Institutional Repository - Research Portal
Dépôt Institutionnel - Portail de la Recherche

researchportal.unamur.be

RESEARCH OUTPUTS / RÉSULTATS DE RECHERCHE

Evolution of a population of protocells: Emergence of species

Carletti, Timoteo; Fanelli, Duccio

Published in:
Proc. of the 2006 International Symposium on Mathematical and Computational Biology

Publication date:
2007

Document Version
Early version, also known as pre-print

Link to publication

Citation for pulished version (HARVARD):
Carletti, T & Fanelli, D 2007, Evolution of a population of protocells: Emergence of species. in R Mondaini & R Dilão (eds), *Proc. of the 2006 International Symposium on Mathematical and Computational Biology: BIOMAT 2006.* E-papers editora, Brasil.

EVOLUTION OF A POPULATION OF PROTOCELLS: EMERGENCE OF SPECIES

T. CARLETTI [*†]

Département de Mathématique, FUNDP
8 rempart de la vierge
B5000 Namur, Belgium
tel +32 81724903   fax +32 81724914
E-mail: timoteo.carletti@fundp.ac.be

D. FANELLI

Dipartimento di Energetica "S. Stecco", University of Florence and INFN
via S. Marta 3
50139 Firenze, Italy
E-mail: Duccio.Fanelli@ki.se

Keywords: Protocell; Chemoton model; Evolution; Speciation; Chaos

The Gánti Chemoton represents a minimalist description of a protocell unit based on three coupled chemical networks, namely a proto–metabolism, a template duplication and the membrane growth. We here propose an improved version of Gánti's "unit of life" where the effects of the volume changes, due to the membrane growth, are explicitly taken into account. The model is further completed by postulating a stochastic mutation mechanism that acts on the template duplication. Within this framework we investigate the evolution of a population of protocells and demonstrate that our hypothesis translates into an open–ended Darwinian evolution, under the pressure of the environment. This observation enables us to conclude that the Chemoton is also a "unit of evolution": differentiation into species is indeed an emergent property of the model.

1. Introduction

Almost all life forms known today are composed by cells, fundamental constituting units which are able to self–replicate and evolve through changes

2

in genetic information, once selection is active, for instance due to the competition for resources.

These highly sophisticated devices are the product of about four billion–years evolution, and, in this respect, represents the relic of primordial life bricks, the *protocells*. It is customarily believed that the latter were most probably exhibiting only few simplified functionalities – nonetheless necessary to consider the protocells alive [1,14,18] – that required a *primitive embodiment structure*, a *protometabolism* and a *rudimentary genetics*, so to guarantee that offspring were "similar"to their parents.

Artificial protocells have not yet been reproduced in laboratory and intense research programs [a] are being established aiming at developing reference models [9,13] that capture the essence of the first protocells appeared on earth and enable one to monitor their eventual subsequent evolution.

In 1971, Gánti [6,7] proposed a pioneering theory that provides a minimalist description of a protocell, termed *Chemoton*, i.e. a simplified model to describe growing and multiplying microspheres controlled by a template duplication process. The original minimal Chemoton hypothesis [6] includes a *membrane* which protects the inner bulk, while filtering the access of generic high energetic "food", which is made available in the surrounding environment. The food is thus *metabolized* through chemical reactions and transformed into basic chemical constituents that are needed to stimulate both template duplication and membrane growth. Finally, in the simplest scheme, the template–polymer is assembled from one specific type of monomer (homopolymer), which acts as an effective *carrier of information*: its length determines in fact the division time, a phenotype property of paramount importance. After such a time, the protocell attains a critical size, above which a division into two perfect halves occurs. It is thus assumed that cell division occurs as a result of a purely physical process [9]. Further, we also assume that each daughter cell contains an identical amount of chemical material, which is equally shared from the mother constituents. Thus, according to the definition proposed in [18,1,14,16], the Chemoton hypothesis defines a *unit of life*.

In [7] Gánti raised the following natural question: can one assume the Chemoton to represent also a *unit of evolution* [16] ? To provide a definite answer to the above question, one has to monitor the dynamical evolution of a population of Chemoton, instead of focusing on the single protocell

---

[a]For instance PACE – Programmable Artificial Cell Evolution – an European Integrated Project in the EU FP6-IST-FET Complex Systems Initiative.

case. Moreover, the influence of the environment on each protocell needs to be properly incorporated into the model.

With respect to the original Chemoton picture, we here develop an improved formulation (see also [2,3]): chemical reactions occur in a varying volume, the latter being intrisically controlled by membrane growth. This volume effect is explicitly accounted for and shown to significantly contribute to the underlying dynamical process. Moreover, the original deterministic model is further modified to accommodate for an effective mutation mechanism: during the template duplication extra monomers are added (removed) according to a pre-assigned probability, to mimic a stochastic source of errors in the duplication process.

Within this novel scenario, the time evolution of a single protocell is monitored as functions of a number of selected variables of key relevance. This knowledge translates into a State Function which unambiguously predicts the ultimate fate of the protocell, once the specifications of the protocell, e.g. the length of the polymer (N) and its ability to self–replicate (V^*), and the environmental condition, e.g the external available food ($\bar{X}$), are assigned.

In particular, assuming that the external food is periodically available, each cycle being for instance associated to tides in the pond were the protocells supposedly live, or the metabolic process to be driven by a periodic source of energy, e.g. the light from the sun, then the *regular functioning* of the model, i.e. periodic growth and division in time of the protocell, relies on the *synchronization* of the three chemical networks [2].

Motivated by this observation, we previosuly [2,3] assumed that only protocells with regular behavior give rise to next generation offspring, while non–synchronized protocells will eventually die. Within this approximation, we provided clear evidence that the model exhibits a simple speciation mechanism, which truly appears as an emergent property. We here adopt a definition of species for asexual beings, based on differences between their genomes [4,11,8], which in our setting is equivalent to different lengths of the double–stranded templates.

The validity of our former strategy, "conservative" form an evolutionary point of view (i.e. non–synchronization implies death), is in turn confirmed by the absence of intermittency (data not shown) in the protocell dynamics.

In this paper we take one step forward by relaxing the above hypothesis: protocells can attain the duplication threshold at different times. The period of the cycle is hence a characteristic associated to each unit.

Let us stress that the computations here reported are made possible

4

because of the use of the State Function concept, a technical solution that allows us to simulate the long time evolution of a large family of protocells [b]. Being the protocells subjected to the pressure of the environment, here exemplified through the amount of available food, which changes in time and depends self–consistently on the size of the population: the larger the number of protocells, the lower the amount of available food, and vice–versa. This open-ended mechanism reproduced in this context mimics therefore an effective Darwinian evolution.

As we shall see, protocells with different templates can develop from a common ancestor, due to the combined action of two Darwinian forces: the mutation mechanism and the competition for the available food. The occurrence of a speciation mechanism in a Chemoton–like population is here demonstrated with reference to the above, generalized, setting. This result contributes to shed new light into the important issue of protocells evolution and dynamics: The Chemoton hypothesis defines not only a reliable unit of life, but also a unit of evolution, once the protocell shape is taken into account. Observe in fact that this remarkable feature cannot be obtained in the *spherical–shaped* Chemoton model [7,12], when both volume dependence and stochastic mutation are accounted for.

2. The model

For the sake of completeness, let us start by briefly introducing the original Chemoton model. The interested reader can refer to to [2,6,7] for further details.

Gánti's Chemoton model is composed by three chemical networks coupled together and schematically outlined in Table 1.

The metabolic autocatalytic cycle is decomposed in five elementary steps, each involving one of the chemicals $A_1, \ldots, , A_5$. Its role is to transform the highly energetic, externally available, "food", $\bar{X}$, into internal materials (i.e. the precursor of membrane molecules, T', and the monomers V') which sustain the growth and self–reproduction processes. The food is here supposed to be buffered into the surrounding environment and therefore assumed constant inside the protocell. The metabolism produces also

---

[b]Each protocell behavior is described by more than one hundred of equations, computing the State Function once for all, enables to formally assimilate each protocell belonging to the population under scrutiny to a black box, which produces specific output, once selected input values are provided. This reduces considerably the computational costs and allows to significantly enhance the statistics over previous investigations.

some "waste", $\bar{Y}$, which is progressively eliminated so to maintain its level constant inside the membrane. Each reaction is reversible and therefore involves direct and inverse *rate constants* labeled, respectively, k_i and k_i'. The latter are set to the values proposed by Gánti in [7] and reported in Appendix A.2.

The second subsystem involves the *double–stranded template*, i.e. the carrier of the information. Since the model deals with homopolymer species, then the information is carried by the *length* of the double–stranded template. As a first step, we assume a constant length of the template, identified by a sequence of N of monomers. In the subsequent analysis, we shall instead introduce a mutation mechanism acting on the template, which, in particular, can alter the polymer length. Due to the thermical fluctuations the bonds at the extremities of the doubled stranded template can open, an event that occurs with high probability. Consequently, when a large enough number of free monomers is available in solution ($[V'] \geq V^*$, V^* being the *polymerization threshold* [c]), the duplication reaction initiates and is further sustained (*strong replication condition* [12]). If the polymer has length N, after $2N$ steps these nested reactions will produce a second identical copy of the double–stranded template and, as by product, R molecules which will allow the transformation of T' into a membrane molecule T.

The last subsystem governs the dynamical expansion of the *growing membrane*, which defines the container delimiting the portion of volume where all reactions occur. It avoids the dispersion of useful chemicals, allows the entrance of food and expells the waste. Once a membrane molecule is produced from a precursor T' and a molecule R, via an intermediate state T^*, it is then incorporated into the membrane, thus contributing to its growth. The increase of the membrane is proportional to the membrane size and to the concentration of membrane molecules [17]: $\frac{dS}{dt} = k_{10}TS$, k_{10} being a positive proportionality factor.

The three subsystems are hence coupled together and the effective functioning of the Chemoton machinery relies on the precise synchronization of the cycles, which should be thefore maintained in time.

We adopt here a deterministic approach to solve the chemical equations. A system of kinetic differential equations is introduced to describe the time evolution of the chemical concentrations. Such a system is reported in

---

[c]This quantity provides an indirect measure of the attitude for polymerization once monomers have been selected. The lower V^*, the more pronounced the ability of the polymer to self–replicate.

6

Table 2. As usual, upper dot stands for time derivative while, to lighten notations, capital letters refer to the concentrations.

Moreover, we assume that once the membrane size has doubled, with respect to its initial value, the protocell halves into two smaller offspring, due to physical instabilities. This event defines the *protocell cycle.*

In Fig. 1 two typical behaviors of the division time as function of the generation number are displayed. The simulations refers to one individual protocell. The interested reader can refer to Appendix A.2 for a detailed account on the numerical scheme here adopted. The plot on the left reports a periodic, regular, orbit: after a short transient the division time converges to an fixed asymptotic value. This in turn implies that a population composed by such individuals could be ideally synchronized and share an identical genetic content at the division event. Moreover the population size will increase exponentially opening the way to possible a Darwinian selection [5,15]. The right panel of Fig. 1 shows an irregular orbit: the division time depends on the generation number. In this case, a virtual family of protocells would be composed by units that divide faster than other, thus possibly resulting in different amount of the inner chemicals, including also the double–stranded template, propagated to the offsprings.
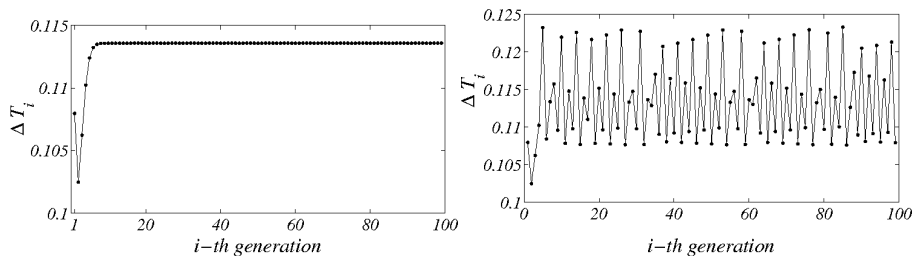

Figure 1.   The Division Time. On the left panel we plot the division time as a function of the generation number, for a regular, i.e. periodic, behavior. After a short transient, the division time attains a fixed value: protocells belonging to different generations employ the same amount of time to take the division step to completion. On the right panel an irregular, i.e. non-periodic, behavior. The division time depends on the generation number and it never achieves an asymptotic constant value.

In the original model proposed by Gánti the Chemoton always keeps a spherical shape when growing: this is an important assumption because the behavior of the protocell is determined by the concentrations of the chemicals. Since the volume of the protocell changes in time, one has to properly insert this information into the relevant chemical equations.

Assume the rate of variation (with respect to time) of a given concentration, c_i, at *fixed* volume, to be given by $f_i(\bar{c})$, where $\bar{c} = (c_1, \ldots, c_k)$. Then allowing the volume Vol to change, yields:

$$\frac{dc_i}{dt} = f_i(\bar{c}) - c_i \frac{1}{Vol} \frac{dVol}{dt} \, , \tag{1}$$

and each chemical equation in Table 2 has to be modified according to eq. (1).

From the chemical equations, one deduces the time evolution of the outer surface: assuming a specific dependence of the volume vs. the surface one can eventually monitor the dynamical variation of Vol. We here hypothesize that the protocells pass from an almost spherical shape to a typical hourglass profile close to the duplication event. The existence of an inner pinch is also confirmed by inspection of real cells' behaviour. The typical evolution of a protocell is schematically depicted in the left panel of Fig. 2. The mathematical details are instead discussed in Appendix A.1.

As already anticipated, the deviations from the original, spherical, profile play a central role. For a given surface, in fact, the shape determines the volume enclosed by the membrane and consequently affects the relative concentrations. The sphere maximizes the inner volume which in turn implies that other, more realistic shapes would delimit smaller portions of space, hence resulting in higher concentration. This observation is exemplified in the right panel of Fig. 2 where the ratio $V(t)/[S(t)]^{3/2}$ is displayed as function of the surface size for the case considered in this study. Observe that, for the case of a sphere such a ratio, is constant and equal to $1/(6\sqrt{\pi})$. Let us observe that already in [12], a first attempt has been made to esplicitly account for volume changes. In that work, however, the authors still assimilate the protocell to a sphere, during all stages of its evolution. Following our approach, the protocell shape gradually progresses towards a smooth splitting, that contributes to effectively reduces the functional discontinuities observed within the original Chemoton picture [12], in conjunction with a duplication event. (See also Appendix A.1)

To make the model more realistic we further introduce a *mutation mechanism*, acting on the double–stranded template. Observe that the genome of each protocell is completely specified through its length, N, and the associated ability to polymerize, V^*. It is thefore quite natural to directly act on the length of the polymer. A *mutation* is an error occurring during the template reproduction: instead of one monomer, two (or more) monomers are added or removed, with a prescribed probability. Alternatively, we can imagine to introduce a second mutation mechanism: as
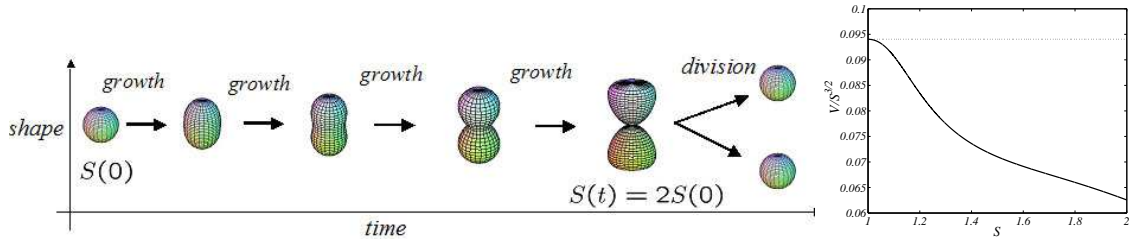
8

Figure 2.   Behavior of the protocell shape in time. *left panel* Time snapshots of the protocell evolution. At $t = 0$ the protocell is assimilated to a sphere whose surface is $S(0) = 1$. At the division time, the surface has doubled, $S(T) = 2S(0)$ and the shape resembles a hourglass. *right panel* The ratio $V/S^{3/2}$ against the surface size, S. The black line stand for our model, while the dotted line refers to the case of a sphere. Note that, by assumption, at $S = 1$ the two curves coincide.

already observed, the polymerization threshold V^* provides a measure of the tendency to polymerize. One can therefore assume that the spatial 3D configuration (e.g. right and left helices of ammino–acids) attained by the monomers in the template cooperatively enhances, or viceversa contrast, the polymerization, which practically translate in a modification of the related polymerization threshold. This effect can be mimicked by a second mutation mechanism responsible for altering the initial value of V^*.

The action of our mutation mechanisms is then algorithmically summarized as follows: the protocell starts its division cycle with some specification of the double–stranded template, say (N_0, V_0^*); then, it grows, produces the needed inner chemicals and duplicates the double–stranded template. Once the membrane size reaches the threshold value, the protocell halves into two offspring, splitting the inner elements and also the polymers, which can in principle be different from the initial ones due to possible mutations occurred during the template duplication. Each daughter is therefore characterized by new specifications of the double–stranded template, respectively (N_1, V_1^*) and (N_2, V_2^*).

In the following we shall adopt our improved theoretical framework to numerically investigate the time evolution of a family of protocells.

3. Results and Conclusions

Interestingly enough, no complete stochastic simulations of a Chemoton–like model has been attempted so far, except [2,3], the reason being ascribed mainly to the huge computations involved, when tracking the evolution of each single protocell while the population size increases exponentially due

to the combined effect of the duplication and mutation mechanisms. As already mentioned, we overcome these technical difficulties by defining an innovative numerical strategy, briefly recalled below.

The time evolution of a *single protocell* is monitored as functions of a number of selected variables of key relevance, e.g. the amount of food initially available, the template length and its polymerization threshold at start. This information is used to build up a *State Function* which enables one to predicts the ultimate fate of the protocell under scrutiny once the initial conditions are specified and without resorting to additional numerical integration. More precisely, for each choice of parameters values (N, V^*), defined over a large grid, and distinct amount of available food [d], we integrate the kinetic equations, according to the protocol discussed in Appendix A.2. After a time transient, generally 100 division cycles, we record the last, typically 10, division times. If these values are all identical to an hypothetic value T_0, then the protocell is developing on a T_0–period orbit. Otherwise, the protocell is visiting an irregular orbit, and each of the recorded division times assumed equally probable. When infering the future fate of the protocell we can therefore procede as follows: in the former case, its division time will be set to T_0, while for the latter case the division time is selected from one of the stored 10 division values, with uniform probability.

The last remark concerns the competition for the resources. Let us suppose that each protocell burns the same type of food. Hence, when the population is larger, a lower quantity of food will become available to each protocell. This competition tends to favour the selection of protocells which are charaterized by shorter division times, the latter being optimized with respect to the genome parameters and external resources. On the other, hand too short division times could be not enough to produce all the inner chemicals needed for the next generation to survive (death by dilution): for instance, the protocell could eventually divide, because the membrane size has reached the threshold value, while two copies of the double–stranded template have not yet been formed. In that case, one of the two offspring would clearly die, being not equipped with the necessary genetic material. These two opposite forces will interact and give rise to a speciation mechanism. It is worth emphasizing that the speciation mechanism here outlined is neither trivial nor a priori predictable. It is indeed resulting from a complex interplay between the postulated mutation scheme and the

---

[d]Here, N ranges from 1 and 55, while V^* covers the interval $(0, 100)$. As concerns the available food $\bar{X}$, the following values have been used 0.1, 1, 10, 50, 100 and 1000.

10

effect of the resonances, an intrinsic feature of our modified Gánti's model that manifests as forbidden regions in the State Function space, see Fig. 6 in Appendix A.2.

The typical simulation is performed as follows.

(1) We start with M_0 initial protocells, each one being described by given genome specification $(N_j, V_j^*)_{1 \leq j \leq M_0}$. The protocell lives in an environment with a specific amount of available food $\bar{X}_0$. Typically, $\bar{X}_0 = 100$ and the protocells are all exact copies of the same ancestor unit;

(2) We extract the division times for all protocells: for each (N_j, V_j^*) we look at the State Function and randomly select one of the 10 possible division times, which is hence assumed as the current division time;

(3) The protocells with the shortest division times will divide first, giving birth to two offspring with new genome, resulting from the stochastic mutation mechanism applied to the mother genome specification.

(4) Because the population size increases, the available food is rescaled with respect to the present number of protocells (the available food is now lesser than before)

(5) The dynamics proceeds further repeating steps (1), (2) , (3) [e] and (4).

Remark that an intermediate steps, between (4) and (5), can be inserted which consists in keeping the total population size below an assigned, critical, threshold. This is achieved by randomly eliminating living protocells exceeding the aforementioned threshold. Such effect can be invoqued to simulate a population of protocells in a pond on the prebiotic earth, exposed to drastic changes of the environment due to storms, floods, or tides.

Several simulation runs are performed, corresponding to different class of initial conditions for the genomes specifications. In a large number of cases, a *speciation* manisfest as an emergent property of the model. From an initial uniform population of protocells, the system progressively evolve towards an asymptotic state being characterized by two, stable, mutant families each one possessing a well defined genome distribution. Such families correspond to new species according to the general definition reported

---

[e]Remark that once new protocells are produced, the older ones have already advanced in the division cycle and thus they possibly divide before the newborn offsprings: we coherently reset the origin of time.

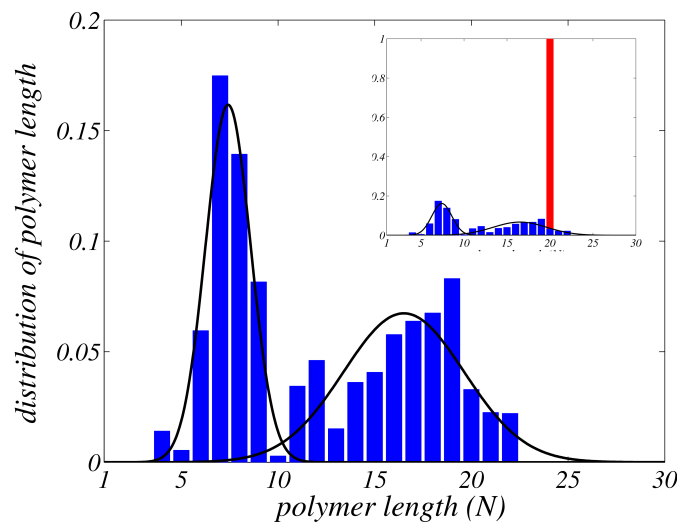in [4,8,11]. Our findings are clearly displayed in Fig. 3.

Figure 3.   Distribution of polymers length after 4500 divisions (blue), starting with 5 identical protocells with polymer length of 20 monomers (red) (shown in the inset). Two species are clearly identified corresponding to the 47.3 % and 52.7 % of the whole population, resulting from the selection of the fittest mutated protocells, in the competition for food. The solid black lines are the results of numerical interpolations: one Gaussian curve has mean 7.4 and standard deviation 1.1, while the second one is centered at 16.5 with standard deviation 3.1. This picture represents a generic evolution: distinct runs relative to different choices of the parameters results in similar behaviors. Our findings are also shown to be robust to small changes of the parameters involved.

The occurrence of a speciation mechanism in a Chemoton–like population is here demonstrated and contributes to shed new light into the important issue of protocells evolution and dynamics. In conclusion, and back to the initial question mutuated by Gánti, our modified Chemoton hypothesis defines not only a reliable unit of life, but also a *unit of evolution*. Importantly, this scenario does not apply to the original Chemoton description, the geometrical aspects related to the process of division playing a fundamental role.

Acknowledgments

12

Appendix A.

The aim of this Appendix is to provide further insight into a number of technical issues that have not been included in the main text to lighten the presentation. In particular, we shall focus on two specific topics: (i) mathematical representation of the membrane profile, computation of the associate volume and discussion of its role in our numerical experiments; (ii) protocols adopted in the numerical simulations.

Appendix A.1. *Mathematical model of the protocell membrane*

The aim of this section is to provide a detailed mathematical description of the membrane envelope that we have assumed in our study. First, let us recall the constraints that have to be verified:

(1) Immediately after a division event has occurred, each offspring has a small surface S_0 and it can be ideally assimilated to a *sphere*. In our simulations we set $S_0 = 1$.

(2) The successive division takes place once the protocell under scrutiny has essentially doubled its surface with respect to its original value. In other words, there exists a specific time T such that $S(T) = 2S_0$. As soon as this steric condition is fulfilled, the protocells divides and generates two identical offspring.

(3) From the chemical equations of Table 2, one can infer the concentration of the membrane molecules, thus enabling us to calculate the corresponding size of the surface, through an *ad hoc* proportionality constant. The volume enclosed can be estimated once a specific shape of the container is imposed, the latter providing a unique relation between the surface size and the inner volume.

(4) To mimic the effect played by the division mechanism on the inner concentrations, we assume that a pinch region will eventually develop (as observed in real cells), the latter being related to the physical instability that initiates the division process. Practically, we shall hypothesize that the protocell progressively deform itself

and finally assume an *hourglass* shape at the division time. This evolution is pictorially represented in the left panel of Fig. 2.

Mathematically we adopted the following description. Let a be a real parameter belonging to the interval $[0, 1]$ and introduce the following functions:

$$h_a(z) = \sqrt{-az^2 + \sqrt{(1-a)(z^2 - z^4) + \frac{a^2 S_0^2}{16\pi^2}}} \text{ and } r(a) = \sqrt{\frac{1 - a + \sqrt{(1-a)^2 + \frac{a^2 S_0^2}{4\pi^2}(a^2 - a + 1)}}{2(a^2 - a + 1)}} \,.$$

Then as a varies in between $[0, 1]$, the rotation body with Cartesian equations:

$$x^2 + y^2 = (h_a(z))^2 \text{ and } |z| \leq r(a) \,,$$

exhibits the desired properties. For $a = 1$, it corresponds to a sphere with surface S_0; conversely, when $a = 0$ an hoursand–like profile is obtained. Intermediate values of a enable to achieve a continuous transition in between the two aforementioned states. This is in turn confirmed by inspection of Fig. 4, where vertical sections are plotted for several values of a.
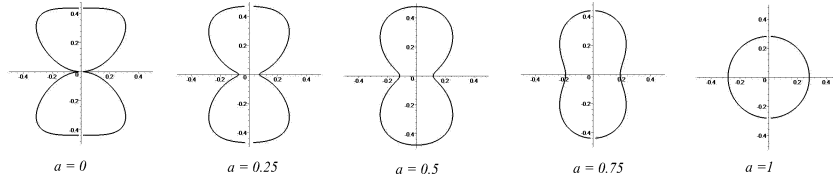

Figure 4.   Vertical sections of the rotated body corresponding to distinct values of the parameter a.

Thanks to the Pappus Guldino Theorems [f], for any given a, one can calculate the surface and the volume associated to the 3D profile obtained above. The following explicit formula are derived:

$$S_a = 4\pi \int_0^{r(a)} h_a(z)\sqrt{1 + (h_a'(z))^2} \, dz \text{ and } V_a = 2\pi \int_0^{r(a)} (h_a(z))^2 \, dz \,.$$

---

[f]The first Theorem of Pappus Guldino states that the area of a surface of revolution generated by rotating a plane curve γ about an axis external to γ and on the same plane is equal to the product of the arc length ℓ of γ and the distance d_1 traveled by its centroid. The second Theorem of Pappus Guldino states that the volume of a solid of revolution generated by rotating a plane figure F about an external axis is equal to the product of the surface area of F and the distance d_2 traveled by its geometric centroid.

14

To fulfill the above requirements we further impose:

$$a = 2 - \frac{S(t)}{S_0}\,.$$

Hence, at $t = 0$, when $S(0) = S_0$, a is equal to 1 and the corresponding shape is a sphere, with unitary surface area. On the contrary, when $S(t) = 2S_0$, then one gets $a = 0$, thus resulting in a hoursand profile. Still, the surface area associated to the final hoursand-like state is not equal to $2S_0$. To satisfy this latter condition, we can eventually deform the rotation body following the prescription reported below. Let

$$l(a) = \sqrt{\frac{S_a}{2 - a}} \text{ and } f_a(z) = \frac{h_a\,(l(a)z)}{l(a)}\,,$$

then it can be shown that the rotation body defined in Cartesian coordinates by:

$$x^2 + y^2 = (f_a(z))^2 \text{ and } |z| \leq \frac{r(a)}{l(a)},$$

matches the required conditions.

The obtained analytical expression of the membrane volume can be directly used into the relevant kinetic equations Eq. 1. Since chemical equations provides us with the time evolution of the outer surface, we choose to explicitly introduce the dependence of the volume vs. the surface and therefore write: $V = f(S)$, where f represents the so called *shape function* [g]. The kinetic equations are hence modified as:

$$\frac{dc_i}{dt} = f_i(\bar{c}) - c_i \frac{f'(S(t))}{f(S(t))} \frac{dS(t)}{dt}\,,$$

where f' stands for the derivative of f with respect to S.

This term plays an essential role in our model: it enables in fact to significantly reduce the discontinuities arising at the division, as we shall discuss in the following. In Fig. 5 we report the results of two typical runs respectively relative to our improved formulation and to the original Ganti model. We focus in particular on the appearance of the discontinuities at division. Both runs refer to the same initial data and parameters choice (for the details of the integration scheme please refer to the next section). As clearly displayed, large jumps of the concentration are found in the sphere–shaped model, the gaps being almost eliminated when our picture

---

[g]Let us note that for the case of a sphere $f(S) = (6\sqrt{\pi})^{-1}S^{3/2}$.

is assumed to hold. This observation suggests that our formulation is indeed more realistic, hence correct, than the spherical–shaped one.
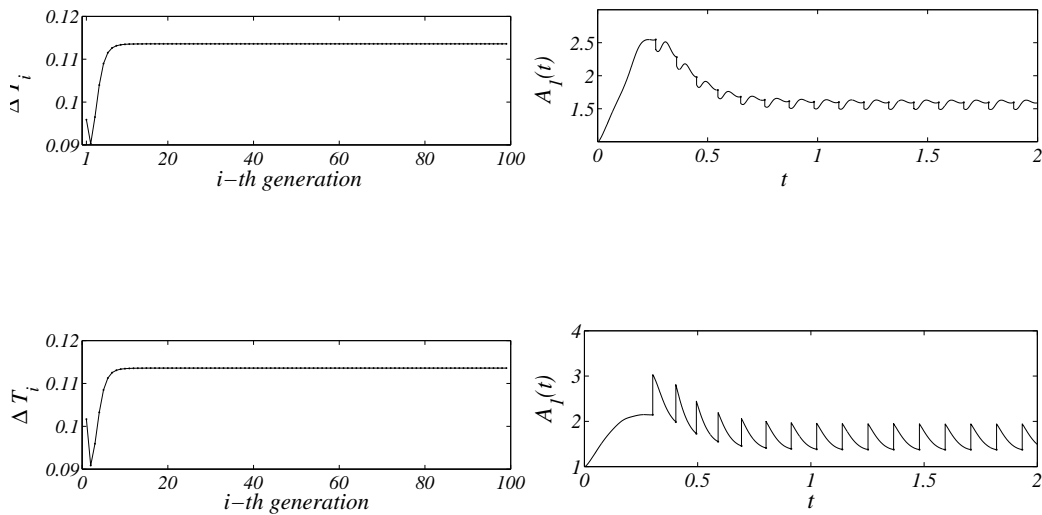

Figure 5. **Upper panels: The hourglass shape model**. On the left: the division time in function of the generation number for 100 successive divisions; observe that after a short transient the division time becomes independent on the division number and eventually reaches its asymptotic value: the division period (here 0.11359 units of time). On the right: the concentration of A_1 vs. time, during the first 16 divisions. Once again after a short transient,the initial concentration reaches an asymptotic constant value (here, 1.5972 units of concentrations). **Lower panels: the sphere shape model**. On the left: the division time in function of the generation number, which exhibits almost the same behavior as before. In fact, it reaches the same asymptotic value 0.11359 in time–units. On the right: the concentration of A_1 vs. time. Again, the qualitative behavior resembles the former one, but the asymptotic value is lower: 1.3747 conc.–units. Observe the large jumps in the concentration value at the division step.

Appendix A.2. *Protocol for a typical numerical simulation*

The numerical results presented in Fig. 5 are obtained by integrating the kinetic equations of Table 2. The aim of this section is to present the typical protocol used in our numerical experiments.

16

First, the parameters have to be assigned. For the reactions constants we adopted the values proposed by Gánti [7] and always used in the past literature. Their nominal values are reported in Table 3. Observe that Gánti introduced these values being inspired to the following, general, criteria: (i) the direct reactions constants must be larger that the inverse ones, to emphasize a privileged direction for the chemical reactions; (ii) for these specific values Gánti demonstrated numerically the existence of periodic solutions. As concerns our current investigation, the first motivation is indeed important. On the contrary, we will show that by varying the other parameters involved, while keeping the same values for the reactions constants, one can eventually observe irregular behaviors.

Secondly, one needs to fix the parameters related to the double–stranded template (namely the associated length and polymerization threshold) and the available external food. The simulations reported in Fig. 5, and other reported in the previous sections are obtained by imposing:

$$N = 25\,, \; V^* = 35.0 \text{ and } \bar{X} = 100.0\,. \qquad (A.1)$$

Finally, the initial concentrations of all the chemicals have to be set. Once again, and motivated by a preliminary analysis [2] of the configurations space, we decided to resort to the original values proposed by Gánti and reported in Table 4.

Once all the above parameters and initial concentrations are assigned, and the maximal number of cycles specified, the kinetic equations are solved using the *ode45* code in Matlab [10] V.6.5.1.199709 Rel. 13.1. When the surface size reaches the critical value $S(T) = 2.0$, we "operate" the division and the halving of the inner materials accumulated during the evolution: the surface size is reset to the initial value $S(T^+) = 1.0$, the concentrations are rescaled by a shape factor, in our case ~ 0.9408, which accounts for the fact that we re-initiate the subsequent evolution from a sphere-like state (the discontinuities seen in the top right panel of Fig. 5). The method proceeds with the subsequent iteration step.

A last remark concerns the construction of the State Function. The procedure just outlined allows us to numerically construct an orbit of a single protocell. Hence we are in a position to analyze the sequence of successive division times (see Fig. 1 or top left panel of Fig. 5) and determine whether the associated behavior is regular or not. Fixing all the quantities but the ones of Eq. A.1, one can construct a *State Function* as follows. For several values of $\bar{X}$ and for each couple (N, V^*) defined over a large grid, $[1, 55] \times [0, 100.0]$, we integrate the kinetic equations and record the last

10 division times after a large transient of about 100 division cycles. For instance referring to Fig. 1 we associate to the values $\bar{X} = 100.0$, $N = 25$ and $V^* = 35.0$ the value for the State Function:

$$F_{100.0}(35.0, 25) = (0.1136, 0.1136, 0.1136, 0.1136, 0.1136, 0.1136$$
$$, 0.1136, 0.1136, 0.1136, 0.1136) \,,$$

namely the orbit is regular: all the division time are equal. While in correspondence to values $\bar{X} = 100.0$, $N = 25$ and $V^* = 42.9$ the State Function is now:

$$F_{100.0}(42.9, 25) = (0.1205, 0.1081, 0.1168, 0.1091, 0.1209, 0.1080$$
$$, 0.1163, 0.1093, 0.1213, 0.1079) \,,$$

the last 10 divisions have all a different duration, hence the orbit is irregular.

We can obtain a rough idea of the behavior of the State Function, by plotting for fixed value of $\bar{X}$, N and V^*, the mean value of the last 10 computed division times. This is reported in Fig. 6 where a color code has been associated to the division time, to emphasize the regular and non–regular behavior we draw in blue each non–regular orbit.
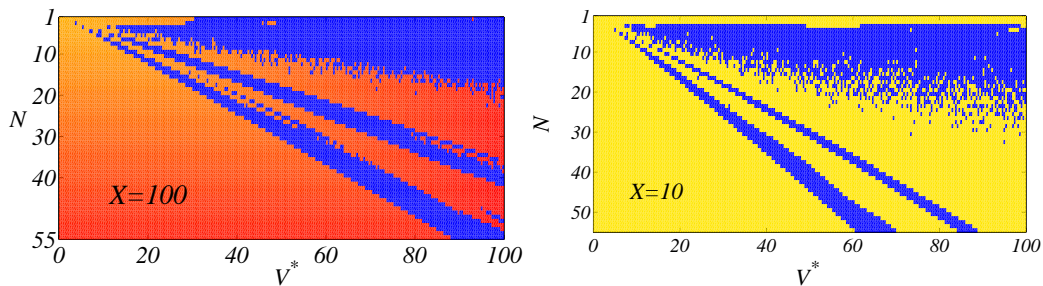

Figure 6.   **State Function**. We report the State Functions obtained for different values of $\bar{X}$ – on the left $\bar{X} = 100.0$ on the right $\bar{X} = 10.0$ – in a typical setting. Colors are assigned according to the classification into regular and irregular behaviors: blue dots refer to irregular ones, whereas regular orbits are represented in red–yellow, the faster the division period the lighter the color-coding. Some key qualitative features are always detected: domains of parameters (N, V^*) corresponding to regular and irregular behaviors coexist, being particularly well mixed in correspondence of specific zones. This peculiar pattern is of paramount importance and intrinsically relates to the *speciation mechanism*.

18

Once several State Functions are stored, we can eventually simulate an open–ended Darwinian evolution, monitoring the dynamical response of a population under the pressure of the surrounding environment.

References

1. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts & P. Walter , *Molecular Biology of the Cell*, New York, Garland, (2002).
2. T. Carletti, *submitted*, (2006).
3. T. Carletti & D. Fanelli, *submitted*, (2006).
4. F.M. Cohan, *Annu. Rev. Microbiol.*, **56**, pp. 457–487, (2002).
5. M. Eigen and P. Schuster, *Naturwiss.*, **64**, pp. 541–565, (1977).
6. T. Gánti, *Biogenesis itself*, J. theor. Biol., **187**, pp. 583–593, (1997).
7. T. Gánti, *I) Theory of Fluid Machineries & II) Theory of Living Systems*, New York, Kluwer Academic/Plenum Publishers, (2003).
8. T.K. Konstantinidis & J.M. Tiedje, *PNAS*, **102**, 7, pp. 2567–2572, (2005).
9. P.L. Luisi, F. Ferri & P. Stano, *Naturwiss.*, **93**, pp. 1–13, (2006).
10. Matlab, *The language of technical computing*, http://www.mathwork.com, (2006).
11. E. Moreno, *Rev. Biol. trop.*, **50**, 2, pp. 803–821, (2002).
12. A. Munteanu & R.V. Solé, *J. Theor. Biol.*, **240**, pp. 434–442, (2006).
13. S. Rasmussen, L. Chen, B. Stadler & P. Stadler, *Origins Life & Evol. Biosph.*, **34**, pp. 171–180, (2004).
14. S. Rasmussen, L. Chen, D. Deamer, D.C. Krakauer, N.H. Packard, P.F. Stadler & M.A. Bedau, *Science*, **303**, pp. 963–965, (2004).
15. E. Szathmáry and J. Maynard Smith, *J. Theor. Biol.*, **187**, pp. 555–571, (1997).
16. E. Szathmary, M. Santos & C. Fernando, *Topics in Current Chemistry*, **259**, pp. 167–211, (2005).
17. J. Singer & G.L. Nicolson, *Science*, **175**, pp. 720–731, (1972).
18. D. Szostak, P.B. Bartel & P.L. Luisi, *Nature*, **409**, 387–390, (2001).

Table 1.   Chemoton Hypothesis. The chemical reactions describing the three coupled chemical subsystems.

Metabolism	Membrane	Template
$A_1 + \bar{X} \underset{k_1'}{\overset{k_1}{\rightleftharpoons}} A_2$	$T' \xrightarrow{k_8} T^*$	$pV_{2N} + V' \underset{k_6'}{\overset{k_6}{\rightleftharpoons}} (pV_{2N}pV_1) + R$
$A_2 \underset{k_2'}{\overset{k_2}{\rightleftharpoons}} A_3 + \bar{Y}$	$T^* + R \underset{k_9'}{\overset{k_9}{\rightleftharpoons}} T$	$(pV_{2N}pV_i) + V' \xrightarrow{k_7} (pV_{2N}pV_{i+1}) + R$
$A_3 \underset{k_3'}{\overset{k_3}{\rightleftharpoons}} A_4 + V'$		$(pV_{2N}pV_{2N}) \longrightarrow pV_{2N} + pV_{2N}$
$A_4 \underset{k_4'}{\overset{k_4}{\rightleftharpoons}} A_5 + T'$		
$A_5 \underset{k_5'}{\overset{k_5}{\rightleftharpoons}} A_1 + A_1$		

Note: In the first column the Autocatalytic Cycle $A_1 \to 2A_1$ is reported, the second one focuses on the the membrane growth and the latter accounts for the template duplication. Here pV_{2N} denotes the double stranded template constituted of $2N$ monomers V', whereas $(pV_{2N}pV_i)$, $i = 1, \ldots, 2N-1$, denotes a partially duplicated polymer where i–monomers have been copied. Direct reaction constants, k_j, are larger than the inverse ones, k_j'. We also assume the so-called strong replication condition [12] to hold true: k_6, k_6' and k_7 are zero if the concentration of V' is below a given threshold V^*. The numerical values for the kinetic constants are set equal to the original values proposed by Gánti [7] and reported in Appendix A.2.

Table 2.   Kinetic differential equations. From the chemical reactions one deduces the differential equations describing the time evolution of the concentrations of the involved chemicals. To lighten notations, we denote the concentrations with capital letters.

Metabolism	Membrane	Template
$\dot{A}_1 = 2\left(k_5 A_5 - k_5' A_1^2\right) - k_1 A_1 \bar{X} + k_1' A_2$	$\dot{T}' = k_4 A_4 - k_4' A_5 T' - k_8 T'$	$p\dot{V}_{2N} = 2k_7(pV_{2N}pV_{2n-1})V'$
		$+ k_6'(pV_{2N}pV_1)R - k_6 pV_{2N}V'$
$\dot{A}_2 = k_1 A_1 \bar{X} - k_1' A_2 - k_2 A_2 + k_2' A_3 \bar{Y}$	$\dot{T}^* = k_8 T' - k_9 T^* R + k_9' T$	$(pV_{2N}\dot{p}V_1) = k_6 pV_{2N}V' - k_6'(pV_{2N}pV_1)R$
		$- k_7(pV_{2N}pV_1)V'$
$\dot{A}_3 = k_2 A_2 - k_2' A_3 \bar{Y} - k_3 A_3 + k_3' A_4 V'$	$\dot{T} = k_9 T^* R - k_9' T - k_{10} T S$	$(pV_{2N}\dot{p}V_{i+1}) = k_7(pV_{2N}pV_i)V'$
		$- k_7(pV_{2N}pV_{i+1})V'$
$\dot{A}_4 = k_3 A_3 - k_3' A_4 V' - k_4 A_4 + k_4' A_5 T'$	$\dot{S} = k_{10} T S$	$\dot{V}' = k_3 A_3 - k_3' A_4 V' + k_6'(pV_{2N}pV_1)R$
		$- k_6 pV_{2N}V' - k_7 \sum_{i=1}^{2N-1}(pV_{2N}pV_i)V'$
$\dot{A}_5 = k_4 A_4 - k_4' A_5 T' - k_5 A_5 + k_5' A_1^2$		$\dot{R} = k_6 pV_{2N}V' - k_6'(pV_{2N}pV_1)R + k_9' T$
		$- k_9 T^* R + k_7 \sum_{i=1}^{2N-1}(pV_{2N}pV_i)V'$

Note: The last equation in the middle column describes the growth of the surface size as a result of the progressive attachment of the membrane molecules [17]. All equations, but the one for $\dot{S}$, have to be modified according to Eq. 1.

Table 3.   Reaction Constants.

$k_1 = 2.0$	$k_2 = 100.0$	$k_3 = 100.0$	$k_4 = 100.0$	$k_5 = 10.0$	$k_6 = 10.0$	$k_7 = 10.0$	$k_9 = 10.0$
$k_1' = 0.1$	$k_2' = 0.1$	$k_3' = 0.1$	$k_4' = 0.1$	$k_5' = 0.1$	$k_6' = 1.0$	$k_8 = 10.0$	$k_9' = 0.1$
$k_{10} = 10.0$							

Note: The reactions constants used in our simulations are reported. They refer to the original choice in [7].

Table 4.   Initial concentrations of involved chemicals.

$A_1(0) = 1.0$	$A_2(0) = 1.8$	$A_3(0) = 1.9$	$A_4(0) = 1.7$	$A_5(0) = 10.0$	$V'(0) = 26.0$
$T^*(0) = 14.0$	$T(0) = 0.0$	$R(0) = 0.0$	$\bar{Y}(0) = 0.1$	$pV_{2N}(0) = 0.01$	$T'(0) = 17.0$
$S(0) = 1.0$					

Note: The table reports the initial concentrations of the chemicals involved in the chemical reactions. $A_i(0)$, with $i = 1, \ldots, 5$ are the chemicals taking part to the metabolic cycle; V' is the concentration of free monomers; T, T' and T^* denote respectively the membrane molecules and two kinds of precursors; R is a byproduct of the polymerization; $\bar{Y}$ is the "waste" produced by the metabolic cycle; pV_{2N} is the concentration of the double–stranded template and $S(0)$ stands for the initial size of the membrane.