the reaction vessel: within each time step, first the positions and velocities of all beads are updated, and second, reaction rules are applied subsequently to all beads in the system.

To summarize, the coarse-grained models described in this chapter draw on a set of (possibly interconnected) beads whose motion is governed by bead-bead interactions expressed by potential functions and by the action of a thermostat. Additionally, transition rules between beads can be defined to describe chemical reactions. The following sections will detail on each of these aspects.

A simple example of the overall setup is given in Fellermann and Solé [81, chapter 4 of this work] where the system consists of beads of three molecules: water, fatty acid surfactants, and oily fatty acid esters. Water is represented explicitly by beads of type **W**, esters are represented as dimers of two interconnected **T** beads, and surfactants are modeled as a **T** bead with a surfactant head **H** bead attached to it. The potential interactions are chosen such that the model qualitatively reproduces the phases of these binary and ternary systems. A catalytic reaction is defined by which ester molecules

are transformed into surfactant molecules in the vicinity of other surfactant molecules:

$$\mathbf{TT} + \mathbf{HT} \longrightarrow 2\mathbf{TT} \tag{2.3}$$

This reaction models the hydrolysis of the ester bond under the assumption that the produced alcohol is small enough to be neglected.

In this setup, the ester forms oil droplets in water which are coated by surfactants. The size distribution of these droplets is determined by the water/oil/surfactant ratio. The relatively slow reaction constantly transforms ester into new surfactants, thereby changing the oil/surfactant ratio of the system. The change in concentrations is accompanied by a shift in surface to volume ratio of the droplets which respond by a shape change from spherical to elongated, rod-like aggregates. Once a critical threshold ratio is reached, the oil core is not sufficiently big to stabilize the aggregate which divides into two spherical aggregates. When ester is supplied constantly, the growth and division process continues and leads to exponential growth of the aggregates.

2.2.2 Interaction potentials

In the picture of Newtonian mechanics, each bead has a potential energy that results from its interactions with other beads. This energy corresponds, for example, to pressure in thermally fluctuating fluids, but could also have other (e.g. electrostatic origins). Alternatively, one might say that the bead "feels" the potential energy of its neighborhood and responds to it by moving away from high energy states – known as the principle of least constraints.

As commonly done in particle based simulations, BD and DPD assume that the potential of a bead i can be expressed as the sum of pairwise interactions with neighboring beads:

$$V_i = \sum_{j \neq i} V_{ij}(\mathbf{r}_i, \mathbf{r}_j).$$

Since the space is assumed to be isotropic, the potential between particles i and j can only depend on their distance rather than their absolute position. Using

$$V_{ij}(\mathbf{r}_i, \mathbf{r}_j) = \phi_{ij}(r_{ij}),$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$, the conservative force in Eq. 2.2 becomes

$$\mathbf{F}_i^{\mathrm{C}} = -\boldsymbol{\nabla} \sum_{i \neq j}^{N} \phi_{ij}(r_{ij}). \tag{2.4}$$

For further computational convenience, ϕ_{ij} is usually truncated after a certain distance r_{c}, such that $\phi_{ij}(r) = 0$ for $r \geq r_{\mathrm{c}}$.

It has to be pointed out that the potential functions used in coarse-grained simulation techniques must not be taken for mere mechanical energies as in atomistic (MD) simulations. Instead, the potential functions used in coarse-grained simulations are rather to be understood as a free energy that captures systemic properties of the system such as heat, pressure, pH, or the concentration of cosolvents [82]. This is particularly important for the study of entropy-driven processes underlying lipid systems, since entropy is included in the free energy. Although there is no closed theory to derive the functional form of the potential energy employed in coarse-grained simulations from first principles, it is still possible to relate the potential functions to structure properties of the system under consideration. For example, there is a one to one mapping

between the potential function and the radial distribution function which expresses the average probability for two beads to be separated by a certain distance. Measuring the radial distribution function of detailed atomistic MD simulations allows to construct energy functions of a coarse-grained representation thereof (see Lyubartsev et al. [83] and references therein). A further difficulty arises from the fact, that the dynamics in coarse-grained simulations should represent a time-average of the assumed underlying atomistic motion, since the fast degrees of motion like most vibrational modes of covalent bonds are meant to be removed in the coarse-graining process and instead comprised in the noise term of the thermostat.

In general, the spatial and temporal averaging of the coarse-graining process motivates the use of much smoother interaction potentials than the ones commonly applied in atomistic MD simulations. In particular, they typically do not possess a singularity at $r = 0$, meaning that two beads are allowed to sit right on top of each other (although under an energy "penalty" that may render this event practically impossible). This can be justified by the observation that (i) the bead position merely identifies the center of mass of a collection of molecules, and (ii) even if the bead represents a single molecule or part thereof, it denotes the average position of the molecule during some short time interval; for both of these mean positions it is perfectly legal for two positions to coincide. Whichever justification one might assent to, the resulting *soft core* potential allows to run the numerical integrator of Eq. 2.1 with a significantly wider time step compared to functional forms that include a singularity.

Given these difficulties, many studies (that treat the solvent explicitly) content themselves with simple potentials of the form

$$\phi_{ij}(r) = \begin{cases} \frac{1}{2} a_{ij} \left(1 - \frac{r}{r_c}\right)^2 & \text{if } r < r_c \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

where $a_{ij} > 0$ denotes the mutual repulsion strength between two beads of type i and j. Note that there is no theoretical foundation for this function other than being the most simple confined function with continuous derivative. It has to be pointed out, however, that Eq. 2.5 relies on the explicit treatment of solvent, since the absence of an energy minimum would not allow for the formation of structures.

Having fixed the functional form of the potential, the matrix (a_{ij}) is the premier place for model calibration. A standard calibration procedure to mesoscopic observables has been suggested by Groot and Rabone [80, 84, 85] (see however Füchslin et al. for issues on its scaling behavior [86, chapter 3 of this work]). Generally, the procedure starts by matching the diagonal interaction parameters a_{ii} to the compressibility of the physical fluid. The off-diagonal elements $a_{ij}(i \neq j)$ express the mutual solubility of substances. and are calibrated in a later step to parameters obtained from mixing theories, in particular the Flory-Huggins theory of polymer mixing where mutual interaction parameters are derived from free energy considerations in a lattice model of polymers. For specific systems, other mesoscopic observables have been suggested for calibration (e.g. [87, 88, 89, 85]).

Additional terms can be added to the "ground" potential if needed. Most prominently, mechanical potentials between bonded beads have been use to model extended molecules. Venturoli and Smit [90] were the first to introduce harmonic spring potentials into DPD simulations, where bonded beads feel the additional potential

$$\phi_{ij}^{\mathrm{S}} = k_{\mathrm{S}} \left(r_{ij} - r_{\mathrm{S}}\right)^2 ,$$

with r_{S} being the optimal bond distance and k_{S} the spring constant. Shillcock and Lipowsky [91] have added a three-body angular potential to model stiffness in extended polymer chains. Higher order potentials used in MD simulations (e.g. torsion potentials) are barely needed in coarse-grained simulations due to the simplistic representation of molecules. Other notable add-ons are Coulomb forces for electrostatic interactions [92], gravity [93], and shear forces [94, 95].

2.2.3 Thermostats

As pointed out in Sect. 2.2.1, thermostats in coarse-grained simulations are an integral part of the system description: whereas MD simulations utilize thermostats merely to keep the system at constant temperature, and generally try to minimize their impact on simulation results, coarse-grained simulations interpret the energy exchange with a heat bath as physically meaningful. As a consequence, coarse-grained models employ entirely different thermostats than the ones found in MD simulations.

The theoretical foundation of thermostats for coarse-grained simulations is the Langevin formalism for diffusive motion [96]: a spherical particle in a steady medium will experience a viscous drag proportional to its velocity and stochastic kicks from other particles randomly hitting it. In the presence of a possible external force $\mathbf{F}$, the equation of motion reads

$$\dot{\mathbf{v}}(t) = \frac{1}{m}\left(\mathbf{F}(t) - \gamma\mathbf{v}(t) + \boldsymbol{\xi}(t)\right), \tag{2.6}$$

where $\gamma > 0$ is a friction coefficient and $\boldsymbol{\xi}$ an uncorrelated random vector with zero mean and finite variance: $\left\langle \boldsymbol{\xi}_i(t), \boldsymbol{\xi}_j(t')\right\rangle = 2\sigma\delta_{ij}\delta(t-t')$. In the context of Eq. 2.2, the first term of the sum represents $\mathbf{F}^{\mathrm{C}}$, the second one $\mathbf{F}^{\mathrm{D}}$, and the last one $\mathbf{F}^{\mathrm{R}}$. The equilibrium temperature is given by γ and σ and resolves to σ/γ. Off-lattice models that employ this thermostat are commonly referred to as Brownian dynamics (BD). They are particularly suited for implicit simulations, since the effect of the solvent is already accounted for in the thermostat. It has to be emphasized, however, that Eq. 2.6 models motion in a steady medium, as the friction is proportional to the absolute velocity of the bead, rather than the velocity relative to the surrounding medium. Thus, the dynamics of the system is purely diffusive and neglects hydrodynamic modes.

To incorporate hydrodynamics into coarse-grained simulations, Hoogerbrugge and Koelman introduced the method of dissipative particle dynamics (DPD) [97] whose thermostat has become increasingly popular. The DPD thermostat decomposes all forces into pairwise contributions:

$$F_i = \sum_{j\neq i}\mathbf{F}_{ij} = \sum_{j\neq i}\mathbf{F}^{\mathrm{C}}_{ij} + \mathbf{F}^{\mathrm{D}}_{ij} + \mathbf{F}^{\mathrm{R}}_{ij}, \tag{2.7}$$

which are required to be central

$$\mathbf{F}_{ij} = -\mathbf{F}_{ji} \quad \text{and} \quad \mathbf{F}_{ij} \propto \hat{\mathbf{r}}_{ij}, \tag{2.8}$$

where $\hat{\mathbf{r}}_{ij}$ is the unit vector pointing from bead j to i. The centrality asserts that the linear and angular momentum of the system is preserved, therefore allowing to analyze hydrodynamic flows. In fact, it can be shown that the DPD thermostat implements a numerical solver for the Navier-Stokes equations [98]. With these requirements, the

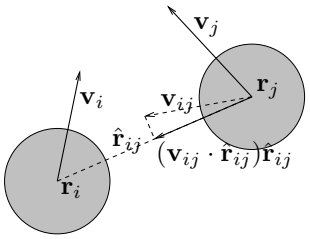


Figure 2.3: Schematic of the dissipative force (friction) in DPD. The force acts central on the line given by the bead positions $\mathbf{r}_i$ and $\mathbf{r}_j$ (parallel to $\mathbf{r}_{ij}$). Its magnitude depends on the relative velocity $\mathbf{v}_{ij}$: friction is maximal when beads approach each other directly and zero when they move parallel.

only (local and memory-free) way to achieve an Uhlenbeck-Ornstein process whose equilibrium distribution is a Gibbs ensemble is to set [99]

$$\begin{aligned} \mathbf{F}_{ij}^{\mathrm{D}} &= -\frac{\sigma^2}{2k_bT}\omega^2(r_{ij})\left(\mathbf{v}_{ij}\cdot\hat{\mathbf{r}}_{ij}\right)\hat{\mathbf{r}}_{ij}, \\ \mathbf{F}_{ij}^{\mathrm{R}} &= \sigma\omega(r_{ij})\zeta_{ij}\hat{\mathbf{r}}_{ij}, \end{aligned}$$

where r_{ij} is the Euclidean distance between beads i and j, and $\mathbf{v}_{ij}$ is the relative velocity between bead j and i. ζ_{ij} is an uncorrelated random variable with zero mean, Gaussian statistics and a variance of $\Delta t^{1/2}$ for the numerical time step Δt. In order to preserve linear momentum, it has to hold that $\zeta_{ij} = \zeta_{ji}$. σ is the friction coefficient of the medium (related to its Reynolds number), k_bT defines the energy unit where T denotes the temperature in Kelvin and k_b is the Boltzmann constant. ω is a dimensionless weighing function which is not specified by the general formalism. Most studies employ a weight function similar to the soft core interaction potential:

$$\omega = \begin{cases} \left(1-\frac{r}{r_{\mathrm{c}}}\right)^2 & \text{if } r < r_{\mathrm{c}} \\ 0 & \text{otherwise.} \end{cases}$$

While it is known that ω has impact on the overall system dynamics, like viscosity [100] and temperature conservation [101], no proposal has yet been made on how to derive the weight function from first principles.

The scalar product in Eq. 2.9 ensures that $\mathbf{F}^{\mathrm{D}}$ is maximal when two beads approach each other and zero, when the particles move parallel (see Fig. 2.3). This leads to the alignment and collaborative motion of nearby beads. As a result, the pairwise coupling of the dissipative and random forces in DPD results in faster dynamics than the ones observed in BD simulations. However, the accelerated dynamics of the DPD thermostat are actually an artifact of the method: in DPD mass diffusion is too high when compared to travel of momentum. The Schmidt number (the ratio of kinematic viscosity vs. self-diffusion coefficient) is about three orders of magnitude lower in the DPD simulation of water than in the real system [84, 100].

2.2.4 Chemical Reactions

Probably the first approach to extend spatially resolved reactor vessels with chemical reactions was the work by Ono [78] who incorporated a stochastic process for reactions

into a BD simulation. This approach has later been put on a more rigorous theoretical foundation (the Smoluchowski model for diffusion-limited reactions) by Andrews and Bray [102] and Monine and Haugh [103]. Similar approaches are the work of Buchanan et al. [104, 105], Fellermann et al. [88, 81, chapters 4 and 5 of this work] and Liu et al. [106], who extended DPD by similar means to incorporate reactions.

Most chemical reactions can be classified as uni- or bimolecular reactions – depending on the number of molecules that participate as reactants. Higher order reactions barely occur in nature due to the unlikelihood of three molecules hitting each other concurrently. Following this formalism, inflow of chemicals can be regarded as a zeroth-order reaction that has no educt, whereas outflow of chemicals can be regarded as a first-order reaction that has no product. Reactions can be further classified as syntheses or analyses, depending on whether covalent bonds are formed or broken by the reaction. The formation and braking of covalent bonds can result in a change of physico-chemical properties of the molecular species. In the coarse-grained representation of BD and DPD, this is expressed as a type change of the reacting beads. Depending on the coarse-graining level and representation of molecules, bond braking and formation might even happen completely below the resolution of the model such that chemical reactions are mere type transformations that do not affect explicitly modeled covalent bonds. In the formalism presented here, chemical reactions only occur between beads that represent a single molecule or part of a molecule – no attempt has been made yet to define chemical reactions between beads that group together several molecules. Examples of possibles reactions are

	$\xrightarrow{k}$	X	inflow (zeroth-order)
A	$\xrightarrow{k}$		outflow (first-order)
A	$\xrightarrow{k}$	X	molecular reconfiguration (first-order)
AB	$\xrightarrow{k}$	$X + Y$	unimolecular analysis (first-order)
$A + B$	$\xrightarrow{k}$	XY	bimolecular synthesis (second-order)

Here, A, B, X, and Y denote beads that may be part of an extended molecule. Concatenated symbols refer to bonded beads. For simplicity, all reactions are written as irreversible reactions. Reversible reactions can be represented as pairs of irreversible reactions, where each member represents one of the two directions.

Evidently, bimolecular reactions occur only when the two reagents are in close vicinity. For diffusion limited reactions, i.e. in the absence of an activation energy barrier, this distance is given by the sum of the molecular radii of the two reagents. If there is an activation barrier, one possibility is to define an effective binding radius smaller than the sum of molecular radii. The size of this binding radius can be related to the effective reaction rate by the Smoluchowski equation

$$r_{\mathrm{b}} = \frac{k}{4\pi(D_{\mathrm{A}} + D_{\mathrm{B}})}, \tag{2.9}$$

where D_A and D_B are the diffusion coefficients of the two reagents and k is the effective (measured) reaction rate [102]. Following Collins and Kimball [107], one can alternatively define a probability for reactions to occur within the unaltered reaction radius. Likewise, unimolecular reactions either occur with a certain probability

$$1 - e^{-k\Delta t},$$

or – if explicit bond breaking is involved in the reaction – once the reacting beads separate by more than a certain unbinding radius $r_\mathrm{u} > r_\mathrm{b}$. r_u can be related to r_b by the equation:

$$r_\mathrm{u} = \frac{r_\mathrm{b}}{1 - k/k_\mathrm{r}}, \tag{2.10}$$

where k_r is the rate of geminate recombination [102]. When a pair of bond forming and breaking reactions is used to model a reversible reaction, care must be taken that the reactants are separated by more than r_b after bond breaking. Note that (2.9) and (2.10) are only valid for numerical time steps Δt smaller than the root mean square displacement of the reacting particles. Andrews and Bray [102] give corrected equations for larger simulation time steps, as well as a comparison of the deterministic and stochastic method. Zeroth-order reaction do not depend on any educt concentration, and their occurrence is simply determined by comparing the reaction rate to a Poisson distributed random number with average $k\Delta t$ for the simulation time step Δt.

Ono [108] and Fellermann et al. [88, 81, chapters 4 and 5 of this work] use an extended variant of the Collins and Kimball scheme that accounts for the presence of nearby catalysts. Other than modeling the exact reaction mechanism by which the catalyst participates in the reaction scheme, it is assumed that the mere presence of catalytic molecules enhances the effective reaction rate. The equation to calculate this effective rate is

$$k = k_i + \sum_C k_C \left(1 - \frac{r}{r_\mathrm{cut}}\right) \tag{2.11}$$

where r is the distance between the catalyst and any of the reactants and the sum runs over all catalysts C with $r < r_\mathrm{cut}$.

All of the stochastic approaches assume that reactions occur independently, which is violated when a single bead might react with one of several nearby other reagents. A proper mathematical framework to deal with this stochastic independence would be provided by the Master equation [96]. However, its computation is too expensive to be redone in each updating step, which is why current studies commonly ignore the problem of interdependence and merely try to avoid it by using small reaction and catalyst rates and constantly reshuffling the list of reactions and reagents to avoid biases.

Having defined the set of possible reaction rules, the integrator for the equations of motion is intertwined with a stochastic process such that in each time step particle motion and chemical reactions are taken care of sequentially.

Further method development needs to be done in connecting the potential, kinetic, and chemical energy reservoirs: the change in internal (chemical) energy associated with bond formation and breaking and change of potential energy due to bead type transformation would need to be accounted for in a velocity change of the reacting beads leading to local heating (A first step in this direction has been taken by Yingling and coworkers [109]). This would e.g. allow to capture effects that rely on activation energy barriers.

2.2.5 Updating schemes and spatial organization

At the heart of all BD and DPD simulators lies an integrator for the stochastic differential equation 2.1. Owing to the special form of the Newtonian dynamics, an inexpensive high-order integrator can be derived from two Taylor expansions:

$$\begin{aligned}\mathbf{r}_i(t+\Delta t) &= \mathbf{r}_i(t) + \dot{\mathbf{r}}_i(t)\Delta t + \frac{1}{2}\ddot{\mathbf{r}}_i(t)\Delta t^2 + \frac{1}{6}\dddot{\mathbf{r}}_i(t)\Delta t^3 + \mathcal{O}(\Delta t^4)\\ \mathbf{r}_i(t-\Delta t) &= \mathbf{r}_i(t) - \dot{\mathbf{r}}_i(t)\Delta t + \frac{1}{2}\ddot{\mathbf{r}}_i(t)\Delta t^2 - \frac{1}{6}\dddot{\mathbf{r}}_i(t)\Delta t^3 + \mathcal{O}(\Delta t^4).\end{aligned}$$

Adding these two equations and substituting $\ddot{\mathbf{r}}_i(t) = \mathbf{a}_i(t)$ leads to a very simple forth-order integrator known as *Verlet* integrator [110]:

$$\mathbf{r}_i(t+\Delta t) = 2\mathbf{r}_i(t) - \mathbf{r}_i(t-\Delta t) + \mathbf{a}_i(t)\Delta t^2 + \mathcal{O}(\Delta t^4).$$

Verlet-based integrators outperform standard procedures like Runge-Kutta not only because of their computational inexpensiveness but also because they preserve phase space volume in conservative systems, which reduces artifacts such as gradual temperature increase.

If velocities are to be known explicitly, e.g. to compute the kinetic energy of the system, the substitution $\mathbf{v}_i(t) = \left(\mathbf{r}_i(t) - \mathbf{r}_i(t-\Delta t)\right)/\Delta t + \frac{1}{2}\mathbf{a}_i(t)\Delta t + \mathcal{O}(\Delta t^2)$ leads to a variant called *Velocity Verlet* algorithm

$$\mathbf{r}_i(t+\Delta t) = \mathbf{r}_i(t) + \mathbf{v}_i(t)\Delta t + \frac{1}{2}\mathbf{a}_i(t)\Delta t^2 \tag{2.12}$$

$$\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t) + \frac{1}{2}\left(\mathbf{a}_i(t) + \mathbf{a}_i(t+\Delta t)\right)\Delta t \tag{2.13}$$

which is forth order in positions $\mathbf{r}_i$ and second order in velocities $\mathbf{v}_i$. This is a commonly used integrator for BD simulations. There is a subtlety, however, when accelerations $\mathbf{a}_i$ do not only depend on positions but also on velocities, as is the case for $\mathbf{F}^{\mathrm{D}}$ in the DPD equations. Eq. 2.13 then becomes

$$\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t) + \frac{1}{2m_i}\left(\mathbf{F}_i(\mathbf{r}(t),\mathbf{v}(t)) + \mathbf{F}_i(\mathbf{r}(t+\Delta t),\mathbf{v}(t+\Delta t))\right), \tag{2.14}$$

where the term $\mathbf{v}_i(t+\Delta t)$ appears in both sides of the equation. To overcome this problem, Groot and Warren [84] have suggested a predictor-corrector like integrator which has become the de facto standard in DPD simulations. In their integrator, the positional update is unchanged (Eq. 2.12). To compute the velocities, the algorithm first makes a prediction (Eq. 2.15) followed by a correction based on the force field of the predicted state (Eq. 2.16):

$$\tilde{\mathbf{v}}_i(t+\Delta t) = \mathbf{v}_i(t) + \frac{\lambda}{m_i}\mathbf{F}_i(\mathbf{r}(t),\mathbf{v}(t)) \qquad \lambda \in [0,1] \tag{2.15}$$

$$\mathbf{v}_i(t+\Delta t) = \mathbf{v}_i(t) + \frac{1}{2m_i}\left(\mathbf{F}_i(\mathbf{r}(t),\mathbf{v}(t)) + \mathbf{F}_i(\mathbf{r}(t+\Delta t),\tilde{\mathbf{v}}(t+\Delta t))\right). \tag{2.16}$$

If $\mathbf{F}_i$ does not depend on $\mathbf{v}$, the original Velocity Verlet integrator is recovered for $\lambda = 0.5$. Due to the stochastic nature of the force field, the order of the integrator is unclear. There is no imperative on how λ should optimally be chosen. Reported values differ between $\lambda = 0.5$ [111] and $\lambda = 0.65$ [84]. Thus, λ clearly depends on other systems parameters and needs to be fine-tuned to the system under consideration.

Subsequent studies [112, 111, 113] have identified artifacts of the Verlet based algorithms, predominantly in the radial distribution function of the system [113]. Building upon work by Pagonabarraga et al. [114], Vattulainen et al. [113] have proposed a

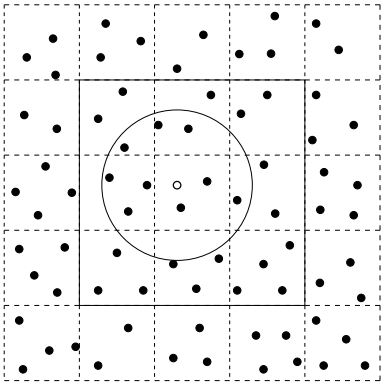


Figure 2.4: Internal partitioning of a two dimensional space (dashed grid) filled with beads (black dots). The size of the cells is at least the cut-off radius of all involved interactions (radius of the solid circle). To find all interaction partners for a given bead (white dot) it is sufficient to consider only beads within a 3×3-Moore neighborhood around the cell that contains the bead (solid square). Note that the grid representation is only internal and does not restrict the actual position of beads within each cell.

more elaborate so-called *self-consistent* integrator in which the system temperature is constantly measured and compared to its target value. The deviance is used to fine-tune the dissipation rate of the system. In general, the performance of DPD integrators is still an active area of research due to the stochastic nature of the interactions.

Beside well-suited integrators, simulators for coarse-grained simulations need to provide efficient means to access neighbors of beads. Fast neighbor look-up is essential for the computation of $\mathbf{F}^{\mathrm{C}}$ in the case of BD (Eq. 2.4) and $\mathbf{F}^{\mathrm{C}}$, $\mathbf{F}^{\mathrm{D}}$, and $\mathbf{F}^{\mathrm{R}}$ in the case of DPD (Eq. 2.7). If all particles where naively held in a simple list or array, look-up time would scale quadratic with the number of particles. By taking advantage of the limited cutoff range r_{c} for all forces, careful bead management can reduce the number of look-ups to scale linear with the number of beads. To achieve this, an algorithm called *domain decomposition* or *linked lists* is commonly used: the space is partitioned in cells with a side length of at least the cutoff radius r_{c}. Each cell holds a list of particles it contains, as well as information about neighboring cells (dashed grid in Fig. 2.4). For each bead, neighbor look-up can now be restricted to the 3×3-Moore neighborhood of the cell that holds the current bead (solid square in Fig. 2.4). If bead interactions are symmetric (e.g. due to Eqs. 2.4 and 2.8) the number of cells to check can even be reduced by one half. With this management, an estimated 62.8% of all tested pairs will be within the cutoff radius (solid circle in Fig. 2.4) and 29.9% in three dimensions. It should be emphasized that the grid representation of the space is only an algorithmic organization principle and does not confine the actual position of beads within each cell. Implementation can either use one linked list for all beads or separate lists for each bead type, which can improve the performance when bead interactions are highly type specific (as in case of chemical reactions).

2.2.6 Applications

Pure BD and DPD models that do not incorporate chemical reactions have been used extensively in areas as diverse as rheology, material sciences, soft matter studies, and

molecular biology. Concerning the latter, the study of lipid strctures such as membranes and vesicles, their self-assembly, fission, fusion, rapture, and dissolution has arrested particular attention [80, 115, 54, 90, 116, 117]. The recent incorporation of chemical reactions into these models has initiated several simulations in the area of artificial chemistry and artificial life.

Gazzola et al. [105] have used reactive DPD to study the catalytic effect of self-assembled lipid structures due to effective upconcentration of hydrophobic reagents. Simulating abstract chemistries, they observed that the affinity of reagents to reside either in the solvent or in the interior of self-assembled lipid structures (micelles), alters the effective reaction networks that emerge in the reaction vessel.

Ono [108] and Fellermann and Solé [81, chapter 4 of this work] have studied the self-assembly, growth, and replication of minimal life-like entities: lipid aggregates equipped with a minimal artificial metabolism. Ono presents two-dimensional BD simulations of vesicular structures driven by a catalytic turnover of precursor molecules, Fellermann and Solé performed three-dimensional DPD simulations of surfactant-coated oil droplets that replicate by means of an auto-catalytic nutrient turnover (described in Sect. 2.2.1).

Complementary to the above work are qualitative simulations on non-enzymatic replication of short genetic polymer sequences. Bedau et al. [118] have analyzed the ability of complementary polymer strands to spontaneously replicate in solution, whereas Fellermann et al. [88, chapter 5 of this work] have studied a similar process at the oil-water interface of the aforementioned lipid aggregates. Taken together, these works present integrated simulations of the spontaneous emergence of artificial life-like aggregates, or “protocells” [65].

2.3 Available Software and Tools

Several BD and DPD simulators exist both in commercial as well as open source applications. The following list is not meant to be complete, but tries to identify those programs that stand out for maturity, extensibility, or interoperability. The programs are similar in that they use configuration files to set up a simulation. The configuration file is sent to a non-interactive simulator that calculates the system trajectory and writes out the result of various analyzers.

2.3.1 ESPresSo

ESPResSo is a fairly developed and feature-rich open source simulation package for soft matter research (BD and DPD among others) developed by H. Limbach et al. [119] and licensed under the GNU General Publishing License (GPL). The software is available for Windows, Unix/Linux, and Mac OS. ESPresSo is written in C and can operate in parallel mode via MPI, and can be controlled via Tcl scripts [120]. Online visualization is delegated to VMD, an independent open source software for molecular visualization [121]. Out of the box, ESPresSo does not include chemical reactions.

In ESPresSo, simulations are set up via configuration scripts that define system size, boundary conditions, bead types, interactions, an initial condition, analyzers, a.s.o. These configuration files are actual Tcl scripts that offer powerful means for arbitrarily complex simulation setups. An example script to set up and run a simulation is shown in Fig. 2.5. The system consists of 1780 **W** beads (water), 600 **TT** dimers (oil), and 10 **HT** dimers (surfactants) in a box of size 10^3. Bead interaction parameters are:

$a_{\mathbf{W,W}} = 25, a_{\mathbf{W,H}} = 15, a_{\mathbf{W,T}} = 80, a_{\mathbf{H,H}} = 35, a_{\mathbf{H,T}} = 80, a_{\mathbf{T,T}} = 15$. The DPD thermostat is chosen with a temperature of 1 and friction parameter 4.5. Integration step size and number of steps is set to 0.01 and 1000, respectively, and the system is finally integrated. Various analyzers could further be incorporated.

The package offers support for various boundary conditions, a variety of potentials for bonded and non-bonded short range interactions, electrostatics (Coulomb forces), rigid bonds, as well as bond-angle and dihedral interactions. ESPresSo also provides tabulated interactions to incorporate interaction potentials measured e.g. from more detailed interactions (see Sect. 2.2.2).

ESPresSo comes with a variety of predefined analyzers for common observables. Among the most common ones are

- statistical physics
  - energies (total, kinetic, Coulomb, non-bonded, bonded)
  - pressure (total, ideal, Coulomb, non-bonded, bonded)
  - stress tensor (total, ideal, Coulomb, non-bonded, bonded)
  - mean square displacement
- distribution and correlation functions
  - radial distribution function
  - structure factor
  - van-Hove autocorrelation function
  - aggregate size distribution
  - free volume distribution
- analyzers for polymer chains
  - end to end distance
  - radius of gyration

In addition, custom analyzers can be defined with ease. Analyzers are ordinary functions that can be used anywhere in the configuration script. ESPresSo provides a few built-in commands for statistics (averaging, errors, etc.) and plotting, the latter by delegation to gnuplot [122].

Owing to the flexibility of the Tcl based design, it is easy to incorporate chemical reactions into ESPreSso. Fig. 2.6 shows an exemplary algorithm that implements the reaction (2.3) based on a simplified version of Eq. 2.11. The implementation is only meant for the purpose of illustration: for a productive setup, the algorithm should be written in a compilable computer language and wrapped into Tcl.

2.3.2 Spartacus

Spartacus is an experimental open source framework for BD and DPD simulations developed by the author and licensed under the GNU General Publishing License (GPL) The software is available for Mac Os, Linux. Spartacus is similar in function to EsPresSo, but focuses on chemistry rather than physics. It incorporates all algorithms for chemical reactions that have been described in Sect. 2.2.4. The framework consists

```
# System parameters
set box_l    10.            ;# box size
set n_W     1780            ;# number of W beads
set n_TT     600            ;# number of TT dimers
set n_HT      10            ;# number of HT dimers
set n_total 3000            ;# total number of particles

# Integration parameters
set cut 1.0                 ;# cut off range r_c
set int_steps    100        ;# micro-integration steps (between analysis)
set int_n_times  1000       ;# macro-integration steps

setmd time_step 0.01
setmd skin      0.4
setmd box_l $box_l $box_l $box_l
thermostat dpd 1.0 4.5 $cut

# Interaction setup
inter 0 harmonic 100 0.75

inter 0 0 soft-sphere 25 2 $cut 0
inter 0 1 soft-sphere 80 2 $cut 0
inter 0 2 soft-sphere 15 2 $cut 0
inter 1 1 soft-sphere 15 2 $cut 0
inter 1 2 soft-sphere 80 2 $cut 0
inter 2 2 soft-sphere 35 2 $cut 0

# Particle setup

# water beads
for {set i 0} { $i < $n_W } {incr i} {
    set posx [expr $box_l*[t_random]]
    set posy [expr $box_l*[t_random]]
    set posz [expr $box_l*[t_random]]
    part $i pos $posx $posy $posz type 0
}

# TT dimers
for {} { $i < [expr $n_W+$n_TT] } {incr i 2} {
    set posx [expr $box_l*[t_random]]
    set posy [expr $box_l*[t_random]]
    set posz [expr $box_l*[t_random]]
    part $i pos $posx $posy $posz type 1
    part [expr $i+1] pos $posx $posy [expr $posz+0.8] type 1 bond 0 $i
    part $i bond 0 [expr $i+1]
}

# HT dimers
for {} { $i < [expr $n_W+$n_TT+$n_HT] } {incr i 2} {
    set posx [expr $box_l*[t_random]]
    set posy [expr $box_l*[t_random]]
    set posz [expr $box_l*[t_random]]

    part $i pos $posx $posy $posz type 1
    part [expr $i+1] pos $posx $posy [expr $posz+0.8] type 2 bond 0 $i
    part $i bond 0 [expr $i+1]
}

# Integration
set j 0
for {set i 0} { $i < $int_n_times } { incr i} {
    integrate $int_steps
    # [...] do some analysis here
}
```

Figure 2.5: Example Tcl script to control ESPresSo for the example system of Sect. 2.2.1. See text for explanation.

```
proc react {reaction}
{
    # reaction is { educt product rate catalyst cat_rate cat_range}
    # educt and product are list of 2 elements, specifying bead types


    set educt     [lindex $reaction 0]
    set product   [lindex $reaction 1]
    set rate      [lindex $reaction 2]
    set catalyst  [lindex $reaction 3]
    set cat_rate  [lindex $reaction 4]
    set cat_range [lindex $reaction 5]


    foreach {part_a} [part]
    {
        # iterate over all possible educts
        set a_id    [lindex $part_a 0]
        set a_type  [lindex $part_a 6]
        set a_bonds [lindex $part_a 33]
        if { $a_type != [lindex $educt 0] } { continue }
        set b_id    [lindex [lindex $a_bonds 0] 1]
        set b_type  [part $b_id print type]
        if { "$a_type $b_type" != $educt } { continue }


        # compute effective reaction rate
        set k $rate
        foreach {cat} [analyze nbhood $a_id $cat_range]
        {
            set c_type [part $cat print type]
            if { $c_type != $catalyst } { continue }
            set k [expr $k+$cat_rate]
        }


        # exchange bead types if a reaction occurs
        if { [expr $k*[setmd time_step]] > [t_random] }
        {
            part $a_id type [lindex $product 0]
            part $b_id type [lindex $product 1]
            continue
        }
    }
}


# [...]
# This part replaces 'integrate $int_steps' from the example in Fig. 4

    for {set t 0} { $t < $int_steps } { incr t}
    {
        integrate 1
        react "{1 1} {1 2} 0 2 0.5 $cut"
    }
# [...]
```

Figure 2.6: Algorithm for reconfiguration reactions in ESPresSo that implements a simplified version of Eq. 2.11 (not implementing the linear decrease with catalyst distance). The function `react` is called with 6 arguments – the reactant, product, spontaneous reaction rate, catalyst bead type, the catalyst's rate enhancement, and range. Reactant and product are assumed to be Tcl lists of 2 elements. Together with the system setup in Fig. 2.5, this implements the example given in Sect. 2.2.1.

```
# set system size, bead number, and density
from simulation.grid import Space3D
size = 10
space = Space3D


# define bead types
from simulation.chemistry import Particle
class W(Particle) :
    pass


class H(Particle) :
    color = "green"


class T(Particle) :
    color = "yellow"


H.binds = [T]
T.binds = [T,H]
particles = [W,H,T]


# reactions
from simulation.chemistry import Reaction
reactions = [
    Reaction(
        [T,T], [H,T],       # educt, product
        1.0, 0.0,           # range (unused), spontaneous rate
        H, 1.0, 0.5         # catalyst, range, and rate
    )
]


# random initial condition
initial_condition = RandomInit(n_W=1780, n_T=600, n_H=10)


# define interactions
import simulation.physics.dpd as dpd


cut = 1
class Physics(dpd.Physics) :
    matrix = {
        (W,W) : (25,cut),
        (W,H) : (15,cut),
        (W,T) : (80,cut),
        (H,H) : (35,cut),
        (H,T) : (80,cut),
        (T,T) : (15,cut),
    }


# thermostat variables and integration time step
gamma = 4.5
sigma = 3.0
time_step = 0.01
physics = Physics(cut, gamma, sigma, time_step)
```

Figure 2.7: The same system setup as in Fig. 2.5 implemented for Spartacus in the python programming language. Unlike ESPresSo, Spartacus uses configuration files only for the system setup but not for the actual integration.

of a core simulation engine written in C and python [123], and can be controlled and extended via python scripts, and inspected at runtime by a graphical interface. Fig. 2.7 shows the example configuration of Fig. 2.5 implemented for Spartacus.

Spartacus offers predefined analyzers comparable to ESPresSo but with emphasis on chemistry instead of physics:

- statistical physics
  - energies (kinetic, potential)
  - pressure (virial and excess pressure)
  - mean square displacement
- distribution and correlation functions
  - velocity correlation and autocorrelation
  - radial distribution function
  - aggregate size distribution
- system chemistry
  - bead numbers
  - reactivity
  - compositional entropy (of aggregates)

Additional analyzers can be defined in the configuration file. Spartacus prints out the result of analyzers for further processing. Alternatively, the system state or screen shots can be saved along a trajectory. Scripts are provided to assemble movies from this data.

2.3.3 Smoldyn

Smoldyn is a spatially resolved simulator for chemical reaction networks with focus on molecular biology. It has been implemented by S. Andrews and licensed under the GNU Lesser Publishing License (LGPL). Smoldyn does not consider bead interactions but merely diffusive motion (implemented as momentumless random walk). Consequently, molecules can only be represented by single beads and their internal structure cannot be modeled – the framework is therefore most appropriate for higher coarse-graining parameters.

The main focus of Smoldyn is its accurate incorporation of chemical reactions [102]. Reactions can be zeroth-, first-, and second-order (type change) reactions. Smoldyn implements reactions based on the Smoluchowski formalism (see Sect. 2.2.4). Smoldyn further allows to define surfaces and compartments for membranes and vesicles by geometrical shapes. These structures can reflect, absorb, or bind beads, and may have an influence on their reactivity and diffusion.

Simulations in Smoldyn are set up by plain text configuration files, implying that the software is not extensible by scripting. The time evolution of a system can be inspected at runtime via a graphical interface and can be captured as movie. Additionally, analyzer results can be printed out for further analysis. Analyzers focus on bead numbers, either in total, or located at specific surfaces or compartments.

2.3.4 LAMMPS

LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) is a fast and feature-rich molecular (MD) dynamics software that includes code for coarse-grained DPD simulations. The software has been developed by S. J. Plimpton, P. Crozier, and A. Thompson [124] and licensed under the GNU General Publishing License (GPL). LAMMPS is implemented in C++ and runs on single processor desktop machines but is designed with massively parallel architectures in mind (employing MPI). LAMMPS does not support chemical reactions and is not extensible by scripting but defines a clear interface for C++ extensions.

LAMMPS consists of a core-engine that is controlled by a textual configuration file. It can print out the system's time evolution. Alternatively, LAMMPS allows to print out the results of custom analyzers that can be defined in the configuration file. Since the configuration file does not define a full-featured programming language, the range of possible analyzers is limited (e.g. to time and ensemble averages of per-atom quantities). More complex observables need to be calculated from the system's trajectory files. The package includes scripts to generate movies and can produce output for VMD [121] and other visualization softwares.

2.4 Conclusion

This chapter has introduced modeling techniques and software for spatially resolved, coarse-grained Artificial Chemistry – primarily Brownian dynamics (BD) and dissipative particle dynamics (DPD) with incorporated chemical reactions. The chapter has derived the underlying mathematical models, and has reviewed implementation principles, means of calibration. The presented methods allow to simulate molecular motion, reaction, and spatial organization in an integrated framework, which makes them valuable tools in the study of emergent physico-chemical structures (e.g. lipid aggregates) and processes that influence or are influenced by those structures, as the existing examples from the areas of artificial chemistry, artificial life, and systems chemistry indicate.

Four existing software packages (ESPresSo, Spartacus, Smoldyn, and LAMMPS) have been presented. These applications differ in focus and features, and vary in their requirements on hardware and user-knowledge from pedagogical tools to high-performance computing applications. Most of the presented programs are scriptable/ extensible, so that flexible simulation setup and incorporation into existing simulation environments can be achieved easily.

Chapter 3

Coarse-Graining and scaling in dissipative particle dynamics (DPD)

Abstract

Dissipative particle dynamics (DPD) is now a well-established method for simulating soft matter systems. However, its applicability was recently questioned because some investigations showed an upper coarse-graining limit that would prevent the applicability of the method to the whole mesoscopic range. This article aims to reestablish DPD as a truly mesoscopic method by analyzing the problems reported by other authors and by presenting a scaling scheme that allows one to apply DPD-simulations directly to any desired length scale.

3.1 Introduction

Dissipative particle dynamics (DPD) was introduced in 1992 by Hoogerbrugge and Koelman [97] as a novel method for performing mesoscopic simulations of complex fluids. Since then, the method has gained significant theoretical support and refinement [84, 99, 125, 126], and has been applied to fluid dynamics in numerous research areas such as rheology [95, 97, 127, 128], material sciences [129, 130], and molecular biology, where membranes [80, 90], vesicles [116, 117], and micellar systems [81, 89, 92] have been modeled.

Initially, DPD was understood to be a truly mesoscopic method able to bridge the whole gap between the underlying atomistic scale (in the range of nanometers and nanoseconds) that is accessible by molecular dynamics (MD) simulations and the macroscopic scale (in the range of micrometers and milliseconds) considered by continuum descriptions. To fulfill this promise, it is crucial that the method is scalable,

meaning that its coarse-graining level can be adjusted without introducing serious artifacts that would render the method worthless.

In fact, it was originally stated that the DPD method is scale-free, meaning that the parameters used in the simulation do not depend on the level of coarse-graining [84]. In a later publication this earlier finding was declared erroneous [80], and it was proposed that interaction parameters determining the conservative forces between DPD particles scale linearly with the coarse-graining level.

Based on this linear scaling relation, the performance of DPD was analyzed for various coarse-graining levels [87, 131]. It was found that there exists an upper coarse-graining level above which the simulated fluid freezes. Trovimof reported that this coarse-graining limit is disappointingly low and only allows up to about 10 water molecules to be grouped together into one DPD particle [131]. This limit would prevent DPD from covering the whole mesoscopic range and confines its applicability essentially to the order of magnitude of MD simulations. Dzwinel and Yuen even concluded that the DPD method would be best suited for the simulation of vapors and gases (where the freezing artifact would happen only for much higher coarsening levels) [87].

The usual method of expressing dynamics in DPD represents the according equations in reduced units, as it is done in several other branches of physics. Using reduced units has considerable advantages: one and the same equation(s) represent a whole family of physical systems which means that qualitative and quantitative statements about the behavior of one system can be translated into the behavior of another one. Nonetheless, different members of this family, e.g. parameterized by some typical length scale, may exhibit different physical properties. An important example for the usage of reduced units is given by the continuous field description of hydrodynamic flow, the reduced Navier-Stokes equations. These reduced equations have the same form for all systems with identical Reynolds and Euler number, though these systems differ with respect to physical parameters, such as the viscosity.

In contrast to the Navier-Stokes equations, which are basically determined by conservative laws, dynamic equations describing a particle-based approach are subject to additional, method related features. For DPD, the relation between different physical systems represented by the same reduced dynamic equation critically depends on the coarse-granining procedure that one uses. Changing the coarse-graining level affects various properties of the "particles" of a particle-based method. Whereas the scaling of some of these features is motivated by physical considerations (for example, the number of DPD particles times the particle mass should be constant), the choice of the scaling of others (e.g. the cut-off radius, see below) is subject to some freedom.

In this article, we study the consequences of different coarse-graining procedures with respect to the reduced dynamical equations the according system defines. The purpose of this work is threefold:

1. As our main result, we argue that the original statement that DPD is scale free can be upheld for equilibrium systems by the usage of an appropriate scaling scheme for the usual conservative DPD interactions (c.f. Eqns. (3.4)). Our coarse-graining procedure renders all interaction parameters, including the elastic conservative one, to scale in such a manner that expressing in reduced units the dynamics of physically equivalent but differently sized systems leads to completely scale free equations.

2. For non-conventional interactions, e.g. surface terms, independence of length cannot be achieved. This is physically plausible: such systems tend to exhibit a typical length scale (e.g. in domain formation) and consequently, systems on

different length scale will look different even when expressed in reduced units. However, we will demonstrate that our coarse-graining procedure does not result in an unfavourable scaling of the interaction parameters. This means: if a system can be simulated with DPD on small scales, at least the scaling of the interactions does not prevent a simulation on a larger scale (though there may be other, system specific issues to be dealt with).

3. DPD is also used for the simulation of transport processes and certain non-equilibrium situations. Whereas, e.g. for the formation of phase diagrams, the time scale plays a somewhat minor role and is closely related to the choice of the energy scale, the simulation of transport phenomena requires a proper calibration of time. We will discuss some of the consequences of this calibration.

In order to present our argument, we need to define the nomenclature used and introduce some notation. By "coarse-graining" we understand the operation of coalescing ν physical particles into one DPD particle [80]. By N, we denote the total number of DPD particles in a simulation and it holds that $\nu N = N_{\text{phys}}$, with N_{phys} being the number of physical molecules the simulation represents. The main objective of this article is the comparison of DPD simulations with different coarse-graining levels ν and ν', which motivates the introduction of the scaling ratio $\phi = N/N' = \nu'/\nu$. In what follows, functions of ϕ will be used to describe the scaling of various quantities at different coarse-graining levels.

By "scaling" we refer to the functional relation between the respective parameters of two systems with different coarse-graining resolutions ν, ν'. In general, simulation quantities and parameters X will be functions of their respective coarse-graining levels; we write $X = X(\nu)$ and $X' = X(\nu')$.

Conventionally, DPD operates in reduced units, such that energy is measured in units of ϵ, length in units of a cut-off radius r_{c}, and mass in units of m, the mass of a single DPD particle; in these units, length, mass, time and energy are dimensionless. We adhere to this practice with the sole difference that we use a time τ instead of an energy as basic unit. That a quantity is expressed in reduced units (with respect to its coarse-graining level) is denoted by a tilde; we use $\tilde{X}, \tilde{X}'$.

In DPD, one usually sets the time unit to

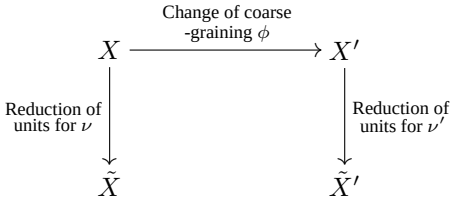
$$\tau = r_{\text{c}}\sqrt{m/\epsilon} \tag{3.1}$$

with r_{c} and m defined as above and ϵ determining the unit of energy, most often set to $\epsilon = k_B T$. Choosing $\tau = r_{\text{c}}\sqrt{m/k_B T}$ is especially convenient for investigations of equilibrium states (e.g., amphiphilic phases), but other choices are possible as well, such as comparing simulated with experimental transport coefficients is a way to determine time scales in non-equilibrium settings. In principle, one has complete freedom to choose the time scale; however, if the unit of time, length, and mass are fixed and the system is governed by a dynamics according to the equations of motion (Eqn. 3.4 below), the energy unit is also given (Eqn. 3.1).

Restating the objective of this work in more technical terms, we will construct a family of DPD-calculations such that its members share physical properties, namely temperature and compressibility, but represent physical systems on different length and time scales. Further, we will show that using this procedure, which is based on an appropriate scaling of length and time together with a specific coarse-graining procedure, the velocity increments calculated from integrating the equations of motion of the DPD particles expressed in reduced units (c.f. Eqns. (3.4) below) do not depend on the

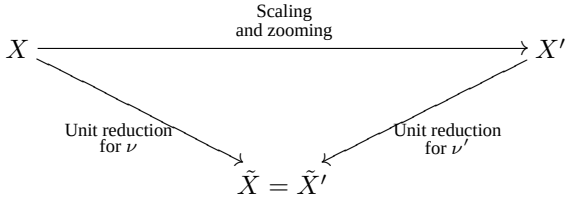
coarse-graining value, i.e. are the same for all members of the constructed family. This is in contrast to earlier publications [80, 132]. Our result implies that, in reduced units, a DPD-calculation performed for a system with small extensions and over a small time interval is numerically identical to one for a much larger system and covering a longer time range.

Using the above notational conventions, one may summarize the combined relative coarse-graining (i.e. scaling by a factor ϕ) and change of units with the following diagram:



(3.2)

The non-trivial fact that two physically equivalent systems, e.g. only differing in their size, exhibit the same reduced equations then alters the above scheme to:



(3.3)

The article is structured as follows: in Sec. 3.2.1, we discuss the results of Groot and Rabone [80] and specify where we deviate from their analysis. We show that their approach of decreasing the number of DPD particles (namely the particle density) while keeping relevant properties (in particular the particles' radius of interaction) constant, is not appropriate. The alternative scaling process we employ is schematically shown in Fig. 3.1A and 3.1B. When we change the level of coarse-graining for the DPD-particles, we accordingly scale their number and adjust their size (radius of interaction).

Fig. 3.1C depicts the main result to be shown in this article, namely that by employing the correct scaling relations and unit reduction one has a complete equivalence of a simulation performed at the scale of Fig. 3.1A with the zoomed version in Fig. 3.1C. This demonstration is split into two parts. In Sec. 3.2.2 and 3.2.3, we change the interaction parameters according to the necessities of the adopted scaling procedure (upper arrow of diagram 3.2). The interaction parameters have to be changed such that when a system with many DPD-particles is mapped onto one with fewer, but larger and heavier particles, the overall system properties are maintained. Sec. 3.2.4 presents simulation results that corroborate the derived scaling relations.

In Sec. 3.3, we analyze the behavior of the DPD algorithm when the rescaled system is expressed in its own set of reduced units and the system size effectively is increased to recover the original number of particles. The change of units affects the natural energy scale (given by $\epsilon = mr_c^2/\tau^2$ and we will show that as a result of our chosen scaling the numerical values that appear in the reduced unit system (Fig. 3.1C) are identical to those of the original one and, as a consequence, their dynamics are equivalent, thereby establishing diagram 3.3, in particular $\tilde{X} = \tilde{X}'$.

Also in Sec. 3.3, the question about the time scales are taken up again; we show that the presented scaling behaves consistently under different methods of time calibration, using different transport processes as measure. In the concluding discussion, we

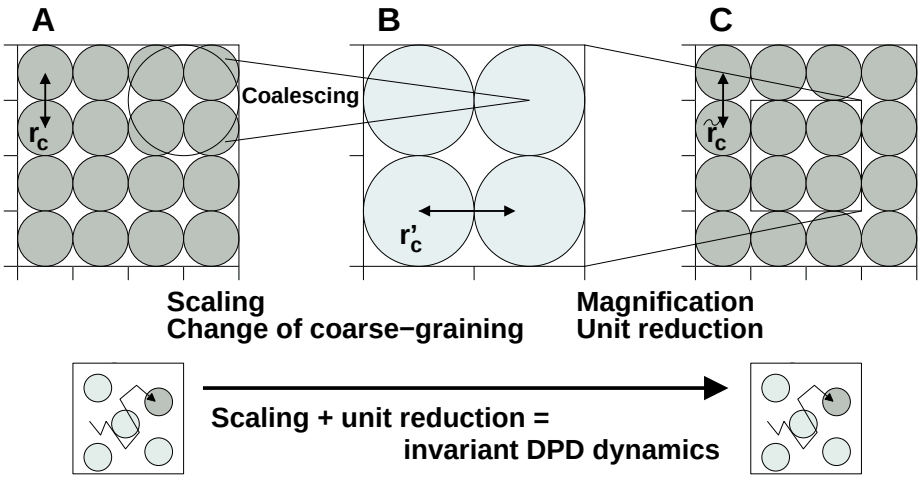


Figure 3.1: Starting from Frame A, a coarse-graining is performed that in this 2D picture coalesces four of the original DPD-particles into one coarse-grained particle, leading to Frame B. Associated with this coarse-graining, a renormalization or scaling is performed that changes the interaction parameters, but leaves the units unchanged. This step also involves scaling some of the interaction parameters in order to maintain relevant physical properties of the system. In transitioning to Frame C, the overall system is expressed in terms of reduced units. The dynamics in the reduced unit system (Frame C) is numerically exactly the same as in Frame A. In fact, Frame C is effectively a zoomed version of Frame A.

comment on conservative interactions with a more complex physical scaling behavior than the repulsion given in Eqns. (3.4) and the resulting consequences for upscaling DPD.

3.2 Scaling DPD

Excellent descriptions of the DPD-method are given in various articles; we will not recapitulate the method itself, but instead refer to Groot and Warren [84]. Here, we only give the definitions of the conservative, dissipative, and random forces in order to define the notation of the parameters:

$$\begin{aligned}
\mathbf{F}_{ij}^{C} &= a_{ij}\chi_{ij}(1-\frac{r_{ij}}{r_c})\hat{\mathbf{r}}_{ij}, \\
\mathbf{F}_{ij}^{D} &= -\gamma\omega^{D}(r_{ij})\left[(\mathbf{v}_i-\mathbf{v}_j)\cdot\hat{\mathbf{r}}_{ij}\right]\hat{\mathbf{r}}_{ij}, \\
\mathbf{F}_{ij}^{R} &= \sigma\omega^{R}(r_{ij})\zeta_{ij}\hat{\mathbf{r}}_{ij},
\end{aligned} \tag{3.4}$$

where r_{ij} is the Euclidean distance between particles i and j, $\hat{\mathbf{r}}_{ij}$ is the unit vector pointing from particle j to particle i, and χ_{ij} equals one for pairs of particles separated by distances less than the force cut-off radius r_c and equals zero otherwise. The parameter a_{ij} determines the magnitude of the conservative interaction and will be regarded in this work as being the same for all pairs of particles: $a_{ij} = a$ (for other types of interaction, see the discussion), ζ_{ij} is a random variable with Gaussian statistics, a vanishing mean and a variance of $1/\Delta t$ for the numerical time step Δt (see e.g. Ref. [96]).

As a consequence, the unit of ζ_{ij} is time$^{-1/2}$. Furthermore, the values of ζ_{ij} in two different time intervals are uncorrelated. In general, if (i, j) and (k, l) are different pairs of particles, ζ_{ij} and ζ_{kl} are independent; however, in order to guarantee the centrality of all forces, one must require that $\zeta_{ij} = \zeta_{ji}$.

The following dissipation-fluctuation relation [99] leads to a thermal equilibrium at a given temperature T:

$$2k_BT\gamma\omega^D(r) = \sigma^2[\omega^R(r)]^2. \tag{3.5}$$

Without loss of generality, we may take $\omega^D(r) = [\omega^R(r)]^2$ for the dimensionless weighing functions $\omega^D(r)$ and $\omega^R(r)$. The dissipation-fluctuation relation (3.5) then reduces to

$$\sigma^2 = 2k_BT\gamma. \tag{3.6}$$

One is free to choose either $\omega^D(r)$ or $\omega^R(r)$ without changing the thermodynamic equilibrium, but it is customary in the literature to take $\omega^R(r) = 1 - r/r_c$ mimicking the conservative force $\mathbf{F}^C$. One also notes that through Eqns. (3.5) and (3.6), the relations between the parameters therein are dependent on the temperature.

In this paper, the physical molecules will refer to bulk fluid particles (such as water), which is in accordance with the cited literature. Other systems including surfaces, such as binary fluids, will be commented on in the discussion.

3.2.1 Compressibility and equation of state

Following Groot and Warren [84], we analyze the scaling behavior of the conservative interaction parameter a by relating the thermodynamic definition of the isothermal compressibility κ_T to the equation of state (involving a) of a system of DPD-particles. The isothermal compressibility is defined as the fractional change in the volume V that results from a change in the pressure P of the system, in a process where the temperature T is constant:

$$\kappa_T = -\left.\frac{1}{V}\frac{\partial V}{\partial P}\right|_T = \left.\frac{1}{\rho}\frac{\partial \rho}{\partial P}\right|_T. \tag{3.7}$$

Here ρ denotes the number density of particles, i.e. N/V. It is convenient to define the dimensionless parameter κ^{-1} by

$$\kappa^{-1} = \frac{1}{k_BT\rho\kappa_T} = \frac{1}{k_BT}\frac{\partial P}{\partial \rho}. \tag{3.8}$$

which we require to be invariant under scaling.

$$\left.\frac{1}{k_BT\rho\kappa_T}\right|_{\text{simulation}} = \left.\frac{1}{k_BT}\frac{\partial P}{\partial \rho}\right|_{\text{simulation}} = \left.\frac{1}{k_BT}\frac{\partial P}{\partial n}\right|_{\text{experiment}} = \text{const.} \tag{3.9}$$

Here, $n = \nu\rho$ denotes the molecular number density of the physical system.

The equation of state relates the pressure with the particle number density ρ. From simulation results, Groot and Warren [84] evaluated the virial expression

$$\begin{aligned} P &= \rho k_BT + \frac{1}{3V}\left\langle \sum_{j>i} r_{ij}F_{ij}^C \right\rangle \\ &= \rho k_BT + \frac{2\pi}{3}\rho^2 \int_0^{r_c} ra\left(1 - \frac{r}{r_c}\right) g(r) r^2 dr. \end{aligned} \tag{3.10}$$

Here, $g(r)$ denotes the radial distribution function. For densities $\rho > 2$ in reduced units (i.e. more than 2 particles in a cubic box with linear dimensions of r_c), the following equation of state is a good approximation to the numerical simulations [84]:

$$P = \rho k_B T + a\alpha\rho^2, \qquad (\alpha = 0.101 \pm 0.001). \tag{3.11}$$

(Note that α has the dimension of *length* to the power of 4.) From this, one concludes that the part of the pressure caused by the conservative interaction scales linearly in a. Further, from Eqn. (3.9), we obtain that at constant temperature

$$\frac{1}{k_B T}\frac{\partial}{\partial\rho}\left(k_B T\rho + a\alpha\rho^2\right) = \text{const.} \tag{3.12}$$

Using $\rho = n/\nu$ where n is the molecular number density and ν the chosen coarse-graining parameter, it follows that

$$1 + \frac{2\alpha a n}{\nu k_B T} = \text{const.} \tag{3.13}$$

Since n and T are constant, Groot and Rabone concluded that a must scale linearly with $\phi = \nu'/\nu$ in order to maintain isothermal compressibility under a change of the coarse-graining level [80].

Several authors regard this scaling as an inherent drawback of the DPD method, since on the micrometer scale the method would appear to be effectively thwarted. We claim that this scaling – and hence its implications – is wrong. The problem is rooted in the construction of the equation of state (Eqn. 3.11). In their computer simulations, Groot and Rabone [80] decreased the density of DPD particles while keeping the interaction cutoff radius r_c constant. This approach allows one to keep the system behavior invariant by scaling the interaction parameters while changing the DPD particle density, without altering their properties. However, we claim that proper scaling means lowering the number of employed DPD particles while simultaneously enlarging their interaction radius. The difference is illustrated in Fig. 3.2. Frame 3.2A is taken to be a system with fine coarse-graining. Frame 3.2B represents a scaled system with a lower DPD particle density but unchanged particle diameters. The result is that the mutual overlap of the soft particles is smaller (as seen in Frame 3.2B). Hence it is intuitively clear that the interaction parameter has to be increased in order to keep the system properties constant; formally, this argument is reflected in Eqn. (3.13). In contrast, Frame 3.2C shows the system with the same scaling ratio as for Frame 3.2B, but with the *relative* overlap of the interacting particles kept constant, which is accomplished by scaling r_c. A closer examination of Frame 3.2C shows that it is part of a magnified version of Frame 3.2A, namely a system where all the lengths associated with a single DPD particle have been uniformly scaled by a factor ϕ while keeping the system size constant ($L' = L$). This results in the following scaling relations for the coarse-graining level, number, mass, and size of DPD particles:

$$\begin{aligned} \nu' &= \phi\nu \\ N' &= \phi^{-1}N \\ m' &= \phi\, m \\ r_c' &= \phi^{1/d}\, r_c, \end{aligned} \tag{3.14}$$

where d is the number of dimensions of the system (see Fig. 3.2C).

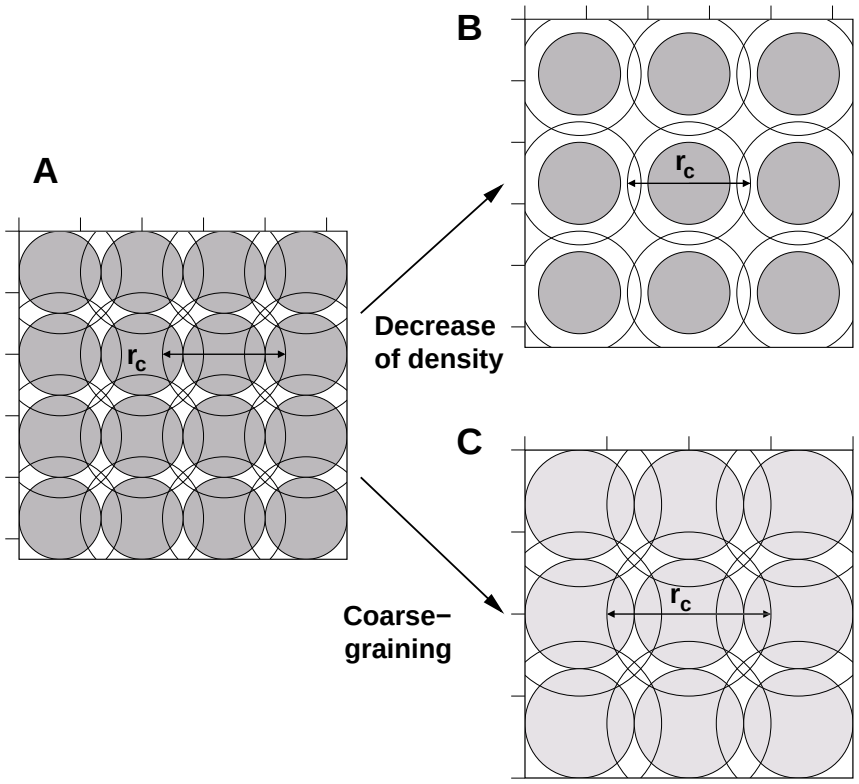


Figure 3.2: Schematic of the scaling process: Frame A shows a DPD simulation with a cutoff radius of r_c. Frame B depicts the coarse-graining procedure performed in Groot and Rabone [80]. With changing particle density, the particle diameter is kept constant while the interparticle force is increased to maintain the system pressure. However, in order to properly conserve systemic parameters like compressibility, both the interaction parameter and the interaction cut-off radius need to be increased as measured in physical units. Frame C depicts the proper scaling in coarse-graining. Along with a decrease of the particle density (in physical units), the interaction range is increased. In this case, the interaction parameter a scales differently than in Frame B in order to preserve systemic properties.

3.2.2 Scaling of the potential energy

We start by calculating the change of potential energy U of a system of DPD particles enclosed in a box that undergoes compression. This change is related to the compressibility of the system and is required to be invariant under scaling. In practice, we require the dependence of a on ϕ, such that the chosen coarse-graining level does not affect the compressibility. For the uncompressed system, we have

$$U_0 = \sum_{i>j} \frac{\chi_{ij} a}{2r_c}(r_{ij} - r_c)^2. \tag{3.15}$$

This equation holds for soft core repulsions which are used throughout the literature. In general, the potential may be viewed as a harmonic approximation of any potential close to an energy minimum. For an isotropically compressed system with box length $(1-\delta)L$, where $\delta \ll 1$ is the relative compression parameter, the change in the interparticle distance $\Delta r_{ij}(\delta)$ is not assumed to be the same for all pairs of particles. However, we require that

$$\Delta r_{ij}(\delta) = \delta r_{ij} + \mathcal{O}(\delta^2), \tag{3.16}$$

which means that we rule out (first order) phase transitions under compression. The total potential energy of the compressed system is then given by

$$U_\delta = \sum_{i>j} \frac{\chi_{ij} a}{2r_c}(r_{ij} - \Delta r_{ij}(\delta) - r_c)^2. \tag{3.17}$$

To first order in δ, we obtain for the change of internal energy

$$\Delta U = U_\delta - U_0 = \sum_{i>j} \chi_{ij}\, a\, (1 - \frac{r_{ij}}{r_c})\, \delta\, r_{ij}. \tag{3.18}$$

Because the change in potential energy of the system as a whole has to be invariant under scaling, we have

$$\sum_{i>j}^{N} \chi_{ij} a \left(1 - \frac{r_{ij}}{r_c}\right) \delta r_{ij} = \sum_{i>j}^{N'} \chi_{ij} a' \left(1 - \frac{r'_{ij}}{r'_c}\right) \delta r'_{ij} \tag{3.19}$$

Due to the scaling of N, the number of terms in the sum of the left hand side of Eqn. (3.19) is proportional to ϕ^{-1}. Since we require ΔU to be invariant under scaling, the force constant a has to scale as

$$a' = \phi^{1-1/d} a, \tag{3.20}$$

the ϕ^1 coming from the change in the number of terms in the sum and the $\phi^{-1/d}$ from the change in length scale. This scaling is the result of our requirement of maintaining the fractional particle overlap during the change of the coarse-graining level. Everything else being equal, this scaling would seemingly still imply an upper coarse-graining limit.

3.2.3 Scaling of time and energy, dissipation and fluctuations

Beside compressibility, the temperature in the simulation should also be unaffected by the coarse-graining. Hence, we require

$$T' = T. \tag{3.21}$$

As detailed in the introduction, there is a gauge freedom for the choice of the time unit:

$$\tau' = \Gamma_{\text{time}}(\phi)\tau. \tag{3.22}$$

We employ this freedom by scaling the time unit equal to the length unit, i.e.

$$\Gamma_{\text{time}}(\phi) = \phi^{1/d} \tag{3.23}$$

and thus $\tau' = \phi^{1/d}\tau$ for which the motivation will become clear in the following. As also described in the introduction, the freedom to set the time unit also affects the energy unit for which we get

$$\epsilon' = m' \frac{r_c'^2}{\tau'^2} = \phi\epsilon. \tag{3.24}$$

The behavior of γ and σ under scaling remains to be determined. Examining Eqn. (3.10), we note that the pressure is independent of γ and σ. This means that, with respect to static compressibility, we have significant freedom in the choice of the scaling function Γ_{therm}:

$$\begin{aligned} \gamma' &= \Gamma_{\text{therm}}(\phi)\gamma \\ \sigma' &= (\Gamma_{\text{therm}}(\phi)\phi)^{1/2}\sigma \end{aligned} \tag{3.25}$$

the latter equation is a consequence of $\sigma' = \sqrt{2\gamma'(k_BT)'}$ and, due to the scaling of the unit of energy, $(k_BT)' = \phi(k_BT)$.

Dimensional analysis motivates the choice

$$\Gamma_{\text{therm}}(\phi) = \phi^{1-1/d}, \tag{3.26}$$

which in turn implies that

$$\begin{aligned} \gamma' &= \phi^{1-1/d}\gamma \\ \sigma' &= \phi^{1-2/d}\sigma. \end{aligned} \tag{3.27}$$

This specific choice will later be shown to be crucial for establishing the scalability of the method, but investigations with other goals (such that deciding on a coarse-graining level ν with least artifacts) may require alternative gauges.

To summarize what has been established throughout the last two sections, changing the level of coarse-graining in DPD requires the following scaling relations:

$$\begin{array}{rclcrcl} N' &=& \phi^{-1}N & & a' &=& \phi^{1-1/d}a \\ m' &=& \phi\, m & & \gamma' &=& \phi^{1-1/d}\gamma \\ r_c' &=& \phi^{1/d}\, r_c & & \sigma' &=& \phi^{1-2/d}\sigma \\ \epsilon' &=& \phi\epsilon & & & & \end{array} \tag{3.28}$$

3.2.4 Validation by simulations

In order to illustrate the correctness of the above scaling arguments, we have measured the pressure in simulations for different coarse-graining levels (i.e. a box of constant size $L = 50$ with DPD-particles of different radii). Simulation parameters for $\phi = 1$ are $r_c = 1, m = 1, \rho = 3.0, \gamma = 4.5, \sigma = 3.0$, and a ranging from 0 to 50. These parameters have been scaled according to Eqn. (3.28) for other coarse-graining values. We want to show that the pressure of this system is invariant under scaling, for all values of a and T, if we follow the described scaling relations.

First, we measure the pressure of a DPD fluid in a closed box with reflective walls (see Fig. 3.3). When a particle collides with a wall, it is reflected elastically and the instantaneous impulse normal to the wall is measured. The pressure is once measured as the time-average of the normal forces on the walls per unit time divided by the surface area of the cube: $P = \langle m\Delta\mathbf{v}_\perp/(A\Delta t)\rangle$ where A is the area of the box and $\Delta\mathbf{v}_\perp$ is the component of the particle velocity orthogonal to the wall. Second, the pressure is measured from the conservative force via the viral theorem (Eqn. 3.10).

Fig. 3.3 show the pressure measured as a function of $a/\phi^{2/3}$ for $\phi = 1$ (circles), $\phi = 8$ (squares), and $\phi = 125$ (diamonds). Hollow symbols give the pressure calculated by wall collisions, whereas solid symbols give the pressure calculated via the virial theorem. The two coincide within standard deviation (smaller than the size of the symbols). For each coarse-graining parameter the equation of state (Eqn. 3.11) is reproduced accurately for small values of a. On the one hand, for $a = 0$, the case of an ideal gas, the pressure should be given by

$$P|_{a=0} = \phi^{-1} N\phi k_B T/V = \rho k_B T, \tag{3.29}$$

which it indeed is seen to be for all values of ϕ. On the other hand, kinetic gas theory establishes

$$P = \rho\langle m\mathbf{v}^2\rangle/3. \tag{3.30}$$

This constitutes a relationship between the thermostat and the conservative mechanical interactions with the walls, which is non-trivial for the DPD equations of motion. The way we measure the pressure gives a direct relation to $\langle\mathbf{v}^2\rangle$. The pressure calculated from kinetic quantities (Eqn. 3.30) coincides with the value required from the thermodynamic relation Eqn. (3.29). This shows that Eqn. (3.11) holds for our simulation. The slopes of the three curves differ, which indicates that α in Eqn. 3.11 depends on the relative box size L/r_c in hardwall simulations. This is most likely due to a crowding of DPD particles at the boundary of the box which walls lack any repulsion potential.

This problem can be avoided by the use of periodic boundary conditions (again in a box of size $L = 50$) and then measuring the pressure via the virial theorem (which has been shown to coincide with the pressure obtained by wall collisions in the previous simulation). Results are shown in Fig. 3.4 for different repulsion parameters a in $[0, 100\phi^{-2/3}]$ and various coarse-graining ratios ϕ in $[1, 125]$. In the case of periodic boundary conditions, the pressure is found to be independent of the coarse-graining. The result confirms the correctness of the scaling scheme summarized in Eqn. (3.28).

3.3 Reduction of units

Our goal is now to show that the velocity increments $\Delta\mathbf{v}$ obtained from integrating the forces are unchanged when the scaling is combined with the according reduction of units: $\Delta\tilde{\mathbf{v}} = \Delta\tilde{\mathbf{v}}'$, which implies that the relative particle motions are unaffected by scaling in the reduced unit systems.

The random variable ζ_{ij} has the unit $\tau^{-1/2}$, as noted in the discussion following Eqn. 3.4. Given the scaling of τ, it follows that

$$\zeta'_{ij} = \phi^{-1/(2d)}\zeta_{ij}, \tag{3.31}$$

and therefore

$$\sigma'\zeta'_{ij} = \phi^{1-1/d}\sigma\zeta_{ij}. \tag{3.32}$$

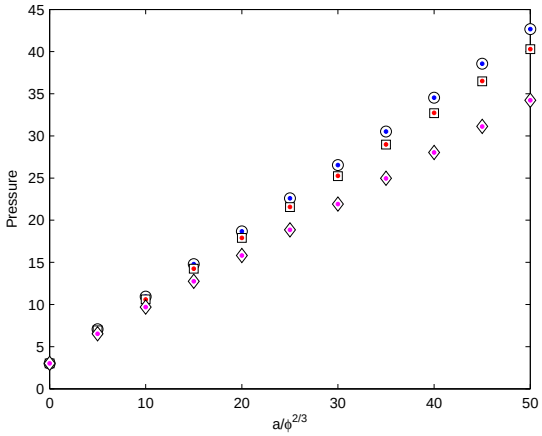


Figure 3.3: Pressure in a DPD simulation with reflective walls for different coarse-graining ratios $\phi = 1$ (circles), $\phi = 8$ (squares), and $\phi = 125$ (diamonds) and various repulsion parameters a in $[0, 50/\phi^{2/3}]$.Hollow symbols give the pressure calculated by wall collisions, whereas solid symbols give the pressure calculated via the virial theorem. The two coincide within standard deviation (smaller than the size of the symbols). For each coarse-graining parameter the equation of state (Eqn. 3.11) is reproduced accurately for small values of a. For the ideal gas scenario ($a = 0$), the theoretical prediction of $P = \rho k_B T$ is recovered independent of ϕ. The slopes of the curves differ, which indicates that α in Eqn. 3.11 depends on the relative box size L/r_c in hardwall simulations.

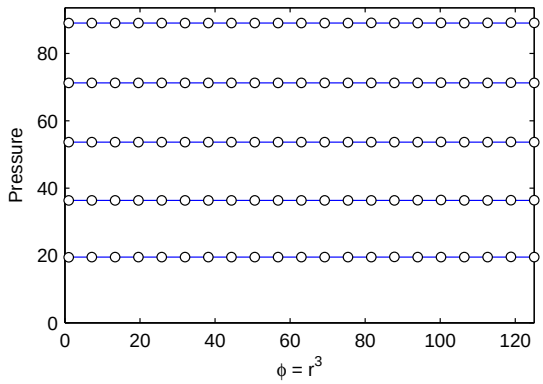


Figure 3.4: Pressure in a DPD simulation with periodic boundary conditions for different repulsion parameters a in $[0, 100\phi^{-2/3}]$ for various coarse-graining ratios ϕ in $[1, 125]$. The pressure (obtained from the virial theorem) is independent of the coarse-graining.

This is the same scaling as seen for a (Eqn. 3.20) and for γ (Eqn. 3.27). As a consequence, all three force components in Eqn. (3.4) scale by a factor $\phi^{1-1/d}$. When velocity increments are calculated during one time step, one finds that the force scaling is canceled by the scaling of mass and time:

$$\begin{aligned}[\Delta\mathbf{v}_i]' &= \sum_{j\neq i} \frac{[\mathbf{F}_{ij}]'}{m'}\Delta t' \\ &= \frac{\phi^{1-1/d}\phi^{1/d}}{\phi}\sum_{j\neq i}\frac{\mathbf{F}_{ij}}{m}\Delta t = \Delta\mathbf{v}_i \end{aligned} \tag{3.33}$$

Since $\Delta\tilde{\mathbf{r}} = \Delta\mathbf{r}'/r_c'$ and $\Delta\tilde{t} = \Delta t'/\tau'$, we get for the velocity increment by considering

$$\Delta\tilde{\mathbf{v}}_i = \Delta\tilde{\mathbf{r}}_i/\Delta\tilde{t} = \Delta\mathbf{v}_i\,\tau/r_c. \tag{3.34}$$

Because time and length scale in the same way we get $\tau'/r_c' = \tau/r_c$. Combining this with Eqn. (3.33) one finally obtains

$$\Delta\tilde{\mathbf{v}}_i = \Delta\tilde{\mathbf{v}}_i', \tag{3.35}$$

which implies that trajectories are numerically equal:

$$\tilde{\mathbf{r}}(\tilde{t}) = \tilde{\mathbf{r}}'(t). \tag{3.36}$$

What remains to be shown is the scaling of the reduced parameters $\tilde{a}$, $\tilde{\gamma}$, and $\tilde{\sigma}$. Since a scales like energy over length, when going to the reduced units of the primed system, we have

$$\tilde{a}' = a'\frac{r_c'}{\epsilon'} = \frac{\phi^{1-1/d}\phi^{1/d}}{\phi}a\frac{r_c}{\epsilon} = \tilde{a}, \tag{3.37}$$

and similarly, since γ scales like energy over length and velocity, from $\tilde{\gamma}' = \gamma'\,r_c'^2/(\epsilon'\tau')$ we get

$$\tilde{\gamma}' = \gamma'\frac{r_c'^2}{\epsilon'\tau'} = \frac{\phi^{1-1/d}\phi^{2/d}}{\phi\phi^{1/d}}\gamma\frac{r_c^2}{\epsilon\tau} = \tilde{\gamma}. \tag{3.38}$$

From the fluctuation-dissipation relation it follows again that

$$\tilde{\sigma}' = \tilde{\sigma}. \tag{3.39}$$

Hence, scaling and unit reduction precisely cancel each other. As a result, the DPD formalism is scale-free for interactions as given in Eqns. (3.4), if space and time are scaled appropriately. This means that the calculation with a single set of parameter values represents systems at arbitrary lengths scales, which establishes the numerical equivalence of zoomed systems as described in Fig. 3.1C.

3.4 Summary and discussion

We have shown that the coarse-graining procedure in DPD can be chosen in such a manner that the different physical systems represented by the reduced DPD dynamics share some physical properties, such as compressibility. This was achieved by the combined effect of a specific scaling of the cut-off radius of DPD particles, a proper

choice of the time scaling, the reduction of units, and finally by requiring a specific but obvious scaling of γ and σ.

The independence of scale shown for bulk fluids cannot necessarily be upheld for other types of interactions; e.g. for binary mixtures of liquids A and B where several conservative interaction parameters occur, say a_{AA}, a_{AB}, a_{BB}. Whereas bulk interactions given by a_{AA}, a_{BB} scale as discussed in this article, a_{AB} is a surface term that determines interfacial energy and therefore scales differently. For bulk interactions, in the presented scheme the parameters determining forces and energies scale in physical units linearly with ϕ, i.e. they are proportional to the number of molecules a DPD particle represents. While this proves to be adequate for simulating elastic properties, non-elastic, surface dependent interaction parameters scale (again in physical units) with $\phi^C, C \leq 1$. Repeating the calculation presented in this work, it then turns out that those interaction parameters effectively shrink with an increase of the coarse-graining. This is physically plausible: assume a system that exhibits domain formation. On a (sufficiently) small scale, the domain boundaries will occur as planar (indicating a, with relation to the bulk energies, large surface energy) whereas on a scale above the typical domain size, minimization of surface does not anymore play a dominant role. This means that if a DPD-calculation can be performed at a small scale, then, at least with respect to the scaling of parameters, also calculations at larger scales will be feasible.

Our calculations reestablish DPD as a method to analyze static properties of equilibrium systems (most notably phase diagrams) over the entire mesoscopic length scale. In particular, our scaling procedure overcomes the reported freezing artifact of DPD. However, the validity of our scaling relation does not necessarily hold for the study of dynamic properties (such as transport processes). To clarify this point, we briefly discuss the scaling of diffusion constants $\tilde{D} = \tilde{D}'$ (being a consequence of the equality of the measured trajectories, see Eqn. 3.36) and the fact that the diffusion constant scales like length squared over time causes an apparent problem: it seemingly implies that relative fluctuations $\tilde{D}/\tilde{L}$ stay constant instead of vanishing. This is not unexpected but a consequence of the coarse-graining. For a general discussion of the renormalization of diffusion, see e.g. Ref. [133]. To what extent the trajectories of the particles in DPD can be understood as representing actual transport processes is discussed by Groot [132] and shown in the approach used by Groot and Warren [84] and Jakobson [134]. These authors resolve the calibration problem by treating the natural time unit τ as undetermined and instead obtain the physical time scale t from calibrating diffusion constants measured in simulation to physical values [84].

Note that there are applications in which the time scales are of minor interest; for example, for the investigation of static properties such as lipid phase diagrams only energy scales are relevant, and a direct physical interpretation is well possible. The study of dynamic phenomena, on the other hand, may either be only possible by changing the interpretation of a DPD particle (and understand it as a mere discretization point in a dynamic lattice and its interaction energies as model parameters without direct physical interpretation), or by using different scaling schemes based on other choices for the gauge function Γ_{therm}.

Acknowledgments

This work was financially supported by the EU project FP6 IST-FET integrated project PACE and by the Los Alamos National Laboratory LDRD-DR grant on “Protocell

Assembly" (PAs). The first author was additionally supported by the BMBF project Systems Biology of the Liver Grant # 3P3137. Most of the work was performed during a workshop at the European Center for Living Technologies (ECLT) in Venice, Italy, organized by Martin Nilsson Jacobi. The stimulating working environment provided by this institution is greatly acknowledged.

Note that in order to understand the according physical time scales we have to comment on transport properties of the method and the scaling of fluctuations. We base our argument on diffusion, but could equally consider viscosity since the two are related by the Schmidt-number, which is dimensionless. The numerical equivalence of the measured diffusion constants $\tilde{D} = \tilde{D}'$ (being a consequence of the equality of the measured displacement) and the fact that diffusion scales like length squared over time causes an apparent problem: it seemingly implies that relative fluctuations $\tilde{D}/\tilde{L}$ stay constant instead of vanishing. This problem disappears when one calibrates the simulation to an actual physical system. Assume that the cutoff radius is related to a physical length by $r_c = l[\mathrm{c}m]$. We then have

$$D\frac{r_c^2}{\tau} = D_{\text{phys}}\frac{\text{cm}^2}{\text{sec}}, \tag{3.40}$$

with $D_{\text{phys}} = \frac{<r(t)^2>}{t}$ referring to the diffusion constant in physical units. We get

$$\tau = \frac{D}{D_{\text{phys}}} l^2[\text{sec}]. \tag{3.41}$$

Consequently, expressed in physical units, it holds for the fluctuations

$$\lim_{l\to\infty} \frac{D_{\text{phys}}}{L_{\text{phys}}} \sim \frac{1}{l} = 0. \tag{3.42}$$

Chapter 4

Minimal model of self-replicating nanocells: a physically embodied information-free scenario

Abstract

The building of minimal self-reproducing systems with a physical embodiment (generically called protocells) is a great challenge, with implications for both theory and applied sciences. Although the classical view of a living protocell assumes that it includes information-carrying molecules as an essential ingredient, a dividing cell-like structure can be built from a metabolism-container coupled system, only. An example of such a system, modeled with dissipative particle dynamics, is presented here. This article demonstrates how a simple coupling between a precursor molecule and surfactant molecules forming micelles can experience a growth-division cycle in a predictable manner, and analyzes the influence of crucial parameters on this replication cycle. Implications of these results for origins of cellular life and living technology are outlined.

4.1 Introduction

The transition from non-living to living systems covers a broad spectrum of increasingly complex organization [135]. One of such first steps in this transition leads from ordinary chemical auto-catalysis to self-replication. The latter can be distinguished from the former by the existence of self-bounded entities which produce copies of themselves, rather than a mere increase in chemical concentration. Therefore, self-replication relies on organization principles unlikely to be found in homogeneous so-

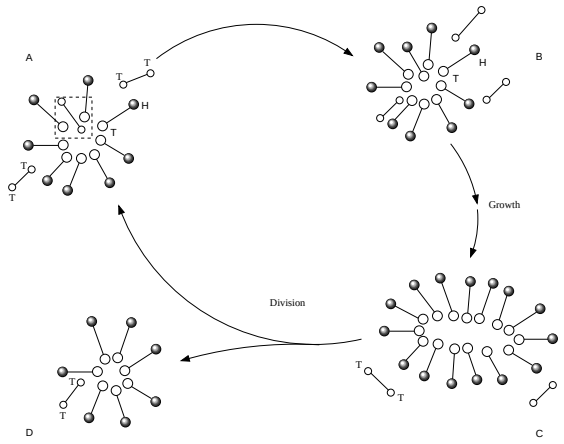


Figure 4.1: The basic model of nanocell replication explored in this paper. Here small-sized micelles are formed by amphiphiles (here indicated as $H-T$ connected pairs of balls). These amphiphiles have a hydrophilic head (H) and a hydrophobic tail (T). Precursor molecules are also shown as two connected, smaller open balls, both of them of hydrophobic character. Under the presence of catalyically active amphiphiles, precursors are transformed into additional amphiphiles. Incorporation of many such building blocks allows the nanocell to grow in size. When a critical size has been reached, the nanocell looses its stability and divides into two smaller aggregates thereby closing the replication cycle.

lutions. Within the last years increasing attention has been payed to the possibility of building small-scale protocells, in particular using a bottom-up approach [136] where the building blocks (not necessarily from biotic origin) would assemble spontaneously and, under appropriate conditions, develop a growth-fission cycle. While extensive research has been performed on the self-reproducing capabilities of biopolymers [137, 63], even much simpler systems can be driven into dynamics that we identify as self-replication. In this context, micelles have been proposed to serve as life-like structures able to undergo self-reproduction [59].

These micellar systems can be considered to be simpler in organization than biopolymers because they lack any genetic information that could be passed from one generation to the next. In this context, protocellular entities lacking information would be under the umbrella of Oparin's views of life origins [138], who suggested that primitive self-replicating vesicles would have predated information-based cells. Similarly, other authors have advocated for this scenario under a computational perspective. Under this view, Dyson [139] indicated that current cells involved both software (DNA and RNA) and hardware (protein machinery). Although hardware can exist under the absence of software, the opposite is not allowed to occur. We can add to this picture of cells that the presence of a container is a very important piece for the hardware to properly work: only when the pieces are close together we can expect the machinery to operate.

The nanocellular system discussed here is based on a micelle that is coupled to a minimal metabolism (figure 4.1). The system is constantly driven away from equilibrium by the supply of precursors, which are supposed to have higher internal energy than the surfactants that build up the micelles. The metabolism transforms precursor

molecules into new building blocks of the nanocell. The metabolic turnover is thereby enhanced by its own outcome or—in other words—the metabolism is an auto-catalytic turnover of precursors into new surfactants. This system resembles one studied experimentally by Bachmann et al. [59]. However, in their work the catalytic effect is a feature of the micelles (probably due to milieu effects), while, in our system, it is a feature of the molecules themselves.

Either way, the replication mechanism of micelles envisioned both in Bachmann's as well as our system is the following: micelles incorporate hydrophobic precursor molecules where they are afterwards transformed into new surfactants. Due to this process, the number of surfactants increases, while at the same time, the volume of the hydrophobic core becomes smaller. It is assumed, that when a critical ratio of surfactants versus core volume is passed, the aggregates become unstable and will divide into two daughter cells. Whether or not the experimental system follows this pathway has, to our knowledge, never be clarified. However, once such a replication mechanism is considered the basis for a more complicated protocell design, knowledge of the exact replication kinetics becomes crucial.

Together with experimental approaches aimed to finding the conditions for protocell replication to occur, there is an increasing need of appropriate, well-grounded theoretical and computational models. Such models can help understanding the constraints that might operate in the self-assembly of micelles and other molecules and how they can properly trigger growth and splitting. Our work is a first step in this direction.

4.2 DPD nanocell model

In this paper we consider a *dissipative particle dynamics* (DPD) approach to modeling embodied protocells employing a physically and chemically simplified formalism. This is one possible approach that can be used in order to tackle the complexities of molecular aggregates. Other approaches include Molecular Dynamics (MD), Brownian Dynamics (BD) and Monte Carlo algorithms. Each of these methods has its own advantages and drawbacks [140, 141] and a compared analysis has been presented elsewhere [142].

Previous work done by Ono and Ikegami involved a lattice-based, protocell dynamics [143]. These structures display a special type of cell-like replication. While they are remarkable in their self-organizing behavior, the underlying rules of interaction and the special properties of the membrane-like structures restrict their relevance to the arena of artificial life. By using more appropriate molecular interactions within a three-dimensional, water-filled environment (see for example Rasmussen and Nilsson's cellular automata approach to self-assembly [71]) we seek to provide the (as far as we know) first computational evidence that such a simple protocellular cycle is feasible.

4.2.1 Dissipative particle dynamics

DPD is a coarse grained, particle based simulation technique comparable to Brownian Dynamics. It was proposed by Hoogerbrugge and Koelmann [97] and gained significant theoretical support in the late 1990s [99, 84, 144]. In the context of biological systems, DPD models have been successfully used to capture the dynamics of membranes [90], vesicles [117, 116] and micelles [89, 145].

A DPD simulation consists of a set of N particles which are described by their type, mass m_i, position $\mathbf{r}_i$, and momentum $\mathbf{q}_i = m_i\mathbf{v}_i$. These particles—usually called *beads* throughout the literature—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.

Newton's Law of motion is used to determine the trajectory of each individual bead:

$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{1}{m}\mathbf{F}_i \tag{4.1}$$

The force $\mathbf{F}_i$ that acts on particle i is expressed as the superposition of pairwise interactions

$$\mathbf{F}_i = \sum_{j=1}^{N} \mathbf{F}_{ij} \tag{4.2}$$

In Newtonian dynamics, the (central) force F_{ij}^C can be expressed as the negative gradient of a potential ϕ_{ij}, namely

$$F_{ij}^C = -\nabla\phi_{ij} \tag{4.3}$$

The resulting dynamics are conservative and obey the Hamiltonian

$$H = \frac{1}{2}\sum_{i=1}^{N} m_i\mathbf{v}_i + \frac{1}{2}\sum_{i,j=1}^{N} \phi_{ij} \tag{4.4}$$

While this approach is undertaken in molecular dynamics simulations, coarse grained simulation techniques try to aggregate some of the molecular degrees of freedom by the use of the so called Langevin formalism: additional forces $\mathbf{F}_{ij}^D$ and $\mathbf{F}_{ij}^R$ are added to the conservative force to express friction and thermal motion. They introduce energy flows between the explicitly modeled mesoscale and an the underlying microscale. Together, they act as a thermostat to regulate the effective temperature, i. e. mean velocity, of the system.

The thermostat used in the DPD formalism is given by the equation

$$\mathbf{F}_{ij}^D + \mathbf{F}_{ij}^R = \left(\eta\omega(r_{ij})(\mathbf{n}_{ij}\cdot\mathbf{v}_{ij}) + \sigma\omega^2(r_{ij})\xi_{ij}\right)\mathbf{n}_{ij} \tag{4.5}$$

where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the distance, $\mathbf{n}_{ij} = (\mathbf{r}_i - \mathbf{r}_j)/r_{ij}$ the (unit) direction, and $\mathbf{v}_{ij} = \mathbf{v}_i - \mathbf{v}_j$ the relative velocity between beads i and j. η is the friction coefficient of the fluid and σ the amplitude of thermal motion. ξ_{ij} is a random variable with $\xi_{ij} = \xi_{ji}$, and otherwise Gaussian statistics. ω is a distance weighing function usually defined as

$$\omega(r) = 2\left(1 - \frac{r}{r_c}\right) \tag{4.6}$$

where r_c is a cutoff range used to limit the maximal range of all interactions for performance reasons. It can be shown that the equilibrium temperature of the system tends towards $k_bT = \sigma^2/2\eta$ [99].

While there is a variety of other thermostats used in coarse grained particle simulations, the above mechanism is unique in that it both conserves linear and angular momenta and fulfills the fluctuation-dissipation theorem. As a consequence of the former, the resulting dynamics are consistent with the Navier-Stokes equations and hence preserve laminar flow properties of the system. The latter property ensures an energy

distribution in the system following Maxwell-Boltzmann statistics. The overall dynamics, therefore, capture both hydrodynamic and thermodynamic traits of the systems.

In almost all DPD studies, the conservative force is derived from a *soft-core potential* of the shape

$$\phi_{ij}(r) = \begin{cases} \frac{1}{2} a_{ij} r_c \left(1 - \frac{r}{r_c}\right)^2 & \text{if } r \leq r_c \\ 0 & \text{if } r > r_c \end{cases} \tag{4.7}$$

The potential energy expressed by ϕ_{ij} should not be understood as the mechanical energy, i. e. enthalpy, of the system, but rather as its free energy contribution [82]. Following this rationale, the interaction parameters a_{ij} are used to express dissimilarities of substances due to high enthalpy as well as entropy contributions, respectively. Therefore, they can be related to Flory-Huggins coefficients known from polymer theory.

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

$$\phi^B_{ij}(r) = \frac{b r_b}{2} \left(1 - \frac{r}{r_b}\right)^2 \tag{4.8}$$

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 equation 4.1 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) with interaction parameters taken from [89], unless otherwise specified:

	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	$80 k_b T$	$80 k_b T$	$15 k_b T$

All beads have mass $1m$. Precursor molecules are modeled as dimers of bonded **T** beads, surfactants as dimers of one **T** and one **H** bead. Here we have: $b = 125 k_b T$, $r_b = 0.5 r_c$ 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 $a_{\mathbf{TT}} < a_{\mathbf{WW}} < a_{\mathbf{HH}}$. Furthermore, surfactant heads have a high affinity to water ($a_{\mathbf{HW}} < a_{\mathbf{WW}}$), which is usually due to charges in the hydrophilic groups of the molecules. This assumption ensures that aggregates with high surface area (spherical micelles) are prefered over aggregates with less surface area (rod-like micelles) in the process of total energy minimization.

4.2.2 Incorporation of chemistry

The metabolic reaction under consideration takes the following form

$$\mathbf{T} - \mathbf{T} \longrightarrow \mathbf{H} - \mathbf{T} \tag{4.9}$$

This reaction is modeled by a stochastic process that has formerly been used in Brownian Dynamics simulations [108]. Inbetween every two steps of the numerical integrator for the DPD equation of motion, each precursor dimer can be transformed into a surfactant molecule with a spontaneous reaction rate k_b. The spontaneous reaction can be catalytically enhanced by nearby surfactants whose catalytic influence decreases linear with the distance to the reactant up to a certain threshold r_{cat}. For simplicity, the effect of several catalysts is modelled as a superposition:

$$k = k_b + \sum_{i \in \mathbf{C}} k_s \begin{cases} \left(1 - \frac{r_{\mathbf{C}}}{r_{cat}}\right) & \text{if } r < r_{cat} \\ 0 & \text{otherwise} \end{cases} \tag{4.10}$$

where $r_{\mathbf{C}}$ is the distance of the catalyst and k_s the maximal catalytic rate per catalyst. For clarity of the results, we set the spontaneous reaction rate in our simulations to $0\tau^{-1}$. In the upcoming simulations, k_{cat} is set to $1.0\tau^{-1}$, 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.

We also introduce particle exchange into the model to mimic the support of new precursors into the system: During the simulation, two water beads can be exchanged by precursor dimers with the probability $2.5 \cdot 10^{-5}\tau^{-1}$ per water bead within a region of radius $2r_c$. Again, bead positions and momenta are preserved.

4.3 Results

We implemented the DPD method using a velocity-Verlet integrator ($\lambda = 0.5$) with a step width of 0.04τ. The spatial domain is three dimensional, with periodic boundary conditions and size $(10r_c)^3$. In all the following simulations, the system is initialized with one surfactant dimer and 2998 water molecules adding up to a mean particle density of $3r_c^{-3}$. Unless otherwise specified, simulations have been run for $0\tau \leq t \leq 1000\tau$ (25000 iterations).

For simulation runs with the parameter set given in the last section, we typically get the following behavior: water beads are successively exchanged by precursors in the exchange region of the system. While diffusing through the system, precursors form droplets due to their hydrophobic trait. Once such a droplet reaches the initial surfactant, the precursors are turned into surfactants. The turnover happens fast compared to supply and diffusion of the precursor. New surfactants quickly arrange into a micellar shape with hydrophobic beads in the interior and hydrophilic beads towards the surface of the assembly. With the parameters introduced before, this rearrangement takes only few time steps and is thus fast compared to the metabolic turnover. Such spontaneously formed nanocells diffuse through the system space as aggregates and eventually incorporate additional precursor droplets in their interior, where the metabolic process is repeated. The evolution of the overall system composition (number of beads per type) traces the different processes on their respective time scales (see figure 4.2 for $a_{\mathbf{HT}} = a_{\mathbf{WT}} = 80k_bT$ (upper panel) and $a_{\mathbf{HT}} = a_{\mathbf{WT}} = 120k_bT$ (lower panel). As one can see, the overall production of surfactants is limited by a linear growth that results from the constant supply of precursors. Locally, however, when a single droplet is consumed by a nanocell, the metabolic turnover exhibits a logistic growth which is suspected from resource limited auto-catalysis. The logistic growth can be best seen in the lower panel of figure 4.2 around $t = 700\tau$.

On the level of individual micelles, the metabolic process increases the aggregate number, i. e. the number of surfactants per micelle. In a pure water-surfactant system,

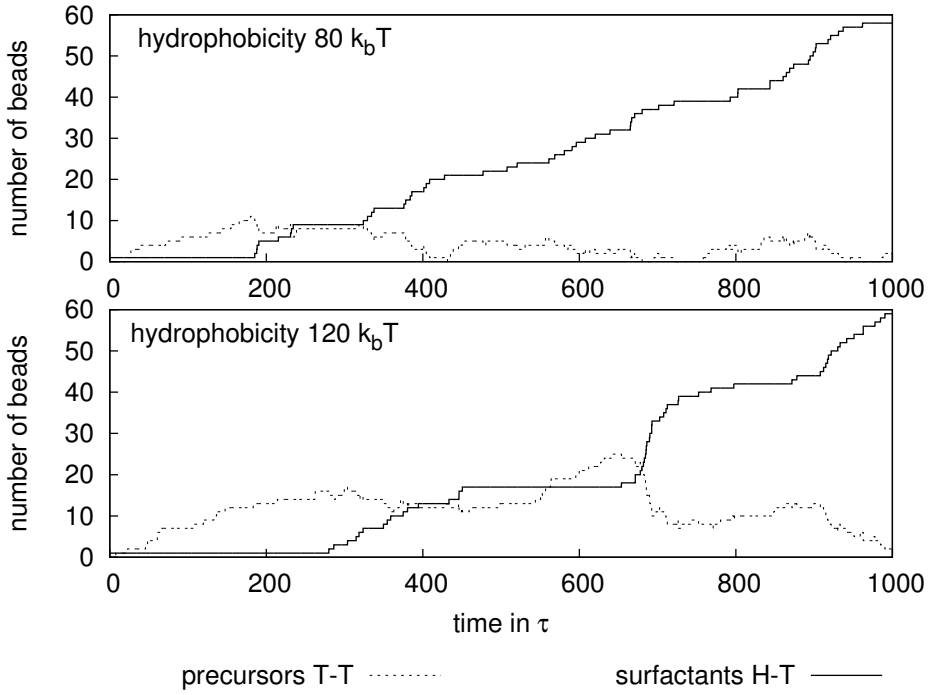


Figure 4.2: Evolution of bead numbers with time for two different hydrophobicity values ($a_{\mathbf{HT}}$ and $a_{\mathbf{WT}}$). The constant supply of precursors is counteracted by their transformation into surfactants. This transformation happens in spurts rather than continuously, as the precursor forms droplets in the aqueous solution. Size and frequency of these spurts depend on the hydrophobicity of the tail beads.

micelles would reject the surplus of surfactants into the bulk phase. In an oil-surfactant-water system, as the one under consideration, the hydrophobic core formed by the precursors, stabilizes the assembly far beyond its original aggregate number. As a consequence, we could observe that nanocells increase in aggregate number when new surfactants are synthesized. While the precursor surfactant ratio shifts, the nanocell changes its shape from a spherical to a rod-like micelle. When all or nearly all of the precursor is turned into surfactant, the nanocell finally becomes unstable and divides into two smaller aggregates (see figures 4.3 and 4.4). The nanocell division occurs in the cylindrical middle-part of the rod-like aggregate by indentation of surfactant heads. It induces vibrating modes into the daughter aggregates as they rearrange back to a spherical shape. Although this vibration is successively dissipated into undirected motion, it sometimes leads back to short series of temporary fusion and fission of the daughter cells.

Elongated micellar structures are well-known from worm-like micelles which usually consist of two surfactants with different curvature. Such worm-like micelles are stable equilibrium aggregates and exhibit an exponential size distribution [146]. In the system studied in this work, however, the elongated aggregate is not stabilized by different curvature values of the components, but by the hydrophobic core. Accordingly, once the precursor is turned over into new surfactant molecules, the elongated structure looses its stability.

There is a second pathway, however, that might jeopardize the above scenario: once

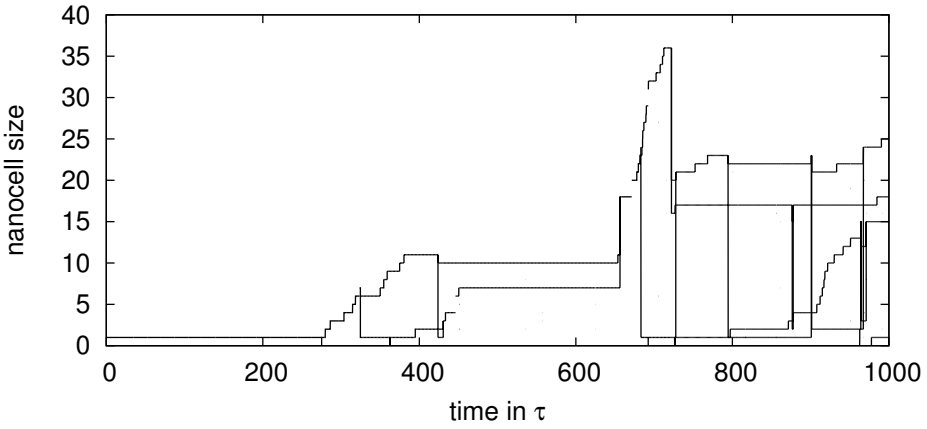


Figure 4.3: Here, the size evolution of figure 4.2 (lower panel) is tracked for individual nanocells: each line designates the size evolution of a single aggregate. Horizontal lines result from fission (or dissociation) events, after which two lines indicate the fate of the daughter cells. Isolated dots denote short term vibrations during which nanocells divide and fuse within less than 1τ. For clarity, such horizontal lines have been suppressed for such vibrations. As one can see, only two daughter cells result from a true fission event (at $t = 722\tau$). The other two result from dissociation of single surfactants that start to turn over precursor droplets found in bulk phase. These surfactant dissociations happen at $t = 324\tau$ and $t = 684\tau$. Furthermore, two nanocells fuse at $t = 655\tau$.

in a while throughout our simulations, nanocells loose individual surfactants into the bulk phase. If this relaxation process happens fast compared to the metabolic turnover, the nanocells might not be able to reach the division size. Surfactants in the bulk phase may however metabolize precursor droplets and spontaneously form nanocells on their own.

Formally, fission events can be written in the form of a chemical reaction:

$$S_{n+m} \xrightarrow{k^-_{m,n}} S_n + S_m \tag{4.11}$$

where S_n, S_m and S_{n+m} are aggregates of size n, m, and $n+m$, respectively, and $k^-_{m,n}$ is the fission rate. For $m = 1$, one obtains dissociations as a special case. Analogously, association and fusion events take the form

$$S_n + S_m \xrightarrow{k^+_{m,n}} S_{n+m} \tag{4.12}$$

In order to quantify micellar fission and surfactant dissociations, nanocells have been identified by a variant of the flood fill algorithm: every two $\mathbf{T}$ beads within a distance of $1r_c$ or less have been considered to belong to the same aggregate. The aggregate number is defined as the number of participating surfactants. This allows for tracking individual aggregate sizes and their transitions through time. Each of the above reaction schemes results in two transitions, given by

$$S_n \to S_{n+m} \quad ; \quad S_m \to S_{n+m} \tag{4.13}$$

and

$$S_{n+m} \to S_n \quad ; \quad S_{n+m} \to S_m \tag{4.14}$$

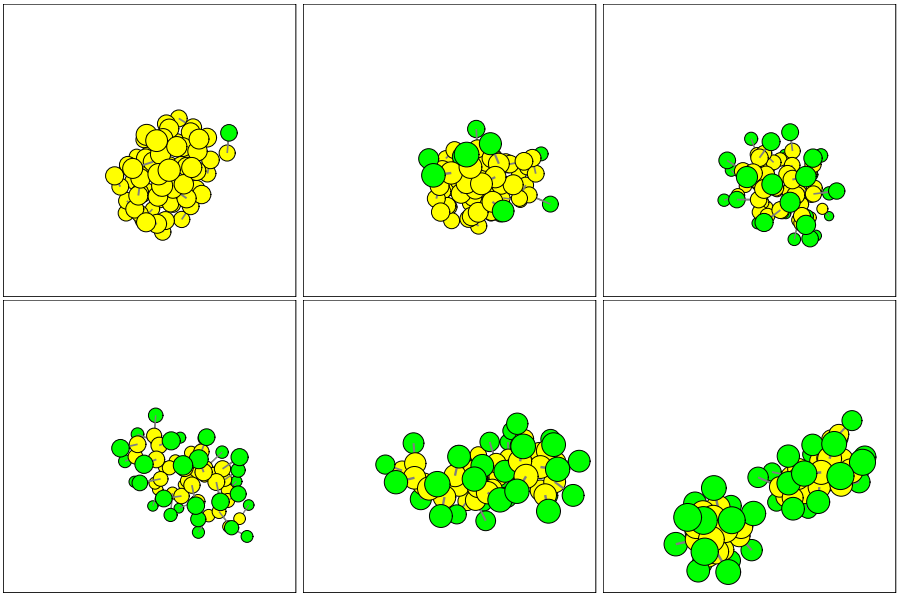


Figure 4.4: Metabolism and fission of a nanocell (**T** beads are shown in light, **H** beads in dark gray—water not shown): The initial surfactant metabolizes a precursor droplet and turns it into a functioning nanocell (panels 1-3). While the precursor is consumed, the nanocell elongates to account for the changing precursor surfactant ratio (panel 4-5). Such elongated structures can be stable for several time units, until—when all precursors are turned into surfactants—the nanocell divides into two daughter cells (last panel).

Furthermore, the turnover of precursors results in the transition

$$S_n \rightarrow S_{n+1} \tag{4.15}$$

Not all transitions can be expressed by the chemical reaction scheme given by 4.11 and 4.12. Those transitions are of the type

$$S_{n+m} + S_l \longrightarrow S_n + S_{m+l}$$

as well as fissions into and fusions from more than two aggregates. For all the simulations performed, such outliers amount to less than 0.3% of the total transitions. They have been neglected for further analysis.

In the remainder of this work, we will analyze two key parameters of the model and their influence on the dynamics of the system. Thereby, we will put our attention on the fission pathways discussed above. The first parameter we analyze is the hydrophobicity, i. e. the values $a_{\mathbf{TW}}$ and $a_{\mathbf{TH}}$. This parameter is related to the dissimilarity between surfactant tails and water. For most surfactants the hydrophobicity is solely a function of the length of the hydrocarbon chains [49]. The value is thus easily adjustable in experimental setups.

Second, we analyze the influence of the catalytic rate k_c— hence, the speed of metabolic turnover—on the division pathway of the nanocell. As we did not specify the molecular implementation of precursor and surfactant, it is conceivable that in an experimental setup, one can choose from a wide range of catalytic efficiencies.

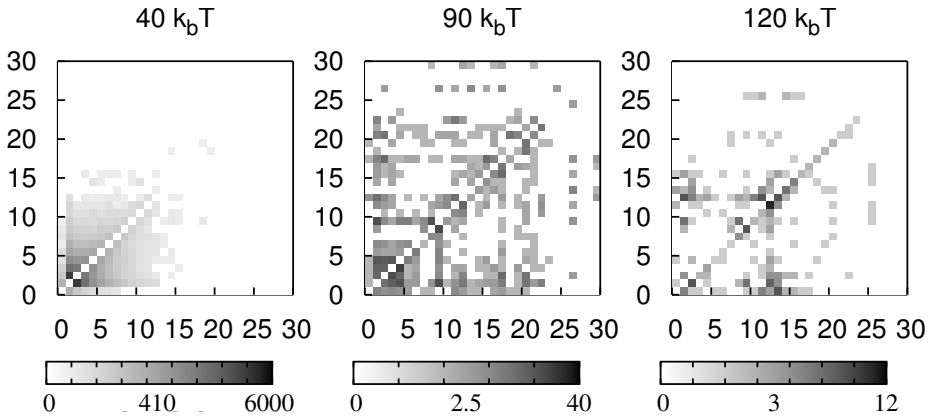


Figure 4.5: Histograms of nanocell size transitions for different hydrophobicities (see text on how transitions are defined). For a transition $S_n \rightarrow S_m$, n is ordered along the vertical, m along the horizontal axis. Colors indicate the number of occurences. Note that colors have been scaled by root functions to emphasize seldom events.

4.3.1 Influence of hydrophobicity on nanocell dynamics

For three hydrophobicities, 40, 90 and $120k_bT$, histograms of such transitions are shown in figure 4.5. The figures reveal a clear trend both in aggregate numbers as well as transition types. For hydrophobicity $40k_bT$, the system is almost entirely composed of single surfactants and small aggregates in bulk phase. 33.4 % of the transitions are dissociations and associations of two single surfactants. For the few bigger aggregates, transitions are distributed more or less homogeneously, i. e. surfactant dissociation is as likely as proper aggregate fission. Thus, for weak hydrophobicities, the system resembles a homogeneous solution without significant formation of structures. For $120k_bT$, on the other hand, the transition histogram looks completely different. Associations or dissociations of two isolated surfactants make only 3.7% of the transitions, for this parameter. The most prominent transition type is the turnover of a precursors within nanocells that range in size from 1 to 25 surfactants (20.6% of all transitions), represented by high values in the lower secondary diagonal. The absence of an upper secondary diagonal reveals that there are no surfactant dissociations except some between aggregate numbers 8 and 16. Due to the higher stability of aggregates, there are far less overall transitions than in the previous case (321 compared to 24743), which poses a problem when trying to obtain smooth histograms. Although there are distinct cases of proper fission events (e. g. $S_{25} \rightarrow S_{14} + S_{11}$, $S_{20} \rightarrow S_{11} + S_9$) it is hard to tell from the histograms whether such fission is more likely than single surfactant dissociation. Varying the hydrophobicity between these two extrema yields traits of both other histograms. One example is given in figure 4.5 for $90k_bT$.

As new precursors are constantly supplied and nanocells grow and divide over time, it is somewhat difficult to capture mean aggregate numbers of the assemblies. Nevertheless, these values are prominent characteristics in the study of micellar systems and their knowledge can help to get insight into the system under consideration. To gain aggregate numbers we compute the average value $m + n$ for all transitions of the form $S_{m+n} \rightarrow S_n + S_m$, i. e. we average aggregate numbers only in moments of fission or dissociation. The results can be regarded as the mean maximal aggregation number of nanocells. Figure 4.6 shows results for hydrophobicity values from $40k_bT$ to $120k_bT$

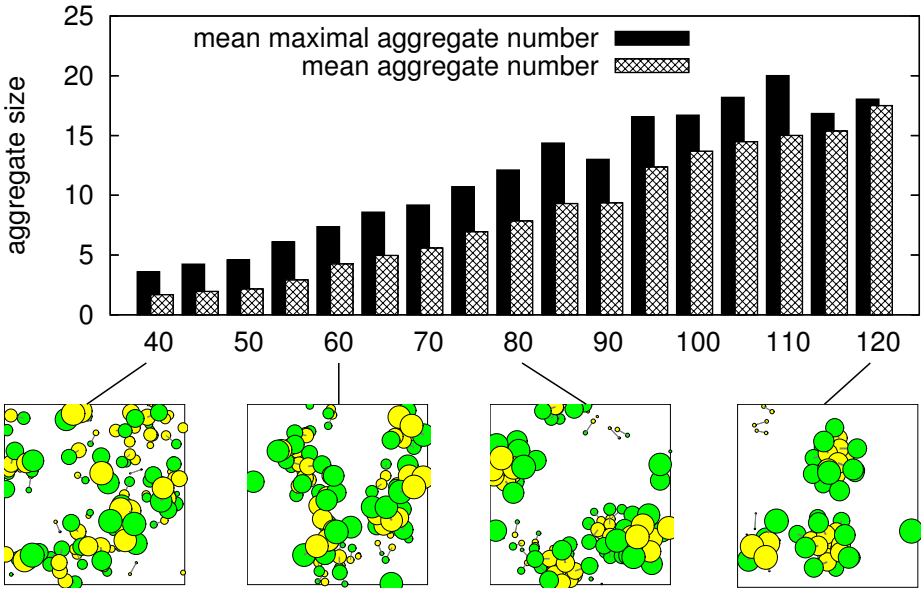


Figure 4.6: Mean aggregate and mean maximal aggregate numbers for hydrophobicity values between 40 and $120k_bT$. The latter averages only the size of nanocells which are actually going to divide, while the former averages all aggregates. System states have been averaged for $500\tau \leq t \leq 1000\tau$. Earlier states have been considered as transient. Below, parts of the final simulation states are shown for selected parameters.

together with a simple average of all aggregate sizes in the system. Both maximal and average values increase constantly from 1.68 (3.59) for $40k_bT$ to 18.04 (17.51) for $120k_bT$. It becomes apparent, that for very weak hydrophobicities most of the surfactants are either isolated in bulk phase or in very small assemblies. For strong hydrophobicities, aggregates are very distinct and single surfactants in bulk phase are rare. There is, however, no sharp boundary or phase transition between small sub-micellar assemblies and proper micelles, but rather a continuous transition.

Going back to the issue of nanocell division, we want to distinguish proper fission into nanocells of approximately equal size from dissociation of sub-micellar aggregates. The previous analysis revealed that one cannot use absolute aggregate numbers, as they exhibit a trend for stronger hydrophobicity. Therefore, we characterized each event of the form of equation 4.11 by the function

$$Q(m,n) = 1 - \frac{|m-n|}{m+n} \tag{4.16}$$

which denotes the relative fission quality. $Q(m,n) = 1$ for $m = n$, i. e. when the resulting nanocells are equal in size. Q becomes smaller as daughter cells become less alike. We have averaged Q as a function of the hydrophobicity over all fission events in the simulation runs discussed before. Mean and standard deviation are shown in figure 4.7. $\langle Q \rangle$ varies between 0.41 and 0.65 with no significant trend for weak or strong hydrophobicities. Moreover, standard deviation is very high. This reveals that fission into any two daughter cells is equally probable, no matter the ratio of their sizes. In terms of fission rates this finding can be written as

$$k^-_{m,n} =: k^-(m+n) \tag{4.17}$$

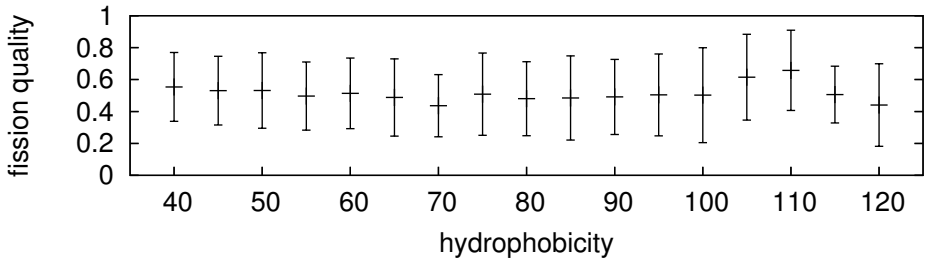


Figure 4.7: Mean quality of nanocell divisions. A value of 1 signifies division into daughter cells of equal size, while a value close to 0 results from single surfactant dissociation. Error bars denote standard deviations.

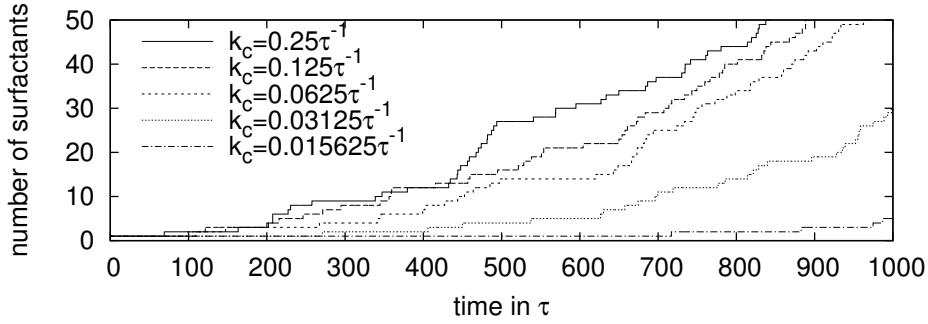


Figure 4.8: Number of surfactants as a function of time for different catalytic rates k_c. For slow metabolic turnover, the exponential shape of the auto-catalysis becomes apparent.

for the system under consideration.

4.3.2 Influence of the catalytic rate on nanocell dynamics

The catalytic rate has been varied from $2^{-6}\tau^{-1}$ to $4.0\tau^{-1}$ in exponential steps. Hydrophobicity has been set to $80k_bT$. Global surfactant dynamics are shown in figure 4.8. For slow metabolic turnover ($< 0.125\tau^{-1}$), the exponential shape of surfactant production becomes apparent, i. e. the constant precursor supply of precursors does not limit surfactant production over the simulated time span. This is tantamount to saying that unmetabolized precursor droplets are present throughout the whole simulation. Deceleration of the global dynamics is reflected in slower fission rates of individual nanocells. For example, we observed that the first fission event is retarded by 180τ on the average when catalytic rates are halved. For $k_c < 0.03125\tau^{-1}$, no fission occurs during the simulated time span.

Figure 4.9 shows mean aggregation numbers as a function of the catalytic rate. As one can see, slow metabolic turnover increases both maximal and average aggregation numbers (from 9.45 (5.27) for $0.5\tau^{-1}$ to 18.55 (9.66) for $0.03125\tau^{-1}$). The trend becomes less apparent for faster metabolic turnover ($1.0\tau^{-1}$ to $4.0\tau^{\cdot 1}$). For $k_c = 0.125\tau^{-1}$, the maximal aggregation number is slightly higher than a monotonic trend would imply—a fact that we relate to statistical deviations, as fission events are considerably rare for small catalytic rates. The increase in both aggregation numbers is a natural consequence of the decelerated metabolism: when the precursor is me-

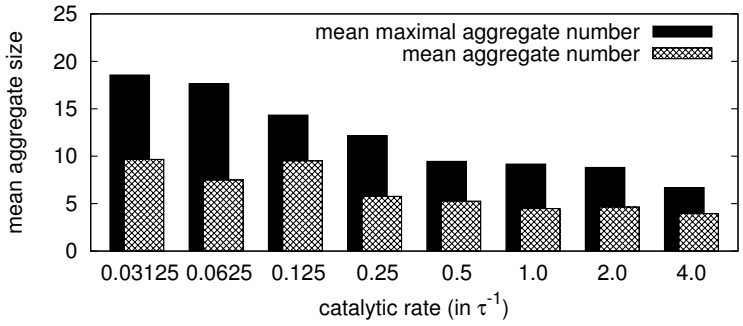


Figure 4.9: Mean aggregate and mean maximal aggregate numbers for catalytic rates between 0.03125 and $4.0\tau^{-1}$. Consult the caption of figure 4.6 on how aggregation numbers have been obtained.

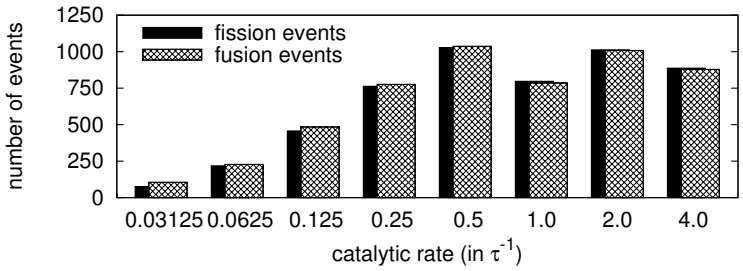


Figure 4.10: Number of overall fission and fusion events (transitions of the form of equations 4.13 and 4.14) as a function of the catalytic rate k_c.

tabolized slowly while its supply is held constant, the size of the hydrophobic core increases, and offers a bigger area for surfactants to attach. Hence, the maximal aggregate numbers increase.

For the above runs, the number of fission and fusion events has been measured (see figure 4.10). For all simulation runs, fusion and fission events are more or less balanced. This reveals that most of these events result from surfactant exchange with the bulk phase or from series of temporary fission and fusion during a single division process rather than from proper nanocell divisions. For low catalytic rates ($0.03125\tau^{-1} \leq k_c \leq 0.25\tau^{-1}$) the number of such balanced transitions falls significantly from 1028 for $k_c = 0.5\tau^{-1}$ to only 8 for $k_c = 0.03125\tau^{-1}$. As in the case of aggregation numbers, the trend in transition numbers can be related to the hydrophobic core: the more hydrophobic particles in the interior of a nanocell, the less dissociations occur on its surface. A strong anti-correlation between aggregation numbers and the number of fission/fusion events (with a correlation coefficient of -0.917) justifies this hypothesis.

It has to be pointed out, however, that the catalytic rate might affect nanocellular dynamics only during a certain transient time. It has been shown how the nanocellular dynamics depend on the ratio between metabolic turnover and precursor supply. Precursors are supplied by diffusion. Therefore, the rate of their incorporation into an individual cell depends on the overall concentration of nanocells. Once a critical cell concentration is reached, precursor incorporation might be slower than its metabolic turnover, which would undermine the above discussed effect. While the duration of such transient will depend on the ratio of precursor supply and turnover, dynamics after the transient might be little affected again. Ideally, simulations would be performed

in a homeostatic system, with an influx of precursor solution at one side and an outflux of reaction products at the opposite site of the system. Up to now, however, little is known about the performance of DPD in such open systems.

4.4 Discussion

In this paper we have presented an information-free nanocell based on a micellar system and a single auto-catalytic reaction that serves as metabolism. This simple system can be understood as a minimal self-replicating chemical system. As such, it denotes the boundary between pure auto-catalysis and a more complicated self-reproducing system which would also include inheritable information. We analyzed the dynamics of this nanocell using a dissipative particle dynamics approach. This simulation technique can cover the relevant time scale, while it has been shown to be still physically accurate compared to other simulation techniques. As a consequence, we have been able to perform analyses of the system in a level of detail, that has—as far as we know—not been reached before in the study of self-replicating entities.

The general replication cycle of micellar nanocells by metabolic turnover and division is very robust against changes in hydrophobicity and catalytic rates. It has been shown that the mean aggregation number of nanocells depends on the hydrophobicity of the surfactant (and precursor) as well as on the catalytic rate of the metabolism. For increasing hydrophobicity, a monotonic change in aggregation number with no sudden phase transition has been observed, ranging from a nearly homogeneous solution with only submicellar aggregates for weak hydrophobicities to the formation of distinct micelles in surfactant-free water for a very high hydrophobicity. The same monotonic increase in aggregation number could be observed for increasing catalytic rates, i. e. fast metabolic turnover.

It has been 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. There is neither a trend in the average size ratio of fission products nor in its variance.

Our work shows that the envisioned replication-cycle of nanocells—namely incorporation and turnover of precursor droplets followed by eventual aggregate division—is achievable over a wide range of parameters. In fact, there is no parameter combination for which the general replication cycle has been rendered impossible: although we have been able to decrease mean aggregation numbers and increase dissociation rates up to a point where the system obeys no clear structures, we could not observe that dissociation of single surfactants jeopardizes the growth of a otherwise stable nanocell.

We have demonstrated the feasibility of a self-replicating system in the absence of genetic information. Although such a system is not able to evolve towards higher complexity, it could have served as a functioning subsystem providing metabolism and embodiment for subsequent protocells of higher evolutionary complexity. 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 orchhestrated by coupling, such that each daughter cell of the dividing container is loaded with exactly one copy of the genomic information, one would obtain a true self-reproducing protocell with the ability to metabolize, divide and evolve.

Apart from prebiotic scenarios in which micelles are considered as possible ances-

tors of subsequent vesicle based organisms, such systems are explored in the context of so-called living technology, i. e. artificial systems that mimic life-like behavior. Our results suggest that the generic replication cycle of micellar nanocells is a robust basis for artificial life forms. We are currently exploring one design of such an artificial protocell in which genomic information is coupled to a micellar self-replicating system similar to the one presented here.

Acknowledgments

The authors would like to thank the members of the Complex Systems Lab for useful discussions. This work has been supported by grants FIS2004-0542, IST-FET PACE project of the European Community founded under EU R&D contract FP6002035 and by the Santa Fe Institute.

Chapter 5

Life-cycle of a minimal protocell

The final version of this article has been published, in Artificial Life, Vol. 13, Issue 4, published by The MIT Press.

Abstract

Cross-reactions and other systematic issues 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 (DPD) simulation method we are able to address the coupled diffusion, self-assembly, and chemical reaction processes, required to model a full life cycle of the protocell, the protocell being 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.

5.1 Introduction

The twilight zone that separates nonliving matter from life involves the assembly of and cooperation among different sub-components, 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, as well as intricate membrane structures [37]. 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 [147].

Several complementary designs of protocells have been proposed that differ in the actual coupling between their various internal components [148, 149, 69, 71, 52]. 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 nano-scale stochasticity can jeopardize appropriate molecular interactions or even what are the effects of molecular information carriers in helping accurate replication to occur. In this paper we present such a minimal 3D model that 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 a their interaction with water, the surfactants spontaneously self-assemble into micelles with the hydrophobic end 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 Fig. 5.1.2). 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 (Fig. 5.1.3). 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 (Fig. 5.1.4). 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 (Fig. 5.1.5), until it finally splits into two daughter cells (Fig. 5.1.6). 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.

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

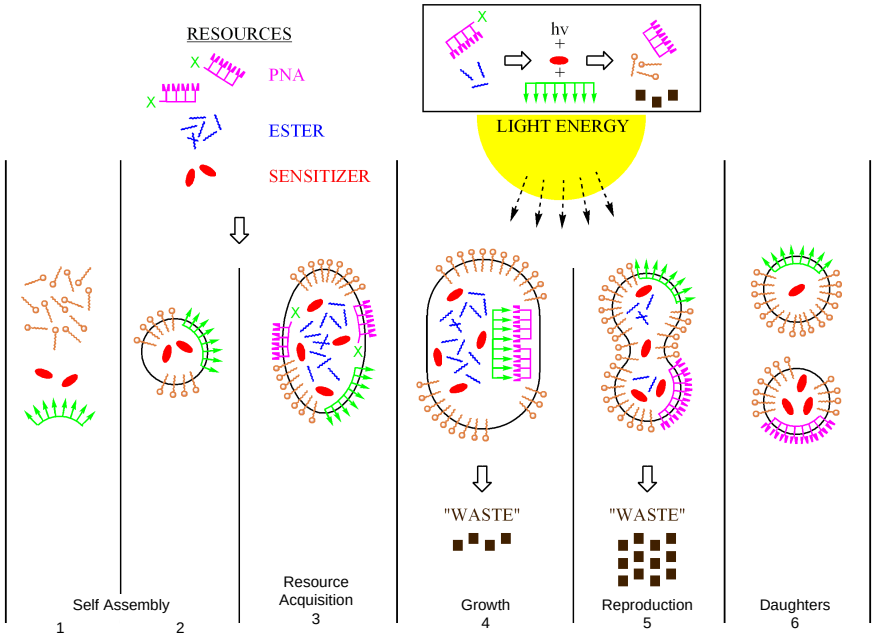


Figure 5.1: Schematic of the life cycle of the Los Alamos Bug: The system consists of surfactants, sensitizers, and a biopolymer that acts as a genome (1). The surfactants spontaneously self-assemble into a micellar container within which the sensitizer resides while the biopolymer sticks at the surface of the container—this forms a complete protocell (2). Resources (genomic oligomers, sensitizers and surfactant precursors in the form of esters) are added to the system and get incorporated into the container (3). The existing information carrier acts as a template for supplied oligomers to hybridize and effectively replicate the genome. 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 (5). Once the container reaches a critical size, it becomes unstable and divides into two daughter cells. This completes the life cycle of the protocell (6).

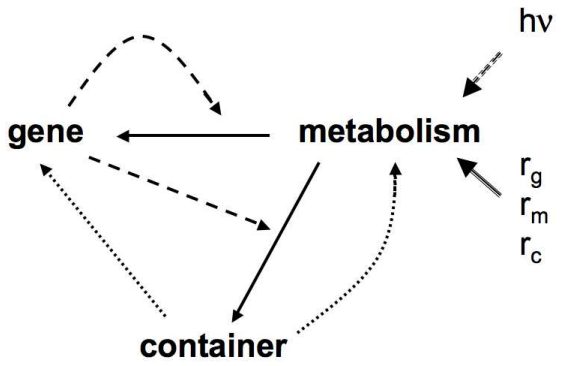


Figure 5.2: Functional coupling between container, metabolism and genome. Note how the gene catalyzes (dashed arrows) the metabolic production (solid arrows) of both gene and container building blocks. The container ensures a high local concentrations (proximity) and facilitates thermodynamic reaction 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, as well as sensitizers r_m.

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 in 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/free" interior of a micelle, the thermodynamics should change sufficiently to allow a dehydration reaction to occur whereby the oligomers become polymerized. Alternatively the water-lipid interface could either 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 Fig. 5.2.

5.2 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 various theoretical support, revision, and expanded capabilities [99, 144, 84], and has been applied to a number of biological systems such as membranes [90, 80], vesicles [117, 116], and micelles [89, 145]. Also chemical reactions have been incorporated into the DPD method [118, 104]. In the context of protocells, DPD has recently been applied to study a self-replicating micellar system [81, chapter 4 of this work]. The DPD formalism used in this work is the revised version from Groot and Warren [84] 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 individual atoms, but instead a group of atoms in a large molecules or several small molecules like water. This grouping allows one to average over fast de-

grees of freedom [150] 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.

5.2.1 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:

$$\frac{d^2\mathbf{r}_i}{dt^2}(t) = \frac{1}{m_i}\mathbf{F}_i\left(\mathbf{r}_i(t)\right) \tag{5.1}$$

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

$$\mathbf{F}_i = \sum_{j\neq i}\mathbf{F}_{ij} = \sum_{j\neq i}\left(\mathbf{F}^C_{ij} + \mathbf{F}^D_{ij} + \mathbf{F}^R_{ij}\right), \tag{5.2}$$

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

$$\mathbf{F}^C_{ij} = -\boldsymbol{\nabla}\phi_{ij} \tag{5.3}$$

$$\mathbf{F}^D_{ij} = -\eta\omega^D(r_{ij})\left(\mathbf{n}_{ij}\cdot\mathbf{v}_{ij}\right)\mathbf{n}_{ij} \tag{5.4}$$

$$\mathbf{F}^R_{ij} = \sigma\omega^R(r_{ij})\xi_{ij}\mathbf{n}_{ij} \tag{5.5}$$

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 $\mathbf{F}^X_{ij}$ now follows:

The conservative force $\mathbf{F}^C_{ij}$ 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

$$V_{ij}(r) = \begin{cases} \frac{a_{ij}}{2}(r-r_c)^2 & \text{if } r < r_c \\ 0 & \text{otherwise} \end{cases} \tag{5.6}$$

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 $\left|\mathbf{F}^C_{ij}(0)\right| = a_{ij}$ to $\left|\mathbf{F}^C_{ij}(r_c)\right| = 0$. The a_{ij}'s depend on the type of interacting particles—and are therefore the appropriate location to parameterize the model. In addition, different particles pairs could be given different values of r_c if one wants 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

$$V^s_{ij}(r) = \begin{cases} \frac{b}{2}\left(r-r_b\right)^2 & \text{if } (i,j) \text{ are bonded} \\ 0 & \text{otherwise} \end{cases} \tag{5.7}$$

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

$$V_{ijk}^{\theta}(\theta_j) = \frac{1}{2} c_{ijk} \left(\theta_j - \theta_{eq}\right)^2 , \tag{5.8}$$

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 eq. (5.4) 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. 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 allows the derivation of thermodynamic properties. As shown by Español and Warren [99], DPD satisfies the fluctuation dissipation theorem if and only if the weighting functions ω^D and ω^R obey the relation

$$\omega^D = (\omega^R)^2. \tag{5.9}$$

In agreement with the DPD standard, we set

$$\omega^D(r) = (\omega^R(r))^2 = \left[2(1 - \frac{r}{r_c})\right]^2 . \tag{5.10}$$

If relation (5.9) 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

$$k_bT = \frac{\sigma^2}{2\eta}. \tag{5.11}$$

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 [131]. 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 [84]. 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 (5.1) 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 5.3.

5.2.2 Incorporation of chemical reactions

A rigorous way to extend the DPD formalism to account 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 a way that reactions result in local heating or cooling. This approach is highly non-trivial 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 employed between every two DPD updates. This algorithm was used in earlier Brownian Dynamics simulations [108]. Our choice can be legitimated by the observation that chemical reactions are relatively rare events in our simulations and thus their impacts are rapidely equilibrated by the DPD thermostat.

Chemical reactions in our system occur between two reactants and fall into two different classes: Each reaction has a given rate for spontaneous occurrence k_s.

$$\begin{aligned} \text{transformation: } \mathbf{U} &\longrightarrow \mathbf{V} \\ \text{polymerization (of different bases or oligomers): } \mathbf{X}+\mathbf{Y} &\longrightarrow \mathbf{XY} \end{aligned}$$

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

$$k = k_s + \sum_{\mathbf{C}} f_{cat}(r_{\mathbf{C}}) \tag{5.12}$$

with

$$f_{cat} = \begin{cases} k_{cat}\left(1 - \frac{r_{\mathbf{C}}}{r_{cat}}\right) & \text{if } r_{\mathbf{C}} < r_{cat} \\ 0 & \text{else} \end{cases} \tag{5.13}$$

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} is 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 Sec. 5.6.

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 non-equilibrium 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.

5.2.3 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_bT) for the water and amphiphiles have been taken from [89] (where surfactants are modeled as dimers as well):

a_{ij}	**W**	**H**	**T**
W	25	15	80
H	15	35	80
T	80	80	15

Bond parameters are $b = 150k_bT$ and $r_b = 0.5r_c$. These parameter values were originally used to analyze polymer surfactant interactions. Later, the phase diagram for varying surfactant concentrations was analyzed [145].

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: $a_{\mathbf{Z}j} = a_{\mathbf{T}j}$.

Genes

The gene is modeled as a strand of covalently bound monomers (**A** and **B**) with hydrophobic anchors (**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 because we want to have a non-charged backbone for the gene molecule to enhance its lipophilic properties. For details, see [71]. 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 **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 and cooperative 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 in the former section. We now implement and test several alternative representations of such base affinities as discussed below.

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

$$\mathbf{F}_{\mathbf{AB}}^{C_1}(\mathbf{r}) = \mathbf{F}_{\mathbf{AB}}^{C}(\mathbf{r}) + \begin{cases} a_2\,(r_{c_2} - r)\,\mathbf{n} & \text{if } r < r_{c_2} \\ 0 & \text{else} \end{cases} \tag{5.14}$$

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 5.3.4). 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 $\mathbf{F}^{C_1}_{\mathbf{AB}}$ compared to $\mathbf{F}^{C}_{\mathbf{AB}}$ 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}.

directed "radial" 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 rather than radial interactions. Hence, in the second representation, we introduce directed attraction parallel to the $\mathbf{A}-\mathbf{T}$ and $\mathbf{B}-\mathbf{T}$ axes, respectively. Here, we replace $\mathbf{F}^{C}_{\mathbf{AB}}$ by

$$\mathbf{F}^{C_2}_{\mathbf{AB}}(\mathbf{r}) = \mathbf{F}^{C}_{\mathbf{AB}}(\mathbf{r}) \; + \; \begin{cases} a_2(r_{c_2} - r)\,(\mathbf{z}\cdot\mathbf{r})\,\mathbf{n} & \text{if } r < r_{c_2} \\ 0 & \text{else} \end{cases} \tag{5.15}$$

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 $a_1 = 35k_bT = a_{\mathbf{AA}} = a_{\mathbf{BB}}$ We call this interaction "radial", because the strongest attraction will be radial towards the center of the micelle, once the PNA strand is attached to the surface of the micelle.

directed "tangential" attraction: The third representation is similar to the second, except that attraction is now perpendicular to the backbone and to the $\mathbf{AT}$ (or $\mathbf{BT}$) axis. The force is attractive towards one side of the PNA and repulsive towards the other—hence, it is the only implementation that catches the directionality of the molecule:

$$\mathbf{F}^{C_3}_{\mathbf{AB}}(\mathbf{r}) = \mathbf{F}^{C}_{\mathbf{AB}}(\mathbf{r}) + \begin{cases} a_2(r_{c_2} - r)\left(\frac{(u+\mathbf{v})\times\mathbf{z}}{|(u+\mathbf{v})\times\mathbf{z}|}\cdot\mathbf{r}\right)\mathbf{n} & \text{if } r < r_{c_2} \\ 0 & \text{else} \end{cases} \tag{5.16}$$

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 $35k_bT$.

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 (eq. 5.8) within the PNA strand: angles of interconnected strands prefer to be stretched out ($\theta_0 = 180^o$, $c_{ijk} = 10k_bT$). 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 5.1 summarizes the chosen set of parameters.

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

$$\mathbf{TT} + \mathbf{Z} \longrightarrow \mathbf{HT} + \mathbf{Z} \tag{5.17}$$

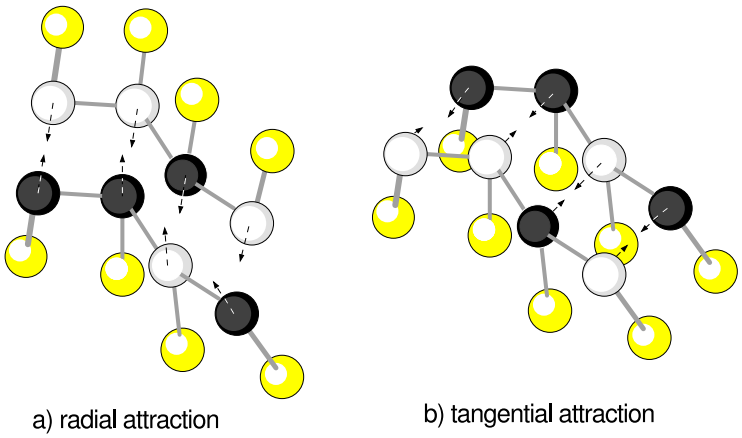


Figure 5.3: Hybridization complexes for radial (a) and tangential (b) attraction between complementary bases. Bases are shown as black and white beads, hydrophobic anchors in yellow. Arrows denote the direction of strongest attraction.

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
A	15	35	80	(*)	35	80
Z	80	80	15	80	80	15

Table 5.1: 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 $a_{\mathbf{AB}}$. Three different force fields have been considered for such interactions. See the text for details.

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 photo-activity of the sensitizer.

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

$$\begin{aligned} \mathbf{A} + \mathbf{B} &\longrightarrow \mathbf{AB} \\ \mathbf{A} + \mathbf{A} &\longrightarrow \mathbf{AA} \\ \mathbf{B} + \mathbf{B} &\longrightarrow \mathbf{BB} \end{aligned} \tag{5.18}$$

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

$$k_{ij} = \frac{1}{2} k_{max} \left(\frac{\mathbf{u}_i + \mathbf{v}_i}{2} \cdot \frac{\mathbf{u}_j + \mathbf{v}_j}{2} + 1 \right) \tag{5.19}$$

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

5.3 Results

We use the model discussed above to study various aspects of the life cycle of the Los Alamos Bug as depicted in Fig. 5.1. In particular, our simulations address the spontaneous self-assembly of protocells (Fig. 5.1, frames 1&2), the incorporation of resources (frames 2&3), the metabolic growth of the protocell (frames 4&5), template reproduction, and finally fission into two daughter cells (frames 5&6). 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 is noted for each individual simulation run. We integrate equation (5.1) numerically with the DPD variant of the leapfrog Verlet integrator discussed in [84] with $\lambda = 0.5$ and a numerical step width of $\Delta t = 0.04\tau$.

5.3.1 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 5.1 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 Fig. 5.4, 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 sub-micellar aggregates in the bulk phase. These observations qualitatively fit theoretical and experimental results [see e.g. 49].

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 [80]. Our calculation is based on sodium alkanesulfates as these are well studied surfactants with properties similar to the fatty acids used in the real chemical implementation. Table 5.2 lists the critical micelle concentration (CMC), i.e. the minimal concentration at which micelles spontaneously form. The table also gives the mean aggregation number and the volume of these molecules. 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 ($V_{\mathrm{H_2O}} = 30\text{Å}^3$) – we get the so-called *coarse graining parameter*

$$N_m = \frac{\frac{1}{2}V_{\mathrm{surf}}}{V_{\mathrm{H_2O}}} \tag{5.20}$$

that 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

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

Table 5.2: Data for sodium alkanesulfate surfactants with varying tail length. For each surfactant, CMC and mean aggregation number are listed from [151]. The molecular volume is estimated from the number n of carbon atoms using the formula $V = 27(n+1)$Å^3 [49] plus a constant 88.51Å^3 for the sulfate group (whose value is derived from the molecular mass ($98.08 g/mol$) and density ($1.84 g/cm^3$) of sulfuric acid). The coarse graining parameter N_m, the physical length scale r_c, and the total surfactant concentration are the interpretation of model parameters in case that the model dimer represents the respective surfactant. Finally, the fraction of micellized surfactant is the prediction of the closed association model for the respective surfactant and the calculated concentration [49].

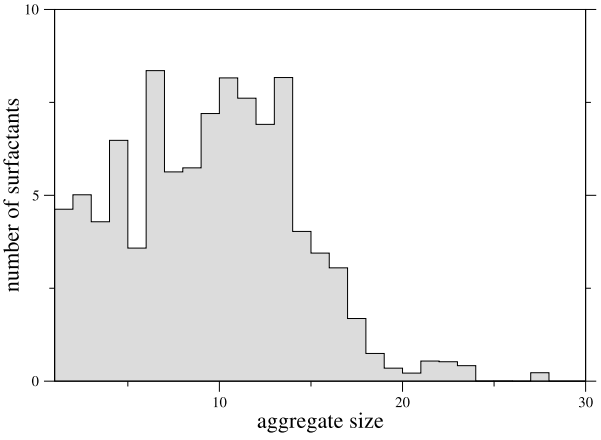


Figure 5.4: Micellar size distribution for a system containing 2.9 water beads and 0.05 surfactant dimers per unit cube. To obtain the aggregate size histograms from a system state, every two surfactants whose **T**-beads are separated by less than r_c are considered to belong to the same aggregate. 20000 systems states of an equilibrated system ($200\tau < t < 1000\tau$) are averaged in the shown distribution.

N_m molecules. Therefore, the physical length scale r_c resolves to

$$r_c \equiv (\rho N_m V_{\mathrm{H_2O}})^{1/3}. \tag{5.21}$$

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

$$1 r_c^{-3} \equiv 55.56 mol/\rho N_m. \tag{5.22}$$

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 to estimate the concentration of free lipids in the bulk phase, because our simulations do not yield a sharp distinction between free lipids—i.e. 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 Fig. 5.4 we get an average of 22.9 free surfactants in the bulk phase out of 98 lipids in the total volume, i.e. 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 to the prediction of the closed association model [49]. 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 $\frac{d[S]}{d[S]_{\mathrm{total}}}\Big|_{\mathrm{CMC}} = \frac{dN[S_N]}{d[S]_{\mathrm{total}}}\Big|_{\mathrm{CMC}} = 0.5$, one can calculate the fraction of micellized surfactant for any total surfactant concentration $[S]_{\mathrm{total}} = [S] + N[S_N]$. The respective ratio $N[S_N]/[S]_{\mathrm{total}}$ is also given in table 5.2.

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 pre-

dictions 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 alkanesulfate surfactant parameters, which are the most well studied surfactants in the scientific community. Given the easy availability of relevant parameters for alkanesulfate 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 [89, 145, 81].

Next, we analyzed a ternary mixture of water, surfactant, and oil. In the system described above, we exchanged an additional 0.1 water beads 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 useful and some less useful as a protocellular container substrate, see e.g. the recent summary discussion in [152].

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 falls by a factor of four or more. This is consistent with simulation results from earlier studies of a similar surfactant-precursor-water system [81, chapter 4 of this work]. However, a more systematic DPD investigation is necessary to address the dynamics, stability, and size distribution issue in this context.

5.3.2 Self-assembly of the protocell

In this section, 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, 4 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, i.e. a micelle that is loaded with sensitizer and which has PNA attached to its surface and whose nucleotides are exposed to the aqueous phase (see Fig. 5.5). 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

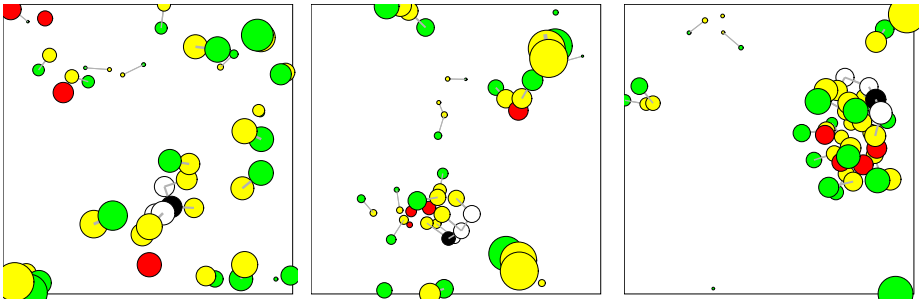


Figure 5.5: Self-assembly of the protocell from a random initial condition. The diagrams show the state of the system at times a) $t = 0\tau$, b) $t = 4\tau$, and c) $t = 10\tau$. Surfactants are shown in green (head bead) and yellow (tail), the sensitizers in red, the PNA backbone in yellow and the PNA monomers in black and white.

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.

5.3.3 Replication of the Container

The dynamics of a surfactant-precursor-water system similar to the one under consideration has been studied in detail in [81, chapter 4 of this work]. Considering precursor and surfactant kinetics, the formerly 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. [153] in 1996 reproduced the micellar self-replication experiments of Bachmann et al. [59]. In 1998 and 2000 Mayer and Rasmussen developed an extended lattice polymer approach [154, 155] 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 [59]. 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 4 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 slower 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 systems

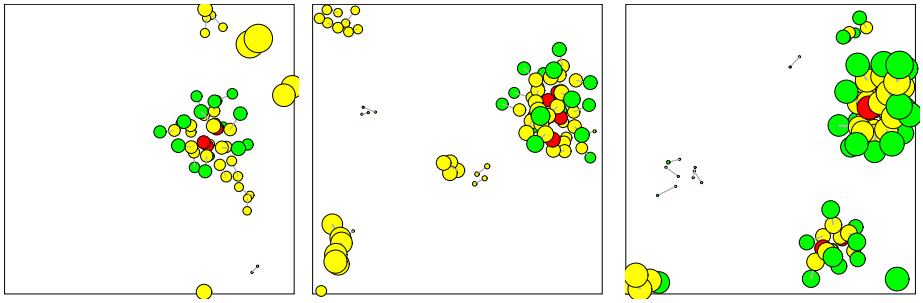


Figure 5.6: Replication dynamics of the container: precursors are fed into the system far from the micelle at the (periodically reflected) edge of the system space. They form droplets in the aqueous phase, which are absorbed by the protocells as a whole. Protocells grow by incorporation of precursors. After a critical amount of precursor is transformed into surfactant, the assembly loses its stability and splits in two daughter cells (right frame).

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.

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 rod-like. 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 Fig. 5.6). 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 to 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 4 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 precur-

sors happened significantly faster than the precursor supply, nearly each droplet was transformed separately. When only few precursors were absorbed at once (i.e. a small droplet), the micelle responded by rejecting several surfactants into the bulk phase. Such loose surfactants then formed sub-micellar 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 4 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 3 sensitizers and the other with 9 surfactants and 1 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.

5.3.4 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 illdefined than the ligation process, which is why our discussion of the replication of the genome is divided in two consecutive steps: hybridization and ligation. Please 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.

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 2 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 $\mathbf{T}$ 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 towards 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/orientation for proper hybridization. The location/orientation was varied to match the different hybridization cases studied. In the case of directed radial 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

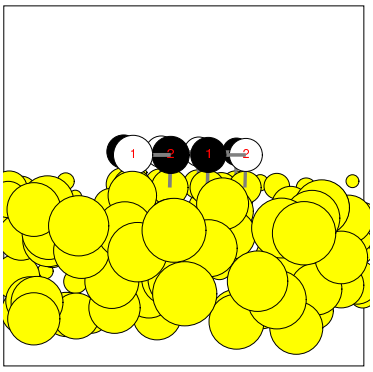


Figure 5.7: Initial setup of the hybridization simulations. The system is initialized with an oil-layer that mimics the oil-water interface of a two-phase system. A four-mer template and two complementary dimers are placed at the interface so that they form a hybridization complex. The association time of such hybridization complexes is measured for different PNA implementations and attraction forces.

the case of tangential attraction, both the template and the oligomers span the interface region as shown in Fig. 5.7.

In the system modeled, we only had two different types of monomers ($\mathbf{A}$, $\mathbf{B}$) with $\mathbf{A}$ and $\mathbf{B}$ being complementary to each other, but not self-complementary. All different 4-mer templates excluding symmetric configurations were used (e.g. $\mathbf{AAAA}$, $\mathbf{AAAB}$, $\mathbf{AABB}$, $\mathbf{ABAB}$, and $\mathbf{ABBA}$) and for each different template only the proper complementary dimer oligomers were used. The different 4-mer 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—i.e. 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 sections 5.2.3 *Genes*, 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 $a_{\mathbf{AB}}$ 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 Fig. 5.8.

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 ($\mathbf{AAAA}$ is the most, while $\mathbf{ABBA}$ is the least stable sequence). However, these differences were rather small.

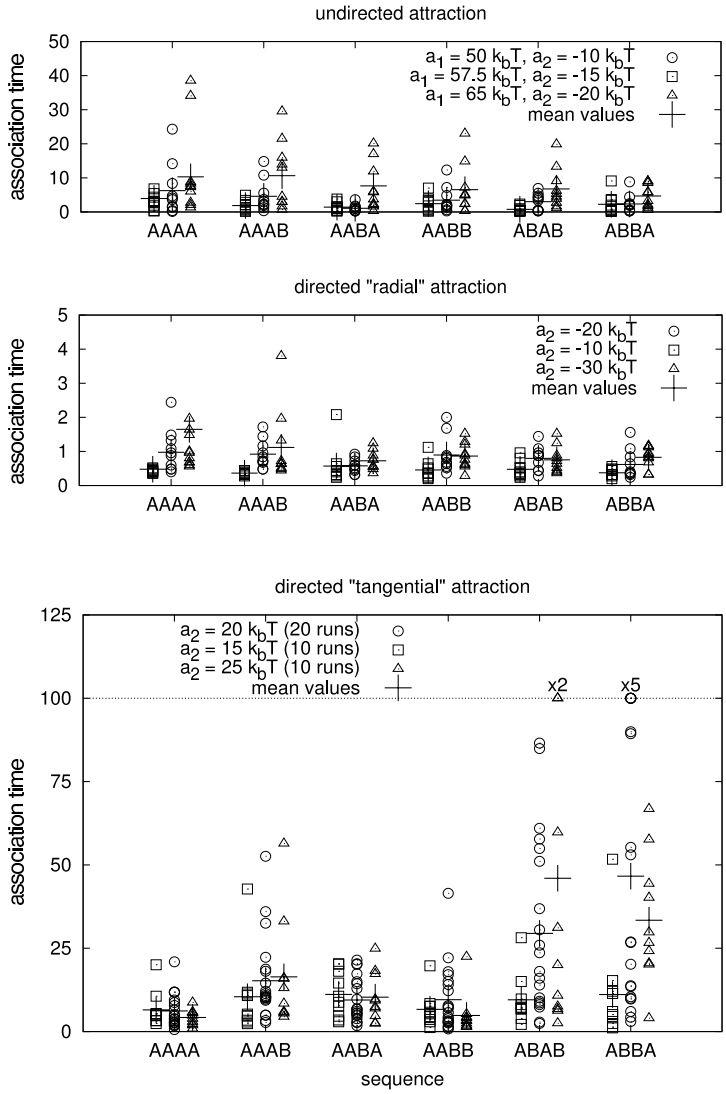


Figure 5.8: The association times (i.e. the time until the initially hybridized complex becomes dehybridized) for different PNA template sequences of length four using a) undirected, b) radial, and c) tangential attraction. For each implementation, three different attraction strengths are compared, as given in the legend for each figure. a_1 denotes the coefficient of the repulsive part, 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 minimal values 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.

directed radial attraction: For directed radial attraction, the mean association times ranged from 0.45τ for $a_2 = -10k_bT$ to 0.98τ for $a_2 = -30k_bT$ ($a_1 = a_{\mathbf{AB}} = 35k_bT$ 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 nature of the hybridization of the PNA for the radial 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 (5.15), the attraction between two PNA molecules on the interface is marginal and the association time is essentially a matter of diffusion.

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. $\mathbf{AAAA}$ and $\mathbf{AABB}$), hybridization is usually less stable than for sequences without equal bases at terminal positions ($\mathbf{ABBA}$ and $\mathbf{ABAB}$). The association time of sequences with only one such dimer lies between the values of 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 $\mathbf{AAAA}$ for the undirected, and $\mathbf{ABBA}$ for the tangential attractions (interaction parameters are given in the caption of Fig. 5.9). Simulations are performed for $0\tau \leq t \leq 1000\tau$. The resulting time series are shown in Fig. 5.9.

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 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 Fig. 5.9). 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 Fig. 5.9 that the undirected attraction performs better on average, it is only during the "locked in" period that the desired

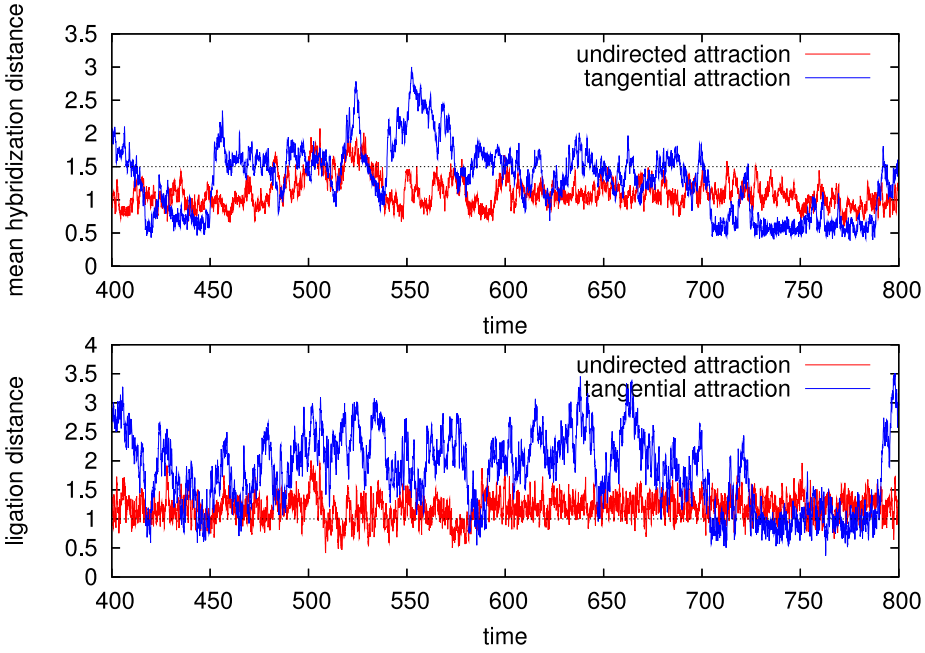


Figure 5.9: Mean hybridization (upper panel) and ligation distance (lower panel) for the PNA templates (and corresponding oligomers) **AAAA** using undirected attraction with $a_1 = 65k_bT$, $a_2 = -20k_bT$ (red) and for **ABBA** using tangential attraction with $a_1 = 35k_bT, a_2 = 40k_bT$ (blue). By 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.

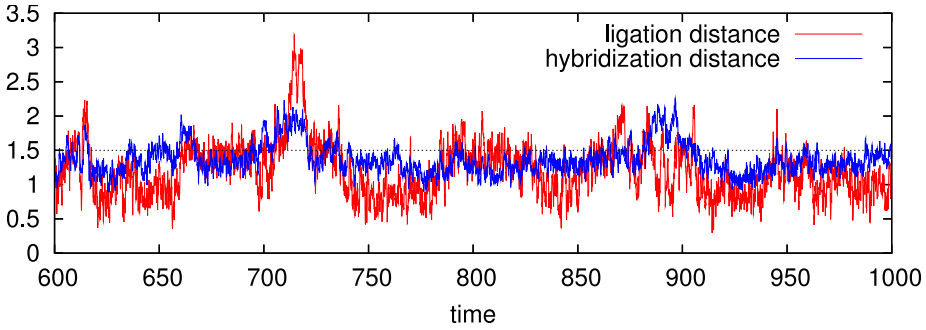


Figure 5.10: Hybridization and ligation distances of PNA template and complementary oligomers in water. For PNA, tangential directed attraction with $a_2 = 40k_bT$ has been used. The nucleotide sequence is **ABBA**.

reactions 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) that the interface contains a lower water concentration than the bulk phase; (iii) that the interface might directly act as a catalyst for the amide bond formation; and (iv) that 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 Fig. 5.10. 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 $0.34r_c$. 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 mainly attract each other due to the hydrophobic interactions between the tail beads of these strands rather than 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, rather than 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 vary the ligation probability depending 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 very

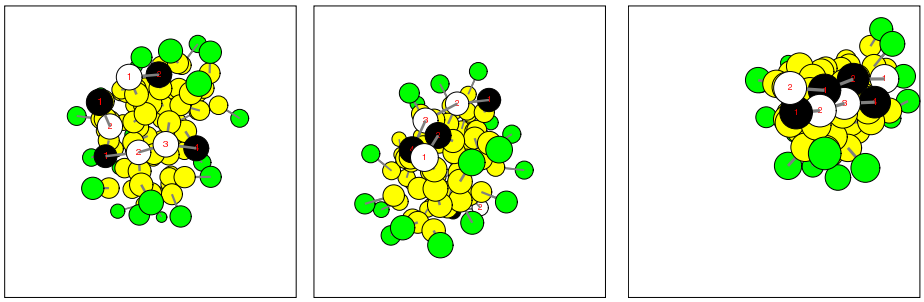


Figure 5.11: The three steps of template directed replication: a) Template (**ABBA**) and oligomers (**BA** and **AB**) diffuse over the surface of the micelle, b) oligomers form a hybridization complex with the template strand, and c) oligomers polymerize to yield a complementary copy of the template.

few sequences that do not offer any possibility for loop formation (such as **AAAA**).

Ligation

To study the polymerization reaction, a four-mer template strand and two complementary dimers are placed randomly on the surface of a loaded micelle (20 surfactant, 20 precursors) within a system of size $(10r_c^3)$ and total density $\rho = 3.0r_c^{-3}$. As the last section identified **ABBA** to form the most stable hybridization complex, we restrict polymerization experiments to this particular sequence using the PNA representation with tangential directed attraction (see Fig. 5.11).

Of the performed simulations, 8 out of 10 generated proper template directed ligation, while the remaining 2 reactions occur spontaneously in the absence of the template strand and define the expected background reaction [156]. 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, 6 out of 8 runs lead 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 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.

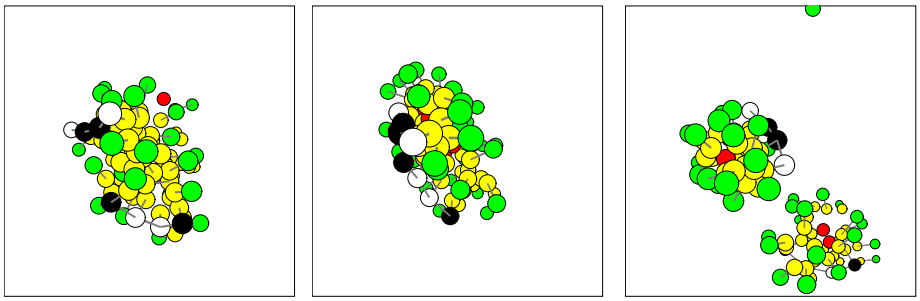


Figure 5.12: The division of the whole protocell completes the life cycle of the Los Alamos Bug. A mature protocell is loaded with precursor molecules, sensitizers, and two complementary PNA strands. During the metabolic turnover of precursors, the aggregate elongates and divides. Both PNA strands and sensitizer molecules tend to distribute evenly among the daughter cells, when only few sensitizers are present.

5.3.5 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 Fig. 5.12. In addition to what was discussed in section 5.3.3, 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, 4 to 8 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 Fig. 5.12.In all cases, metabolic turnover initiated the division of the aggregate at times of between 50 to 100 τ 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 rod-like 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 (eq. (5.8)) of the PNA strands, the aggregate tends to elongate in a direction that is parallel to the PNAs long axis.

Using only 4 sensitizers, the distribution of sensitizers and PNA among the daughter cells was rather diverse: in one out of 10 runs, all sensitizers and templates remained in one of the fission products, while the other consisted of only 11 surfactants. In 7 of

the runs the partition was nearly even: both sensitizers and templates were equally distributed among the two daughter 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 8 sensitizers instead of 4 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 that 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 6 sensitizers, proper division can still be observed, but the results are less reliable than in the case of 4 sensitizers. For 6 sensitizers, equipartition of sensitizers was only achieved in one out of five simulations. The other runs lead to empty micelles or a situation where one of the daughter micelles has only one sensitizer bead. Equipartition of PNA could not be achieved for the cases with either 6 or 8 sensitizer beads.

5.4 Discussion

Because of the inherent simplifications of the aggregated DPD simulation technique and due to the inherent complexity of our protocell system, accurate predictions of neither the detailed kinetic nor 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, reaction kinetics, combined with these and other DPD studies, hopefully can address the quantitative prediction issues in a more complete manner [147].

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 3-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, volume-wise

they will tend to grow ever more rapidly the larger they become prior to being absorbed. The spurt-like 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 introduce an impact on 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 [157, 158].

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 counter-intuitive 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: (i) What is the effect of heating the whole system in order to de-hybridize the gene templates? Obviously, the lipid aggregate has to be more heat tolerant than the gene duplex. (ii) What is the effect of defining the gene duplex as the photo-catalyst as in the originally proposed protocell design [71]? In our simulations, the sensitizer has been assumed to do the photo-fragmentation 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, e.g. as one of the lipophilic anchors? (iii) 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. [159]? 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? (iv) What new issues arise when the protocell goes through more than one generation of its life cycle, e.g. 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 [160] and reaction kinetics studies [161] as well as experimental findings as they arise.

5.5 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 finds are: (a) 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, which was not anticipated. (b) As anticipated, there is a tradeoff between the lipophilic strength of the genetic backbone that makes it stick to the aggregate and its ability to easily hybridize with a complementary string. (c) 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 pf hybridization attraction. (d) 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. (e) As anticipated, the stability of two full complementary gene strings is much higher than a gene template and two complementary unligated gene pieces. (f) We observe that the template directed replication rate is dependent on the monomer component sequence and not only on the monomer component composition. (g) 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

The authors would like to thank the members of the Barcelona Complex Systems Lab as well as members of the Los Alamos protocell team for useful discussions. This work is supported by the Programmable Artificial Cell Evolution (PACE) project funded by the 6^{th} European Union Framework Program under contract FP600203 and the Los Alamos sponsored LDRD-DR Protocell Assembly (PAs) project.

5.6 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 to a number taken from a suitably normalized pseudo-random 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
```

Chapter 6

Toward integrated simulations of a minimal protocell

6.1 Introduction

Earlier studies have shown that surfactant-coated oil droplets are able to divide into two daughter droplets, if the internal oil is successively transformed into additional surfactant under the action of a metabolic reaction [81, chapter 4 in this work]. It has also been shown that template directed replication of single stranded biopolymers at the oil-water interface of such droplets is achievable [88, chapter 5 in this work]. It was found that even a very simple model representation of the biopolymer gives rise to the emergence of a molecular fitness function: due to numerous geometric interactions, different base sequences differ significantly in their ability to form hybridization complexes that are stable enough for template directed replication. It has further been shown that reliable gene hybridization occurs only a the oil-water-interface provided by the container. This is an example of the catalytic couplings of subsystems in the design of this particular protocell. The reverse coupling of the container dynamics to the container of the cell had not been shown in the earlier studies [88, chapter 5 in this work].

This work has shed light on the dynamics of the container and the genome of such minimal protocells and analyzed crucial steps of the protocellular life-cycle in detail. However, an integrated simulation that combines all steps the life-cycle in a single simulation had not been achieved so far. Here, we present work that leads to such integrated simulations and analyze the additional problems that occur when the individual subsystems are integrated.

In particular, the following issues have not been addressed yet:

1. The replication of the biopolymer needs to be completed by melting the double stranded genome after replication, to allow for their reuse as templates. As indicated in chapter 5, the temperature cycle needs to be gentle enough, to preserve the integrity of the aggregate. Section 6.2 deals with temperature changes in DPD and analyzes the system response to temperature changes.

2. The catalytic rate of the sensitizer molecule needs to be coupled to the sequence and configuration of the biopolymer. Specifically, the sensitizer should catalyze the metabolic turnover reaction only in the presence of a double stranded

biopolymer. This is achieved in section 6.3

These studies raise an unforeseen issue in the original design of the protocell as presented in Refs. [71] and [88, chapter 5 of this work]: As will be seen, the protocell needs to employ double stranded polymers instead of single strands, or several copies of the biopolymer (depending on the position in the life-cycle) in order to produce viable offspring. Section 6.4 presents a revised life-cycle based on these findings.

As we will show in section 6.5, this introduces product inhibition during gene replication in our simulation.

6.2 Temperature scaling in DPD

The DPD algorithm has been comprehensively described in chapter 2. Here, we only give the stochastic differential equation of motion. In order to define parameters and units for the following derivation:

$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{1}{m_i}\sum_{j\neq i}\left(\mathbf{F}^{\mathrm{C}}_{ij} + \mathbf{F}^{\mathrm{D}}_{ij} + \mathbf{F}^{\mathrm{R}}_{ij}\right), \tag{6.1}$$

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

$$\mathbf{F}^{\mathrm{C}}_{ij} = \begin{cases} a_{ij}(1-\frac{r_{ij}}{r_c})\hat{\mathbf{r}}_{ij} & \text{if } r_{ij} < r_c \\ 0 & \text{otherwise} \end{cases} \tag{6.2}$$

$$\mathbf{F}^{\mathrm{D}}_{ij} = -\gamma\omega^{\mathrm{D}}(r_{ij})\left(\hat{\mathbf{r}}_{ij}\cdot\mathbf{v}_{ij}\right)\hat{\mathbf{r}}_{ij} \tag{6.3}$$

$$\mathbf{F}^{\mathrm{R}}_{ij} = \sigma\omega^{\mathrm{R}}(r_{ij})\zeta_{ij}\hat{\mathbf{r}}_{ij}. \tag{6.4}$$

DPD commonly works in reduced units where $m = 1$, $r_c = 1$, and $k_bT = 1$, denote the mass, length, and energy scale. From these, the natural unit of time follows as

$$\tau = r_c\sqrt{m/k_bT}. \tag{6.5}$$

According to Ref. [84], the temperature in the DPD method is given by the relation

$$k_bT = \frac{\sigma^2}{2\gamma} \equiv 1. \tag{6.6}$$

In the above equations, the terms ω^{R}, ω^{D}, r_{ij}/r_{c}, and $\hat{\mathbf{r}}_{ij}$ are all dimensionless. $(\hat{\mathbf{r}}_{ij}\cdot\mathbf{v}_{ij})$ has the units of a velocity. Since the random parameter ζ_{ij} has the unit $\tau^{-1/2}$, we get the following units for the individual force parameters a, γ, and σ:

$$[a] = k_bT/r_c \tag{6.7}$$

$$[\gamma] = (k_bT)^{1/2}m^{1/2}/r_c \tag{6.8}$$

$$[\sigma] = (k_bT)^{3/4}m^{1/4}/r_c^{1/2} \tag{6.9}$$

6.2.1 A simple method for temperature scaling in DPD

As interaction potentials in DPD are effective potentials and as such influenced by the temperature. A rigorous way of implementing a temperature cycle in DPD would require to calibrate system interactions, namely $\mathbf{F}^{\mathrm{C}}(r)$, to systems at each temperature of interest. While this is quite an endeavor on its own, one would then need to also derive

a temperature scaling for the dissipative and random interactions and their parameters σ and γ which are usually treated as (uncalibrated) simulation parameters, but know to influence fluid properties [126]. Improved calibration procedures [100, 162, 163] will hopefully be developed further to address this issue. In the meantime, we employ a simple though less rigorous temperature scaling procedure.

We intend to increase the temperature T by a factor of ϕ, such that $T' = \phi T$ (we employ a primes to refer to the heated system in physical units). The most intuitive way to achieve tempering in the DPD framework is to simply change the amplitude of the random force by setting:

$$\sigma' = \phi^{1/2}\sigma. \tag{6.10}$$

According to Eq. (6.6), this will result in a temperature $T' = \phi T$. Since DPD commonly works in reduced units, one might want to keep the temperature in the simulation fixed to $k_B\tilde{T} = 1$, and change all units accordingly:

$$\tilde{T} = \phi^{-1}T' = T \tag{6.11}$$

In this procedure, we first scale σ as given in the previous paragraph, and then change the energy unit $k_b\tilde{T}$ by a factor of ϕ^{-1}. The units of length and mass are unaffected by this scaling, but the unit of time scales as

$$\tilde{\tau} = r_c\sqrt{m/k_b\tilde{T}} = r_c\sqrt{m/k_b\phi^{-1}T} = \phi^{1/2}\tau. \tag{6.12}$$

Note that these scaling relations are easy to incorporate in the DPD algorithm, since they do not affect the length scale of the method. If they would, one would need to change the particle number as well as the model representation of molecules during the course of the simulation. In contrast, to implement the above scaling relations, one simply needs to alter the DPD parameters a, γ, and σ, according to

$$\tilde{a} = \phi^{-1}a \tag{6.13}$$

$$\tilde{\gamma} = \phi^{-1/2}\gamma \tag{6.14}$$

$$\tilde{\sigma} = \phi^{-3/4}\sigma' = \phi^{-1/4}\sigma. \tag{6.15}$$

It has to be warned that this simplistic temperature scaling is likely to introduce artifacts when extrapolating over phase transitions present in the modeled systems. We validate our method by comparison with experimental data.

Comparison with experimental data

The left panel of Fig. 6.1 shows simulation results of the self-diffusion constant of water as a function of the temperature for both the scaling of σ (squares) according to Eq. (6.10), as well as the rescaled system given by Eq. (6.13) (circles). For low temperature values, the two curves coincide. For increased values of T, artifacts of the numerical solver render the self-diffusion of water slightly lower in the rescaled system as in the original one. One might compensate this behavior with a step-width correction, which has not been done here.

The self-diffusion of water has been measured experimentally by Krynicki et al. [164] (cf. right panel of Fig. 6.1). It has been found to be best described by the theoretical predictions of the free volume model by Cohen and Turnbull [165]:

$$D = AT^{1/2}e^{-E/R(T-T_0)}, \tag{6.16}$$

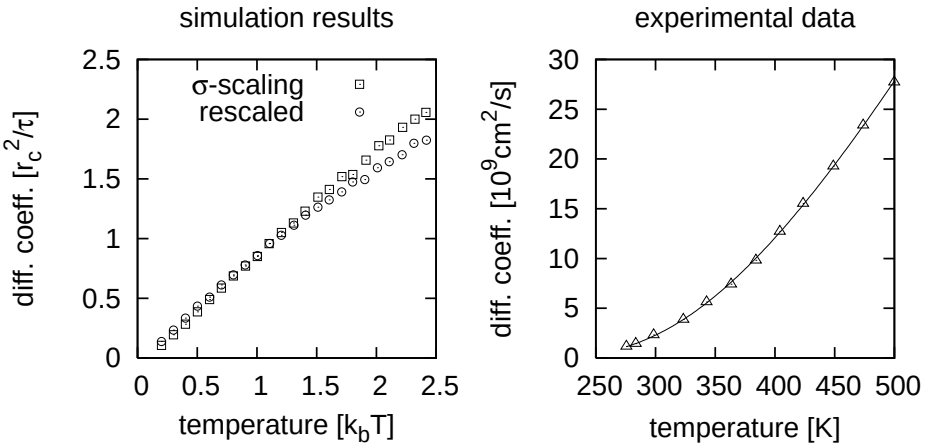


Figure 6.1: Self-diffusion of water D as a function of the temperature. Left: simulation results for two different temperature scaling methods (see text). Simulations have been performed in a box of size $(20r_c)^3$ with particle density $3r_c^{-3}$ and have been truncated once the displacement of particle exceeds half the box size. $t < 5.0\tau$ has been taken as transient. Right: experimental results by Krynicki et al. [164] and a fit of the free volume model to the data points (see text).

where T_0 is the (pressure dependent) temperature at which the free volume disappears (the glass transition temperature), E is an activation energy, A a proportionality constant with units $\mathrm{cm^2 s^{-1} K^{-1/2}}$, and $R = 8.314472 \mathrm{JK^{-1} mol^{-1}}$ the ideal gas constant. A numerical fit of the free volume model to the measured data reveals a transition temperature of $T_0 = 117.5\mathrm{K}$, $E = 6.50\mathrm{kJ\ mol^{-1}}$, and $A = 9.61\mathrm{cm^2 s^{-1} K^{-1/2}}$.

Naturally, we cannot except our temperature scaling (applied to structureless DPD particles) to render an accurate picture of all the anomalities of water – the melting point anomality being one of them. In particular, phase transitions such as freezing and glas transition cannot be accounted for by our method. This significantly limits calibration of the temperature unit to physical values. Apart from both increasing monotonically, the temperature dependence of the diffusion constant in our simulation is qualitatively different from these data.

The closest we can get toward a calibration of the temperature scale is the following approach: We identify the freezing point of water (273K) with the freezing point in the DPD simulation ($0.02 k_B T'$) and $1 k_B T'$ with room temperature about 300K (for which the repulsion value a=25 has been obtained from the compressibility of water). A temperature difference of $\Delta T =$ then corresponds to about 30K, which is within a physically meaningful range. Note, however, that this calibration introduces an affine mapping (as opposed to a linear proportionality) between the physical energy scale and the one of the method.

6.2.2 Temperature response of system components

Temperature dependence of lipid aggregates

We analyze the micellar size distribution of water-surfactant systems of different amphiphile representations for various temperatures. The system size in these simulations is 12^3 and the surfactant concentration is 0.0167 surfactant molecules per water bead.

Fig. 6.2 shows the aggregate size distribution of **HT** amphipihles as a function of the temperature for hydrophobicity values of 80 and 120. For $a_{\mathbf{HT}} = a_{\mathbf{WT}} = 80$, aggregates tend to dissolve, when the temperature is raised. For $k_BT = 1.5$ and higher, the only maximum in the micellar size distribution is the monomer peak at one. Inlays in Fig. 6.2 show the center of gravity of the distribution as a function of temperature. They reveal an exponential dependency between the temperature and the center of gravity. The thermal stability of aggregates can be significantly increased by the use amphiphiles with longer hydrophobic chains: Fig. 6.3 shows simulation results for a aggregates formed by **HTT** amphiphiles. For these systems, statistics are necessarily noisy due to the bigger size of aggregates. Comparing these results with Fig. 6.2 (upper panel), one finds that such aggregates are bigger by a factor of up to 5 for a temperature of $T' = 1T$ and still exhibit a mean aggregate size number of 10 at $T' = 2.5T$. While the envisioned container replication cycle has been shown to perform for short chained surfactants with the hydrophobicity values analyzed in Fig. 6.2 [81, chapter 4 in this work], the performance of container replication for longer hydrophobic chains would need to be verified in simulation.

Temperature response of the biopolymer

Here and in the following we simulate nucleotide sequences with a tangential attractive force as introduced in chapter 5 which has been identified as the best performing gene representation.

To analyze the effect of a temperature change on the genome, we perform the following computer simulations: a cubic box of size $(6r_c)^3$ is initialized with a two phase oil water system with two complementary 4-mer strands forming a hybridization complex at the oil-water interface. The chosen template in these simulations is the self-complementary sequence **ABAB** which has been found to form stable hybridization complexes with its oligomers at unit temperature (see section 5.3.4 in chapter 5). The simulation is run for 4000τ and the distances between the complementary bases are measured for each time step. When none of the base distances exceeds $1.5r_c$ (the interaction distance of the attractive force) the genome is considered to be hybridized.

The simulations of complementary 4-mer double strands are compared to simulations of one 4-mer template and two complementary dimers. The template-oligomer complex is considered hybridized only if the bases of both templates are within the attraction range. Taken together, these simulations model the genome dynamics prior to and after ligation. The ligation reaction itself is turned off in the course of the simulations.

The fraction of time within which the genome is completely hybridized is shown in Fig. 6.4 as a function of the temperature. The results reveal a phase transition between the mainly hybridized configuration ($k_bT < 1$) and the mainly molten configuration ($k_bT > 2.5$) of complementary 4-mers (line with black circles). The center of this phase transition - where the strands are hybridized 50% of the time - can be varied by changing the stiffness of the nucleic acid backbone: The upper panel of Fig. 6.4 shows hybridization behavior for a stiffness of 15 (which corresponds to the choice in chapter 5). For this value, the melting curves suggest a temperature cycle around 1.5 ± 0.5. The lower panel reveals that a more flexible backbone (stiffness set to 5) lowers the melting temperature and suggests a cycle around 1.1 ± 0.5. The time fraction of complete hybridization for one 4-mer strand plus oligomers (lines with white circles) is significantly lower and unaffected by the backbone stiffness.

We further simulate the dynamics of a 6-mer template (sequence **ABBABB**) with

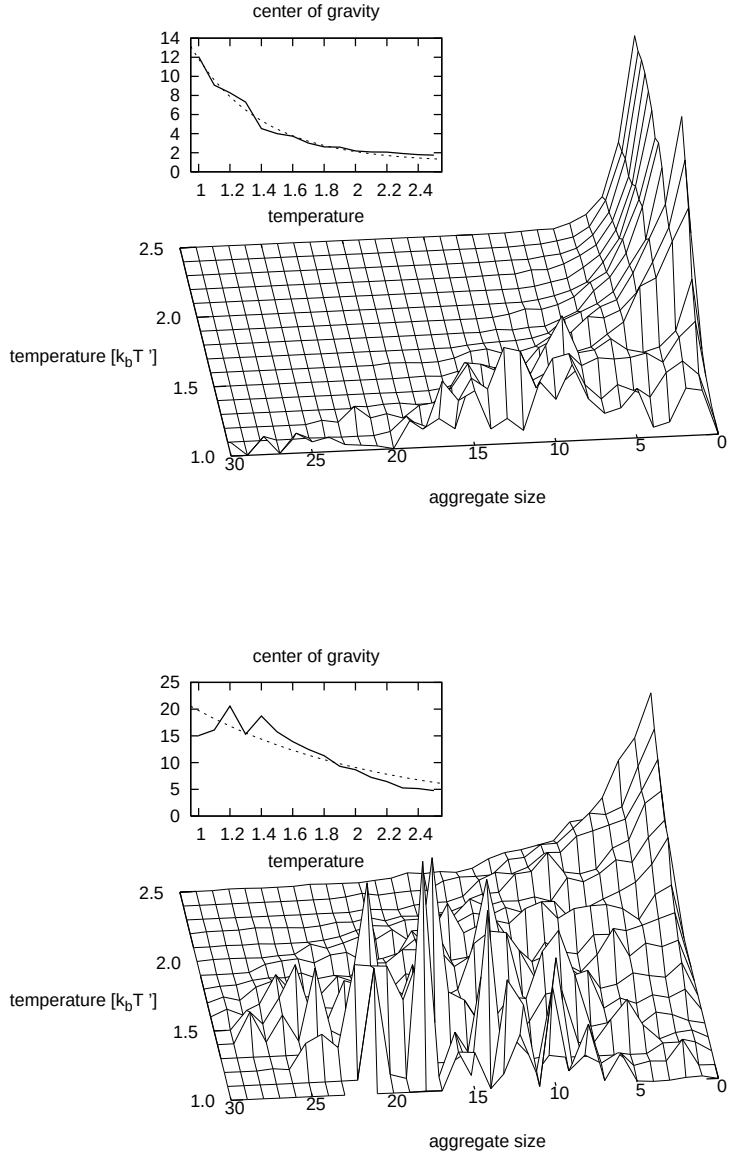


Figure 6.2: Time-averaged micellar size distributions for different temperatures for $a_{\mathbf{HT}} = a_{\mathbf{WT}} = 80$ (upper panel) and $a_{\mathbf{HT}} = a_{\mathbf{WT}} = 120$ (lower panel) for aggregates composed of **HT** amphiphiles. The inlays show the motion of the center of gravity of the distributions (solid line) and its least squares fit to an exponential (dotted line). See text for details.

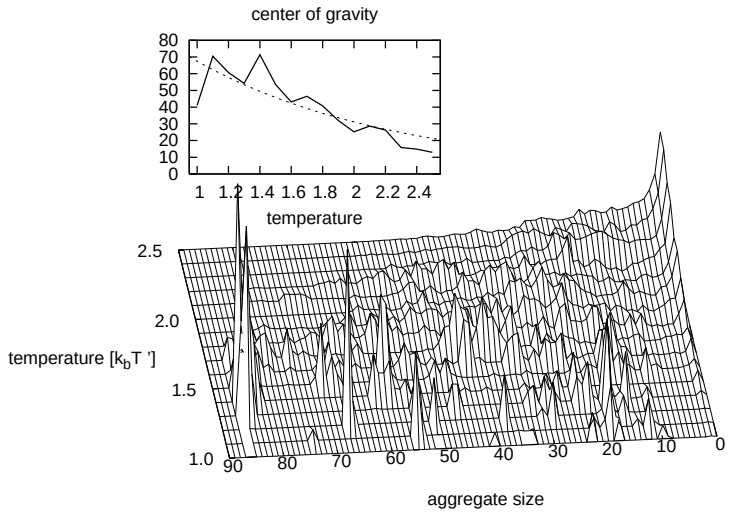


Figure 6.3: Time-averaged micellar size distributions for different temperatures for $a_{\mathbf{HT}} = a_{\mathbf{WT}} = 80$ for aggregates formed by **HTT** amphiphiles.

two complementary 3-mers, and a backbone stiffness of $15k_bT$. Melting curves are shown in Fig. 6.5. It is found that by moving from 4-mer to 6-mer templates, the melting point slightly increases from around $1.5T$ to $1.7T$, whereas the stability of the oligomer/template complex increases by a factor of about two. This shows that the synergetic effect (that individual complementary base pairs exert on each other by geometrically constraining their position) decreases with the template length.

In all simulations, it is conceivable to reduce the attractive force between complementary bases in order to lower the melting temperature of the double strand configuration. However, the previous analysis in chapter 5 has revealed that lowering the attractive force disfavors the attachment of short oligomers to the template strand.

Summary

The performed simulations indicate, that a temperature cycle is feasible that is strong enough to melt 4-mer double strands while keeping the container essentially intact. To achieve this, hydrophobicity of the surfactants can be increased while the backbone stiffness is simultaneously lowered. A temperature cycle around 1.1 ± 0.5 with hydrophibicity 120 and stiffness 5 is expected to fulfill the requirements.

6.3 Coupling information, container, and genome

In order to be a veritable information carrier, the genome has to somehow affect the functioning of the protocell. In the Los Alamos minimal protocell, this is done by a direct coupling of nucleotides in the genome to the catalytic activity of the photo sensitizer: a nucleotide with a suitably chosen redox potential (oxo-guanine) can serve

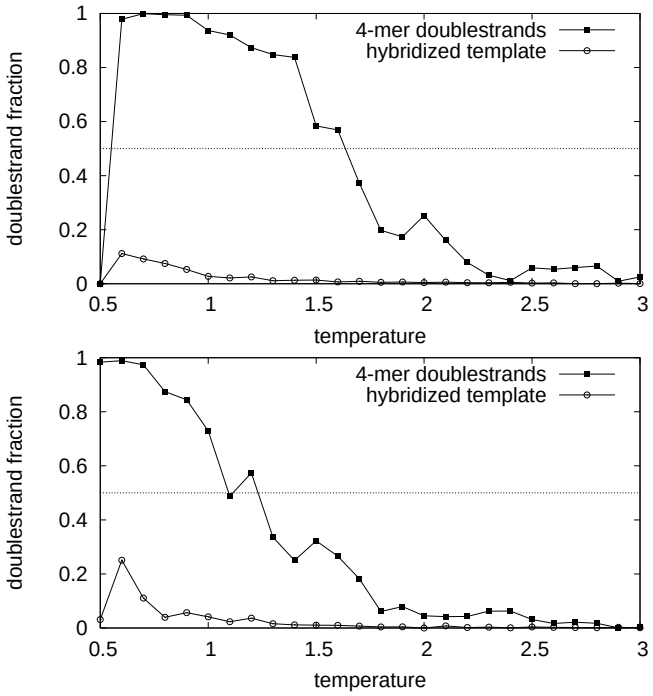


Figure 6.4: Time fraction of double strand configurations of two complementary 4-mer strands (line with black circles) and of one 4-mer strand with complementary dimers (line with white circles) as a function of temperature. In case of template and dimers, the genome is only considered hybridized when both dimers attach. The upper panel shows melting curves for strand with a backbone stiffness of 15 (corresponding to previous simulations), the lower panel corresponds to a stiffness of 5.

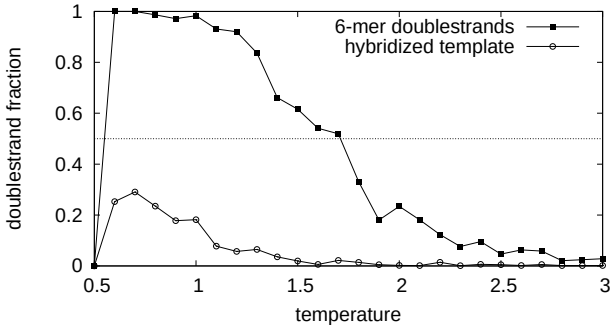


Figure 6.5: Time fraction of double strand configurations of two complementary 6-mer strands with sequence **ABBABB** (line with black circles) and of one 6-mer strand with complementary 3-mers (line with white circles) as a function of temperature (backbone stiffness set to 15). See caption of Fig. 6.4 for descriptions.

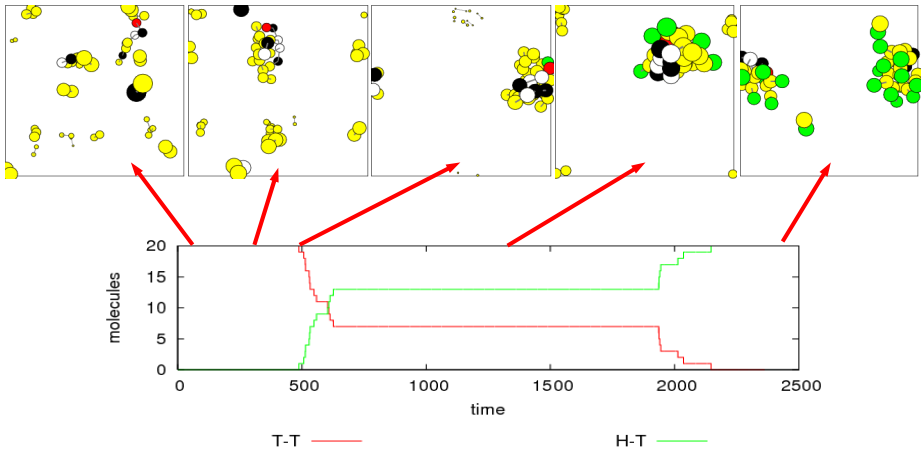


Figure 6.6: Coupling of gene replication and metabolic turnover: A solution of lipid precursors, dimers, and a 4-mer template with attached sensitizer (panel one) self-assembles into an oil droplet with the information molecules attaced at the interface (panel two). Once hybridization and ligation of the dimers occurs, the sensitizer catalyzes the metabolic turnover (panel 3 through 5). The graph below shows the time evolution of precursors and ampiphiles in the system.

as eletron donor in the photo cleavage reaction of the ester which again is catalyzed by the sensitizer (see Ref. [72] for details and an experimental confirmation). In order to allow for the electron transfer to occur, sensitizer and oxo-guanine must be in relative proximity. It is hypothesized that delocated π-electrons in stacked aromatic rings of nucleotides in the biopolymer can propagate this electron transfer. This extends the required proximity over the biopolymer and possibly its hybridized complement [72, 156].

The design of the Los Alamos minimal protocell exploits the electrochemical properties of the metabolism to introduce a coupling of gene replication and container replication: covalently binding the sensitizer to the template strand and oxo-guanine to a complementary oligomer – or vice versa – will result in a gene-metabolism complex that drives container growth and division only after successful gene replication.

Naturally, electrocemistry is far beyond the scope of our simulation technique. In our system level oriented toy model, we therefore implement this coupling by a simple hybridization check: when two complementary polymer strands are hybridized (i.e. the distance of each base pair is smaller than $1.5r_c$), the catalytic rate enhancement of a sensitizer bead that is covalently bound to one of the strands will be set to 1.0, otherwise it is turned to 0.0.

Snapshots of a simulation of this coupling are shown in Fig. 6.6: a solution of lipid precursors, dimers, and a 4-mer template with attached sensitizer (panel one) self-assembles into an oil droplet with the information molecules attaced at the interface (panel two). Once hybridization and ligation of the dimers occurs, the sensitizer catalyzes the metabolic turnover (panel 3 through 5). However, the concentration of precursors is too low for the resulting aggregate to divide.

6.4 Revision of the protocellular life-cycle

The protocellular design presented in Ref. [71] envisions single stranded nucleic acid strands that would only be in a double strand configuration during gene replication, i.e. after ligation of hybridized oligomers and before melting of the strands. During the container replication, the polymers are thought to be single stranded, such that each daughter cell receives one of the complementary strands after container replication (see Fig. 5.1 in chapter 5).

However, if the metabolic turnover rate of the sensitizer depends on the genome to be in double strand configuration (see section 6.3), the container division process can only be triggered prior to melting. Having only one double strand leads to only one of the two daughter cells being fertile, such that the system as a whole has not replicated.

We therefore suggest a revised system that employs double stranded polymers in the first place (see Fig. 6.7). Melting of the double strand only occurs during gene replication: prior or parallel to the supply of oligomers, the double strand is melted. Cooling of the system after intake of the oligomers then allows for template directed gene replication, after which the aggregate contains two identical double stranded copies. Each of these copies is able to mediate the metabolic reaction. Since the two double strands are more likely to be spatially separated, it is possible for them to end up in different daughter cells. As in the previous design, ligation and lipid precursor turnover are allowed be triggered simultaneously if the polymerization reaction is based on the same photo reaction as lipid turnover.

Note that in this design, the genetic information of the two daughter cells is identical, rather than complementary as in the original design. This can avoid complications during repeated protocell replication.

Furthermore, our simulations suggest that lipid precursor supply and intake should precede melting in order to stabilize the container prior to heating (see section 6.2). Oligomers can either be supplied simultaneously or after melting.

6.5 Product inhibition during template replication

With the modifications described in sections 6.2 through 6.4, an integrated simulation of the entire protocellular life-cylce can be approached. First, however, it is advisable to validate that the individual steps in the life-cycle still perform in the modified system. In particular, the hybridization of gene precursors (oligomers) to template strands after cooling (rightmost arrow in Fig. 6.7) needs to be tested for a system that uses a duplicated set of polymers compared to the previous simultations in section 5.3.4.

We perform simulations on hybridization and ligation of oligomers at the surface of surfactant-coated oil droplets employing two 4-mer template strands and four complementary oligomers. Different base sequences in the templates are simulated and scenarios that use identical template strands (e.g. $\mathbf{ABAB}$) as well as different template strands (e.g. $\mathbf{AAAB}$ and $\mathbf{BBBA}$) are compared. Although the two templates will be complementary in the final setup (as they result from a molten double strand), simulations are also performed on non-complementary template strands (such as two $\mathbf{AAAB}$ strands) for comparison.

Fig. 6.8 shows snapshots that illustrate typical template-oligomer interactions as they appear in the simulations. In most cases, it is found that the two template strands hybridize to each other, thereby preventing oligomer attachment. In some cases (lower left panel) two templates and one oligomer form less defined hybridization complexes.

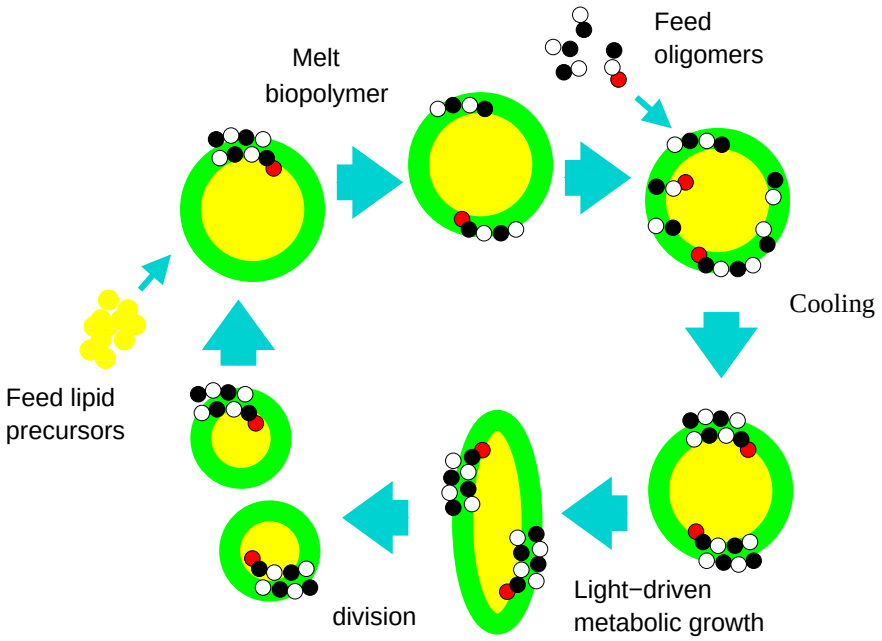


Figure 6.7: Revised version of the protocellular life-cycle proposed in Ref. [71] (shown in Fig. 5.1): The main design change is to have the biopolymer in double strand configuration, rather than as suggested as a single strand. This allows for equipartition of the replicated double strand into two fertile daughter cells after countainer division (see text).

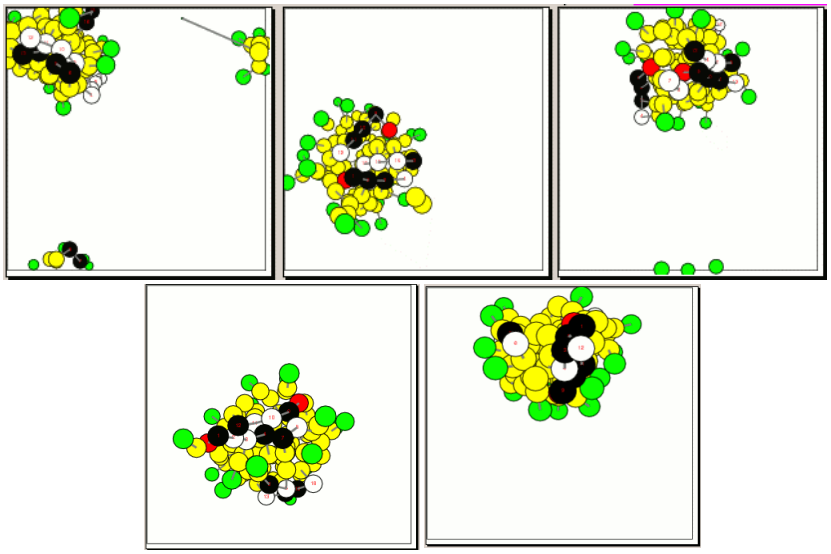


Figure 6.8: Product inhibition of 4-mer template strands on the surface of the aggregates. The panels show results for different base sequences in the templates (both for identical and different sequences in the template strands). In most cases, it is found that the two template strands hybridize to each other, preventing oligomer hybridization to occur. In some cases (lower left panel) two templates and one oligomer form a less defined hybridization complex that equally inhibits proper template directed replication. We further find interactions between two original and a newly formed template strand (upper central panel).

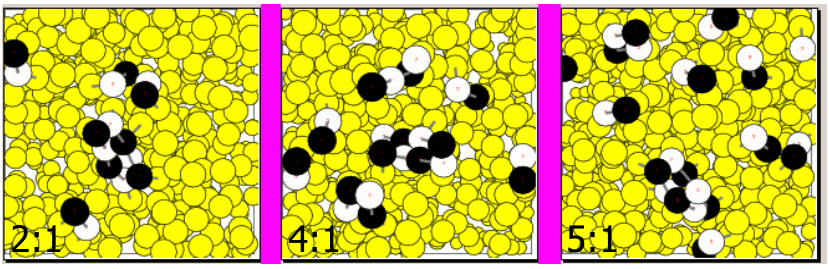


Figure 6.9: Competing template/oligomer interactions for oligomer concentrations ranging from 2:1 to 5:1 oligomers per template. Product inhibition occurs for all tested concentration ratios.

We also find interactions between two original and one newly formed template strand (upper central panel). Only in trivial cases (e.g. two **AAAA** templates) such product inhibition does not occur. However, even for non-complementary template strands (e.g. two **AAAB** sequences in the lower right panel and **ABBA** with **AAAB** in the upper right panel), partial complementary attraction can still cause product inhibition. Our preliminary investigations suggest that product inhibition is more likely for sequences that were found to form stable oligomer/template complexes (such as **ABAB**) if only one strand is used (see Fig. 5.8 in chapter 5). In all cases, the reason for product inhibition is the significantly stronger binding between two templates as opposed to template/oligomer binding, as reported in sections 5.3.4 and 6.2.2.

The problem of product inhibition is also known from experimental studies of non-enzymatic template-directed DNA replication [61, 62], and it has been shown that the resulting growth dynamics can prohibit proper Darwinian evolution [166]. While this issue has been solved for the Los Alamos minimal protocell on a global level [159, 167], it still hinders gene replication on a local level – at least in our simulation.

We try several attempts to resolve product inhibition, some of them motivated by experimental findings. To isolate the problem, these simulations are preformed on oil/water interfaces as in section 6.2.2. System size has been choosen small ($(6r_c)^3$) in order to relate findings to dynamics on the equally small surface of our surfactant-coated oil droplets. Preliminary results are listed:

1. **Increasing oligomer concentration:** We increase the oligomer concentration in order to move the equilibrium of the competing template/template and template/oligomer binding reactions toward template/oligomer complexes. Fig. 6.9 shows simulation results with 2, 4, and 5 oligomers per template. Oligomer concentration has not been increased higher, as this favors the spontaneous background reactions of non-hybridized oligomers (as can be seen in the central panel of Fig 6.9 where two dimers ligated in the absence of a template). Product inhibition occurs in all simulated scenarios. Visual inspection of the system dynamics suggests that this is due to an effectively irreversible binding of the template/template complex (as can also be deduced from Fig. 6.4).

2. **Altering attractive forces between complementary bases:** In order to turn the template/template hybridization into an effectively reversible reaction, we lower the attractive force between complementary bases. This, however, also lowers the (already low) affinity of oligomers to bind to the template, such that stable oligomer/template complexes cannot be observed over longer periods of time.

Equally, increasing the attractive force to stabilize oligomer hybridization simultaneously increases the stability of template/template complexes. In summary, altering the attractive forces of complementary bases cannot overcome product inhibition.

3. **Changing the preferred backbone angle of template strands:** We test the introduction of a natural curvature in the backbone of template strands. This is achieved by changing the equilibrium angle θ_{eq} in Eq. (5.8) from 180° to 135°. Since such bending only applies to polymer strands longer than 3 bases, we hope that ligation of hybridized oligomers would induce a reconfiguration of the newly formed strand, which would eventually induce a separation of the strands. It is found by our simulations, however, that bending of the strands occurs perpendicular to the preferential direction of base attraction: both strands curve toward the oil phase, leaving complementary base distances unaltered. As a consequence, the modification has no effect on the stability of template/template complexes. We note that a more elaborate force field could be employed to constrain the relative orientation of backbone curvature and direction of attraction by combining Eqns. (5.8) and (5.16).

4. **Employing longer polymer stands:** As a main problem in the dynamics of genome replication in our model is the relatively low hybridization affinity of dimers, we test the performance of longer templates and oligomers. Using 6-mer templates in combination with 3-mers enhances the stability of oligomer/template hybridization by a factor of two (see Fig. 6.5) with comparably minor stability increase of the 6-mer/6-mer complex. Visual inspections of simulations results confirm that this is a viable solution that deserves to be studied more carefully in the future.

5. **Repeated temperature cycle:** If relatively stable oligomer/template hybridization can be achieved (by a combination of the above mechanisms), we hope that product inhibition can be overcome with the help of a repeated temperature cycle where melting and cooling is repeated to undo irreversible template/template hybridization until oligomer hybridization and ligation turned out successful. This has not been tested by us in simulation yet.

It is hoped that a combination of the above modifications will allow for replication of the molten double strand. If the actual solution requires elongated information polymers and/or a surplus of oligomers, the performance of container replication and equipartition (section 5.3.5 in chapter 5) need to be confirmed for the additional load.

6.6 Summary

We have presented work on the Los Alamos minimal protocell toy model that builds upon the results of chapters 4 and 5. A simple procedure to alter temperature in the DPD method has been established and compared to experimental findings. A mechanism to couple the information, container, and metabolic subsystems have been added to the model. This has revealed a necessary design change of the originally proposed system, namely it was shown that a double stranded information carrier is required for successful replication of the entire aggregate.

Steps toward an integrated simulation that employ double stranded genes and a temperature cycle have been presented. It is found that the modified protocell design

is prone to product inhibition which currently prevents successful gene replication. A careful balance of system parameters is needed to fulfill partially competing constraints in an integrated simulation of the complete life-cycle. We suspect similar complications to occur also in experimental work toward the implementation of the Los Alamos minimal protocell. To what extend these complications overlap with the ones found in our simulations cannot be predicted from the model.

Conclusion

Self-replication is a fundamental feature of all living organisms, yet has only been accomplished to limited extend in artificial systems.

In the presented work, we have developed a physically motivated simulation framework based on the method of dissipative particle dynamics. The method has been extended over the course of our studies to cope with chemical reactions, scaling of the spatial resolution, complex intermolecular interactions, and temperature variations.

The framework enabled us to study design proposals of life-like, artificial, molecular aggregates. In particular, we focussed on the design of the Los Alamos minimal protocell, and a simplified information-free predecessor. We have performed system-level simulations of the design which attempt to account for theoretical, and experimental knowledge, as well as results from other computational models to the extend possible. This allowed us to address key issues of the replicating subsystems – container, genome, and metabolism – both individually and when mutually coupled. We have analyzed each step in the life-cycle of the molecular aggregate, and a final integrated simulation of the entire life-cycle has been prepared.

Our simulations confirmed most assumptions of the theoretical design. In particular, we could support (i) the feasibility of surfactant-coated oil droplets as a self-replicating container subsystem; (ii) the mutual catalytic coupling between container and genome, container and metabolism, metabolism and genome, as well as the final coupling of all components; and (iii) the necessity of a temperature cycle for the replication of the information subsystem. Various minor confirmations have been summarized at the end of each chapter.

The model has also revealed unanticipated system-level findings, most notably (i) a significant influence of the oil core on both the dynamics of the whole system as well as its subsystems; (ii) the emergence of a molecular fitness function from geometric properties of the genome; (iii) disfavored equipartition of aggregates that are loaded with a substantial undigestable oil core; (iv) the unfeasibility of a single stranded information carrier; (v) the problematic of product inhibition on the limited surface area of small oil droplets. More work has to be dedicated on the last item of this list.

We emphasize that these findings could only be achieved by the use of a physically motivated simulation framework. We also emphasize that the above findings are qualitative and do not claim predictive power.

Taken together, these findings support our hypothesis that self-replication and similar life-like features can be achieved in systems of formerly unanticipated simplicity – if these systems exploit physicochemical principles that are immanent to their physical scale.

Bibliography

[1] J. von Neumann. Theory and organization of complicated automata. In W. Aspray and A. Burks, editors, *Papers of John von Neumann on Computing and Computer Theory*. MIT Press, Cambridge, MA, USA, 1987.

[2] A. Lindenmeyer. Mathematical models for cellular interaction in development. *J. Theo. Biol.*, 18:280–315, 1968.

[3] P. Godfrey-Smith. The replicator in retrospect. *Biol. Phil.*, 15:403–423, 2000.

[4] C. Langton. *Artificial Life - The proceedings of an interdisciplinary workshop on the synthesis and simulation of living systems*. Santa Fe Institute, Santa Fe NM, USA, 1989.

[5] M. Sipper. Fifty years of research on self-replication: An overview. *Artif. Life*, 4(3):237–257, 1998.

[6] R. A. Freitas Jr. and R. C. Merkle. *Kinematic Self-Replicating Machines*. Landes Bioscience, Georgetown, TX, USA, 2004.

[7] B. McMullin. What is a universal constructor? In *Proceedings of the Second European Conference on Artificial Life, Brussels*, 1993.

[8] J. von Neumann. *Theory of Self-Reproducing Automata*. University of Illinois Press, Champaign, IL, USA, 1966.

[9] U. Pesavento. An implementation of von Neumanns self-reproducing machine. *Artif. Life*, 2:337–354, 1995.

[10] R. Nobili and U. Pesavento. Generalised von Neumann's automata. In E. Besussi and A. Cecchini, editors, *Artificial Worlds and Urban Studies*, 1996.

[11] R. Nobili, 2008. currently unpublished, a working implementation can be found at `http://www.pd.infn.it/ rnobili/wjvn/index.htm` .

[12] W. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.*, 7:115–133, 1943.

[13] R. A. Freitas Jr. and W. P. Gilbreath. Advanced automation for space missions. *J. Astronaut. Sci.*, 30:1 – 11, 1982.

[14] J. Thatcher. Universality in the von Neumann cellular model. In A.W. Burks, editor, *Essays on Cellular Automata*, pages 132–186. University of Illinois Press, Urbana, IL, USA, 1970.

[15] G. Tempesti, D. Mange, and A. Stauffer. Self-replicating and self-repairing multicellular automata. *Artif. Life*, 4(3):259–282, 2006.

[16] C. Salzberg and H. Sayama. Heredity, complexity, and surprise: Embedded self-replication and evolution in CA. In *Proceedings of the 6th International Conference on Cellular Automata for Research and Industry*, 2004.

[17] E. F. Codd. *Cellular Automata*. Academic Press, New York, USA, 1968.

[18] C. G. Langton. Self-reproduction in cellular automata. *Physica D*, 10:135–144, 1984.

[19] J. Byl. Self-reproduction in small cellular automata. *Physica D*, 34:295–299, 1989.

[20] J. A. Reggia, S.L. Armentrout, H.-H. Chou, and Y. Peng. Simple systems that exhibit self-directed replication. *Science*, 259:1282–1287, 1993.

[21] H. Sayama. Introduction of structural dissolution into Langton's self-reproducing loop. In C. Adami, R. K. Belew, H. Kitano, and C. E. Taylor, editors, *Proceedings of the Sixth International Conference on Artificial Life*, pages 114–122, Los Angeles, CA, USA, 1998. MIT Press.

[22] H. Sayama. Spontaneous evolution of self-reproducing loops implemented on cellular automata: A preliminary report. In Y. Bar-Yam, editor, *Proceedings of the Second International Conference on Complex Systems*, Nashua, NH, USA, 1998. Perseus Books.

[23] H. Sayama. Toward the realization of an evolving ecosystem on cellular automata. In M. Sugisaka and H. Tanaka, editors, *Proceedings of the Fourth International Symposium on Artificial Life and Robotics*, pages 254–257, Beppu, Oita, Japan, 1999.

[24] D. Mange, A. Stauffer, E. Petraglio, and G. Tempesti. Self-replicating loop with universal construction. *Physica D*, 191:178–192, 2004.

[25] L.S. Penrose and R. Penrose. A self-reproducing analogue. *Nature*, 179:1183, 1957.

[26] H. Jacobson. On models of reproduction. *American Scientist*, 46:255–284, 1958.

[27] J. Suthakorn, Y.T. Kwon, and G.S. Chirikjian. A semi-autonomous replicating robotic system. In *Computational Intelligence in Robotics and Automation*, volume 2, pages 776–781, 2003.

[28] J. Suthakorn, A.B. Cushing, and G.S. Chirikjian. An autonomous self-replicating robotic system. In *Advanced Intelligent Mechatronics*, volume 1, pages 137– 142, 2003.

[29] K. Lee and G.S. Chirikjian. Robotic self-replication. *IEEE Robot. Automat.*, 14(4), 2007.

[30] V. Zykov, E. Mytilinaios, M. Desnoyer, and H. Lipson. Evolved and designed self-reproducing modular robotics. *IEEE Trans. Robot.*, 23(2), 2007.

[31] C. Paul. Morphology and computation. In *Proceedings of the International Conference on the Simulation of Adaptive Behaviour*, pages 33–38, Los Angeles, CA, USA, 2004.

[32] C. Paul. Morphological computation. *Robotics and Autonomous Systems*, 54(8):619–630, 2006.

[33] R. Pfeifer and F. Iida. Morphological computation: Connecting body, brain and environment. *Jap. Sci. Monthly*, 58(2):48–54, 2005.

[34] *Proceedings of the International Conference on Morphological Computation*, 2007.

[35] T. McGeer. Passive dynamic walking. *Int. J. Roboti.Res.*, 9(2):62–82, 1990.

[36] D. L. Nelson and M. M. Cox. *Lehninger Principles of Biochemistry*. W. H. Freeman and Company, New York, USA, 4^{th} edition, 2005.

[37] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Watson. *Molecular Biology of the Cell*. Garland Science Publishing, 2002.

[38] C. Woese. *The Genetic Code*. Harper & Row, 1968.

[39] W. Gilbert. The RNA world. *Nature*, 319:618, 1986.

[40] R.l A. Hughes and A. D. Ellington. Ribozymes and the evolution of metabolism. In J. Tze-Fei Wong and A. Lazcano, editors, *Prebiotic Evolution and Astrobiology*. Landes Bioscience, Austin, TX, USA, 2008.

[41] P.J. Unrau and D.P. Bartel. RNA-catalysed nucleotide synthesis. *Nature*, 395:260–263, 1998.

[42] W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner, and D. P. Bartel. RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. *Science*, 292:1319, 2001.

[43] H. S. Zaher and Peter J. Unrau. Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA*, 13:1017–1026, 2007.

[44] N. Ban, P. Nissen, J. Hansen, P. Moore, and T. Steitz. The complete atomic structure of the large ribosomal subunit at 2.4 A resolution. *Science*, 289(5481):90520, 2000.

[45] F. Schluenzen, A. Tocilj, R. Zarivach, J. Harms, M. Gluehmann, D. Janell, A. Bashan, H. Bartels, I. Agmon, F. Franceschi, and A. Yonath. Structure of functionally activated small ribosomal subunit at 3.3 angstroms resolution. *Cell*, 102(5):61523, 2000.

[46] P. L. Luisi, P. Waldea, and T. Oberholzer. Lipid vesicles as possible intermediates in the origin of life. *Curr. Opin. Coll. Int. Sci.*, 4(1):33–39, 1999.

[47] D. Deamer, J. P. Dworkin, S. A. Sandford, M. P. Bernstein, and L. J. Allamandola. The first cell membranes. *Astrobiology*, 2(4), 2002.

[48] W.R. Hargreaves and D.W. Deamer. Liposomes from ionic, single-chain amphiphiles. *Biochemistry*, 17:37593768, 1978.

[49] D. F. Evans and H. Wennerström. *The Colloidal Domain - Where Physics, Chemistry, Biology, and Technology Meet*. Wiley-VCH, New York, 1999.

[50] P.-A. Monnard. The dawn of the RNA world: RNA polymerization from monoribonucleotides under prebiotically plausible conditions. In J. Tze-Fei Wong and A. Lazcano, editors, *Prebiotic Evolution and Astrobiology*. Landes Bioscience, Austin, TX, USA, 2008.

[51] B. J. Frisken, C. Asman, and P. J. Patty. Studies of vesicle extrusion. *Langmuir*, 16(3):928 –933, 2000.

[52] M. Hanczyc and J. W. Szostak. Replicating vesicles as models of primitive cell growth and division. *Current Opinion in Chemical Biology*, 8:660?664, 2004.

[53] T. Baumgart, S. T. Hessm, and W. W. Webb. Imaging coexisting fluid domains in biomembrane models coupling curvature and line tension. *Nature*, 425:821–824, 2003.

[54] H. Noguchi and M. Takasu. Adhesion of nanoparticles to vesicles: A Brownian dynamics simulation. *Biophys. J.*, 83:299–308, 2002.

[55] J. Macía and R. V. Solé. Protocell self-reproduction in a spatially explicit metabolism-vesicle system. *J. Theo. Biol.*, 245(3):400–410, 2007.

[56] R. V. Solé, J. Macía, H. Fellermann, Munteanu A, J. Sardanyés, and S. Valverde. Models of protocell replication. In Rasmussen et al. [65], pages 213–231.

[57] P. A. Bachmann, P. Walde, P. L. Luisi, and J. Lang. Self-replicating reverse micelles and chemical autopoiesis. *J. Am. Chem. Soc.*, 112:8200–8201, 1990.

[58] P. A. Bachmann, P. L. Luisi, and J. Lang. Self-replicating micelles: aqueous micelles and enzymatically driven reactions in reverse micelles. *J. Am. Chem. Soc.*, 113:8204–8209, 1991.

[59] P. A. Bachmann, P. L. Luisi, and J. Lang. Autocatalytic self-replicating micelles as models for prebiotic structures. *Nature*, 357:57–59, 1992.

[60] A. Patist, J. R. Kanicky, P. K. Shukla, and D. O. Shah. Importance of micellar kinetics in relation to technological processes. *J. Coll. Int. Sci.*, 245:1–15, 2002.

[61] D. Sievers and G. von Kiedrowski. Self-replication of complementary nucleotide-based oligomers. *Nature*, 369:221 – 224, 1994.

[62] B. G. Bag and G. von Kiedrowski. Templates, autocatalysis and molecular replication. *Pure & App. Chem.*, 68(11), 1996.

[63] G. Kiedrowski. A self-replicating hexadeoxynucleotide. *Angew. Chem.*, 25(10):932–935, 1986.

[64] M. Hanczyc. The early history of protocells: The search for the recipe of life. In Rasmussen et al. [65], pages 3–18.

[65] S. Rasmussen, M. Bedau, L. Chen, D. Deamer, D. Krakauer, N. Packard, and P. Stadler, editors. *Protocells: Bridging Nonliving and Living Matter*. MIT Press, Cambridge, USA, 2008.

[66] D. G. Gibson, G. A. Benders, C. Andrews-Pfannkoch, E. A. Denisova, H. Baden-Tillson, J. Zaveri, T. B. Stockwell, A. Brownley, D. W. Thomas, M. A. Algire, C. Merryman, L. Young, andJ. I. Glass V. N. Noskov, J. C. Venter, IIl C. A. Hutchison, and H. O. Smith. Complete chemical synthesis, assembly, and cloning of a mycoplasma genitalium genome. *Science*, 319(5867):1215 – 1220, 2008.

[67] P.-A. Monnard, H. Fellermann, G. Goranovic, M.H. Hanczyc, H.-J. Ziock, J.M. Boncella, and S. Rasmussen. Assembly of a minimal protocell: towards self-replicating systems. *Rep. Prog. Phys.*, 2009. in preparation.

[68] J. Griesemer and E. Szathmáry. Gántis chemoton model and life criteria. In Rasmussen et al. [65], pages 481–512.

[69] T. Ganti. *The Principles of Life*. Oxford University Press, Oxford, UK, 2003.

[70] A. Munteanu and R. V. Solé. Phenotypic diversity and chaos in a minimal cell model. *J. Theo. Biol.*, 240(3):434–442, 2006.

[71] S. Rasmussen, L. Chen, M. Nilsson, and S. Abe. Bridging nonliving and living matter. *Artificial Life*, 9:269–316, 2003.

[72] S. Rasmussen, J. Bailey, J. Boncella, L. Chen, G. Collis, S. Colgate, M. DeClue, H. Fellermann, G. Goranovic, Y. Jiang, C. Knutson, P.-A. Monnard, F. Mouffouk, M. Nielson, A. Sen, A. Shreve, A. Tamulis, B. Travis, P. Weronski, J. Zhang, X. Zhou, H.-J. Ziock, and W. Woodruff. Assembly of a minimal protocell. In S. Rasmussen, M. Bedau, L. Chen, D. Deamer, D. Krakauer, N. Packard, and P. Stadler, editors, *Protocells: Bridging Nonliving and Living Matter*, pages 125–156. MIT Press, Cambridge, USA, 2008.

[73] K. Ruiz Mirazo and A. Moreno. On the origins of information and its relevance for biological complexity. *Biol. Theo.*, 1(3):227–229, 2006.

[74] P. Dittrich, J. Ziegler, and W. Banzhaf. Artificial chemistries – a review. *Artif. Life*, 7:225–275, 2001.

[75] F. J. Varela, H. R. Maturana, and R. Uribe. Autopoiesis: The organization of living systems. *BioSystems*, 5(4):187–196, 1974.

[76] B. Mayer, G. Köhler, and S. Rasmussen. Simulation and dynamics of entropy-driven, molecular self-assembly process. *Phys. Rev. E*, 55(4):4489–4500, 1997.

[77] B. Mayer and S. Rasmussen. The lattice molecular automaton (LMA): A simulation system for constructive molecular dynamics. *Int. J. Mod. Phys. C*, 9(1):157–177, 1998.

[78] N. Ono and T. Ikegami. Self-maintenance and self-reproduction in an abstract cell model. *J. Theor. Biol.*, 206(2):243–253, 2000.

[79] D. Bedrov, G. D. Smith, K. F. Freed, and J. Dudowicz. A comparison of self-assembly in lattice and off-lattice model amphiphile solutions. *J. Chem. Phys.*, 116(12):4765–4768, 2002.

[80] R. D. Groot and K. L. Rabone. Mesoscopic simulation of cell membrane damage, morphology change and rupture by nonionic sufactants. *Biophys. J.*, 81:725–736, 2001.

[81] H. Fellermann and R. Solé. Minimal model of self-replicating nanocells: a physically embodied, information-free scenario. *Phil. Trans. R. Soc. Lond. Ser. B*, 362(1486):1803–1811, 2007.

[82] I. Pagonabarraga and D. Frenkel. Dissipative particle dynamics for interacting systems. *J. Chem. Phys.*, 115:5015–5026, 2001.

[83] A. P. Lyubartsev and A. Laaksonen. On the reduction of molecular degrees of freedom in computer simulations. *Lect. Notes Phys.*, 640:219–244, 2004.

[84] R. D. Groot and P. B. Warren. Dissipative particle dynamics: Bridging the gap between atomistic and mesoscale simulation. *J. Chem. Phys.*, 107(11):4423–4435, 1997.

[85] A. Maiti and S. McGrother. Bead-bead interaction parameters in dissipative particle dynamics: Relation to bead-size, solubility, and surface tension. *J. Chem. Phys.*, 120(3):1594–1601, 2003.

[86] R. M. Füchslin, H. Fellermann, A. Eriksson, and H. Ziock. Coarse-graining and scaling in dissipative particle dynamics. *J. Chem. Phys.*, 130(1), 2009.

[87] W. Dzwinel and D. A. Yuen. Matching macroscopic properties of binary fluids to the interactions of dissipative particle dynamics. *J. Mod. Phys. C*, 11(1):1–25, 2000.

[88] H. Fellermann, St. Rasmussen, H.-J. Ziock, and R. Solé. Life-cycle of a minimal protocell: a dissipative particle dynamics (DPD) study. *Artif. Life*, 13(4):319–345, 2007.

[89] R. D. Groot. Mesoscopic simulation of polymer-surfactant aggregation. *Langmuir*, 16:7493–7502, 2000.

[90] M. Venturoli and B. Smit. Simulating self-assembly of model membranes. *PhysChemComm*, 10, 1999.

[91] J. C. Shillcock and R. Lipowsky. Equilibrium structure and lateral stress distribution of amphiphilic bilayers from dissipative particle dynamics simulations. *J. Chem. Phys.*, 117(10):5048–5061, 2002.

[92] R. D. Groot. Electrostatic interactions in dissipative particle dynamics-simulation of polyelectrolytes and anionic surfactants. *J. Chem. Phys.*, 118(24):11265–11277, 2003.

[93] P. B. Warren. Vapor-liquid coexistence in many-body dissipative particle dynamics. *Phys. Rev. E*, 68:066702, 2003.

[94] S. Chen, N. Phan-Thien, X. Fan, and B. Cheong Khoo. Dissipative particle dynamics simulation of polymer drops in a periodic shear flow. *J. Non-Newtonian Fluid Mech.*, 118(1):65–81, 2004.

[95] J. S. Sims and N. S. Martys. Simulation of sheared suspensions with a parallel implementation of QDPD. *J. Res. Natl. Inst. Stan.*, 109(2):267–277, 2004.

[96] C. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry, and Natural Sciences*, volume 13 of *Springer Series in Synergetics*. Springer, New York City, 3rd edition, 2004.

[97] P. Hoogerbrugge and J. Koelman. Simulating microscopic hydrodynamic phenomena with dissipative particle dynamics. *Europhys. Lett.*, 19:155–160, 1992.

[98] P. Español. Hydrodynamics from dissipative particle dynamics. *Phys. Rev. E*, 52(2):1734–1742, 1995.

[99] P. Español and P. Warren. Statistical mechanics of dissipative particle dynamics. *Europhys. Lett.*, 30:191–196, 1995.

[100] A. Eriksson, M. Nilsson Jacobi, J. Nyström, and K. Tunstrøm. Effective thermostat induced by coarse-graining of SPC water. *J. Chem. Phys.*, 129(2):024106, 2008.

[101] C. Pastorino, T. Kreer, M. Müller, and K. Binder. Comparison of dissipative particle dynamics and Langevin thermostats for out-of-equilibrium simulations of polymeric systems. *Phys. Rev. E*, 76:026706, 2007.

[102] S. S. Andrews and D. Bray. Stochastic simulation of chemical reactions with spatial resolution and single molecule detail. *Phys. Biol.*, 1:137–151, 2004.

[103] M. I. Monine and J. M. Haugh. Reactions on cell membranes: Comparison of continuum theory and Brownian dynamics simulations. *J. Chem. Phys.*, 123(7):074908, 2005.

[104] A. Buchanan, G. Gazzola, and M. A. Bedau. *Systems Self-Assembly: multidisciplinary snapshots*, chapter Evolutionary design of a model of self-assembling chemical structures. Elsevier, London, 2006.

[105] G. Gazzola, A. Buchanan, N. Packard, and M. Bedau. Catalysis by self-assembled structures in emergent reaction networks. In *Advances in Artificial Life*, volume 4648 of *Lecture Notes in Computer Science*, pages 876–885. Springer, 2007.

[106] H. Liu, H. Qian, Y. Zhao, and Z. Lua. Dissipative particle dynamics simulation study on the binary mixture phase separation coupled with polymerization. *J. Chem. Phys.*, 127:1449031–1559038, 2007.

[107] F. C. Collins and G. E. Kimball. Diffusion-controlled reaction rates. *J. Colloid Sci*, 4(425), 1949.

[108] N. Ono. *Artificial Chemistry: Computational Studies on the Emergence of Self-Reproducing Units*. PhD thesis, Univ. Tokyo, 3-8-1 Komaba Meguro-ku, Tokyo 153-8902, Japan, 2001.

[109] Y. G. Yingling and B. J. Garrison. Coarse-grained chemical reaction model. *J. Phys. Chem. B*, 1008:1815–1821, 2004.

[110] Loup Verlet. Computer “Experiments” on classical fluids. I. thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, 159:98–103, 1967.

[111] K. E. Novik and P. V. Coveney. Finite-difference methods for simulation models incorporating nonconservative forces. *J. Chem. Phys.*, 109(18):7667–7677, 1998.

[112] G. Besold, I. Vattulainen, M. Karttunen, and J. M. Polson. Towards better integrators for dissipative particle dynamics simulations. *Phys. Rev. E*, 62:7611–7614, 2000.

[113] I. Vattulainen, M. Karttunen, G. Besold, and J. M. Polson. Integration schemes for dissipative particle dynamics simulations: From softly interacting systems towards hybrid models. *J. Chem. Phys.*, 116:3967–3979, 2002.

[114] I. Pagonabarraga, M. H. J. Hagen, and D. Frenkel. Self-consistent dissipative particle dynamics algorithm. *Europhys. Lett.*, 377(42), 1998.

[115] H. Noguchi and G. Gompper. Dynamics of vesicle self-assembly and dissolution. *J. Chem. Phys.*, 125:164908, 2006.

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

[117] S. Yamamoto, Y. Maruyama, and S. Hyodo. Dissipative particle dynamics study of spontaneous vesicle formation. *Journal of Chemical Physics*, 116(13):5842–5849 5842–5849, 2002.

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

[119] H. Limbach, A. Arnold, B. A. Mann, and C. Holm. ESPResSo – an extensible simulation package for research on soft matter systems. *Comput. Phys. Commun.*, 174(9):704–727, 2006.

[120] Tcl Core Team. The tool command language (Tcl). `http://www.tcl.tk/`.

[121] W. Humphrey, A. Dalke, and K. Schulten. VMD - visual molecular dynamics. *J. Molec. Graphics*, 14:33–38, 1996. `http://www.ks.uiuc.edu/Research/vmd/`.

[122] T. Williams and C. Kelley. gnuplot - an interactive plotting program. `http://www.gnuplot.info/`.

[123] Python Software Foundation. The Python programming language. `http://www.python.org/`.

[124] S. J. Plimpton. Fast parallel algorithms for short-range molecular dynamics. *J. Comp. Phys.*, 117:1–19, 1995.

[125] S. Y. Trofimov, E. L. F. Nies, and M. A. J. Michels. Thermodynamic consistency in dissipative particle dynamics simulations of strong ideal liquids and liquid mixtures. *J. Chem. Phys.*, 117(20):9383–9394, 2002.

[126] P. Español and M. Serrano. Dynamical regimes in the dissipative particle dynamics model. *Phys. Rev. E*, 59:6340–6347, 1999.

[127] J. M. V. A. Koelman and P. J. Hoogerbrugge. Dynamic simulations of hard-sphere suspensions under steady shear. *Europhys. Lett.*, 21(3):363–368, 1993.

[128] N. S. Martys. Study of a dissipative particle dynamics based approach for modeling suspensions. *J. Rheol.*, 49(2):401–424, 2005.

[129] Y. Kong, C. W. Manke, W. G. Madden, and A. G. Schlijper. Effect of solvent quality on the conformation and relaxation of polymers via dissipative particle dynamics. *J. Chem. Phys.*, 107:592, 1997.

[130] Y. Kong, C. W. Manke, W. G. Madden, and A. G. Schlijper. Modeling the rheology of of polymer sulotions by dissipative particle dynamics. *Tribology Letters*, 3:133, 1997.

[131] S. Y. Trofimov. *Thermodynamic consistency in dissipative particle dynamics*. PhD thesis, Technische Universiteit Eindhoven, 2003.

[132] R.D. Groot. *Novel Methods in Soft Matter Simulations*, volume 640 of *Lecture Notes in Physics*, chapter Applications of Dissipative Particle Dynamics. Springer, 2004.

[133] L.P. Kadanoff. *Statistical Physics: Statics, Dynamics, and Renormalization*. World Scientific Publishing Company, 2000.

[134] A.F. Jakobsen, O.G. Mouritsen, and M. Weiss. Close-up view of the modifications of fluid membranes due to phospholipase A2. *J. Phys.: Condens Matter*, 17:S4015–S4024, 2005.

[135] J. M. Smith and E. Szathmary. *The Major Transitions in Evolution*. W. H. Freeman Press, New York, 1995.

[136] W. Szostack, D. P. Bartel, and P. L. Luisi. Synthesizing life. *Nature*, 409:387–390, 2001.

[137] T. Tjivikua, P. Ballester, and J. Rebek. A self-replicating system. *J. Am. Chem. Soc.*, 112(3):1249–1250, 1990.

[138] A. I. Oparin. *The Origin of Life*. Macmillan, New York, 1936.

[139] F. J. Dyson. *Origins of Life*. University Press, Cambridge, 1999.

[140] D. W. Heermann. *Computer Simulation Methods in Theoretical Physics. 2nd edition*. Springer, 1990.

[141] K. Binder and D. W. Heermann. *Monte Carlo Simulation in Statistical Physics. 3rd edition.* Springer, 1997.

[142] I. A. Chen and J. W. Szostak. A kinetic study of the growth of fatty acid vesicles. *Biophys. J.*, 87:988–998, 2004.

[143] N. Ono and T. Ikegami. Model of self-replicating cell capable of self-maintenance. *Lect. Notes in Comp. Sci.*, 1674:399–406, 1999.

[144] C. Marsh. *Theoretical Aspects of Dissipative Particle Dynamics*. PhD thesis, Lincoln College, University of Oxford, 1998.

[145] Shi-Ling Yuan, Zheng-Ting Cai, and Gui-Ying Xu. Dynamic simulation of aggregation in surfactant solution. *A. Chim. Sin.*, 60(2):241–245, 2002.

[146] M. Kröger and R. Makhloufi. Wormlike micelles under shear flow: A microscopic model studied by nonequilibrium-molecular-dynamics computer simulations. *Phys. Rev. E*, 53(3):25312536, 1996.

[147] An integrated multiscale computational and experimental approach is applied in the Los Alamos National Laboratory sponsored Protocell Assembly (PAs) project and the European Commission sponsored Programmable Artificial Cell Evolution (PACE) project.

[148] P. L. Luisi, P. Walde, and Oberholzer T. Enzymatic RNA synthesis in self-reproducing vesicles: An approach to the construction of a minimal synthetic cell. *Ber. Bunsenges. Phys. Chem.*, 98:1160–1165, 1994.

[149] A. Pohorille and D. Deamer. Artificial cells: prospects for biotechnology. *Trends in Biotechnology*, 20:123–128, 2002.

[150] B. M. Forrest and U. W. Suter. Accelerated equilibration of polymer melts by time-coarse-graining. *J. Chem. Phys.*, 102(18):7256–7266, 1995.

[151] E. A. G. Aniansson, S. N. Wall, M. Almgren, H. Hoffmann, L. Kielmann, W. J. Ulbricht, R. Zana, J. Lang, and C. Tondre. Theory of the kinetics of micellar equilibria and quantitative interpretation of chemical relaxation studies of micellar solutions of ionic surfactants. *J. Chem. Phys.*, 80:905, 1976.

[152] J. McCaskill, N. Packard, S. Rasmussen, and M. Bedau. Evolutionary self-organization in complex fluids. *Phil. Trans. R. Soc. Lond. B*, 2006. in press.

[153] P. V. Coveney, A. N. Emerton, and B. M. Boghosian. Simulation of self-reproducing micelles using a lattice-gas automaton. *J. Amer. Chem. Soc.*, 118:10719–10724, 1996.

[154] B. Mayer and S. Rasmussen. Self-reproduction of dynamical hierarchies in chemical systems. *Complex Adaptive Systems*, pages 123–129, 1998.

[155] B. Mayer and S. Rasmussen. Dynamics and simulation of micellular self-reproduction. *Int. J. Mod. Phys. C*, 11:809–826, 2000.

[156] P. Nielson. Peptide nucleic acid (PNA) as prebiotic and abiotic genetic material. In S. Rasmussen, M. Bedau, L. Chen, D. Deamer, D. Krakauer, N. Packard, and P. Stadler, editors, *Protocells: Bridging Nonliving and Living Matter*. MIT Press, Cambridge, USA, 2008.

[157] G. F. Joyce. Non-enzyme template-directed synthesis of RNA copolymers. *Orig. Life Evol. Biosph.*, 14:613–620, 1984.

[158] O. L. Acevedo and L. E. Orgel. Non-enzymatic transcription of an oligodeoxynucleotide 14 residues long. *J. Mol. Biol.*, 197(2):187–193, 1987.

[159] T. Rocheleau, S. Rasmussen, P. E. Nielson, M. N. Jacobi, and H. Ziock. Emergence of protocellular growth laws. *Philos. Trans. R. Soc. B*, 2007. in press.

[160] P. Weronski, Y. Jiang, and S. Rasmussen. Molecular dynamics (MD) study of small PNA molecule in lipid-water system. *J. Biophys.*, 2006. in press.

[161] C. Knutson, G. Benkö, T. Rocheleau, F. Mouffouk, J. Maselko, A. Shreve, L. Chen, and S. Rasmussen. Metabolic photo-fragmentation kinetics for a minimal protocell. *Artif. Life*, 2006.

[162] A. Eriksson, M. Nilsson Jacobi, J. Nyström, and K. Tunstrøm. Using force covariance to derive effective stochastic interactions in dissipative particle dynamics. *Phys. Rev. E*, 77(1):016707, 2008.

[163] A. Eriksson, M. Nilsson Jacobi, J. Nyström, and K. Tunstrøm. A method for estimating the interactions in dissipative particle dynamics from particle trajectories. 2008. prerint at arXiv number 0811.1017.

[164] K. Krynicki, C. D. Green, and D. W. Sawyer. Pressure and temperature dependence of self-diffusion in water. *Faraday Discuss. Chem. Soc.*, 66:199 – 208, 1978.

[165] M. H. Cohen and D. Turnbull. Molecular transport in liquids and glasses. *J. Chem. Phys.*, 31:1164, 1959.

[166] E. Száthmary. Simple growth laws and selection consequences. *Trends Ecol. Evol.*, 6:366–370, 1991.

[167] A. Munteanu, C. Stephan-Otto Attolini, S. Rasmussen, H. Ziock, and R. V. Solé. Generic darwinian selection in catalytic protocell assemblies. *Phil. Trans. R. Soc. B*, 362(1486), 2007.

Erklärung über die Eigenständigkeit der erbrachten wissenschaftlichen Leistung

Ich erkläre hiermit, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Bei der Auswahl und Auswertung folgenden Materials haben mir die nachstehend aufgeführten Personen in der jeweils beschriebenen Weise unentgeltlich geholfen.

1. Anders Eriksson Ph. D.: Kollaborative Erstellung der Simulationen in Kapitel 3
2. Dr. Rudolf M. Füchslin: Kollaborative Erarbeitung der Skalierungsprozedur in Kapitel 3
3. Prof. Dr. Steen Rasmussen: Gemeinsame Konzeption und hilfreiche Anregungen zu den Kapiteln 5 und 6

Weitere Personen waren an der inhaltlichen materiellen Erstellung der vorliegenden Arbeit nicht beteiligt. Insbesondere habe ich hierfür nicht die entgeltliche Hilfe von Vermittlungs- bzw. Beratungsdiensten (Promotionsberater oder andere Personen) in Anspruch genommen. Niemand hat von mir unmittelbar oder mittelbar geldwerte Leistungen für Arbeiten erhalten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Ort, Datum Unterschrift