amcs

MATHEMATICAL MODELLING OF MOLECULE EVOLUTION IN PROTOCELLS

DARIUSZ MYSZOR, KRZYSZTOF A. CYRAN

Institute of Informatics
Silesian University of Technology, Akademicka 16, 44-100 Gliwice, Poland
e-mail: {dariusz.myszor,krzysztof.cyran}@polsl.pl

In this article, we analyse the process of the emergence of RNA polynucleotides located in an enclosed environment, at an early stage of the RNA world. Therefore we prepared a mathematical model, composed of a set of differential equations, which simulates the behaviour of an early biological system bounded by a protocell membrane. There is evidence that enclosed environments were available on the primordial Earth. There are also experimental proofs that RNA strands can develop in these formations. The proposed model allows analysis of the influence of membrane permeability on the composition of internal material. It takes into account phenomena that lead to the elongation of an RNA strand (ligation), fission of molecules (phosphodiester bond breakage) and replication of polynucleotides. Results obtained from the model point out that the existence of protocells might support concentration of material and creation of longer molecules.

Keywords: RNA world, proto-cell, mathematical model.

1. Introduction

Life, defined by seven pillars: program, improvisation, compartmentalization, energy, regeneration, adaptability and seclusion (Koshland, 2002), is a common phenomenon on the surface of our planet. Living entities exist in many possible conditions, even the most extreme ones (Cary *et al.*, 1998; Schleper *et al.*, 1995; Stetter, 2007); however, regardless of the ubiquitous nature of life, we still do not know how this process began. The problem of the emergence of life is difficult to solve; up to now direct evidence, of the earliest life formation, has not been found, and because life emerged over 4 billion years ago all traces of this process probably have vanished (Lazcano, 2010). Moreover, the exact conditions on the early Earth are unknown (Lunine, 2006; Sleep, 2010; Mojzsis *et al.*, 1999).

There are many speculations and theories concerning the values of environmental parameters such as chemical composition of pre-biotic atmosphere (Zahnle *et al.*, 2010), oceans (Feig and Uhlenbeck, 1999) and rocks (Sleep, 2010), distribution of temperature (Gaucher *et al.*, 2010) and pressure. One way of researching the origin of life is laboratory experiments, although they usually require proper substrates of chemical reactions and reproduction of specific conditions, thus being

quite expensive. Much cheaper are mathematical and simulation models. Over the years many of them were utilized in biological applications (such as extensions of Lotka–Vottera equations (Xu *et al.*, 2011) or applications of Markov chains (Wieczorek, 2010)) and in the service of investigation of the emergence of life (Niesert, 1987; Myszor and Cyran, 2009; Ma *et al.*, 2007; Szabó *et al.*, 2002; Cyran, 2009). By using these models we can easily simulate the behaviour of a sample for a wide range of values of parameters. In some cases results obtained from artificial models might, at the first sight, look counter-intuitive, although often after detailed analysis they prove to be correct. Therefore outcomes of simulation and mathematical models might inspire new laboratory experiments (Kimmel and Axelrod, 2002).

1.1. RNA world. Currently, the most popular theory concerning the emergence of life is called the RNA world (Woese, 1968; Gilbert, 1986). It claims that, in a certain period, RNA played a main role in the field of life. RNA strands acted as chemical reaction catalysts and genetic information storage. Both features are important for life. Catalytic abilities are needed to speed up, or even enable, some chemical reactions such as creation of nucleotides, as well as formation and replication of informational strands. The storage of genetic information is required for

heredity, defined as the ability to pass features between generations.

There are a number of possible evidences supporting the RNA world hypothesis, e.g., the discovery of ribozymes by S. Altman and T. Cech, riboswitches (Breaker, 2010) and ribosomes (Noller, 2006). However, there are also many unanswered questions and problems to solve, e.g., how to create efficiently all types of nucleotides in the prebiotic conditions (Sutherland, 2010). One of the main challenges of the RNA world is a limitation of the length of RNA strands. Longer molecules might be able to catalyse chemical reactions like polymerization of polynucleotides. These abilities can lead to an increase in the system complexity and, in a longer time perspective, to emergence of life. However, in the prebiotic conditions, recreated in a laboratory, we can obtain strands up to 10 nucleotides when there are no catalysts and up to 50 nucleotides when there are available mineral catalysts (Ferris *et al.*, 1996). The main obstacle on the way leading to long strands is the weakness of the phosphodiester bond between nucleotides in a water environment. Laboratory experiments revealed that this bond might be broken up in a variety of processes like hydrolysis (De Lucrezia *et al.*, 2007) or UV radiation (Fishkis, 2007). Therefore there is a constant search for conditions that protect long chains from dissolution and facilitate the process of strands elongation.

1.2. Compartmentalization. Compartmentalization is denoted as one of the seven pillars of life. In contemporary living creatures, compartmentalization occurs at many levels. There is a division at the level of organisms, tissues, cells. There are many possible manners of primordial compartmentalization, for example, aerosols (Dobson *et al.*, 2000), shallow pools of primordial solutions and protocells (Chen *et al.*, 2006). Protocells (also denoted as compartments or packages) are primitive cell formations, composed of chemical compounds separated from the outer world by a membrane (wall). Inside these entities, chemical reactions might take place. Nowadays there is a discussion about the influence of package formations on the processes that operated during origins of life such as elongation and replication of strands. Membranes could be formed from amphiphilic molecules (Schrum *et al.*, 2010; Rasmussen *et al.*, 2008). Amphiphilic molecules were available on the primordial Earth because they were brought to our planet by meteorites (Deamer, 1985; Deamer and Pashley, 1989).

There is also a possibility to synthesize amphiphilic molecules under a wide variety of conditions (Hanczyc *et al.*, 2003). Moreover, there are RNA molecules that catalyse amphiphiles formation. Laboratory experiments confirmed that amphiphilic membranes can self-assembly spontaneously in prebiotic environments (Chen and Walde, 2010). Minerals and specific ions that existed on the primordial Earth can speed up the process of formation of membranes. First packages might not possess any advanced and efficient mechanisms that could control internal conditions. Therefore, physicochemical properties of membranes are very important.

One of the most important property of primordial membranes is permeability. The perfect wall of a protocell allows the passage of small molecules (like nucleotides) and other substances necessary for replication of strands. At the same time it should be able to maintain larger formations like polynucleotides within the protocell. Permeability is closely related with the length of strands forming amphiphilic compounds. Shorter strands allow higher permeability, although they reduce stability and ratio of self-assemblage (Pohorille and Deamer, 2002).

The membrane of a protocell might carry out many functions (Pohorille and Deamer, 2009; Matsuura *et al.*, 2002), for example, it

- separates genetic material from the external, potentially dangerous, environment (Szostak *et al.*, 2001),

- maintains specific conditions inside packages as concentration of specialized chemical compounds (ions, mineral particles) that might be needed during the process of replication,

- allows diffusion of small molecules (such as nucleotides),

- can incorporate chemical compounds that modify its properties (for example, create ion channels (Pohorille and Deamer, 2009)), which might help in the achievement of better evolutionary properties.

Even in the simplest packages, the empirically proven cooperation can occur between the content of a protocell and its behaviour. With the growth of internal material, the pressure inside the protocell is increasing, the package is also growing and might cannibalize on other compartments that have weaker material (Chen *et al.*, 2004; 2006). Hence, selection at the RNA level might govern the selection at the cell level. An increase in the internal pressure may lead to fission of the cell. Cells containing strands with better catalytic activities might split more often than other compartments and overtake the supremacy in the population.

2. Model description

2.1. Template model. The original mathematical model was described by Nowak and Ohtsuki (2008), Manapat *et al.* (2009), as well as Ohtsuki and Nowak (2009). It was created in order to simulate competition between life (represented as replication processes) and prelife (represented as the process of elongation and decay

of strands). The authors of the template model needed the simplest possible dynamic system that can produce information and complexity. Therefore, without loss of generality, the model assumes that, instead of four nucleotides, there are two types of binary monomers in the system (0 and 1). All nucleotides in the system are chemically activated. The nucleotides build the strands. Longer polynucleotides are created from shorter molecules in a stepwise manner by addition of a single nucleotide (0 or 1). During elongation of a molecule, the nucleotide is always added to the end of the precursor strand, for example, strand 01001 is created from precursor strand 0100 by addition of 1 nucleotide and strand 01110 is created from precursor strand 0111 by addition of 0 nucleotide.

This can be schematically presented in the following way:

$$i^{\text{Prec}} + 0 \rightarrow i^{\text{Prec}}0$$
$$i^{\text{Prec}} + 1 \rightarrow i^{\text{Prec}}1, \tag{1}$$
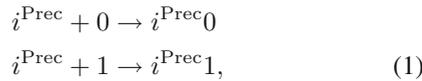
where i^{Prec} represents a sequence, e.g., 0010 or 0110, $i^{\text{Prec}}0$ and $i^{\text{Prec}}1$ are concatenations of sequence i^{Prec} with 0 and 1, respectively. The dynamics of the system are described by an infinite set of differential equations:

$$X_i' = A_i X_{i^{\text{Prec}}} - (D + A_{i0} + A_{i1})X_i$$
$$+ RX_i(F_i(1 - Q)^{\text{Len}_i} - \theta), \tag{2}$$

$$X_i' = \frac{dX_i}{dT}, \tag{3}$$

$$\theta = \frac{\sum_i F_i X_i}{\sum_i X_i}, \tag{4}$$

where i represents a particular sequence (for example, $i = 0010$), X_i denotes the abundance of sequence i, i^{Prec} stands for the precursor sequence of i (it means that after the addition of a proper nucleotide to the end of i^{Prec} it becomes sequence i), $X_{i^{\text{Prec}}}$ is the abundance of the sequence i^{Prec}. A_i is the rate of the elongation process which leads to the creation of polynucleotide i from the sequence i^{Prec}. D denotes the ratio of the process which eliminates the entire molecules from the environment (decay). T stands for time. F_i is the replication fitness of sequence i and R scales relative rates of template directed replication and non-template sequence elongation. Note that the influence of replication rises with R. R might be influenced by the frequency of template separation (when the frequency of separation rises, R rises, too) or the availability of activated nucleotides in the system. The parameter θ scales down molecule frequencies in order to keep population size constant. It might be interpreted as the existence of the system in the environment in which concentration of molecules is at a constant level

(for example, chemostat). Q is the probability of wrong monomer incorporation during the replication process and finally Len_i is the length of strand i. There are three parts in each equation (2):

- $A_i X_{i^{\text{Prec}}}$ is responsible for creation of strand i in the process of elongation. Longer molecules are created from shorter ones through addition of activated nucleotides (0 or 1) to the right side of the precursor sequence (see Eqn. (1)). The rate of this reaction for a given strand i is determined by coefficient A_i.

- $(D + A_{i0} + A_{i1})X_i$ is responsible for a decrease in the number of molecules of type i as a result of creation of longer strands (polynucleotide i through addition of nucleotide (0 or 1) becomes a different polynucleotide ($i0$ or $i1$, respectively), therefore the number of i molecules decreases) and a decay process (whole molecules disappear from the environment).

- $RX_i(F_i(1 - Q)^{\text{Len}_i} - \theta)$ is responsible for template direct replication; this process is represented by the standard selection equation of evolutionary dynamics (Nowak and Ohtsuki, 2008). Incorporation of this equation leads to the increase in the abundance frequency of selected sequences (these with higher F_i) and decrease in the abundance frequency of other sequences which lower replication fitness.

The number of free activated nucleotides Nuc^{Free} (nucleotides that are not part of any polynucleotide) is constant during the simulation ($X_0 = \frac{1}{2}\text{Nuc}^{\text{Free}}$, $X_1 = \frac{1}{2}\text{Nuc}^{\text{Free}}$).

2.2. Compartment model. The main purpose of our model was to simulate the elongation and replication of short sequences in conditions which resemble those in the protocells. We introduced a number of modifications and improvements into the system in order to make it more realistic. The permeability of the membrane controls the number of nucleotides in the system. When the membrane is not permeable, the amount of material inside the protocell (free nucleotides and nucleotides composing strands $\text{Nuc}^{\text{Total}}$) is at a constant level and all the time during the experiments the following equation is satisfied:

$$\sum_{i \in W} X_i \text{Len}_i = \text{Nuc}^{\text{Total}}. \tag{5}$$

In the above formula, Len_i is the number of nucleotides in the sequence i, W is set of all sequences shorter than $\text{Len}^{\text{Max}} + 1$ (the parameter Len^{Max} is the maximum length of polynucleotides available in the model). At the beginning of experiments there is a given number of 0 and 1 activated nucleotides in the system ($\text{Nuc}^{\text{Total}}$), $X_0 =$

$X_1 = \frac{1}{2} \mathrm{Nuc}^{\mathrm{Total}}$. The maximal length of molecules has influence on the number of sequences which are analysed in the system $\mathrm{Seq}^{\mathrm{No}}$.

2.3. Diffusion of molecules.

Compared with the template model, we eliminated the simplification caused by the decay process (coefficient D) and we assumed that molecules cannot vanish. The amount of material in the package is determined by the initial conditions and membrane permeability. The protocell wall formed from long chain amphiphilic molecules has very low or even no permeability, so the number of nucleotides inside the compartment is constant. If we assume that packages are represented by aerosols filled with the solution of primordial material, there is also no diffusion. The permeability of the membrane increases as the length of chains forming amphiphilic compounds decreases. In prebiotic conditions it might be easier to create short amphiphilic molecules than long ones. Admittedly, they do not create as stable bilayer structures although, there are certain substances which could support stabilization and emergence of these formations (Deamer, 2008).

Walls composed of short strand amphiphilic are characterized by higher permeability. Different types of molecules have diverse sizes, which is the reason why the ratio of diffusion between particles species is varying. In general, longer polynucleotides are less likely to infiltrate through the amphiphilic barrier, and therefore we assume that only nucleotides might permeate through the membrane. The compartment is surrounded by the solution of primordial material (nucleotides). The number of available activated nucleotides in the neighborhood of the protocell is denoted by U. In most computations we assume that $U + \sum_{i \in W} \mathrm{Len}_i X_i = \mathrm{const.}$

There are two parameters describing the diffusion rate in the model: P^{InDiff} the ratio of nucleotide diffusion from the environment into the protocell, and P^{OutDiff} the ratio of nucleotide diffusion from the protocell into the environment. There are U nucleotides in the surroundings of the compartment but only a fraction of them are located close enough to the membrane in order to be able to infiltrate through it. The parameter P^{InDiff} is responsible for the modelling of the permeability of the membrane and simulating the availability of nucleotides which are in the range to the package that allows diffusion.

Molecule diffusion might be described by the following equations:

$$U' = Out_0^{\mathrm{Diff}} + Out_1^{\mathrm{Diff}} - In_0^{\mathrm{Diff}} - In_1^{\mathrm{Diff}}, \quad (6)$$

$$In_i^{\mathrm{Diff}} = \begin{cases} 0 & \text{for } i \neq 0 \text{ and } i \neq 1, \\ P^{\mathrm{InDiff}} U & \text{for } i = 0 \text{ or } i = 1, \end{cases} \quad (7)$$

$$Out_i^{\mathrm{Diff}} = \begin{cases} 0 & \text{for } i \neq 0 \text{ and } i \neq 1, \\ P^{\mathrm{OutDiff}} X_i & \text{for } i = 0 \text{ or } i = 1, \end{cases} \quad (8)$$

where In_i^{Diff} denotes the number of i molecules which infiltrate through the membrane from the environment into the interior of the protocell and Out_i^{Diff} represents the number of i molecules which cross the membrane from the interior of the protocell into the environment. Equations (7) and (8) state that only nucleotides (sequence 0 or 1) can pass through the membrane, therefore for other sequences In_i^{Diff} and Out_i^{Diff} are equal to zero.

2.4. Molecule elongation.

Equations (10)–(15) describe the formation and elongation of strands. Molecules (nucleotides and polynucleotides) located close enough to each other might, with probability P^{Join}, join spontaneously with each other (Sutherland, 2010) and form polynucleotides. Elongation of molecules might also be acquired in the mineral catalysed ligation process (Ferris, 2005; Ertem, 2004). The first set of equations (10), (11) assumes that the number of molecules in the compartment does have no influence on the elongation probability (we call it model with constant coefficients). Equations (13) and (14) assume that the number of particles in the package has influence on the number of interactions between components and, as a result, on the elongation probability (we call it a model with stochastically changing coefficients). The number of ligated molecules depends on the number of particles inside the protocell.

Two molecules a and b ($a, b \in W, \mathrm{Len}_a + \mathrm{Len}_b \leq \mathrm{Len}^{\mathrm{Max}}$) can become ligated and create longer polynucleotide. Schematically, this process can be represented by the following equations:

$$\begin{aligned} a + b &\rightarrow ab, \\ b + a &\rightarrow ba. \end{aligned} \quad (9)$$

As a result of this reaction, for

- $a \neq b$ one among of two different sequences might be acquired, half of attained sequences will be ab and the other half ba strands;

- $a = b$ one type of sequence is acquired (because $ab = ba$) and only $\frac{1}{2} x_a$ strands might be created.

Equations (10) and (11) model this phenomenon by incorporation of the factor $1/2$ which serves in both cases $a = b$ and $a \neq b$; in addition, in Eqns. (13) and (14) the factor $In_{a,b}^{\mathrm{JoinC}}$ is responsible for the fulfilment of this requirement. Creation of a longer strand (i) from two shorter molecules (a,b) leads to an increase in the number of i strands, and formula In_i^{Join} is responsible for the modelling of this phenomenon. On the other hand, molecule i might become a substrate in the elongation reaction. Therefore the number of these molecules is decreasing, and the formula Out_i^{Join} is responsible for implementation of this case.

Longer strands join with a lower rate because, in order to become connected, a proper end to end alignment of substrates is required. It is modelled through incorporation of formula $\mathrm{Lim}_{a,b}^{\mathrm{Join}}$ (Eqn. (12)) into the elongation equations. Real molecules can also be rotated before the ligation process but this model does not support this phenomenon. Each target sequence might have a different ratio of substrate joins. For example, laboratory experiments point out that the mineral catalysed ligation process acts most efficiently on strands composed of C and G nucleotides (Ferris, 2006). What is more, there might be available molecules that support ligation of certain polynucleotides. In order to introduce this phenomenon into the model, we applied the selection intensity parameter, S_i, which directly modifies the ligation probability of sequences i ($S_i \geq 0$, $P^{\mathrm{Join}} S_i \leq 1$).

The influence of S_i on the number of i sequence, is positive when $S_i > 1$, neutral for $S_i = 1$ and negative when $S_i < 1$,

$$
In_i^{\mathrm{Join}}
$$
$$
= \frac{1}{2\mathrm{Seq}^{\mathrm{No}}} \sum_{a,b \in W_i^{\mathrm{Sub}}} \mathrm{Min}(X_a, X_b) P^{\mathrm{Join}} S_i \mathrm{Lim}_{a,b}^{\mathrm{Join}},
$$
$$(10)$$

$$
Out_i^{\mathrm{Join}} = \frac{1}{2\mathrm{Seq}^{\mathrm{No}}}
$$
$$
\cdot \sum_{a \in W_i^{\mathrm{Join}}} \mathrm{Min}(X_a, X_i) P^{\mathrm{Join}}(S_{ia} + S_{ai}) \mathrm{Lim}_{a,i}^{Join}, \quad (11)
$$

$$
\mathrm{Lim}_{a,b}^{Join} = 2^{\frac{-\mathrm{Len}_a \mathrm{Len}_b + 1}{10}}, \quad (12)
$$

$$
In_i^{\mathrm{Join}} = \frac{1}{2\mathrm{Seq}^{\mathrm{No}}}
$$
$$
\cdot \sum_{a,b \in W_i^{\mathrm{Sub}}} In_{a,b}^{\mathrm{JoinC}}(\mathrm{Min}(X_a, X_b), S_{ab}) \mathrm{Lim}_{a,b}^{\mathrm{Join}},
$$
$$(13)$$

$$
Out_i^{\mathrm{Join}} = \frac{1}{2\mathrm{Seq}^{\mathrm{No}}} \sum_{a \in W_i^{\mathrm{Join}}} (In_{i,a}^{\mathrm{JoinC}}(\mathrm{Min}(X_a, X_i), S_{ia})
$$
$$
+ In_{a,i}^{\mathrm{JoinC}}(\mathrm{Min}(X_a, X_i), S_{ai}) \mathrm{Lim}_{a,i}^{\mathrm{Join}},
$$
$$(14)$$

$$
In_{a,b}^{\mathrm{JoinC}}(V^{\mathrm{Min}}, V^{\mathrm{Sel}})
$$
$$
= \begin{cases} 2 \displaystyle\sum_{\substack{k=\lfloor V^{\mathrm{Min}}+1 \rfloor - 1 \\ k=k-2}}^{k>0} (1 - (1 - P^{\mathrm{Join}} V^{\mathrm{Sel}})^k), \\ \qquad \text{for } a = b \\[4pt] \displaystyle\sum_{\substack{k=\lfloor V^{\mathrm{Min}}+1 \rfloor - 1 \\ k=k-1}}^{k>0} (1 - (1 - P^{\mathrm{Join}} V^{\mathrm{Sel}})^k), \\ \qquad \text{for } a \neq b. \end{cases}
$$
$$(15)$$

where W_i^{Sub} is a set of all pairs of substrates (a,b), which in the elongation process might give the sequence i and W_i^{Join} is a set of all sequences (a), which can join the molecule i and create a strand shorter than or equal to $\mathrm{Len}^{\mathrm{Max}}$ ($\mathrm{Len}_a + \mathrm{Len}_i \leq \mathrm{Len}^{\mathrm{Max}}$). $\mathrm{Min}(X_a, X_b)$ is the smallest number of molecules among substrates of type a and substrates of type b. If there are X_a and X_b substrates of type a and b, respectively, then if $X_a < X_b$ only X_a of ab and ba strand might be created. In the equation $In_{a,b}^{\mathrm{JoinC}}(V^{\mathrm{Min}}, V^{\mathrm{Sel}})$ this is incorporated through V^{Min} parameter, which is the smallest number of molecules among substrate a and substrate b, V^{Sel} is the selection intensity of sequence ab.

2.5. Phosphodiester bond breakage. The phosphodiester bond between nucleotides in a polynucleotide can break as an effect of hydrolysis or UV radiation (Luisi, 2010; Cech and Golden, 1999; Fishkis, 2010). It leads to the creation of two shorter molecules (Eqns. (17) and (18)). Longer strands have more phosphodiester bonds so the probability of fission of the strand increases with the length of the polynucleotide. Each phosphodiester bond might break with probability P^{Split}. W_i^{Split} is a list of all strands that can break and as a result create strand i. Strand j might break in $\mathrm{Len}_j - 1$ locations so the probability that a given sequence i will be created from strand $j \in W_i^{\mathrm{Split}}$ in the phosphodiester bond break reaction is $1/(\mathrm{Len}_j - 1)$.

Schematically, this process can be represented by the following equations:

$$
j \to ia \to i + a,
$$
$$
j \to ai \to a + i. \quad (16)
$$
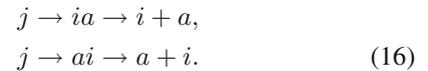
Some strands, if broken in the middle, create two twin sequences. If those twin sequences are equal to the sequence t for which computations are conducted it should be taken into account. Formula $Eq_{t,j}^{\mathrm{Split}}$ (Eqn. (19)) is responsible for the implementation of this condition,

$$
In_i^{\mathrm{Split}}
$$
$$
= \sum_{a \in W_i^{\mathrm{Split}}} \frac{1}{\mathrm{Len}_a - 1} Eq_{i,a}^{\mathrm{Split}}
$$

$$\cdot\ X_a(1\ -\ (1\ -\ P^{\text{Split}})^{(\text{Len}_a-1)}), \quad (17)$$

$$Out_i^{\text{Split}} = X_i(1 - (1 - P^{\text{Split}})^{(\text{Len}_i-1)}), \quad (18)$$

$$Eq_{t,j}^{\text{Split}} = \begin{cases} 1, & \text{if } \neg\exists a, b \in W_j^{\text{Sub}} : a = b, \\ 1, & \text{if } \exists a, b \in W_j^{\text{Sub}} : a = b \neq t, \\ 2, & \text{if } \exists a, b \in W_j^{\text{Sub}} : a = b = t, \end{cases} \quad (19)$$

where In_i^{Split} is responsible for the increase in the number of molecules i created in the fission process from the longer sequences whereas Out_i^{Split} describes the decrease in the number of molecules i broken in this process.

2.6. Molecule replication. Replication of strands is modelled as a one step process. Two molecules a and b attach to the template t ($t = C_{ab}$ or $t = C_{ba}$; C_i is a sequence complementary to sequence i), become ligated and then the created polynucleotide disconnects from the template. This process might be schematically represented by the following equations:

$$a + b + C_{ab} \rightarrow ab + C_{ab},$$
$$b + a + C_{ba} \rightarrow ba + C_{ba}. \quad (20)$$
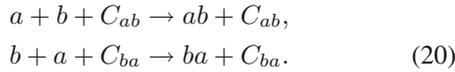
Strand ab or ba must be complementary to the sequence t, $\text{Len}_a + \text{Len}_b = \text{Len}_t$. Every molecule i with $\text{Len}_i > 1$ might be the template. The efficiency of the replication relays strictly on the abundance of templates and substrates in the solution.

Each sequence i is related with the replication fitness F_i ($F_i \geq 0$, $F_i P^{\text{Repl}} \leq 1$) which modifies probability of strand replication P^{Repl}. This can be an effect of the composition of the strand or existence of catalysts supporting replication of certain molecules.

As opposed to the template model, sequences with lower fitness are not removed from the population but have lower possibility of progenies creation. Each strand i has a complementary strand C_i. Similarly to the ligation process, we have two sets of equations: one when the probability of replication is constant, Eqns. (21)–(24) (model with constant coefficients), and one when the number of molecules has influence on the replication rate, Eqns. (27)–(30) (model with stochastically changing coefficients). In this case the equation $In_{a,b}^R(V^{\text{Min}}, V^F)$ is responsible for the modelling of these influence.

Substrates a,b might join to the template if

- $a \neq b$: one among of two different sequences might be acquired as a result $1/2$ of a,b pairs can join to the templates C_{ab} and $1/2$ can join the templates C_{ba};

- $a = b$: then a,b pairs might join only the templates C_{aa} and only $1/2 X_a$ strands can be created.

That is why the factor $1/2$ is incorporated into replication equations. In addition, for the case with stochastically changing coefficients, the factor $In_{a,b}^R$ is also responsible for implementation of the above mentioned conditions.

During the first phase of replication an error might occur and a strand that is not fully complementary to the template might be attached; as a result, a mutated polynucleotide is created. We assume that the attached strand might not be complementary to the template on one position. Mutation rate P^{ReplMut} scales down the faithful replication probability (Eqns. (25) and (31)). Therefore, even if these two processes (faithful replication and mutated replication) are simulated by different formulas, the ratio of replication stays at a constant level. For template sequence C_i there are Len_i types of sequences which might be created in the process of mutant replication, therefore the number of available substrates which might be attached and create a mutant sequence is rising with the template length. In order to keep the mutant replication rate P^{ReplMut} at a constant level for all templates C_i, an additional factor $1/\text{Len}_i$ is incorporated into mutation equations. W_i^{Mut} is set of all pairs a, b that after ligation give a sequence that is not complementary to the template strand i on only one position. In_i^{Repl} (faithful replication) and In_i^{ReplM} (unfaithful replication) are responsible for the increase of the number of molecules formed in the process of replication, creation of new strands leads to a decrease in the number of reaction substrates, therefore Out_i^{Repl} and Out_i^{ReplM} are introduced,

$$In_i^{\text{Repl}} = \frac{1}{2\text{Seq}^{\text{No}}} \sum_{a,b \in W_i^{\text{Sub}}} P^R \text{Min}(X_{C_i}, X_a, X_b)F_i, \quad (21)$$

$$Out_i^{\text{Repl}}$$
$$= \frac{1}{2\text{Seq}^{\text{No}}} \sum_{a \in W_i^{\text{Join}}} P^R[\text{Min}(X_{C_{ia}}, X_i, X_a)F_{C_{ia}}$$
$$+ \text{Min}(X_{C_{ai}}, X_i, X_a)F_{C_{ai}}], \quad (22)$$

$$In_i^{\text{ReplM}} = \frac{1}{2\text{Len}_i\text{Seq}^{\text{No}}}$$
$$\cdot \sum_{a,b \in W_i^{Mut}} P^{RM}\text{Min}(X_{C_i}X_a, X_b)F_i, \quad (23)$$

$$Out_i^{\text{ReplM}}$$
$$= \frac{1}{2\text{Seq}^{\text{No}}} \sum_{a \in W_i^{\text{Join}}} P^{RM}$$
$$\cdot \Big[\frac{1}{\text{Len}_{C_{ia}}}\text{Min}(X_{C_{ia}}, X_a, X_i)F_{C_{ia}}$$
$$+ \frac{1}{\text{Len}_{C_{ai}}}\text{Min}(X_{C_{ai}}, X_a, X_i)F_{C_{ai}} \Big], \quad (24)$$

$$P^R = P^{\text{Repl}} - P^{\text{Repl}} P^{\text{ReplMut}}, \qquad (25)$$

$$P^{RM} = P^{\text{Repl}} P^{\text{ReplMut}}, \qquad (26)$$

$$In_i^{\text{Repl}}$$
$$= \frac{1}{2\text{Seq}^{\text{No}}} \sum_{a,b \in W_i^{\text{Sub}}} In_{a,b}^{RC}(\text{Min}(X_{C_i}, X_a, X_b), F_i), \qquad (27)$$

$$Out_i^{\text{Repl}}$$
$$= \frac{1}{2\text{Seq}^{\text{No}}} \sum_{a \in W_i^{\text{Join}}}$$
$$\cdot \Big[(In_{i,a}^{RC}(\text{Min}(X_{C_{ia}}, X_a, X_i), F_{C_{ia}})$$
$$+ In_{a,i}^{RC}(\text{Min}(X_{C_{ai}}, X_a, X_i), F_{C_{ai}}) \Big], \qquad (28)$$

$$In_i^{ReplM}$$
$$= \frac{1}{2\text{Len}_i \text{Seq}^{\text{No}}}$$
$$\cdot \sum_{a,b \in W_i^{\text{Mut}}} In_{a,b}^{RMC}(\text{Min}(X_{C_i}, X_a, X_b), F_i), \qquad (29)$$

$$Out_i^{ReplM}$$
$$= \frac{1}{2\text{Seq}^{\text{No}}} \sum_{a \in W_i^{Join}}$$
$$\cdot \Big[\frac{1}{\text{Len}_{C_{ia}}} In_{i,a}^{\text{RMC}}[\text{Min}(X_{C_{ia}}, X_a, X_i), F_{C_{ia}})$$
$$+ \frac{1}{\text{Len}_{C_{ai}}} In_{a,i}^{\text{RMC}}(\text{Min}(X_{C_{ai}}, X_a, X_i), F_{C_{ai}}] \Big], \qquad (30)$$

$$In_{a,b}^{RC}(V^{\text{Min}}, V^F)$$
$$= In_{a,b}^R(V^{\text{Min}}, V^F) - In_{a,b}^R(V^{\text{Min}}, V^F) P^{\text{ReplMut}}, \qquad (31)$$

$$In_{a,b}^{RMC}(V^{\text{Min}}, V^F) = In_{a,b}^R(V^{\text{Min}}, V^F) P^{\text{ReplMut}}, \qquad (32)$$

$$In_{a,b}^R(V^{\text{Min}}, V^F)$$
$$= \begin{cases} 2 \sum_{\substack{k=\lfloor V^{\text{Min}}-1 \rfloor+1 \\ k=k-2}}^{k>0} (1 - (1 - P^{\text{Repl}} V^F)^k), \\ \qquad\qquad \text{for} \qquad a = b \\ \sum_{\substack{k=\lfloor V^{\text{Min}}-1 \rfloor+1 \\ k=k-1}}^{k>0} (1 - (1 - P^{\text{Repl}} V^F)^k), \\ \qquad\qquad \text{for} \qquad a \neq b. \end{cases} \qquad (33)$$

Similarly to strand elongation, in formula $In_{a,b}^{RC}$, V^{Min} is the smallest number of molecules, although this time among three types of molecules: template C_{ab}, substrate a and substrate b. V^F is the fitness of template C_{ab}. The equations P^R, P^{RM} (in the case of a model with constant coefficients), In^{RC} and In^{RMC} (in the case of a model with stochastically changing coefficients) scale the influence of mutation in order to keep the replication rate at the constant level and prevent the exceeding of the P^{Repl} value.

2.7. Final equations. When all the above mentioned phenomena are taken into account, the system is defined by an infinite set of differential equations:

$$\begin{aligned} X_i' = & In_i^{\text{Join}} - Out_i^{\text{Join}} \\ & + In_i^{\text{Split}} - Out_i^{\text{Split}} \\ & + In_i^{\text{Repl}} - Out_i^{\text{Repl}} \\ & + In_i^{\text{ReplMut}} - Out_i^{\text{ReplMut}} \\ & + In_i^{\text{Diff}} - Out_i^{\text{Diff}}. \end{aligned} \qquad (34)$$

Each equation might be divided into subsets of composing factors, where:

- $In_i^{\text{Join}} - Out_i^{\text{Join}}$ are responsible for molecule elongation,

- $In_i^{\text{Split}} - Out_i^{\text{Split}}$ are responsible for phosphodiester bond break,

- $In_i^{\text{Repl}} - Out_i^{\text{Repl}}$ are responsible for faithful replication,

- $In_i^{\text{ReplMut}} - Out_i^{\text{ReplMut}}$ are responsible for replication with mutation,

- $In_i^{\text{Diff}} - Out_i^{\text{Diff}}$ are responsible for membrane permeability.

For the case with a limitation of the length of strands we have Seq^{No} differential equations, where

$$\text{Seq}^{\text{No}} = \sum_{k=1}^{\text{Len}^{\text{Max}}} 2^k. \qquad (35)$$

There are two types of nucleotides (0 and 1) in the model. It is quite easy to modify differential formulas in order to support four kinds of nucleotides; however, this increases significantly the number of equations. Instead of $\sum_{k=1}^{\text{Len}^{\text{Max}}} 2^k$ equations for the case with two nucleotide types, there will be $\sum_{k=1}^{\text{Len}^{\text{Max}}} 4^k$ equations.

3. Experiments and results

3.1. Initial conditions. In order to establish the influence of initial values of parameters on the model behaviour, we conducted a set of experiments. The probability that two randomly chosen molecules will join, P^{Join}, and the probability that the phosphodiester bond will break, P^{Split}, depend on many factors such as the presence of chemical agents in the solution or temperature. Unfortunately, the exact conditions on the early Earth are unknown. On the other hand, we might assume that the environment in the place of life emergence should foster its creation. Based on this assumption we conducted a set of computations with P^{Join} and P^{Split} varying in the range from 0 to 1. The initial number of nucleotides in the package was varying in the range from 1000 to 100000 (at $t = 0$, $x_0 = x_1 = \frac{1}{2}\text{Nuc}^{\text{Total}}$).

The graph (Fig. 1) presents the number of molecules as a function of P^{Join}. The number of nucleotides in the compartment was constant and the maximum length of polynucleotides that can be created was set to 6 ($\text{Len}^{\text{Max}} = 6$) or 8 ($\text{Len}^{\text{Max}} = 8$, unpublished data). For a constant value of P^{Split} ($P^{\text{Split}} = 0.001$) and $P^{\text{Join}} > 0.5$, both models settle in the steady state and yield similar results (stabilization is caused by the lack of short molecules). However, the model with stochastically changing coefficients reaches this state much earlier. In this model we observe rapid increase of the number of long molecules when P^{Join} grows.

In most figures in this article a convention is established according to which the length of the molecule is denoted by the colour (The shortest components—nucleotides—have the light grey designated. With the length of the molecule, the line representing component on the graph becomes darker, thus the longest 6-nucleotide-long strands are represented by black lines.) In addition, selected sequences, in the elongation or replication processes, are marked with black dots and sequences complementary to them are marked with vertical lines.

Our previous research (Myszor and Cyran, 2010) indicated that the shape of histograms should be similar to Fig. 2(c). Based on this assumption, we use a model with stochastically changing coefficients and the following parameters: $P^{\text{Join}} = 0.01$, $P^{\text{Split}} = 0.001$, with the number of nucleotides ($\text{Nuc}^{\text{Total}}$) equal to 10000.

We performed an additional set of simulations in order to check how the initial number of nucleotides affects the length of polynucleotides within the package. As can be observed in Fig. 3, a higher number of nucleotides results in more possible interactions between molecules and fosters the creation of polynucleotides.

The maximum length of polynucleotides, Len^{Max}, (see Fig. 2), has strong influence on computational effort. The increase of the maximum length of strands by one (from Len^{Max} to $\text{Len}^{\text{Max}} + 1$) results in additional $2^{\text{Len}^{\text{Max}}+1}$ differential equations and introduces new coefficients to the already existing equations. In most computations we chose $\text{Len}^{\text{Max}} = 6$ as a maximum strand length. This value is also used for a majority of computations in the work of Nowak and Ohtsuki (2008). In order to keep a constant amount of material in the system, we did not allow creation of a strand longer than Len^{Max} in the elongation process.

3.2. Selection intensity. In an actual environment many factors such as strand composition or existence of catalysts promoting the ligation are responsible for different elongation rates. This phenomenon is simulated in our experiments by selection intensity S_i. Therefore each sequence i can have different S_i.

We analysed the influence of selection intensity when only one sequence j was selected ($S_i = 1$ for $i \neq j$, $S_i > 1$ for $i = j$; for results, see Fig. 4). An increase in the selection intensity leads to an increase in the number of selected sequences. For $S_j \in (2, 10)$ we observe a rapid growth of the number of selected molecules. Above, $S_j > 10$, the dynamic of growth is limited. When additional sequences have $S_i > 1$ (Fig. 5), then selection of new sequences leads to a limitation of the influence of this phenomenon on the already selected molecules. The number of selected molecules with equal length tends to converge to a common value. In the next experiment, $S_i > 1$ for all sequences i for which $\text{Len}_i = 6$ ($S_i = 1$ for $i : \text{Len}_i < 6$). Results of simulations are presented in Fig. 6.

Note that in the discussed conditions of the protocell, influence of selection intensity (at the level of a single sequence or many sequences) is strictly limited by the number of available molecules. Relatively small values of selection intensity (in the range (2–10)) lead to a visible increase in the number of selected sequences; however, further increasing of the value of this parameter seems not to increase the number of selected molecules.

3.3. Replication. Polynucleotides might replicate in the template directed ligation process. We experimented with different replication rates when all sequences had equal probability of replication (P^{Repl}, $P^{\text{ReplMut}} = 0$). Results of simulations are presented in Fig. 7. When P^{Repl} is in the range (0–0.03), then the creation of sequences with 5 nucleotides is favoured. Further increasing of this parameter value leads to a rapid increase in the number of longer sequences (composed of 6 nucleotides). A similar effect is also visible when longer strands were allowed in the system (Fig. 7).

Similarly to the elongation process there might be available catalysts promoting the replication of certain strands (for example, an RNA replicase) (Joyce and
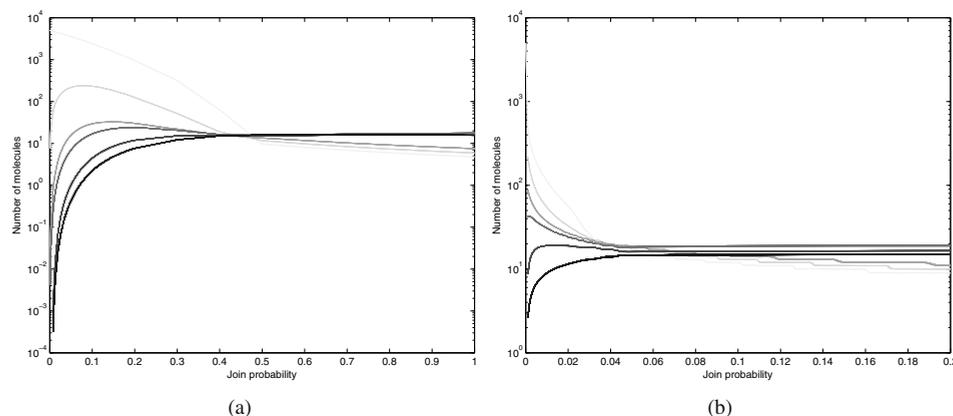
Fig. 1. Number of molecules as a function of P^{Join} for a model with a constant coefficient (a) and a model with stochastically changing coefficients (b). $P^{\text{Split}} = 0.001$, $S_i = 1$, $P^{\text{Repl}} = 0$, $P^{\text{InDiff}} = 0$, $P^{\text{OutDiff}} = 0$, $\text{Len}^{\text{Max}} = 6$.
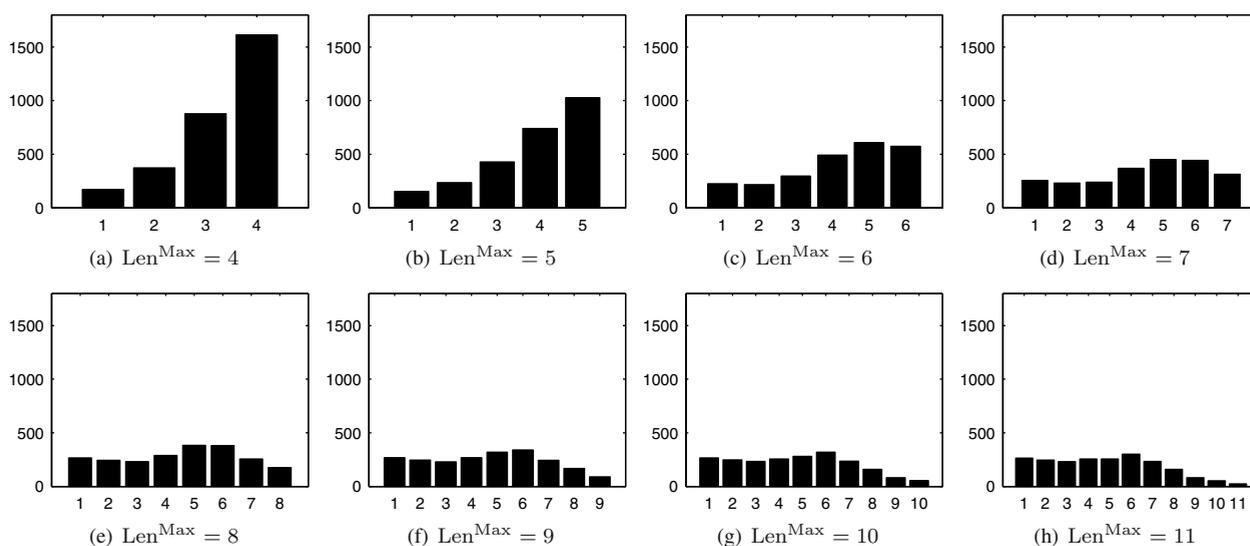

Fig. 2. Histograms of the length of pseudo RNA molecules for different values of Len^{Max}. $P^{\text{Join}} = 0.01$, $P^{\text{Split}} = 0.001$, $S_i = 1$, $P^{\text{Repl}} = 0$, $P^{\text{InDiff}} = 0$, $P^{\text{OutDiff}} = 0$.

Orgel, 1999). When replication intensity acts at the level of only one sequence j ($F_i > 1$ for $i = j$; $F_i = 1$ for $i \neq j$), it might be interpreted as the existence of a very specific replicase that can support the replication of only one type of molecule. An increase in F_i leads to a rapidly growing number of a selected molecules (Fig. 8) and might even cause domination of a selected strand in the environment. The number of its complementary polynucleotides is also rising. However, it seems to be blocked at a certain level and despite further increase of the number of selected molecules it does not go any higher. This phenomenon is visible for different types and lengths of selected templates. It is worth mentioning that the influence of replication intensity in protocell conditions is limited. There is a significant difference in

the numbers of selected sequences i between $F_i = 20$ and $F_i = 30$. However there is only a minor difference when $F_i = 60$ and $F_i = 70$. This is especially visible in Fig. 13(b). This observation is related with an important conclusion, i.e., that at some level the creation of catalysts which has superiority in the field of the reaction ratio (and usually more complex structure) does not bring sufficient advantages to the system. Therefore, such a molecule might easily disappear from the population, replaced by less efficient, although structurally simpler sequences.

Even more interesting results are observed when the replicase acts at the level of one sequence i ($\text{Len}_i = 6$, $F_i = 5$) and the probability of the replication is changing in the range (0.01–0.2) (Fig. 9). For $P^{Repl} \in$ (0.01–0.03), the number of selected polynucleotides is

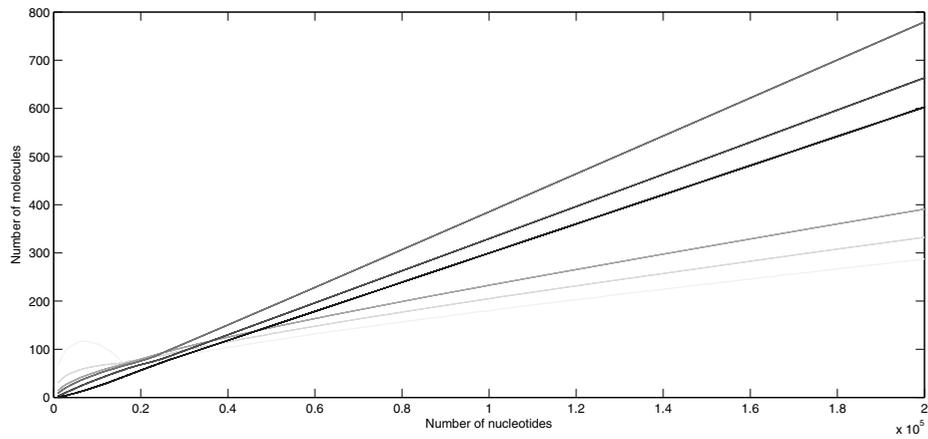Fig. 3. Number of molecules as a function of $\mathrm{Nuc}^{\mathrm{Total}}$. $\mathrm{Len}^{\mathrm{Max}} = 6$, $P^{\mathrm{Join}} = 0.01$, $P^{\mathrm{Split}} = 0.001$, $S_i = 1$, $P^{\mathrm{Repl}} = 0$, $P^{\mathrm{InDiff}} = 0$, $P^{\mathrm{OutDiff}} = 0$.
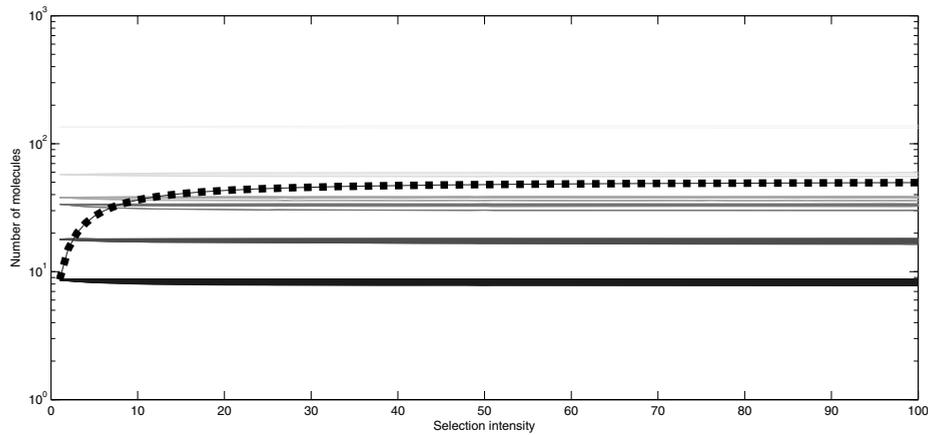

Fig. 4. Number of molecules as a function of selection intensity S_i. There is one selected strand type i ($\mathrm{Len}_i = 6$). $\mathrm{Len}^{\mathrm{Max}} = 6$, $P^{\mathrm{Join}} = 0.01$, $P^{\mathrm{Split}} = 0.001$, $P^{\mathrm{Repl}} = 0$, $P^{\mathrm{InDiff}} = 0$, $P^{\mathrm{OutDiff}} = 0$.
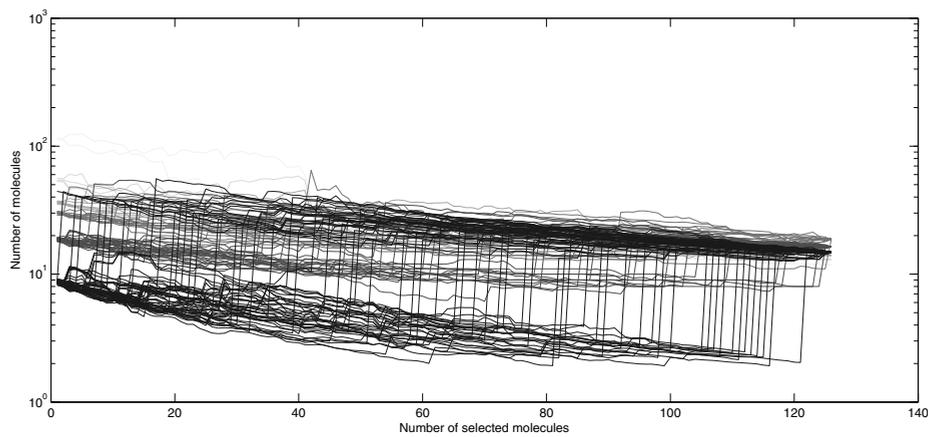
Fig. 5. Number of molecules as a function of the number of selected types of sequences. In each consecutive simulation, 10 additional random types of sequences were added to the selected sequences set S^{Seq}, $\forall i \in S^{\mathrm{Seq}}$ $S_i = 10$. $P^{\mathrm{Join}} = 0.01$, $P^{\mathrm{Split}} = 0.001$, $P^{\mathrm{Repl}} = 0$, $P^{\mathrm{InDiff}} = 0$, $P^{\mathrm{OutDiff}} = 0$, $\mathrm{Len}^{\mathrm{Max}} = 6$.
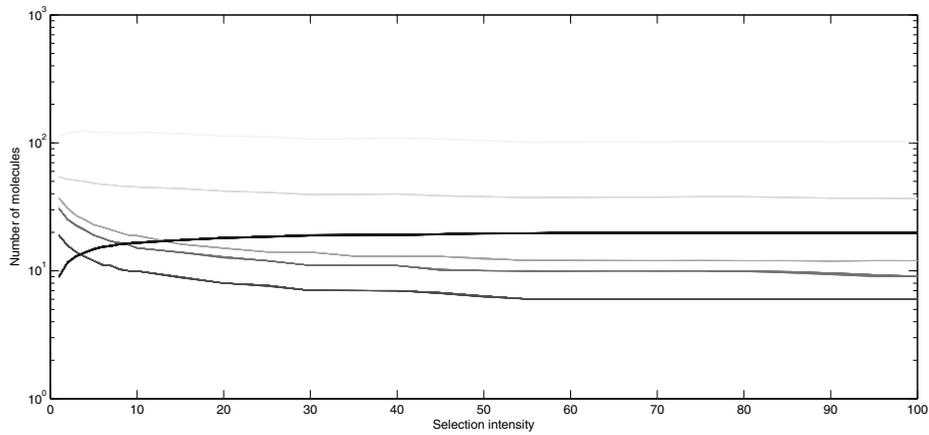
Fig. 6. Number of molecules as a function of selection intensity S_i. $S_i = 1 : \mathrm{Len}_i < 6, S_i > 1 : \mathrm{Len}_i = 6, \mathrm{Len}^{\mathrm{Max}} = 6$, $P^{\mathrm{Join}} = 0.01$, $P^{\mathrm{Split}} = 0.001$, $P^{\mathrm{Repl}} = 0$, $P^{\mathrm{InDiff}} = 0$, $P^{\mathrm{OutDiff}} = 0$.
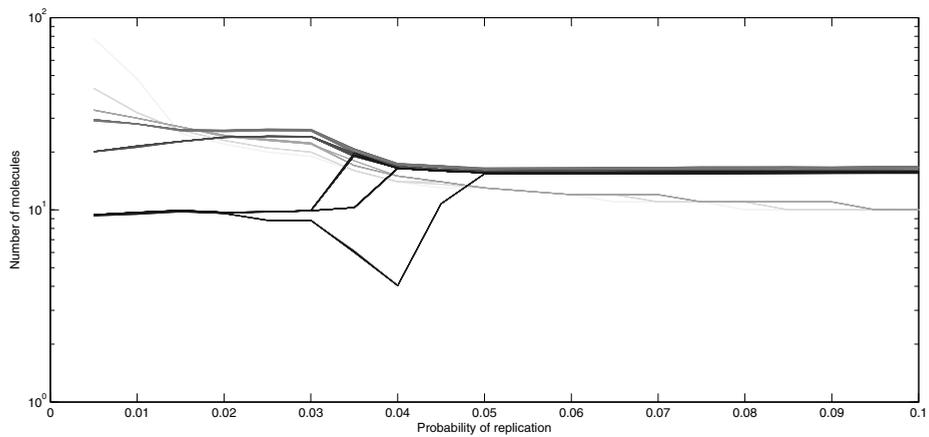

Fig. 7. Number of molecules as a function of probability of replication P^{Repl}. $\mathrm{Len}^{\mathrm{Max}} = 6$, $P^{\mathrm{Join}} = 0.01$, $P^{\mathrm{Split}} = 0.001$, $S_i = 1$, $P^{\mathrm{ReplMut}} = 0$, $P^{\mathrm{InDiff}} = 0$, $P^{\mathrm{OutDiff}} = 0$.
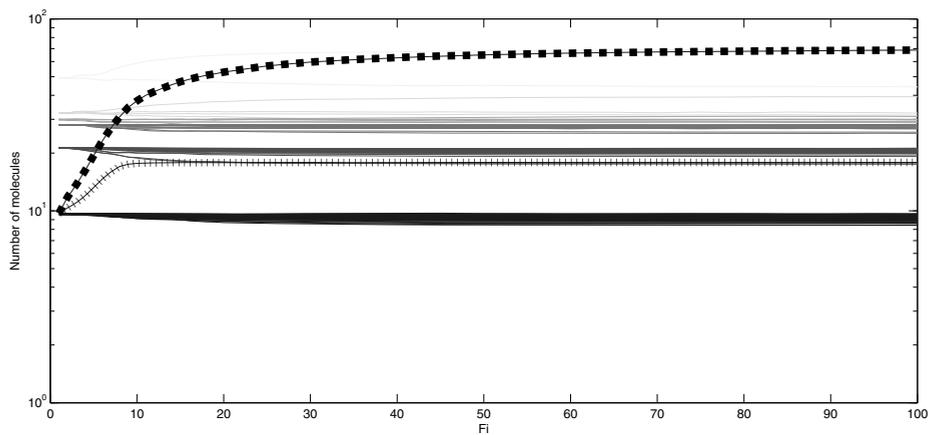
Fig. 8. Number of molecules as a function of replication intensity F_i. One sequence type, i, is selected, $\mathrm{Len}_i = 6$, $\mathrm{Len}^{\mathrm{Max}} = 6$ $P^{\mathrm{Join}} = 0.01$, $P^{\mathrm{Split}} = 0.001$, $S_i = 1$, $P^{\mathrm{Repl}} = 0.01$, $P^{\mathrm{ReplMut}} = 0$, $P^{\mathrm{InDiff}} = 0$, $P^{\mathrm{OutDiff}} = 0$.

increasing. When the replication probability exceeds 0.03, the number of 6-nucleotide-long molecules is increasing, which leads to the limitation of the number of shorter sequences and, as a consequence, to a rapid fall of the number of selected strands.

When replicase supports replication of all sequences with 6 nucleotides (Fig. 10), then the relatively low F_i values (2–10) introduce rapid growth of the number of selected sequences, which is connected with visible decrease of the number of shorter sequences. However, it can be observed that the number of free nucleotides is rising as en effect of phosphodiester bond break in long sequences.

Processes of replication are often subjected to mutations. There are many different types of mutations; Single Nucleotide Polymorphism (SNP) is incorporated into the model. At the early stage of life, the replicase might be very primitive. RNA molecules able to replicate other strands were achieved in laboratory conditions, although these formations are very inefficient and slow and have high mutation rate (above 3% per template nucleotide). As a good approximation we can assume that, in order to be able to maintain the strand i with genetic information during consecutive replications, the maximum mutation rate should not exceed $1/\mathrm{Len}_i$ (Joyce, 2002). For the selected strand with the length equal to 6, this rule tells us that the mutation rate should be smaller than 0.64 per strand.

Results of our experiments have shown that, for the case when all molecules have equal replication intensity ($F_i = 1$), the influence of mutations on the model behavior can be neglected. When there is a catalyst promoting a replication of one type of polynucleotides, then the mutations limit the number of selected molecules and their complementary strands. At the same time, the number of mutated sequences is rising. Figures 12 and 13(a) present results of the experiment with constant P^{ReplMut} and varying replication intensity. The number of selected sequences is rising as a function of F_i. However, there are less selected sequences than in the case when the mutation is not operating. Figures 11 and 13(b) present the influence on mutations on the number of selected molecules. It can be observed that a high mutation rate can efficiently compensate replicase actions.

3.4. Diffusion. Simple package might not possess any advanced mechanisms that allow for acquiring of genetic material from its surroundings however, the membrane might be permeable to molecules. In the model, the nucleotides located close enough to the wall of the protocell, might infiltrate through it. We conducted set of numerical computations in order to observe the influence of a permeability on the content of the package. At the beginning of the simulation, there were 100000 nucleotides in the environment surrounding the protocell.

P^{OutDiff} was constant and equal to 0.1, P^{InDiff} was varying from computation to computation (the results are presented in Fig. 14). With the increase of P^{InDiff} the number of nucleotides in the system increases too. First, the number of nucleotides rises slowly with P^{InDiff}, but when this coefficient reaches values between (0.0002–0.0003), the number of nucleotides inside the protocell grows rapidly. The concentration of molecules within protocell leads to the creation of polynucleotides (which cannot infiltrate the cell's membrane) and to the decrease in the number of free nucleotides. So the number of nucleotides that can diffuse out of protocell is falling because they are used to build strands in the ligation process. It leads to accumulation of material inside protocell. Our computations with $P^{\mathrm{Join}} = 0.001$ and $P^{\mathrm{Join}} = 0.02$ seem to confirm these findings. The reasons for this are as follows:

- lower ligation rate causes that this effect does not occur in analysed conditions (Fig. 15(a)),

- higher ligation rate leads to earlier accumulation of material inside protocell (Fig. 15(b)).

4. Conclusions

In this article, the mathematical model of the behaviour of short RNA molecules in conditions which resemble those in early protocells was described. The outcomes returned by the model show that the number of molecules inside the package might be important, because too low the concentration can limit the possibility of the creation of longer strands, which are essential in the process of the emergence of the RNA world. A limited number of molecules inside the package leads to a decrease in the efficiency of catalysts promoting the ligation or replication processes. Particles supporting these processes have some optimal level of efficiency (which depends on the number of molecules inside the compartment). Above this value, a further increase in the catalyst efficiency does not introduce significant changes in the number of selected polynucleotides. Therefore a constant supply of short molecules is required.

When the membrane is permeable processes acting inside protocell have huge influence on the number of nucleotides inside the package. The diffusion ratio of molecules depends on their size. A larger formation is less likely to permeate through the membrane. The processes of strand elongation and strand replication create larger polynucleotides, thus preventing them from infiltrating the package wall. This phenomenon might lead to accumulation of material inside the protocell and accumulation of material is one of the most important concern at the early stage of the RNA world. Notably this process is very sensitive to the values of parameters; for
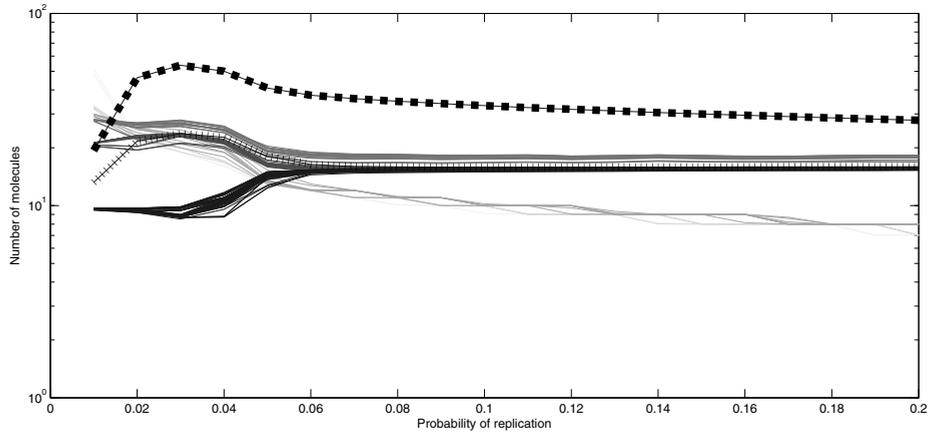
Fig. 9. Number of molecules as a function of probability of replication, when $F_i = 5$. One sequence type, i, is selected, $\text{Len}_i = 6$. $\text{Len}^{\text{Max}} = 6$. $P^{\text{Join}} = 0.01$, $P^{\text{Split}} = 0.001$, $S_i = 1$, $P^{\text{ReplMut}} = 0$, $P^{\text{InDiff}} = 0$, $P^{\text{OutDiff}} = 0$.
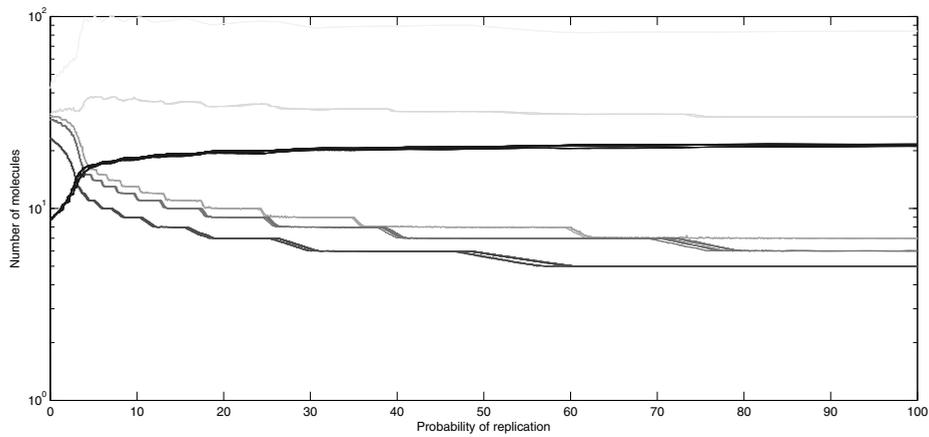

Fig. 10. Number of molecules as a function of replication intensity F_i. Sequences with length equal to 6 are selected. $P^{\text{Join}} = 0.01$, $P^{\text{Split}} = 0.001$, $S_i = 1$, $P^{\text{Repl}} = 0.01$, $P^{\text{ReplMut}} = 0$, $P^{\text{InDiff}} = 0$, $P^{\text{OutDiff}} = 0$, $\text{Len}^{\text{Max}} = 6$.

(a)   (b)
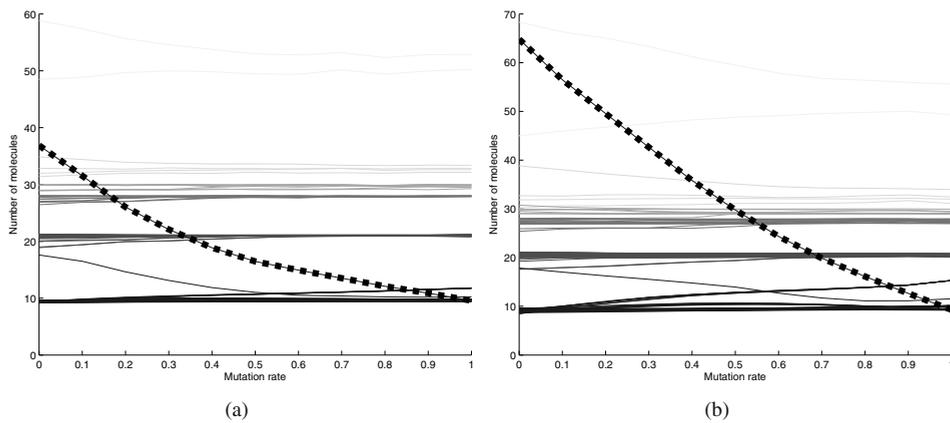
Fig. 11. Number of molecules as a function of mutation rate P^{ReplMut}. One type of sequence, i, is selected ($\text{Len}_i = 6$), $F_i = 10$ (a), $F_i = 50$ (b). $P^{\text{Join}} = 0.01$, $P^{\text{Split}} = 0.001$, $S_i = 1$, $P^{\text{Repl}} = 0.01$, $P^{\text{InDiff}} = 0$, $P^{\text{OutDiff}} = 0$, $\text{Len}^{\text{Max}} = 6$.
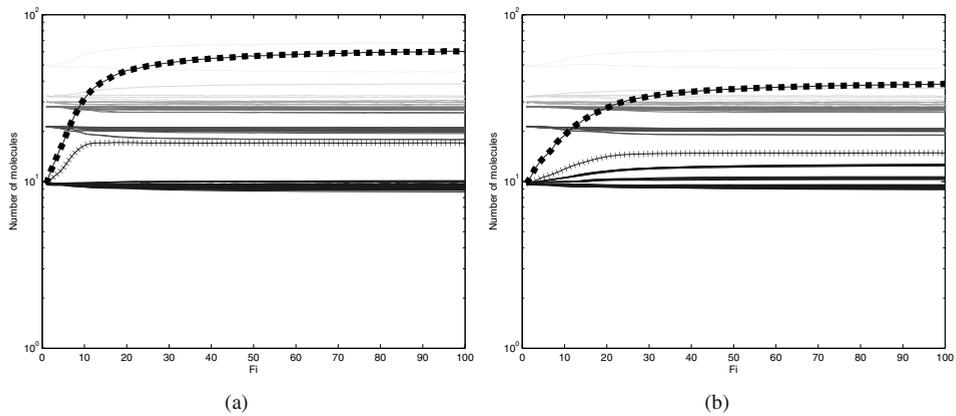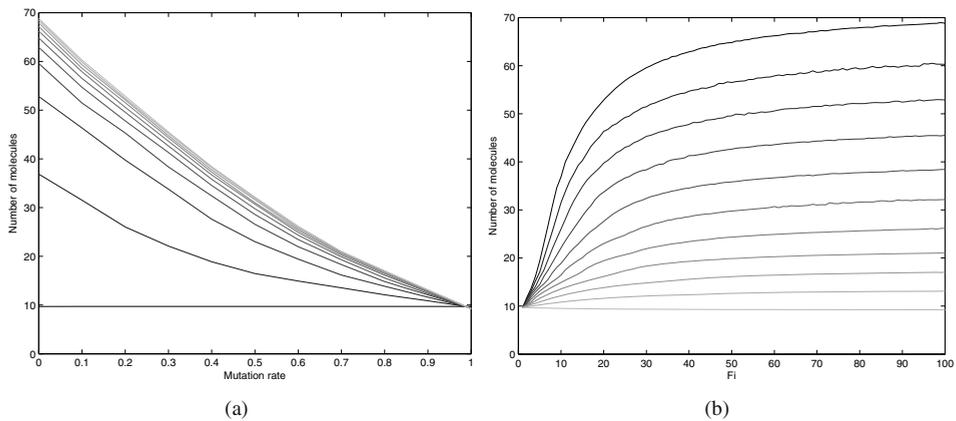
Fig. 12. Number of molecules as a function of replication intensity F_i. One type of sequence, i, is selected ($\text{Len}_i = 6$), $P^{\text{ReplMut}} = 0.1$ (a), $P^{\text{ReplMut}} = 0.4$ (b). $P^{\text{Join}} = 0.01$, $P^{\text{Split}} = 0.001$, $S_i = 1$, $P^{\text{Repl}} = 0.01$, $P^{\text{InDiff}} = 0$, $P^{\text{OutDiff}} = 0$, $\text{Len}^{\text{Max}} = 6$.

Fig. 13. Number of i molecules as a function of replication intensity F_i—each line represents different F_i value (light gray: $F_i = 100$, black: $F_i = 1$ (step is set to 10)) (a), mutation rate—each line represents different P^{ReplMut} value (light gray: $P^{\text{ReplMut}} = 1$, black: $P^{\text{ReplMut}} = 0$ (step is set to 0.1)) (b). One sequence type, i, is selected ($\text{Len}_i = 6$). $P^{\text{Join}} = 0.01$, $P^{\text{Split}} = 0.001$, $S_i = 1$, $P^{\text{Repl}} = 0.01$, $P^{\text{InDiff}} = 0$, $P^{\text{OutDiff}} = 0$, $\text{Len}^{\text{Max}} = 6$.
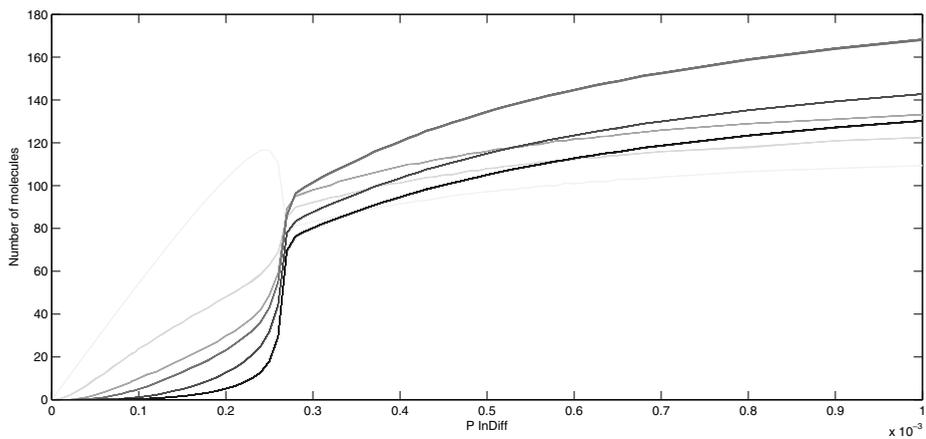

Fig. 14. Number of molecules as a function of diffusion P^{InDiff}. $P^{\text{Join}} = 0.01$, $P^{\text{Split}} = 0.001$, $S_i = 1$, $P^{\text{Repl}} = 0$, $P^{\text{OutDiff}} = 0.1$, $\text{Len}^{\text{Max}} = 6$.
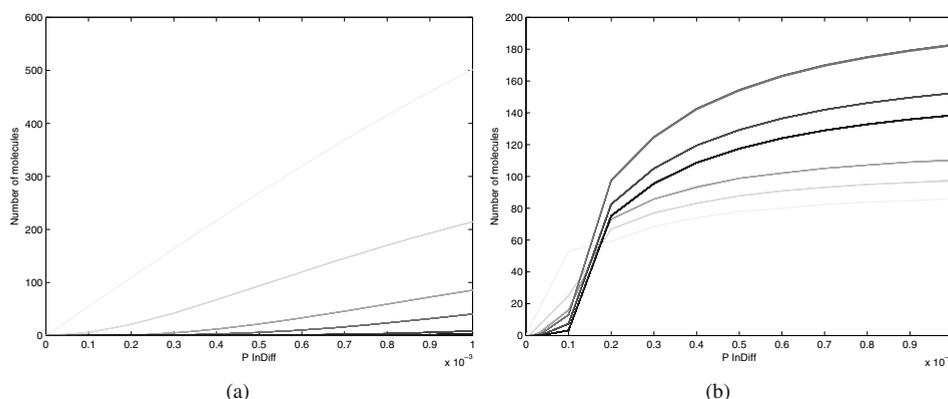
Fig. 15. Number of molecules as a function of diffusion P^{InDiff}. $P^{\text{Join}} = 0.001$ (a). $P^{\text{Join}} = 0.02$ (b). $P^{\text{Split}} = 0.001$, $S_i = 1$, $P^{\text{Repl}} = 0$, $P^{\text{OutDiff}} = 0.1$, $\text{Len}^{\text{Max}} = 6$.

proper parameter configuration there is a step change in the number of molecules in the protocell.

The model presented in this article extends the works of Nowak and Ohtsuki (2008), although there are still areas which could be modelled in a more detailed way, for example, strand replication. Currently it is one step process. In the next version of the model creation of double strand complex, followed by ligation of attached strands and replicated strands separation might be implemented. Each phase might be modelled separately; it will give better insight into the influence of the ratio of strands ligation or strands separation on the polynucleotides obtained in the system. Also new phenomena at the level of protocells, such as fission of packages, might be introduced.

Life's origins should not be perceived as a single event but rather as a gradual process of complexity creation. Protocells might play an important role in such a process, because the interplay between membrane and internal processes could lead to an increase in molecule concentration and thus support creation of RNA world.

Acknowledgment

References

Breaker, R.R. (2010). Riboswitches and the RNA world, *Cold Spring Harbor Perspectives in Biology* **4**(2): 1–15.

Cary, S.C., Shank, T. and Stein, J. (1998). Worms bask in extreme temperatures, *Nature* **391**(6667): 545–546.

Cech, T. and Golden, B. (1999). Building a catalytic active site using only RNA, *in* R.F. Gesteland, T.R. Cech and J.F. Atkins (Eds.), *The RNA World*, 2nd Edn., Cold Spring Harbor Laboratory Press, New York, NY, Vol. 37, pp. 321–349.

Chen, I.A., Roberts, R.W. and Szostak, J.W. (2004). The emergence of competition between model protocells, *Science* **305**(5689): 1474–1476.

Chen, I.A. and Walde, P. (2010). From self-assembled vesicles to protocells, *Cold Spring Harbor Perspectives in Biology* **2**(7): 1–13.

Chen, I., Hanczyc, M., Sazani, P. and Szostak, J. (2006). Protocells: Genetic polymers inside membrane vesicles, *in* I.A. Chen, M.M. Hanczyc, P.L. Sazani and J.W. Szostak (Eds.), *The RNA World*, 3rd Edn., Cold Spring Harbor Laboratory Press, New York, NY, Vol. 43, pp. 57–88.

Cyran, K.A. (2009). Information amount threshold in self-replicating RNA-protospecies: Branching processes approach, *International Journal of Mathematics and Computers in Sumulation* **3**(1): 20–29.

De Lucrezia, D., Anella, F. and Chiarabelli, C. (2007). Question 5: On the chemical reality of the RNA world, *Origins of Life and Evolution of Biospheres* **37**(4): 379–385.

Deamer, D. (2008). Experimental approaches to fabricating artificial cellular life, *in* S. Rasmussen (Ed.), *Protocells: From Bridging Nonliving and Living Matter*, MIT Press, Cambridge, MA, pp. 19–38.

Deamer, D.W. (1985). Boundary structures are formed by organic components of the Murchison carbonaceous chondrite, *Nature* **317**(6040): 792–794.

Deamer, D.W. and Pashley, R.M. (1989). Amphiphilic components of the murchison carbonaceous chondrite: Surface properties and membrane formation, *Origins of Life and Evolution of Biospheres* **19**(1): 21–38.

Dobson, C.M., Ellison, G.B., Tuck, A.F. and Vaida, V. (2000). Atmospheric aerosols as prebiotic chemical reactors, *Proceedings of the National Academy of Sciences* **97**(22): 11864–11868.

Ertem, G. (2004). Montmorillonite, oligonucleotides, RNA and origin of life, *Origins of Life and Evolution of the Biosphere* **34**(6): 549–570.

Feig, A. and Uhlenbeck, O. (1999). The role of metal ions in RNA biochemistry, *in* R.F. Gesteland, T.R. Cech and J.F. Atkins (Eds.), *The RNA World,* 2 nd Edn., Cold Spring Harbor Laboratory Press, New York, NY, pp. 287–319.

Ferris, J.P. (2005). Mineral catalysis and prebiotic synthesis: Montmorillonite-catalyzed formation of RNA, *Elements* **1**(3): 145–149.

Ferris, J.P. (2006). Montmorillonite-catalysed formation of RNA oligomers: The possible role of catalysis in the origins of life, *Philosophical Transactions of the Royal Society, B: Biological Sciences* **361**(1474): 1777–1786.

Ferris, J.P., Hill, A.R., Liu, R. and Orgel, L.E. (1996). Synthesis of long prebiotic oligomers on mineral surfaces, *Nature* **381**(6577): 59–61.

Fishkis, M. (2007). Steps towards the formation of a protocell: The possible role of short peptides, *Origins of Life and Evolution of Biospheres* **37**(6): 537–553.

Fishkis, M. (2010). Emergence of self-reproduction in cooperative chemical evolution of prebiological molecules, *Origins of Life and Evolution of the Biosphere* **41**(3): 261–275.

Gaucher, E.A., Kratzer, J.T. and Randall, R.N. (2010). Deep phylogeny: How a tree can help characterize early life on earth, *Cold Spring Harbor Perspectives in Biology* **2**(1): 1–16.

Gilbert, W. (1986). The RNA world, *Nature* **319**(6055): 618.

Hanczyc, M.M., Fujikawa, S.M. and Szostak, J.W. (2003). Experimental models of primitive cellular compartments: Encapsulation, growth, and division, *Science* **302**(5645): 618–622.

Joyce, G.F. (2002). The antiquity of RNA-based evolution, *Nature* **418**(6894): 214–221.

Joyce, G. and Orgel, L. (1999). Prospects for understanding the origin of the rna world, *in* R.F. Gesteland, T.R. Cech and J.F. Atkins (Eds.), *The RNA World*, 2nd Edn., Cold Spring Harbor Laboratory Press, New York, NY, Vol. 37, pp. 49–77.

Kimmel, M. and Axelrod, D. (2002). *Branching Processes in Biology*, Springer-Verlag, New York, NY.

Koshland, D.E. (2002). The seven pillars of life, *Science* **295**(5563): 2215–2216.

Lazcano, A. (2010). Historical development of origins research, *Cold Spring Harbor Perspectives in Biology* **2**(11): 1–16.

Luisi, P. (2010). *The Emergence of Life: From Chemical Origins to Synthetic Biology*, Cambridge University Press, New York, NY.

Lunine, J.I. (2006). Physical conditions on the early Earth, *Philosophical Transactions of the Royal Society, B: Biological Sciences* **361**(1474): 1721–1731.

Ma, W., Yu, C. and Zhang, W. (2007). Monte Carlo simulation of early molecular evolution in the RNA World, *Biosystems* **90**(1): 28–39.

Manapat, M., Ohtsuki, H., Bürger, R. and Nowak, M.A.A. (2009). Originator dynamics, *Journal of Theoretical Biology* **256**(4): 586–595.

Matsuura, T., Yamaguchi, M., Ko-Mitamura, E.P., Shima, Y., Urabe, I. and Yomo, T. (2002). Importance of compartment formation for a self-encoding system, *Proceedings of the National Academy of Sciences* **99**(11): 7514–7517.

Mojzsis, S., Krishnamurthy, R. and Arrhenius, G. (1999). Before RNA and after: Geophysical and geochemical constraints on molecular evolution, *in* R.F. Gesteland, T.R. Cech and J.F. Atkins (Eds.), *The RNA World*, 2nd Edn., Cold Spring Harbor Laboratory Press, New York, NY, pp. 1–47.

Myszor, D. and Cyran, K.A. (2009). Estimation of the number of primordial genes in a compartment model of RNA world, *in* K.A. Cyran, S. Kozielski, J.F. Peters, U. Stańczyk and A. Wakulicz-Deja (Eds.), *Man–Machine Interactions*, Advances in Intelligent and Soft Computing, Vol. 59, Springer, Berlin/Heidelberg, Chapter 15, pp. 151–161.

Myszor, D. and Cyran, K.A. (2010). Influence of non-enzymatic template-directed RNA recombination processes on polynucleotides lengths in Monte Carlo simulation model of the RNA world, *International Journal of Applied Mathematics and Informatics* **4**(1): 1–8.

Niesert, U. (1987). How many genes to start with? A computer simulation about the origin of life, *Origins of Life and Evolution of the Biosphere* **17**(2): 155–169.

Noller, H.F. (2006). Evolution of ribosomes and translation from an RNA world, *in* I.A. Chen, M.M. Hanczyc, P.L. Sazani and J.W. Szostak (Eds.), *The RNA World*, 3rd Edn., Cold Spring Harbor Laboratory Press, New York, NY, Vol. 43, pp. 287–307.

Nowak, M. A. and Ohtsuki, H. (2008). Prevolutionary dynamics and the origin of evolution, *Proceedings of the National Academy of Sciences* **105**(39): 14924–14927.

Ohtsuki, H. and Nowak, M. A. (2009). Prelife catalysts and replicators, *Proceedings of the Royal Society, B: Biological Sciences* **276**(1674): 3783–3790.

Pohorille, A. and Deamer, D. (2002). Artificial cells: Prospects for biotechnology, *Trends in Biotechnology* **20**(3): 123–128.

Pohorille, A. and Deamer, D. (2009). Self-assembly and function of primitive cell membranes, *Research in Microbiology* **160**(7): 449–456.

Rasmussen, S., Bedau, M.A., Chen, L., Deamer, D., Krakauer, D.C., Packard, N.H. and Stadler, P.F. (2008). *Protocells: Bridging Nonliving and Living Matter*, Vol. 8, MIT Press, Cambridge, MA.

Schleper, C., Piihler, G., Kuhlmorgen, B. and Zillig, W. (1995). Life at extremely low pH, *Nature* **375**(6534): 741–742.

Schrum, J.P., Zhu, T.F. and Szostak, J.W. (2010). The origins of cellular life, *Cold Spring Harbor Perspectives in Biology* **2**(9): a002212.

Sleep, N.H. (2010). The Hadean-Archaean environment, *Cold Spring Harbor Perspectives in Biology* **2**(6): 1–14.

Stetter, K.O. (2007). Hyperthermophilic life on Earth—and on Mars?, *in* R. Pudritz, P. Higgs and J. Stone (Eds.), *Planetary Systems and the Origins of Life*, Cambridge University Press, New York, NY, pp. 135–147.

Sutherland, J.D. (2010). Ribonucleotides, *Cold Spring Harbor Perspectives in Biology* **2**(4):1–13.

Szabó, P., Scheuring, I., Czárán, T. and Szathmáry, E. (2002). In silico simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity, *Nature* **420**(6913): 340–343.

Szostak, J.W., Bartel, D.P. and Luisi, P.L. (2001). Synthesizing life, *Nature* **409**(6818): 387–390.

Wieczorek, R. (2010). Markov chain model of phytoplankton dynamics, *International Journal of Applied Mathematics and Computer Science* **20**(4): 763–771, DOI: 10.2478/v10006-010-0058-7.

Woese, C.R. (1968). *Genetic Code*, Harper & Row, New York, NY.

Xu, C., Liao, M. and He, X. (2011). Stability and Hopf bifurcation analysis for a Lotka–Volterra predator-prey model with two delays, *International Journal of Applied Mathematics and Computer Science* **21**(1): 97–107, DOI: 10.2478/v10006-011-0007-0.

Zahnle, K., Schaefer, L. and Fegley, B. (2010). Earth's earliest atmospheres, *Cold Spring Harbor Perspectives in Biology* **2**(10): 1–17.

Dariusz Myszor received M.Sc. degree in computer science (2007) from the Silesian University of Technology, Gliwice, Poland. Currently he is a Ph.D. student at the same university and is involved in research and teaching activities of the Institute of Informatics at the SUT. His current research interests are focused in such areas as neural networks, computational population genetics, computer simulations, early life and bioinformatics.

Krzysztof A. Cyran, an associate professor in the Institute of Informatics at the Silesian University of Technology, received his Ph.D. degree and habilitation in technical sciences (computer science). He has published about 100 papers, with around 200 citations. He is a reviewer and member of editorial boards and programme committees of scientific journals and conferences. He has been a visiting scholar at Rice University in Houston, USA, and the vice-head of the Institute of Informatics at his home university. Currently, he is the director of the Virtual Flying Laboratory. Professor Cyran, rewarded by the President of Poland, is a coordinator and investigator in numerous funded projects (including four 7FP EU projects).