Stochastic simulations of minimal cells: the Ribocell model

Fabio Mavelli

From Eighth Annual Meeting of the Italian Society of Bioinformatics (BITS)
Pisa, Italy. 20-22 June 2011

Abstract

Background: Over the last two decades, lipid compartments (liposomes, lipid-coated droplets) have been extensively used as in vitro “minimal” cell models. In particular, simple and complex biomolecular reactions have been carried out inside these self-assembled micro- and nano-sized compartments, leading to the synthesis of RNA and functional proteins inside liposomes. Despite this experimental progress, a detailed physical understanding of the underlying dynamics is missing. In particular, the combination of solute compartmentalization, reactivity and stochastic effects has not yet been clarified. A combination of experimental and computational approaches can reveal interesting mechanisms governing the behavior of micro compartmentalized systems, in particular by highlighting the intrinsic stochastic diversity within a population of “synthetic cells”.

Methods: In this context, we have developed a computational platform called ENVIRONMENT suitable for studying the stochastic time evolution of reacting lipid compartments. This software - which implements a Gillespie Algorithm - is an improvement over a previous program that simulated the stochastic time evolution of homogeneous, fixed-volume, chemically reacting systems, extending it to more general conditions in which a collection of similar such systems interact and change over the course of time. In particular, our approach is focused on elucidating the role of randomness in the time behavior of chemically reacting lipid compartments, such as micelles, vesicles or micro emulsions, in regimes where random fluctuations due to the stochastic nature of reacting events can lead an open system towards unexpected time evolutions.

Results: This paper analyses the so-called Ribocell (RNA-based cell) model. It consists in a hypothetical minimal cell based on a self-replicating minimum RNA genome coupled with a self-reproducing lipid vesicle compartment. This model assumes the existence of two ribozymes, one able to catalyze the conversion of molecular precursors into lipids and the second able to replicate RNA strands. The aim of this contribution is to explore the feasibility of this hypothetical minimal cell. By deterministic kinetic analysis, the best external conditions to observe synchronization between genome self-replication and vesicle membrane reproduction are determined, while its robustness to random fluctuations is investigated using stochastic simulations, and then discussed.

Background

In recent years, many researchers have been actively working in the field of the de novo synthesis of the artificial cell [1-5], i.e. a cell made from scratch using both synthetic and natural compounds. This scientific challenge has many relevant aspects: first of all, it can reinforce the theory of abiogenesis in the origins of life debate [6], proving that life can emerge spontaneously in a test tube, at least in suitable experimental conditions. Furthermore, the possibility to create a population of artificial cells programmed for the synthesis of chemical compounds of pharmacological and industrial interest is a significant biotechnological goal [7,8]. Artificial cells can also be envisaged, in the maybe not too distant future, as microscopic diagnostic and pharmacological labs to be delivered into the human body in order to synthetize and release drugs as a response to an external stimulus in presence of a disease [9].
The topic of artificial cells is strongly related to the minimal cell notion that defines the simplest cell to be considered alive and then to be experimentally implemented. A minimal living cell, or protocell, can be defined as the minimum supramolecular bounded structure based on the lowest number of molecular species and metabolic processes that is capable of self-maintaining, self-reproducing and evolving [10]. Self-Maintenance is the necessary condition that must be fulfilled by the protocell to be considered alive, i.e. it must stay in a steady state where all its constituents are continuously synthesized and refurbished [11]. Nevertheless, this is not sufficient to implement cellular life as we actually know it; in fact, real cells are also able to grow and to self-reproduce. On the other hand, Self-Reproduction can be seen as a consequence of the cellular metabolism that can keep the minimal cell in a stationary state or bring it to a continuous growing regime that forces the organism to divide in order to maintain its internal coherence, i.e. its stability. Both these features are individual properties of a single cell that can be observed during its life time. In contrast, Evolvability is a collective property that can be exhibited only by a population of cells and on a time scale of several generations, according to a Darwinian selection mechanism [6,12].

Some years ago, Szostak and colleagues proposed a minimal cell prototype called the Ribocell: the RNA-based cell, that in principle can exhibit all the three properties to be considered a living protocell [13]. This theoretical cellular model consists in a self-replicating minimum genome coupled with the self-reproduction of the lipid vesicular container [14]. These authors envisaged the existence of two hypothetical ribozymes [13], one (R_L) able to catalyze the conversion of molecular precursors (P) into membrane lipids (L) and the other one (R_P) able to duplicate RNA strands. Therefore, in an environment rich in both lipid precursors (P) and activated nucleotides (NTPs), the Ribocell can self-reproduce if both processes, i.e. genome self-replication and membrane reproduction (growth and division), are somehow synchronized.

In previous papers [15,16], we have presented an in silico implementation of the Ribocell based on the internal metabolism reported in Figure 1 and on the recently introduced self-replicating lipid vesicle model [17]. By means of a deterministic analysis, we showed that if the kinetic constant for lipid formation k_L is in the range: 1.7·10^{-5}s^{-1}M^{-1}≤k_L≤1.7·10^{-4}s^{-1}M^{-1} then synchronization between vesicle reproduction and genome replication can spontaneously emerge under the model assumptions and kinetic parameters reported in Table 1. Deterministic calculations were performed for two ribozymes 20 bases long and showed that the Ribocell reaches a stationary growth and division regime, where the cell size remains constant after each division along with the amount of genetic materials. Although the observed cell life time stabilizes after the first 10 generations, it remains very high, at over 80 days for all the k_L values in the synchronization range, making the Ribocell very hard to implement and study experimentally.

In this paper, we first apply this model to 100-base-long ribozymes, in an attempt to find the best experimental conditions to reduce so as Ribocell life time. By using the deterministic approach, the robustness of the stationary growth and division regime will be investigated in terms of external substrate concentrations, vesicle size and initial riboyme amount in order to define optimal external conditions for Ribocell self-reproduction.

Therefore, the influence of riboyme length will also be explored in the optimal external conditions by ranging strand size from 20 to 200 bases in length and keeping all the other kinetic parameters constant. 20 bases is in fact the minimum length required to observe a folded RNA structures, i.e. a structure that can reasonably exhibit catalytic action. On the other hand, entities of about 200 nucleotides have been suggested as plausibly ancient proto-riboosomes [18] even though, more recently, smaller subunits of 60 nucleotides have also been considered as plausible candidates [19]. Moreover, the dependence of Ribocell time behavior on the kinetic constants of RNA dimer formation and dissociation will also be studied.

Finally, stochastic simulations will be performed in order to test the robustness of the ribocell base on 100-base length ribozymes in optimal external conditions, with the aim of elucidating the role of intrinsic and extrinsic stochasticity on the time behavior of a protocell population.

**Methods**

Self-reproducing vesicles are compartmentalized chemically-reacting systems where self-assembly processes are coupled with chemical reactions that produce amphiphilic molecules. To study the time behavior of these systems, we use both a deterministic and a stochastic approach in order to get insights into the average behavior of the protocell population and, at the same time, to elucidate the role of random fluctuations. Given a certain minimal cell model, i.e. a reaction mechanism with all the required parameters (kinetic constants, permeability coefficients, initial concentrations), the deterministic analysis can be done by numerically solving the ordinary differential equation set (ODES) or by analytically integrating an approximated reduced set of differential equations. Examples of the latter approach can be found in our previous works where self-reproducing micelles [20] and vesicles [21] were studied. In order to take into
account the stochastic effects, some years ago we developed a Monte Carlo program [22] based on the Gillespie’s direct method [23] designed for coupling amphiphile self-assembly [24] with chemical reactions and hydrophobic solute absorption [25] in a homogeneous well-stirred macroscopic reactor. More recently, this program has been improved and a computational platform called ENVIRONMENT [26] was later developed to cope with the more general case of a collection of reacting lipid compartments that interact and change over time. In the next sub-sections, the general model for a self-reproducing vesicle will be recalled and discussed [17].

**In silico chemically reacting vesicles**

A chemically reacting vesicle can be described as a homogeneous reacting aqueous domain enclosed by a lipid bilayer. Molecules can be exchanged with the external environment thanks to transport processes through the lipid membrane. A flux of water can also take place through the membrane in order to balance the osmotic pressure, i.e. the difference between the internal and the external overall concentrations. Chemical reactions can occur in the vesicle water core, according to the assumed internal metabolism, and amphiphilic molecules can be absorbed from and released towards both the external and internal aqueous solutions. Hence, the vesicle time state is defined by the following array:

\[
x^T = (n_1^C, ..., n_N^C, n_L^μ, V_C)
\]

where \(n_i^C\) are the molecular numbers of species \(X_i\) \((i = 1, 2, ..., N)\) present in the vesicle aqueous core and \(n_L^μ\) is the number of amphiphiles \(X_L\) \((1 ≤ L ≤ N)\) in the membrane. \(V_C\) is the water internal volume. In the stochastic approach, all \(n_i^C\) and \(n_L^μ\) are discrete integer numbers and there exist as many arrays as vesicles in the systems, while in the deterministic analysis there is a single array with

![Figure 1 The Ribocell internal metabolism. (1) Reversible RNA strand association, (2) catalyzed template transcription \((S = R_p, _pR_p, R_s, \) and \(_L_s\) complexation, (3) lipid synthesis.](image)

### Table 1 Kinetic Parameters for the in silico Ribocell model at room temperature.

| Kinetic Parameters | Values | Process Description | Ref. |
|--------------------|--------|---------------------|------|
| \(k_{SS}\) \(s^{-1}M^{-1}\) | \(8.8 \times 10^6\) | Formation of dimers \(R_pR_p\) and \(R_pR_s\) | [28] |
| \(k_S\) \(s^{-1}\) | \(2.2 \times 10^{-6}\) | Dissociation of dimers \(R_pR_p\) and \(R_pR_s\) | [28] |
| \(k_{RPS}\) \(s^{-1}M^{-1}\) | \(5.32 \times 10^5\) | Formation of \(R@S\) | [29] |
| \(k_{R@SS}\) \(s^{-1}\) | \(9.9 \times 10^{-3}\) | Dissociation of Complexes \(R@S, S\) | [29] |
| \(k_{NTP}\) \(s^{-1}M^{-1}\) | \(0.113\) | Nucleotide Polymerization in Oleic Vesicle | [31] |
| \(k_L\) \(s^{-1}M^{-1}\) | \(1.710^3\) | Lipid Precursor Conversion* | [30] |
| \(k_{NTP}\) \(s^{-1}M^{-1}\) | \(7.6 \times 10^9\) | Oleic acid association to the membrane | [26] |
| \(k_{OL}\) \(s^{-1}M^{-1}\) | \(7.6 \times 10^2\) | Oleic acid release from the membrane | [26] |
| \(P_W\) \(\text{cm} \cdot \text{s}^{-1}\) | \(4.2 \times 10^9\) | Membrane Permeability to Lipid Precursor | [31] |
| \(P_{NTP}\) \(\text{cm} \cdot \text{s}^{-1}\) | \(1.9 \times 10^{-11}\) | Membrane Permeability to Nucleotides | [31] |
| \(P_W = P_S\) | \(0\) | Membrane Permeability to W and genetic staff | | |
| \(P_{OL}\) \(\text{cm} \cdot \text{s}^{-1}\) | \(1.0 \times 10^{-3}\) | Oleic Acid Membrane Permeability to Water | [32] |

*\(k_L\) is \(10^5\) times larger than the value of the splicing reaction, catalyzed by the hammerhead ribozyme.*
The vesicle surface is estimated by the formula: 

\[ S = \frac{\pi}{2} \sqrt{\frac{L}{\rho}} \]

where \( L \) is the head area and \( \rho \) is the molar volume of the amphiphile. This formula accounts for the double layered structure of the membrane.

Monte Carlo simulations according to the Gillespie differential kinetic equations. On the other hand, the deterministic reaction rates are given according to the mass action law and they make it possible to write down the deterministic set of ordinary differential equations. The deterministic reaction rates may bring the vesicles towards unstable conditions. The stability of the vesicle membrane can then be monitored by means of the reduced surface \( \phi \):

\[ \phi = \frac{S}{S^0} = \frac{S_n}{\sqrt{36\pi V_C^2}} \]

that is, the ratio between the actual membrane surface \( S_n \) and the spherical area that would perfectly wrap the actual core volume \( V_C \). Assuming that, for a given size, the spherical shape \( (\phi = 1) \) represents the minimum energy state, swollen \( (\phi < 1) \) and deflated \( (\phi > 1) \) vesicles are in high energetic conditions due to the elastic and the bending tension, respectively. Therefore, vesicles are assumed to be stable only in a small range of \( \phi \) values around 1:

\[ (1 - \varepsilon) \leq \phi \leq 2^{1/3} (1 + \eta) \]

\( \varepsilon \) and \( \eta \) being the osmotic and dividing tolerance coefficients, respectively. In fact, vesicles in hypotonic solutions can swell, stretching the membrane until they reach a critical state: \( \phi = 1 - \varepsilon \). The osmotic tolerance \( \varepsilon \) can be experimentally determined by measuring the maximum difference in osmolite aqueous concentrations -between the internal core (\( C_i^F \)) and the external environment (\( C_E^F \)) - that vesicles can bear. For oleic acid, \( \varepsilon \) was found equal to 0.21 [27]. In our model, at the critical bursting point, vesicles are assumed simply to break in two, giving rise to two new vesicles that are in high energetic conditions due to the elastic and the bending tension, respectively.

### Table 2 Deterministic rates against propensity density probability for reacting and transport events.

| Event                        | Deterministic Rate \( (M_a)^3 \) | Propensity Density Probability \( \phi^n \) |
|------------------------------|----------------------------------|--------------------------------------------|
| Internal Chemical reactions  | \( a_{1,\mu}X_1 + \ldots + a_{N,\mu}X_N \rightarrow b_{1,\mu}X_1 + \ldots + b_{N,\mu}X_N \) | \( \frac{k_r}{(V_C N_A)^{\mu - 1}} \sum_{i=1}^N (n_i^C)^{\mu - 1} \) |
| Solute \( X_n \) membrane transport | \( P_n \frac{S^C_n}{N_A V_C} \left( C_n^E - C_n^C \right) \) | \( D_n S_n \) |
| Membrane Lipid Release       | \( \frac{N_A V_C}{k_{\text{out}} n_i^C} \) | \( k_m \) |
| Membrane Lipid Uptake        | \( \frac{k_{\text{in}} S_n}{N_A V_C} \left( C_n^C \right) \) | \( k_{\text{in}} \) |
| Water Flux \((d)\)           | \( v_{\text{in}} P_{\text{appl}} S_n \) \( (C_n^C - C_n^E) \) | \( V_C \sum_{i=1}^N (n_i^C)^{-1} \) |

\( a \) and \( b \) stoichiometric matrices, \( N_A \) Avogadro’s number, \( k_r \) kinetic constant, \( M_a \) moleularity [22]

- The relationship between the macroscopic permeability \( P_n \) and the molecular diffusion coefficient \( D_n \) is: \( D_n = P_n \lambda_n N_A \lambda_n \) being the membrane thickness.
- The absolute value guarantees that the propensity density probability is positive and the molecules move in the opposite direction from the concentration gradient.
- \( v_{\text{in}} \) is the water molar volume, while \( C_i^E \) and \( C_i^C \) are the total osmotic concentration in the external and internal aqueous solutions, respectively.

Real values that represents the average time state of the entire reacting vesicle population.

Table 2 compares the concentration change rates with the event propensity probability density functions for all the mentioned processes. The deterministic reaction rates are given according to the mass action law and they make it possible to write down the deterministic set of ordinary differential equations. On the other hand, the propensity probability density functions are used to carry out Monte Carlo simulations according to the Gillespie’s direct method by means of the ENVIRONMENT software [26]. Moreover, in the stochastic simulations, the water flux will be considered instantaneous due to the high permeability of lipid membranes to water. At each step throughout the MC run, the aqueous vesicle core volume is rescaled in order to keep the vesicle in an osmotic balanced state by using the formula reported in Table 2. Conversely, deterministic calculation will explicitly take into account the water flux dynamics through the lipid membrane. However, a discrepancy between the two approaches can emerge only in presence of an abrupt change of the osmolite external concentration, due to fast dilution or solute addition of the external solution [26]. Another main difference between the two approaches is how the encapsulated molecules are distributed between the daughters after protocol division. In the deterministic approach, all the molecular content of the mother vesicle is halved while it is randomly distributed between the twin daughters in the stochastic simulations.

### Self-reproducing vesicles

The vesicle surface is estimated by the formula: \( S_n = \frac{a L n_i^P}{2} \) where \( a_i \) is the amphiphile \( X_L \) head area and 1/2 takes into account the double layered structure of the membrane. In presence of the synthesis of fresh amphiphiles, the membrane surface and the aqueous core volume can follow two different dynamics and this may bring the vesicles towards unstable conditions. The stability of the vesicle membrane can then be monitored by means of the reduced surface \( \phi \):

\[ \phi = \frac{S}{S^0} = \frac{S_n}{\sqrt{36\pi V_C^2}} \]

that is, the ratio between the actual membrane surface \( S_n \) and the spherical area that would perfectly wrap the actual core volume \( V_C \). Assuming that, for a given size, the spherical shape \( (\phi = 1) \) is the minimum energy state, swollen \( (\phi < 1) \) and deflated \( (\phi > 1) \) vesicles are in high energetic conditions due to the elastic and the bending tension, respectively. Therefore, vesicles are assumed to be stable only in a small range of \( \phi \) values around 1:

\[ (1 - \varepsilon) \leq \phi \leq 2^{1/3} (1 + \eta) \]

\( \varepsilon \) and \( \eta \) being the osmotic and dividing tolerance coefficients, respectively. In fact, vesicles in hypotonic solutions can swell, stretching the membrane until they reach a critical state: \( \phi = 1 - \varepsilon \). The osmotic tolerance \( \varepsilon \) can be experimentally determined by measuring the maximum difference in osmolite aqueous concentrations -between the internal core (\( C_i^F \)) and the external environment (\( C_E^F \)) - that vesicles can bear. For oleic acid, \( \varepsilon \) was found equal to 0.21 [27]. In our model, at the critical bursting point, vesicles are assumed simply to break in two, giving rise to two new vesicles that are in high energetic conditions due to the elastic and the bending tension, respectively.
down, releasing all their internal content into the external environment and remaining in solution as flat bilayers, since bilayer sealing processes are not considered in this model. On the other hand, deflated vesicles are supposed to be able to divide in order to minimize the bending energy. The dividing condition is reached when they can form two equal volume spherical daughters ($\varphi = 2^{1/3}$). So $\eta$ introduces a tolerance that is linked to the relative flexibility exhibited by any membrane. However, as a simplifying assumption in this work, $\eta = 0$ will be supposed. As already mentioned, after each division, all the molecular content of the mother vesicle is halved in the deterministic approach, while it is randomly distributed between the twin daughters in the stochastic simulations.

The Ribocell model

Figure 1 reports the internal metabolism proposed for the Ribocell. Both pairs of RNA strands reversibly associate (1) and these processes are shifted towards the dimer formation and are strongly dependent on temperature. The replication of any RNA strand is catalyzed by the polymerase $R_P$ according to the steps in bracket (2). This process is described as a catalytic template-directed addition of mononucleotides with high fidelity and processivity. It starts with $R_P$ binding any of the monomeric strands $S$ ($S = R_P, cR_P, R_L$ and $cR_L$) to form the complex $R@S$. This complex will then initiate the polymerization of the conjugate strand $S$, by coupling and iteratively binding the complementary bases and releasing the by-product $W$. When the strand $S$ has been completely formed, the polymerase ribozyme releases the new dimer. Finally, the ribozyme $R_L$ catalyzes the conversion of the precursor $P$ into the lipid $L$ (3).

Table 1 shows the values of all kinetic constants and membrane permeabilities used in this work, with their respective references. In this work, the length of the two ribozymes is assumed to be 100 nucleotides long (20 being the minimum base number for observing a folded RNA conformation). Both $R_L$ and $R_P$ are created with a random sequence of bases, and they are assumed to have similar kinetic behaviors for the sake of simplicity. The kinetic constants of formation $k_{SS}$ and dissociation $k_S$ of both dimers were set equal to the values experimentally measured for a sequence of 10 nucleotides [28]. The kinetic constants for both complex formation $R@S$ and complex dissociation $R@S, S$ ($S = R_P, cR_P, R_L$ and $cR_L$) were set equal to those measured for the human enzyme $\beta$-polymerase [29] that binds a DNA single strand and dissociates from a DNA double helix, respectively. The rate constant for the catalytic synthesis of lipid $k_L$ was taken to be $10^5$ times larger than that of the splicing reaction, catalyzed by the hammerhead ribozyme [30]. The kinetic behavior of different nucleotides is assumed to be the same and a single value is assigned to $k_{\text{NTP}}$ derived from experimental data simulations (De Frenza private communication) of the DNA template directed synthesis in fatty acid mixed vesicles [31]. In the same way, the common value of the membrane permeability to activated nucleotides was also estimated, while the membrane is assumed impermeable to genetic material. The kinetic constants of the membrane/aqueous solution lipid exchange for oleic acid vesicles are taken from a previous work [26] where they were obtained by simulating the competition between isotonic and osmotically swollen oleic vesicles [27]. The amphiphile head area $\alpha_t = 0.3$ nm$^2$ and the osmotic tolerance $\varepsilon = 0.21$ are defined according to data reported in literature for oleic acid vesicles [27]. The only two parameters assigned arbitrarily are therefore the membrane permeability to the byproduct: $P_W = 0.0$ cm/s, based on the assumption that $W$ is a charged species, and the permeability to the precursor: $P_P = 0.42 \cdot 10^{-8}$ cm/s, corresponding to the oleic acid membrane’s permeability to Arabinol [32] and comparable to those of similar organic compounds.

A common simplifying assumption to both approaches is to consider the external concentrations of nucleotides (NTPs), lipid precursor ($P$) and inert compound ($B$) to be constant throughout the time-course of the process, thanks to an incoming flux of material in the reactor vessel: continuous stirred tank reactor approximation. The byproduct ($W$) concentrations is also assumed to be constantly equal to zero outside. For all these compounds, except $B$, the internal aqueous concentration is zero at the beginning. Moreover, all the calculations are performed starting from an initial isotonic condition, so that the internal concentration of $B$ is properly adjusted in order to counterbalance the presence of the lipid precursor and nucleotides outside and the genetic material inside, respectively.

As has been pointed out before, given the set of kinetic parameters reported in Table 1 we are looking for a set of initial concentrations that can allow the Ribocell to reach a stationary regime of growth and division, i.e. a dynamic state where the protocol grows and divides producing two twin daughters the same size as their mother at the beginning of its life cycle.

In order to achieve this in our model, a spontaneous synchronization between membrane and the aqueous volume core of the self-replicating vesicle must take place. By introducing the control growth coefficient as the ratio between the relative change rates of volume $\nu_V$ and surface $\nu_S$:

$$\gamma = \frac{\nu_V}{\nu_S} = \left(\frac{1}{V_C} \frac{dV_C}{dt}\right) / \left(\frac{1}{S_\mu} \frac{dS_\mu}{dt}\right)$$  

(4)
It is easy to show (unpublished observations) for growing protocells that synchronization can take place only if $\gamma = 1$, while if $\gamma > 1$ the volume will increase much faster than the membrane surface and the vesicle can become energetically unstable, leading to an osmotic burst. On the other hand, if $\gamma < 1$ then the vesicle will divide decreasing in size generation by generation. It is beyond the scope of this work to go into the mathematical details of this formula, so the interested reader should refer to an incoming paper for a more detailed discussion.

Nevertheless, it is worthwhile to keep in mind that $v_3$ is essentially proportional to the rate of lipid synthesis, since amphiphile uptake by the membrane is very fast when the concentration of lipids is above the equilibrium value. Instead, having assumed the external value $C_T^{E}$ to be constant and a vesicle being in an osmotic balanced condition: $C_T^{E} = C_T^{C} = V_C N_T^{C}; v_T$ is driven by the overall internal $N_T^{C}$ population rise. Therefore, since $N_T^{C}$ increases essentially owing to the waste production that takes place with ribozyme self-replication and lipid synthesis, this has the effect of coupling the membrane reproduction with the genome replication and with the volume growth: osmotic synchronization.

When the Ribocell reaches a stationary regime, at each division the genetic materials can be randomly distributed between the daughters. If the amount of genetic material is very low, then this can result in a separation of $R_P$ from the other RNA strands. In fact, the Ribocell must contain a minimum genetic kit of three RNA filaments in order to be capable of self-replicating its entire genome: one $R_P$ that catalyzes the RNA base pair transcription, one ($R_L$ or $cR_L$) and one ($R_{cP}$ or $cR_{cP}$) that work as templates for the transcription. Moreover, since $R_L$ is necessary to catalyze lipid precursor conversion, the optimal minimum 3-ribozyme kit must be made up of $2R_P$ and one $R_L$. This minimum kit should be at least doubled before cell division, in order to have a chance that both daughters continue to be active. Therefore, if a random distribution of RNA filaments takes place after vesicle division, ribozyme segregation between the two daughters might occur. Different scenarios can be envisaged as sketched in Figure 2: death by segregation is reached if vesicles are produced without any ribozymes (empty vesicles) or containing one lone $R_P$ or many filaments of $cR_{cP}$ and/or $cR_{cP}$ (inert vesicles). Vesicles that encapsulate $R_L$ strands are self-producing: they are able to synthesize lipids and then can grow and divide producing in turn self-producing and/or empty vesicles. On the other hand, vesicles containing more than one molecule of $R_P$ or both $R_P$ and $cR_P$ filaments are able to self-replicate this reduced genome (self-replicating genome).

Figure 2 Different-reacting protocells and vesicles obtained by RNA segregation due to Ribocell division. Nucleotides (NTPs) and waste (W) have been omitted for the sake of clarity, along with the reversible association of RNA.
vesicles) but they cannot self-reproduce the membrane. So they are destined for an osmotic burst due to an unbalanced increase in waste concentration. Finally, a reduced version of the Ribocell consists in a lipid aggregate that contains one $R_F$ filament and $R_L/R_I$ strands. As a consequence of this, reduced ribocells are able to replicate the $R_L/R_I$ genetic stuff, and at the same time to synthesize lipids. Therefore, they can grow and divide, producing in turn at least one reduced ribocell and/or self-replicating, inert and empty vesicle. With the help of stochastic simulations, we will try to explore all the possible scenarios.

Results and discussion

Deterministic analysis

In the present paper, we firstly explore the dependence of the stationary regime on the external concentrations of substrates for ribozymes 100 nucleotides long setting $k_L$ to $1.7 \cdot 10^3 s^{-1} M^{-1}$, i.e. the minimum value in the previously observed synchronization range. The aim of this preliminary deterministic study is to find the optimal initial conditions in order to achieve a stationary regime with the shortest life time. All the outcomes are reported in additional file 1.

As first, the dependence of Ribocell state on overall external concentration $C_T^E$ is analyzed at the stationary regime, reached after 20 generations. Since $[P_{ex}] = [N_{ex}] = 5.0 \cdot 10^{-4} M$, the overall external concentration can be approximated to $C_T^E = [I_{ex}]$. The upper plots in Figure 3 show that when the overall external concentration $[I_{ex}]$ increases, then the Ribocell radius $r_{20}$ decreases, while the life cycle $\Delta t_{20}$ rises. Thus, vesicles become smaller and more dormant as the overall external concentration rises. This can be ascribed to the mechanism of synchronization itself and is in agreement with what we reported in a recent work (unpublished paper) where an inverse dependence of the vesicle steady size on overall external concentration was explicitly derived from the general stationary condition $\gamma = 1$. On the other hand, the observed increase in Ribocell life time is a direct consequence of the reduction in size, since a smaller membrane surface decreases the transport efficiency of substrates (lipid precursor and nucleotides) from outside. As a consequence of this, all metabolic processes slow down since they are sustained by the transport of external substrates. These two effects, i.e. the increase in lifetime and the slowdown of the metabolism, determine the linear rise in concentration of the

![Figure 3](http://www.biomedcentral.com/1471-2105/13/S4/S10)  
Figure 3: Dependence of the Ribocell stationary regime on the external concentration of the inert compound $[I_{ex}]$. Vesicle radius (left upper plot), division time (right upper plot), overall internal concentrations of RNA strands (left lower plot) and genome composition percentage (right lower plot) were determined after 20 generations.
overall genetic material, see the lower left plot of Figure 3. At the steady regime, the genome composition is almost independent of \( I_{ex} \), as shown by the lower right plot of Figure 3. Two equal fractions of \( R_L \) and \( \cdot R_L \) strands are present, both around 33%, while the percentages for \( R_P \) and \( \cdot R_P \) are lower: 24% and 10%, respectively. Due to the high \( k_{ss} \) value, the ribozymes are mainly present in the form of dimers and this accounts for the equal fractions observed for the lipase ribozymes, while the percentage of \( R_P \) is greater than that of \( \cdot R_P \), since some \( R_P \) strands are involved as catalyzers in template duplication. This also explains why the overall percentage of polymerase ribozymes (\( \approx 34\% \)) is lower than that of lipase ribozymes (\( \approx 66\% \)). In fact, not all polymerase ribozymes are available as templates for duplication. Moreover, this asymmetry is amplified as long as the total concentration of the genetic material increases, see data in additional file 1.

Setting \( I_{ex} = 0.3 \) M, we study the dependence of the stationary division regime on the external concentration of the substrates: lipid precursor \( [P_{ex}] \) and nucleotides. The same concentration value \( [N_{ex}] \) is set for the four different nucleotides since they have been assumed to have the same kinetic behavior. The upper plot of Figure 4 shows the opposite effects of \( [N_{ex}] \) and \( [P_{ex}] \) on the stationary vesicle radius \( \rho_{20} \). Higher \( [N_{ex}] \) concentrations speed up genome self-replication with respect to lipid synthesis, accelerating waste production and leading to larger core volumes. Conversely, increasing \( [P_{ex}] \) reduces \( \rho_{20} \) since membrane self-reproduction becomes faster. For the same reason, the total concentration of genetic material is increased due to the high concentrations of nucleotides and the low concentrations of the lipid precursor, while, both substrate concentrations decrease cell life time when they are increased, since all the metabolic processes are accelerated. If \( [N_{ex}] \geq 0.05 \) M, the Ribocell undergoes an osmotic burst since volume growth is too fast compared to lipid production for any value of \( [P_{ex}] \) in the studied range (see additional file 1).

In the upper plots of Figure 5, the time trends of the core volume on the left, the growth control coefficient \( \gamma \) and the reduced surface \( \varphi \) (on the right) are reported for a Ribocell starting with a number of dimers \( N_0 = 100 \) for both \( R_R \) and \( R_R \). These plots show that after a few generations the steady division regime is reached,

![Figure 4](image-url)
as confirmed by $\gamma$ that tends to 1.0. At the real beginning, the Ribocell undergoes a fast volume increase, as demonstrated by $\gamma > 1.0$ and $\varphi < 1.0$. This is due to the transport process of the lipid precursor and the nucleotides from the external environment that blows up the vesicle. In the lower plot of Figure 5, the division time is reported against the number of generations for ribocells starting with a different initial number of strands $N_0$. In all three cases, the same division time is reached after 10 generations, i.e. 68.2 days, that remains constant for the following generations. The higher $N_0$, the faster the cell division in the first generations.

Having determined optimal external conditions, the influence of ribozyme length is now investigated by keeping all the other kinetic parameters constant. Calculations have been performed changing in turn the length of $R_L$ or $R_P$, and fixing the size of the other ribozyme to 100 bases, or changing the size of both $R_L$ and $R_P$ but keeping the same length. Results are reported in Figure 6. In all the studied cases, an increase in strand length determines at the stationary regime Ribocells with a longer life cycle $\Delta t_{25}$, with a larger radius $r_{25}$ and a lower RNA total concentration [RNA Strands]. The variations observed are quite small compared to those for 100-base ribozymes, except for the [RNA Strands] that show a change about 15%. Furthermore, the Ribocell shows to be much more sensitive to the change in size of the polymerase ribozyme $R_P$ rather than $R_L$. This can be ascribed to the fact that, being longer, $R_P$ requires more time to self-replicate and this decreases

---

**Figure 5** Deterministic time evolution of the Ribocell. Aqueous core volume (left upper plot), growth control coefficient $\gamma$ on the left axis, and reduced surface $\varphi$ on the right axis (right upper plot) are reported against time for a ribocell starting with the initial number $N_0$ of both $R_L$ and $R_P$ dimers equal to 100; on lower plot, division times against generation number for different initial $N_0$ values are displayed. In the upper plots, the vertical dashed lines represent the cell division times that take place when $\varphi = 2^{1/3}$.
the overall concentration of all the polymerase ribozymes and in turn the efficiency of genome self-replication and membrane self-reproduction.

Finally, Figure 7 displays the dependence of $\Delta t_{25}$ on the kinetic constants for RNA dimer formation $k_{SS}$ and dissociation $k_S$. The plot clearly shows that the Ribocell life cycle at stationary regimes does not depend explicitly on the kinetic constant single values $k_{SS}$ and $k_S$ but on their ratio: $k_{SS}/k_S$, that is on the thermodynamic constant of RNA dimerization. The more thermodynamically stable the RNA dimers, the longer it takes to observe Ribocell self-reproduction. For instance, if $k_{SS}/k_S$ is decreased by two orders of magnitude, the Ribocell life time reduces from 68.2 days to 11.8-6.4 days. The study of Ribocell time behavior approaching the stationary regime as a function of $k_{SS}$ and $k_S$ values would require a much deeper analysis that is beyond the scope of this paper.

**Stochastic simulations**

Stochastic simulations were performed by means of the parallel version of ENVIRONMENT, running 32 statistically equivalent simulations of a 10-ribocell solution on different CPUs. Therefore, the outcomes were obtained as averages from a population of 320 vesicles. Kinetic parameters used for simulations are those reported in Table 1 while the initial conditions are shown in bold by additional file 1. At each cell division, only one of the two
offspring was kept while the other was discarded in order to reduce computation time, thus keeping the number of monitored vesicles constant. This is in agreement with the assumption that the external concentrations of all substrates are fixed due to an incoming flux of material, i.e. the substrates cannot ever be exhausted.

On the left in Figure 8, the time evolutions of protocell populations obtained by stochastic simulations are reported for the three studied cases starting with a genome made up of 1, 10 and 100 dimers of both $R_c R_l$ and $R_c R_P$. At the end of the simulations of all three cases, similar compositions of the protocell population are

![Figure 8: Stochastic simulation outcomes](http://www.biomedcentral.com/1471-2105/13/S4/S10)

Figure 8 Stochastic simulation outcomes. Plots on the left show the time evolution of ribocell populations while the legend shows the final compositions; plots on the right report the stochastic $\langle \Delta t_d \rangle$ (circles with error bars) and deterministic $\Delta t_d$ (red triangles) division times on the left axis while the percentage of dividing protocells (dashed gray lines) is reported on the right axis against the generation number $n$. The initial genome was made up as follows: 1 (upper plots), 10 (middle plots) and 100 (lower plots) dimers of both $R_c R_l$ and $R_c R_P$. 
obtained with low percentages of real ribocells (3.3-6.7%) while the most populated fractions are those of empty (40.0-41.7%) self-producing (26.7-33.3%) and broken (18.3-25.0%) vesicles, respectively. Reduced ribocells are present only in the first generations since they very soon decay into self-producing and empty vesicles. Inert vesicles, i.e. vesicles entrapping free chains of \( R_P \) and/or \( R_L \) or a single \( R_P \), are not formed and this can be ascribed to the high stability of RNA dimers and complexes so that the chance of finding free RNA monomers at the time of vesicle division is extremely improbable. Thus, the three studied cases are differentiated by their time behavior rather than by the final protocell population, as confirmed by the plots on the right in Figure 8, where the average division time \( \langle \Delta t_n \rangle \) and the number of dividing protocells are reported against the generation number. In agreement with deterministic predictions, for the first generations the average division times \( \langle \Delta t_n \rangle \) are higher for ribocells starting with a lower initial number \( N_0 \) of dimers although, in all cases, the deterministic \( \Delta t_n \) (red triangles) are greater than the stochastic averages (black circles with error bars). This can be partially ascribed to the fact that the average \( \langle \Delta t_n \rangle \) is calculated on all the protocells that undergo the \( n \)-th division and only some of them are real ribocells. In fact, generation by generation, the protocell population is enriched by self-producing vesicles that can divide more quickly if a free \( R_L \) monomer is present and this lowers the average division time.

As an example, Figure 9 reports the time behavior of a single Ribocell with a starting genome made up of just one \( R_P R_L \) and one \( R_P R_P \). In the upper plot, a comparison between the reduced surface time course determined deterministically and simulated stochastically is shown. As can be seen, the stochastic time trend presents a very irregular time behavior compared to the deterministic one that describes a highly synchronized oscillating regime of growth and division. In contrast, stochastic simulations highlight the alternation of dormant phases, where the reduced surface remains practically constant, both the core volume and the membrane surface being constant (data not shown), to very active steps where protocell growth takes place very fast, leading to a division event.

In order to account for this behavior, in Figure 10 the time course of the total number of lipase \( R_L \) and polymerase \( R_P \) ribozymes present as monomers, complementary strands, dimers and complexes, are reported.

Figure 9 Stochastic time behavior of a single ribocell with \( N_0 = 1 \). In the upper plot, the stochastic time trend (black line) of the reduced surface \( j \) is compared with the deterministic time course (gray line); vertical dashed lines indicate the simulated division times. The lower plot explicitly shows the 14 divisions and the two transformations of the ribocell due to RNA segregation. After the first division, it becomes a reduced ribocell and after 10 generations a self-producing vesicle.
against time, along with the number of free RL and RP monomers that exhibit catalytic activity. As can be seen, the fast growth and division step corresponds to the presence in the vesicle core of a free RL chain while, in the dormant phase, ribozymes are all coupled in the form of dimers or complexes. In fact, during RNA template transcription in the first generation life time, the volume remains practically constant since the amount of waste molecules produced is not sufficient to promote a substantial water flux from the external environment. As a consequence, self-producing vesicles with a genome made up only of RL monomers can reproduce very efficiently since no dormant phase can occur, given that the formation of RCRL dimers is impossible. This is what happens at high generation numbers in the protocell population time evolutions reported in Figure 8. It is clear in both cases with a starting genome \( N_0 \) equal to 1 and 10. In fact, at high generation numbers, the only dividing protocells are self-producing vesicles that present free RL monomers in the core volume. Although there are very few of these protocells, they can divide very efficiently, with a \( \Delta t \) of around 0.81 days, and with division times that are very close to one another, so the population average \( <\Delta t_n> \) seems very low with a small error bar as shown by the upper and middle plots on the left in Figure 8.

**Conclusions**

In this paper, we applied an already published Ribocell *in silico* model [15,16] to the case of two hypothetical ribozymes 100 nucleotides long by using the kinetic parameters reported in Table 2. The kinetic constant for the lipid formation, \( k_L \), was set equal to 1.7E+3s\(^{-1}\)M\(^{-1}\), the lowest value that exhibited a stationary regime of growth and division in a previous work [16] in which ribozymes 20 nucleotide long were assumed. Length of 100 nucleotides was chosen as a compromise between the need to reduce calculation times and the choice of a plausible ribozyme size, keeping in mind that the RNA subunits in ribosomes devoted to protein elongation have a comparable length [18,19].

By means of deterministic analysis, the robustness of the stationary regime was also investigated as a function of the initial conditions, the length of ribozymes and the kinetic constants of the RNA dimerization. For 100-base long ribozymes, the best experimental conditions in
terms of the external concentrations have been found: \([N_{\text{ex}}] = [P_{\text{ex}}] = 1.0 \times 10^{-2} \text{M}\) and \([I_{\text{ex}}] = 0.3 \text{M}\) in order to observe a stationary regime with the lowest division time: \(\Delta t_{20} = 68.3\) days. A so high protocell life time can be mainly ascribed to the RNA reversible association that is shifted towards the dimer formation rendering the concentration of the catalyzers \(R_L\) and \(R_P\) very low. This has been confirmed by analyzing the dependence of the life cycle on the thermodynamic constant of ribozyme dimerization. A small influence on the stationary regime was observed on changing the length of the RNA strands: an increase in filament size determines higher life times and a lower amount of genetic material. This effect is more pronounced when the \(R_P\) length changes. On the other hand, the external concentrations of substrates and inert compounds appear to highly affect the stationary regime in terms of vesicle size, genetic material amount and genome composition. Moreover, deterministic calculations have also shown that the stationary regime can be reached from very different initial genome composition, although when too much free \(R_L\) ribozymes are present at the beginning the death for ribozymes segregation can be observed since the protocell divides too quickly before the genome replication. Stochastic simulations have been done starting from a population of 320 identical ribocells with an initial genome composed by \(N_0 = 1,10\) and 100 dimers of both \(R_LR_L\) and \(R_PR_P\). The analysis of simulations outcomes shows that the ribocell time behaviors is highly influenced by random fluctuations. Since the genetic material is randomly distributed at each cell division, this can produce different type of protocells, ranging from empty vesicles to genuine ribocells, their internal metabolism being highly influenced by the presence of the catalytic RNA strands. In fact, deterministic analysis cannot take into account the disappearance of ribozymes due to a vesicle division, since this approach simply halves the genetic amount and follows the reacting molecule time courses in terms of population averages, i.e. real positive numbers that can be less than one without being zero. On the other hand, the stochastic simulations are more realistic to the random loss of ribozymes from the genome being capable of describing a population of protocells with completely different time behaviors. As a consequence, the simulation outcomes show that ribocells are not enough robust to survive to random fluctuations. In fact only about the 5% of the initial population survive as genuine ribocells after 15-25 generations and on a longer time window they are destined for extinction. Furthermore, the time course of each single protocell is also greatly influenced by intrinsic stochasticity in particular by the time fluctuations of the RNA dimer dissociation. In fact, when all the RNA strands are associated in dimers, protocells remain in a lazy phase, whereas free \(R_L\) monomers induce fast growth and division steps and free \(R_P\) cause the fast RNA replication without changing the vesicle size appreciably. Therefore these two process are synchronized only by chance and this also represents a reason of weakness of this model protocol.

In order to implement experimentally ribozymes-based minimal cells two main improvements are necessary. As first, more free monomers of both \(R_L\) and \(R_P\) must be available in the vesicle core so that the ribocell life cycle will be speeded up and the division time lowered. This can be achieved by increasing the working temperature since it has been recently show that fatty acid vesicles are stable up to \(90^\circ\)C [33] and the efficiency of the self-catalyzed replication of RNA strands increases with a temperature rise. This is also in agreement with our results since the \(k_{SS}\) value used can be considered as the appropriate 100-base long RNA association constant for a higher temperature than \(25^\circ\)C [34]. For a more detailed theoretical analysis at high temperature, it is necessary, of course, to estimate the kinetic constants for all the steps involved in the internal metabolism. This will be the topic of a future work. The second necessary improvement consists in finding a way to really synchronize the genome self-replication and the membrane reproduction. This is a much more complex task to achieve. Working with high concentration of the genetic material, it can avoid, or at least reduce, the ribozymes segregation and this should be compatible with a high working temperature. The best strategy could be to have a fine control of the \(R_LR_L\) dissociation since when free \(R_L\) monomers are present in the aqueous core the membrane growth quickly and the division takes place very soon. Thus the \(R_LR_L\) dissociation can act as a trigger for the membrane growth and division.

Finally rephrasing the George Box famous sentence, we are aware that this in silico Ribocell model is in a some way wrong, but we hope it might inspire researchers involved in the lab implementation of the ribozymes-based minimal cell.

Additional material

**Additional file 1:** Deterministic Outcomes of the Ribocell time behavior

stationary values for different initial conditions

Acknowledgements

This work has been funded by the MIUR (PRIN2008FY7VR4_002). Parallel calculations were done at CINECA Super Computing Italian Centre thanks to the approved ISCRAC project (HP10CVJLGZ). The authors would like to thank Anthony Green for his careful review of the English in this manuscript.

This article has been published as part of BMC Bioinformatics Volume 13 Supplement 4, 2012: Italian Bioinformatics Society (BITS): Annual Meeting.
28. Christensen U: Thermodynamic and kinetic characterization of duplex formation between 2'-O, 4'-C-Methylene-modified oligoribonucleotides, DNA and RNA. Bioscience Reports 2007, 27:327-333.

29. Tsoi PY, Yang M: Kinetic study of various binding modes between human DNA polymerase β and different DNA substrates by surface-plasmon-resonance biosensor. Biochem J 2002, 361:317-325.

30. Stage-Zimmermann TK, Uhlenbeck OC: Hammerhead ribozyme kinetics. RNA 1998, 4:815.

31. Mansy SS, Schrum JP, Kishnamurthy M, Tobé S, Treco DA, Szostak JW: Template-directed synthesis of a genetic polymer in a model protocell. Nature 2008, 454:122-U10.

32. Sacerdote MG, Szostak JW: Semipermeable lipid bilayers exhibit diastereoselectivity favoring ribose. PNAS 2005, 102:6004-6008.

33. Mansy SS, Szostak JW: Thermostability of model protocell membranes. PNAS 2008, 105:13351-13355.

34. Lincoln TA, Joyce GF: Self-Sustained Replication of an RNA Enzyme. Science 2009, 323:1229-1232.

Cite this article as: Mavelli: Stochastic simulations of minimal cells: the Ribocell model. BMC Bioinformatics 2012 13(Suppl 4):S10.