Toward a Self-replicating Metabolic P System

Giuditta Franco, Vincenzo Manca

Department of Computer Science,
University of Verona
{giuditta.franco, vincenzo.manca}@univr.it

Summary. This work concerns the synthesis of a 'minimal cell' by means of a P system, which is a distributed rewriting system inspired by the structure and the functioning of the biological cell. Specifically, we aim to define a dynamical system which exhibits a steady metabolic evolution, resulting in self-maintenance and self-reproduction. Metabolic P systems represent a class of P systems particularly promising to model a minimal cell in discrete terms, since they have already successfully modeled several metabolisms. The main further step is thus to find a simple way to obtain Metabolic P system self-replication.

This paper deals with ideas presented at the BWMC11 (held in Seville, Feb 2011) and opens a new trend in membrane computing, based on computational synthetic biology oriented applications of P systems modeling. The framework is here outlined, and some problems to tackle the synthesis of a minimal cell are discussed. Moreover, an overview of literature and a list of appealing research directions is given, along with several references.

1 Introduction

The idea of synthesizing a minimal cell by mathematical and engineered tools is not new in literature, namely there is a recent trend in synthetic biology which is aimed at building a synthetic endomembrane structure, whose compartments (usually formed by liposomes) contain the minimal and sufficient ingredients to perform the basic function of a biological cell (essentially self-maintenance and self-reproduction).

Such an interest originates from the old wondering about "what is life?", and specifically from the question "*how* was possible for a primitive chemical system to evolve through levels of increasing complexity from disordered and unstructured primordial soup to the cellular life as we know it?" [17]. In scientific research, the bottom-up approach, which looks for a plausible process leading from simple molecules to more complex ones, to protocells, and finally to living cells, has still many open questions, although some progress has been done in our knowledge of prebiotic chemistry [14]. An alternative approach, called top-down, focuses on the

synthesis of minimal forms of life starting from our knowledge of modern cells, that is, from general principles of structure and function organization (matter conservation, anabolism and catabolism, species distribution, enzymatic control, autopoiesis). This has an experimental counterpart, and it can be classified as a constructivist approach for scientific knowledge, according to Feynmans famous motto "What I cannot create, I do not understand".

In principle, several implementations of minimal life are possible [19], namely primitive cells, minimal cells, bioreactors, molecular robots (soft-robots) [14]. Current experimental strategies consider synthetic cells as systems having two main components: compartments (i.e., lipid vesicles) and their content (biomacromolecules, such as DNA, RNA, enzymes, ribozymes, PNA, ribosomes, catalytic peptides) [14, 15]. In this context, aggregation phenomena need to be reproduced in laboratory and a preliminary simulation in silico helps to set the quantity range of RNA-polimerasi and RNA-sintasi, in order to get an efficient synchronization of self-mantainance and self-replication [3].

2 Some models

In the cell, molecules react together according to their biochemical reactivity and environmental conditions, giving rise to complex molecules starting from simpler ones. Also, a higher chemical complexity (usually referred to as "supramolecular chemistry") appears as the result of self-organization of molecules into structures (membranes) and oscillating reactions such as auto-catalytic networks within micro-compartments [15].

Among the most active groups working on creating living cells in the laboratory, we recall David Deamer at the University of California, Jack Szostak at Harvard, Tetsuya Yomo at the Osaka University, Steen Rasmussen at the FLinT (Southern Denmark University). Besides we mention the notable research, both on the construction of self-reproducing vesicles and on synthetic minimal cells, started about twenty years ago in the Luisi's group at the ETH (Zurich). It roots in the concept of *autopoiesis*, the theoretical framework that guides the construction of minimal living cells and accounts for the dynamical process at the basis of living entities [15]. The notion of *autopoietic cell* dates back to the work of H. R. Maturana and F. J. Varela in the seventies [11]. It essentially requires a shell/membrane composed by *i)* building blocks L (representing the lipids and the proteins of cell membranes), that eventually decay to a waste product W, and *ii)* an internal metabolism, a black box E (representing the cellular genetic/metabolic network), able to both generate blocks L (from precursors P entering the membrane from the environment) and maintain a transformation of metabolites Q (arriving from outside) that produces and expels waste product Z. According to this scheme, P and Q are the basic nutrients for cell growth, W and Z the waste materials.

In [15] it has been shown that a supramolecular assembly of L molecules can grow and duplicate at the expenses of matter P from outside without any internal

metabolism. However, if we impose to have an autopoietic mechanism, based on a minimal, existent DNA/RNA/enzyme genetic/metabolic network E, then minimal cells exhibiting living properties (self-mantainance, self-reproduction, and possibility to evolve) have a minimal number of genes (a number between 200 and 300, according to results from literature in comparative genomics), enzymes, robosomes, tRNAs and low molecular weight compounds [1, 15]. In this context, protein synthesis is one of the key function for a living cell. The expression of functional proteins inside lipid vesicles by using a minimal set of enzymes, tRNAs and ribosomes, was also investigated in [15] at the aim of constructing continuous models of functional cells, while an efficient protein-synthesizing system was developed in [13].

In [3] a kinetic model of (autopoietic) ribocell was built by means of a differential equation system, where variations of metabolite and lipid concentrations, as well as membrane volume variations, are established by modeling processes such as RNA strands replication, catalyzed by polymerase ribozyme, pairings of RNA-polymerase and RNA-sintase, and conversions of precursors into membrane lipids, catalyzed by ribozymes. The time evolution is deterministic rather than stochastic, by assuming that in average different membranes have the same time behavior. The expansion measure of membrane surfaces is considered, in such a way that self-replication of the whole cell (in two daughters) is assumed as soon as the membrane surface reaches the area sufficient to form two spherical membranes. This model resulted in synchronized genomic duplication and cell replication, with the kinetic values within ranges suggested by the literature. According to the simulations reported in [3], (at room temperature) cell division occurs every 26.6 days, and may be speeded up by increasing the temperature (for example, up to 42°C). The goal of our research is to reproduce a similar autopoietic deterministic system in discrete terms, where biomolecules are represented by multisets of objects, membranes are compartments where rewriting rules are distributed to work in parallel, and the computation is the dynamics observed in a cell at a "suitable" level of abstraction. von Neumann first conceived a self-replicating computational model, by pioneering cellular automata (CA) able to self-replicate [20], but that "mitosis process" was not supposed to be synchronized with any internal metabolism or with other properties typical of the biological cell.

3 Main questions

According to the top-down approach, building a synthetic cell by means of a computational model is in itself a way to understand (or at least to get more information on) the basic concepts of living systems and of their parts.

A main question here is: what are the minimal components, the simplest form of machinery, to get 'biological universality' (behaviors typical of life)? We can say that the minimal number of life criteria is three: self-mantainance, self-reproduction, and evolution capability [7]. Then, a metabolism (internal dynamics) in compart-

ments has to be realized (self-mantainance), together with a simultaneous replication of main internal components and of all membranes (self-reproduction). This is driven by genomic information, by means of gene expression, which gives rise to (structural or enzymatic) proteins, able to perform functions (such as catalyzing biochemical reactions occurring in metabolism).

From a logical viewpoint, Is it necessary to have the genetic/regulative mechanism at the basis of metabolism? Or maybe the presence of ribozomes (naturally self-replicating substances), RNA polymerase and nucleotides, would be enough to have an RNA-based autocatalyst living system? According to the RNA world hypothesis [5], a set of rybozymes is actually sufficient, because RNA can both store information like DNA and act as an enzyme like protein. In this perspective, DNA polymer is just a product of evolution to have redundancy and robustness to errors, including point mutations. In modern cells indeed, DNA, through its greater chemical stability, took over the role of data storage, while proteins, which are more flexible in catalysis through the great variety of amino acids, play only the role of specialized catalytic RNA molecules. The presence of both genomes and ribozymes is therefore redundant to have a simplest self-replicating metabolic system [6], though most likely an independent storage mechanism is required for systems which adapt to the environment. To have a system with the capability to evolve by adaption, sensitivity and adaption to the environment need to be taken into account in the model, by analyzing the exchange of matter with the environment, and the reaction of the system dynamics to environment changes.

Overall, nature exhibits the two levels, informational (genes) and functional (enzymes) - are they necessary to perform an efficient mitosis, or to realize the cell (Darwinian) evolution? A possible answer is that in cells of complex organisms, which need to store more information and for a longer time, the stability of genomic molecules make their existence necessary to have life. In this case, the genomic level would have turned out necessary in the evolution in order to allow a major and more structured complexity of organisms.

4 Our approach

A self-replicating metabolic system requires a synchronization of its internal dynamics in such a way that the metabolic activity is maintained, while the system exchanges matter with the environment, grows, and replicates its own membrane structure together with the contained metabolic processes. We aim at modeling such a dynamical system by a P system [16], that is, by a computational model inspired by the cell. This approach seems the most natural to reproduce in silico what is observed in the cell. Hence, this research may be framed in a context of membrane computing models, and aims at building a self-replicating metabolic membrane system where molecular and cellular peculiarities are represented in symbolic and algorithmic terms.

A similar work has been developed in [18], where a self-replicating membrane system has been exhibited, which initiates with a process of self-inspection, then

copies the membrane contents (objects and rules), incrementally composes the genome and finally create the mother cell outside the current skin membrane. With respect to our goal, in [18] there is not any metabolism in the cell, and the rules are inspired by artificial rather than biological systems.

This subject had already attracted attention in [2], where, however, entire membrane systems were replicated in one macro-step. A closer look to biology may be found in [12], where a "Dogmatic P system" (inspired by the central dogma of molecular biology), which exhibits transduction and transcription processes in the nucleus, is proved to be universal.

A P system is a multi-compartment structure realizing a parallel, distributed, object multiset rewriting system (for more details see `http://ppage.psystems.eu/`). A metabolic P system (shortly, MP system [10]) is essentially a multiset grammar where multiset transformations are regulated by state functions (called regulators), whose values (at each step) represent the fluxes associated to the rewriting rules [8]. Once we know the regulators, a deterministic Markovian dynamics of such systems may be observed as time series of the substances. On the other hand, there are theories and algorithms developed in the framework of Metabolic P systems, which allow to compute regulators starting from observed time series [9].

We would like to keep our self-duplicating metabolic P model as simple as possible - we do not use extra features (such as priorities, polarization, fluxes, probability) if they are not necessary or biologically motivated.

4.1 A research plan

The point is defining a Metabolic P model, in order to understand which are the general rules and regulations on which the synchronization of genomic duplication and membrane reproduction are based.

One cell divides to produce two genetically identical cells. Eukaryotic cells include a variety of membrane-bound structures, collectively referred to as the endomembrane system. Nuclear division is often coordinated with cell division. As a first approximation, we assume to have only two, nested (external and nuclear) membranes $[_0[_1]_1]_0$.

A possible initial conguration (having all genes in the nucleus) includes a multiset of objects g representing genes, r objects representing enzymes, one ribozyme t, and some metabolites (including lipids). Environment is assumed to have "sufficient" resources m.

The goal is to exchange objects with the environment, while metabolites growing, up to reach an "approximatively" double amount they had initially. The concept of matter duplication should be further defined, for example by asking for having an amount which is double than it was initially within a certain range of error, but in this first attempt we leave it undetermined.

Simultaneously, enzymes are produced from genes, for feeding reactions, which guarantee an internal metabolic dynamics of the system, and generate new membranes. In this respect, let us list in the following paragraph some biologically motivated guidelines to set up our model.

Genes replicate (by polymerase) and produce enzymes (by ribozymes). They do not move across the (nuclear) membrane. Ribozymes are self-producing (by means of genes). Enzymes catalyze and allow the application of rewriting rules, in fact metabolites transform by means of rules. Lipids, produced by the metabolism, generate membranes, when present in sufficient amount. In order to account for the size of a membrane, we assume that more external membranes are, and more lipids are necessary to generate a new copy of them.

The rewriting rules will be meant to realize three main processes: enzyme production, metabolism, and liposome production.

Enzyme production. Nuclear transcription of the genes is realized by means of the ribozyme (we can assume to have only one ribozyme, able to perform transcription) and some nucleotides. It produces the enzyme polymerase in the nucleus, and three enzymes r, r', r'' in the region of the external membrane. In the most internal membrane, polymerase is employed to duplicate the genes by using nucleotides - maybe this could be done more realistically with string rewriting rules.

Metabolism. Matter is taken from outside by both membranes. Enzymes produced in the nucleus (by means of the ribozyme) catalyze transformation reactions: specifically, the rule catalyzed by r increases the quantity of the ribozyme itself at the expenses of matter arriving from outside; the rule catalyzed by r' produces lipids able to form the nuclear membrane, and the rule catalyzed by r'' produces lipids able to form the external membrane. Of course we have also matter which does not need to be duplicated, just working as fuel (coming from outside) and as garbage (expelled outside). In a further refinement of the model enzyme degradation should be also considered.

Liposome production. Lipids transform into membranes (liposome organization), representing their organization in structures as vesicles.

Once given the rules, the choice of an application strategy needs some discussion. Even if a non-deterministic or probabilistic evolution of a proto-cellular system could be interesting to study, here we intend to reproduce the deterministic behaviour typically observed in cell replication. Presumably we cannot avoid to use flux functions associated to the rules (as metabolic systems do), for example to regulate the entrance of objects from outside. Maximal parallelism would imply that the cell gets in one first step the infinite resources we have in the environment! Fluxes instead would allow us to modulate any reaction according to the system state.

As an initial set up, a few equations/constraints need to be given in order to impose that there exists a moment k, where the genes and the ribozymes are approximatively double than they were initially, and the lipids are in sufficient amount to form a second copy of the membranes. We assume that reaching such a state will be enough to have two copies of the initial cell.

5 Future Work

Regarding future activities of this research, it is plenty of ideas and dreams. Once we will have an MP system self-replicating, while exchanging matter with the environment to keep its internal metabolic dynamics on, both the role of energy in such an exchange and a form of adaption to the environment should be studied [14], by analyzing the consequent reactions of the system to different (even energetic) stimuli. Receptivity and reactivity should be investigated to better understand the robustness of the single cell and of cell networks. Communication and interaction among (synthetic and/or real) cells is a crucial task [4], for example to model morphogenesis (e.g., embryogenesis) and tissue organization. From the viewpoint of a tissue system, the process of mitosis of each single cell is limited in time, single healthy cells do not live forever but tissue do, while new cells rise and old ones die. Tissues keep alive under certain boundaries (density, dimension) while single cells produce new cells and eventually die: the cellular and tissue systems have a quite different dynamics and functioning, even if tightly interrelated. In [15] simple autopoietic systems are modeled by vesicles populations; it is shown that simple vesicles may grow and divide according to physical laws, also revealing an unexpected pattern consisting in the conservation of the average size in a population of self-reproducing vesicles.

Phenomena such as cell differentiation and speciation are fundamental to understand and better control many processes of biomedical interest. For example, embryonic cells are interesting as they have illimitable replicative power and the ability to generate any type of tissue, a property they have in common with stem cells. On the other hand, non-controlled proliferation and differentiation of stem cells often denote presence of cancer. Cell migration can be also involved in such kinds of processes, and in our research a way to represent both molecular and cellular migration in the context of P systems should be found. In this framework we should start by modeling the concept of biological gradient, maybe by means of nested membrane localization. Finally, cell Darwinian evolution of synthetic cells could give interesting insights on several controversial issues in population genetics evolution theories (such as the importance of the chance in evolutionary transformations, known as genetic drift).

Aknowledgments

We would like to thank Natasha Jonoska for her helpful comments and stimulating suggestions about the presentation of this work at the Brainstorming on membrane computing held in Seville in February 2011. On the occasion of the meeting, the first author gratefully received interesting, specific highlights by many colleagues, with whom there were fruitful discussions. In particular, she wishes to thank very much George Păun, Alfonso Rodrigez Paton, Mario Pérez-Jiménez, Marian Gheorghe, and José Maria Sempere.

References

1. C. Chiarabelli, P. Stano, P. L. Luisi, *Chemical approaches to synthetic biology*, Current Opinion in Biotechnology 20(4): 492-497, 2009.
2. E. Csuhaj-Varjú, A. Di Nola, Gh. Păun, M. J. Pérez-Jiménez, G. Vaszil, *Editing configurations of P systems*, Third Brainstorming Week on Membrane Computing, pp 131-154, 2004.
3. P. Della Gatta, F. Mavelli, *Ribocell modeling*, Wivace 2009, pp 55-64.
4. P. M. Gardner, K. Winzer, B. G. Davis, *Sugar synthesis in a protocellular model leads to a cell signaling response in bacteria*, Natural Chemistry 1, pp 377–383, 2009.
5. W. Gilbert, *Origin of life: The RNA world*, Nature 319:618, 1986.
6. G. F. Joyce, L. E. Orgel, *The RNA world* (Eds R. Gesteland, T. R. Cech, J. F. Atkins], pp 49-77, Cold Spring Harbor Laboratory Press, New York, 1999.
7. P. L. Luisi, *About various definitions of life*, Origins of Life and Evolution of the Biosphere 28, pp 613-622, 1998.
8. V. Manca, *Fundamentals of metabolic P systems*, In: Gh. Paun, G. Rozenberg, A. Salomaa (eds.), Handbook of Membrane Computing, Chapter 19, Oxford University Press, 2009.
9. V. Manca, *Metabolic P dynamics*, In: Gh. Paun, G. Rozenberg, A. Salomaa (eds.), Handbook of Membrane Computing, Chapter 20, Oxford University Press, 2009.
10. V. Manca, *Metabolic P systems*, Scholarpedia, 5(3):9273, 2010.
11. H. R. Maturana, F. J. Varela, *Autopoiesis and cognition: the realization of the living.* Reidel, Dordrecht, 1980.
12. J. M. Sempere, *"Dogmatic" P systems*, Eighth Brainstorming Week on Membrane Computing, pp 291- 300, 2010.
13. Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, T. Ueda, *Cell-free translation reconstituted with purified components*, Nature Biotechnology 19, pp 751- 755, 2001.
14. P. Stano, *Cellule artificiali: dall'attuale quadro teorico-sperimentale al loro uso come robot molecolari*, Wivace 2009, pp193-198.
15. P. Stano, P. L. Luisi, *Chemical approaches to synthetic biology: from vesicles self-reproduction to semi-synthetic minimal cells*, Proc. of the Alife XII Conference, Odense, Denmark, pp 147-153, 2010.
16. G. Păun, *Membrane computing. An introduction.* Springer, 2002.
17. J. W. Szostak, D. P. Bartel, P. L. Luisi, *Synthesizing life*, Nature 409: 387-390, 2001.
18. C. Teuscher, *From membranes to systems: Self-configuration and self-replication in membrane systems*, Biosystems 87, pp 101- 110, 2007.
19. C. Venter, *Creation of a bacterial cell controlled by a chemically synthesized genome*, Science 329 (5987): 52-56, 2010.
20. J. von Neumann, *Theory of self-reproducing automata*, University of Illinois Press, Urbana,, Illinois, 1966.