How artificial life relates to theoretical biology

Hugues Bersini
IRIDIA – ULB
CP 194/6
50, av. Franklin Roosevelt
1050 Bruxelles – Belgium
bersini@ulb.ac.be

Abstract:

There is a long tradition of software simulations in theoretical biology to complement pure analytical mathematics which are often limited to reproduce and understand the self-organisation phenomena resulting from the non-linear and spatially grounded interactions of a huge number of diverse biological objects. Since John Von Neumann and Alan Turing pioneering works on self-replication and morphogenesis, proponents of artificial life have chosen to resolutely neglecting a lot of materialistic and quantitative information deemed not indispensable and have focused on the rule-based mechanisms making life possible, supposedly neutral with respect to their underlying material embodiment. Minimal life begins at the intersection of a series of processes which need to be isolated, differentiated and duplicated as such in computers. Only software developments and running make possible to understand the way these processes are intimately interconnected in order for life to appear at the crossroad. This paper attempts to set out the history of life as the disciples of artificial life understand it, by placing these different lessons on a temporal and causal axis, showing which one is indispensable to the appearance of the next and how does it connect to the next. These successive stages are the: 1) the appearance of chemical reaction cycles and autocatalytic networks 2) the production by this network of a membrane promoting individuation and catalysing constitutive reactions 3) the self-replication of this elementary cell 4) the genetic coding and open-evolution by mutation, recombination and selection. The task of artificial life is to set up experimental software platforms where these different lessons, whether taken in isolation or together, are tested, simulated, and, more systematically, analysed. Some of these existing software platforms will be sketched: chemical reaction networks, Varela's autopoietic cellular automata, Ganti's chemoton mode and genetic algorithms.

Introduction to Artificial Life

Would a theoretical biologist be surprise to be told that the computer use and software developments should help him make substantial progress in his discipline? It is doubtful. There is a long tradition of software simulations in theoretical biology to complement pure analytical mathematics which are often limited to reproduce and understand the self-organisation phenomena resulting from the non-linear and spatially grounded interactions of a huge number of diverse biological objects. Nevertheless, proponents of artificial life would bet that they can help him further, transcending his daily modelling/measuring recurrent practice by using software simulation in the first instance and, to a lesser degree, robotics, in order to abstract and elucidate the fundamental mechanisms common to living organisms. They hope to do so by resolutely neglecting a lot of materialistic and quantitative information deemed not indispensable. They want to focus on the rule-based mechanisms making life possible, supposedly neutral with respect to their underlying material embodiment, and to replicate them in a non-biochemical substrate. In artificial life, the importance of the substrate is purposefully understated for the benefit of the function (software should "supervene" to an infinite variety of possible hardware). Minimal life begins at the intersection of a series of processes which need to be isolated, differentiated and duplicated as such in computers. Only

software development and running make possible to understand the way these processes are intimately interconnected in order for life to appear at the crossroad.

Artificial life obviously relates to exobiology, this other recent scientific discipline equally centred on life and the study of its origins, not only on the obvious environment of earth, but also throughout the universe. Exobiology cannot restrict itself to a mere materialistic view of life, in order to detect it elsewhere, as the material substrate could be something totally different. This substrate could be as much singular on a distant planet as it could be in the RAM memory somewhere in an university computer lab. The presence of life might be suspected through its functions, much before being able to dissect it. Artificial life does not attempt to provide an extra thousandth attempt definition of life, any more than do most biologists. "Defining" is a sociological endeavour which consists in grounding something semantically rather weak on a stronger semantic support. How to assess the solidity of this semantic support is far from easy. It solidifies with the possibility to connect the expression to define with a perceived reality although, for the philosopher Wittgenstein and for most of the common expressions, the usage will finally decide on the final meaning. As a matter of fact, the concept of "life", as opposed to "gravity" or "electromagnetism" or "quantum reduction of a wave packet", has already been in widespread existence prior to any scientific reading or reification. Its semantic limits are now definitely out of control, allowing James Lovelock (Lovelock, 2000) the latitude to assimilate the planet Earth to a living organism and many protagonists of artificial life to see life everywhere as soon as an amusing little animation appears on their screen or a mechanical dog wags its tail. The rejection of an authoritative definition of "life" is often compensated for by a list of functional properties which never finds unanimity amongst its authors. Some demand more properties, others require fewer of those properties that are often expressed in terms of a vague expression such as "self-maintenance", "self-organisation", "metabolism", "autonomy", "self-replication", "open-ended evolution". A first determining role of artificial life consists in the writing and implementing of software versions of these properties and of the way they do connect, so as to disambiguate them, making them algorithmically precise enough that, at the end, the only reason for disagreement on the definition of life would lie in the length or the composition of this list and on none of its items.

The biologist obviously remains the most important partner; but what may he expect from this "artificial life"? What can he expect from these new "Merlin hackers", whose ambitions seem, above all, disproportionally naïve. These computer platforms could be useful in several ways, presented in the following in terms of their increasing importance or by force of impact. First of all, they can open the door to a new style of teaching and advocating of the major biological ideas i.e. computer software as pedagogical helps like, for example, Richard Dawkins (1986), who, bearing the Darwinian good news, did so with the help of a computer simulation where sophisticated creatures known as "biomorphs" evolve on a computer screen by means of a genetic algorithm. These same platforms and simulations can, insofar as they are sufficiently flexible, quantifiable and universal, be used more precisely by the biologist, who will find in them a simplified means of simulating and validating a given biological system under study. Cellular automata, Boolean networks, genetic algorithms and algorithmic chemistry are excellent examples of software to download, parameterise and use to produce the natural phenomenon required. Their predictive power varies from very qualitative (their results apparently reproduce very general trends of the real world) to very quantitative (the numbers produced by the computer may be precise enough to be compared with those measured in the real world). Although being at first very qualitative, a precise and clear coding is already the guarantee of an advanced understanding accepted by all. Algorithmic

writing is an essential stage in formalising the elements of the model and making them objective. The linguistic and qualitative writing of many biological papers could benefit in clarity by just attempting a software instantiation of such writing. The more the model allows to integrate what we know about the reality reproduced, the detailed structures of objects and relationships between them, the more the predictions will move from qualitative to precise and the easier the model will be validated according to the Karl Popper ideal falsifying process.

Finally, through systematic software experiments, these platforms can lead to the discovery of new natural laws, whose impact will be as greater as the simulated abstractions will be present in many biological realms. In the 1950's, when Alan Turing (Turing, 1952) discovered that a simple diffusion phenomenon, propagating itself at different speed, depending on whether it is subject to a negative or positive influence, produces zebra or alternating motifs, it had a considerable effect on a whole section of biology studying the genesis of forms (animal skins, shells of sea creatures, (Meinhardt, 1998)). When some scientists discover that the number of attractors in a Boolean network or a neural network exhibits a linear dependency on the number of units in these networks (Kauffman, 1993, 1995), these results can equally well apply to the number of cells expressed as dynamic attractors in a genetic network or the quantity of information capable of being memorised in a neural network. Entire chapters of biology dedicated to networks (neural, genetic, protein, immune, hormonal) had to re-written in the light of these discoveries. When some scientists have recently observed a non-uniform connectivity in many networks, whether social, technological or biological, showing a small number of key nodes with a large number of connections and a greater number of nodes with far fewer, and when, in addition, they explain the way in which these networks are built in time (Barabasi 2002) by preferential attachment, again biology is clearly affected. Artificial life is of course at its apogee when it reveals new biological facts, destabilizing biologists' presuppositions or generating new knowledge, rather than simply illustrating or refining the old. At the moment, the fact that this discipline is rather young and has a certain immaturity in comparison with mainstream biology has lead several observers to remain in want of information in the face of the current discrepancy between promises and reality. In my opinion, they have tended to underestimate the importance of the results already obtained, as they are too riveted to their microscope. They should show less reticence, less coldness and arrogance towards these new computer explorers but more curiosity and conviviality for these people who, like them, have set out on the conquest of life while remaining in front of their computer screens.

Beginning at the next chapter, I shall attempt to set out the history of life as the disciples of artificial life understand it, by placing the different landmark steps on a temporal and causal axis, showing which one is indispensable to the appearance of the next and how does it connect to the next. This history will certainly be very incomplete and full of numerous unknowns, but most people involved in artificial life will be in agreement. They will mainly disagree on the number of these functions and on the causal sequence of their appearance, acknowledging however that the appearance of any will have been conditioned by the presence and the functioning of the previous ones. The task of artificial life is to set up experimental software platforms where these different lessons, whether taken in isolation or together, are tested, simulated, and, more systematically, analysed. I shall sketch some of these existing software platforms whose running delivers interesting take home messages to open-minded biologists.

The history of life seen by artificial life

Appearance of chemical reaction cycles and autocatalytic networks

In order for a system to emerge and maintain itself inside a soup of molecules which are potentially reactive and contain very varied constituents (which could correspond to the initial conditions required for life to appear i.e. the primordial soup), this reactive system must form an internally cycled network or a closed organisation, in which every molecule is consumed and produced back by the network. Above all, in order for life to begin, all of the constituent components must have been able to stabilize themselves in time. These closed networks of chemical reactions are thus perfect examples of systems which, although heterogeneous, are capable of maintaining themselves indefinitely, despite the shocks and impacts which attempt to destabilize them. This comes about through a subtle self-regeneration mechanism, where the molecules end up producing those molecules which have produced them. It may be obtained on a basic level in a perfectly reversible chemical reaction but can be obtained more subtly in the presence of a lot of intermediary molecules and catalysts. By this reactions-based roundabout in which they all participate, all molecules contribute to maintaining themselves at a constant concentration, compensating and re-establishing any disruption in concentration undergone by any one of them. The bigger the network, the more stable it should be and the more molecules it will contribute to maintaining in a concentration zone which will vary very little, despite external disruptions.

A network of this kind will be materially closed but energetically open if none of the molecules appears in or disappears from the network as a result of material fluxes, whereas energy, originating in external sources, is necessary for the reactions to start and take place. The presence of such a energy flux, maintaining the network far from the thermodynamic equilibrium, is needed since, without it, no reactive flow would be possible circulating through the entire network. A molecular end of the cycle must be re-energized back in order to start again the whole circular reaction process. This cycle thus acts as a chemical machine, energetically driven from the outside. As soon as one of the molecules is being produced in the network without, in its turn, producing one of the molecules making up the network, it absorbs and thus destroys the network. In the presence of molecules of this kind, produced but non-productive, some kind of waste, the only way of maintaining the network becomes to feed it materially and to make it open to material influx. The network acts on the flow of material and energy as an intermediate ongoing stabilization zone, made up of molecules which may be useful to other vital functions (such as the composition of enclosing membranes or catalyzing self-replication), to be described in the following chapters. It transforms, as much as it "keeps on" all the chemical agents which it recruits. Biologists generally agree that a reactive network must exist prior to the appearance of life, at least to catalyze and make possible the other life processes such as the genetic reading and coding; it is open to external influences in terms of matter and energy, but necessarily contains a series of active cycles. They are most often designated as "metabolism" or "proto-metabolism", the most popular and active advocates of this "metabolism-first" hypothetical scenario of the origin of life being: (De Duve, 2002, Ganti, 2003, Maynard-Smith and Szathmary, 1999, Stuart Kaufmann 1993, Robert Shapiro, 2007 and Freeman Dyson, 2000).

In my lab (Lenaerts and Bersini, 2009), we give priority to the study of chemical reaction networks viewing them as key protagonists in the appearance of life. These chemical reaction networks where the nodes are the molecules participating in the reactions and the connections the reactions linking the reacting molecules to the molecules produced are generally

characterised by fixed point dynamics, the chemical balances during which the producers and the products mutually support each other. The attractors in which these networks fix themselves are as dynamic - the concentrations slowly stabilize – as they are structural – the molecules participating in the network are chosen and "trapped" by the network as a whole. These networks are perfect examples of systems which combine dynamic (the chemical kinetics in this case) and metadynamics (the network topological change), as new molecules may appear as the results of reactions while some of the molecules in the network may disappear if their concentration vanishes in time. Both the structure of the network and the concentration of its constituents tend to stabilize over time. Kauffman (1993, 1995) and Fontana (1992) were the forerunners in the study of the genesis and properties of these networks. Figure 1 illustrates the work of these two artificial life pioneers, dedicated to the study of prebiotic chemistry, limiting the reactions studied to polymerization, such as: aa + bb → aabb or inversely, depolymerisation or hydrolysis : abaa → ab + aa.

Figure 1: Representation of a network of chemical reactions of polymerisation (a+b→ ab) and depolymerisation (ab→ a+b) taking place in a simulated chemical reactor. Molecules are represented by circles and reactions by square. Each reaction can be catalyzed, like the arrows pointing to the squares show, by a molecule of the network (giving rise to an autocalytic network). Some molecules can appear (like the molecule "aab") or simply disappear from the network. Reaction cycles can appear, like the one surrounded in the figure (aa→baaaa→baaaaaab→baaa→aa).

Kauffman showed that provided the probability that a reaction takes place is affected by the presence of a catalyst, which is itself produced by the network (in such case the whole network is said to be autocatalytic), a phenomenon of percolation or phase transition, characteristic of this type of simulation, is produced. For probabilities which are too low, the network does not pop up because the reactions are too improbable, but as soon as a threshold value is reached for this same probability, the network "percolates", giving rise to multiple molecules produced by multiple reactions. Kauffman grants a privileged status to this threshold value and to the giant "explosive" network resulting from it in his scenario of the origin of life, without really arguing the reason why such a status should exist, but passing the immense interest and enthusiasm which phenomena of phase transition arouse amongst physicists on to the world of biology. Fontana for his part is concerned with the inevitable appearance of reaction cycles (such as that illustrated in figure 1). All the molecules produced by these cycles in the network in turn produce molecules of the network. He is amongst those many biologists who see these closed networks or organisations as forming a key stage in the appearance of life, due both to their stability and to the fact that they form structural and dynamic attractors for the system. They cause a stabilization and internal regulation zone together with a energetic motor in a chemical soup which is continually being crossed by a flow of matter and energy. Fontana goes on to show how these networks are also capable of self-regeneration and self-replication.

With Tom Lenaerts (Lenaerts and Bersini, 2009), we have programmed the genesis of these chemical reaction networks by adopting the Object-Oriented (OO) programming paradigm. The OO simulator aims to reproduce a chemical reactor and the reaction network which emerges from it (like shown in figure 2). This coevolutionary (dynamics + metadynamics) model incorporates the logical structure of constitutional chemistry and its kinetics on the one hand and the topological evolution of the chemical reaction network on the other hand. The network topology influences the kinetics and the other way round since only molecules with a sufficient concentration are allowed to participate into new reactions (to avoid a combinatorial explosion of molecules and reactions). Our model is expressed in a syntax that remains as close as possible to real chemistry. Starting with some initial molecular objects and some

initial reaction objects, the simulator allows us to follow the appearance of new molecules, the reactions in which they participate as well as the development of their concentration over a period of time. The molecules are coded as canonical graphs. They are made up of atoms and bonds which open, close or break during the reactions. The result of the simulation consists in various reaction networks, unfolding in time, and whose properties can be further studied (for instance the presence and the properties of reaction cycles or the nature of the network particular topology such as scale-free or random).

Figure 2: The OO chemical simulator developed by Lenaerts and Bersini (2009). On the left, the molecules are represented as canonical graphs. On the right, the outcome of the simulator is an evolving reaction network which can be studied in its own right (the presence of cycles, the type of topology...).

One of these reaction schemes, more than just cycling, can further be autocatalytic, when a product of the reaction cycle is the double in concentration of one of the reactant: a + b -> a + a. This is for instance the case of the so-called formose reaction (that Ganti and Szatmary have discussed at large in (Ganti, 2003)), during which a two-carbon molecule, reacting twice with a monomer composed of one carbon, leads to a four-carbon molecule, which then splits in order to duplicate the original molecule. This is the chemical variant of the genetic self-replication since in both cases an original molecule is duplicated. As will be discussed later, Ganti has been the first to connect and synchronize these two replication processes: chemical and genetic, in order for the cell to simultaneously duplicate its boundary, its metabolism and its informational support. In the presence of autocatalysis, the reaction kinetics amounts to an exponential increase and, more interestingly, when various autocatalytic cycles enter in antagonistic interaction, turns out to be responsible for symmetry breaking (one of the cycle, initially favoured, wins and takes it all). The early origin of life should not be studied without taking account the self-organization of chemical networks, the emergence and antagonism of autocatalytic cycles and how energy flows drive the whole process. Such chemical networks are for instance appealing to understand the onset of biological homochirality as the destabilization of the racemic state resulting from the competition between enantiomers and from amplification processes concerning both autocatalytic competitors (one left-oriented and the other right, see (Plasson et al., 2007)). The chemical reaction network under study (shown in figure 3) is made up of the same type of polymerization and depolymerization reactions as the one studied by Fontana. In the additional presence of epimerization reactions allowing the transformation of a right-hand monomer into a left-hand one and vice-versa, the concentration of one family of monomers (for instance the left one) vanishes in favour of the other. The flux of energy is transferred and efficiently distributed through the system, leading to cycle competitions and to the stabilization of asymmetric states.

Figure 3: The prebiotic chemical reactor system responsible for a homochiral steady state studied by (Plasson et al. 2007). The complete set of reactions is indicated containing activation(the necessary energy source), polymerisation and hydrolysis (which together shapes the cycles) and epimerization (which induces the competition between the enantiometers).

Production by this network of a membrane promoting individuation and catalysing constitutive reactions

The appearance of a reaction network of this kind undeniably creates the stability necessary for exploiting its constituents in many reactive systems such as the ones dedicated to the construction of membranes or the replication of molecules carrying the genetic code. This network also acts as a primary filter as it can accept new molecules within it, but can equally well reject other molecules seeking to be incorporated within it. They will be rejected, as they

do not participate in any of the reactions making up the network. Can we see a primary form of individuation in this network? No, because by definition it can only be unique as no spatial frontier allows it to be distinguished from another network. Although it is roughly possible to conceive of an interpenetration of several chemical networks, establishing a clear separation between these networks would remain a problem.

It would seem fundamental that a living organism of any kind can be differentiated from another. We know that the reproduction of a second organism from a first is a central mechanism of life and can only operate if the "clone" elaborates something to spatially distinguish itself from its "original". The best way of successfully completing this individuation and to be able to distinguish between these networks is to revert to a spatial divide, which can only be produced by some form of container capable of circumscribing these networks in a given space. Biochemists are well acquainted with an ideal type of molecule, a raw material for these membranes in the form of lipidic/amphiphilic molecules or fatty acids, the two extremities of which behave in an antagonistic fashion – the first hydrophilic, attracted to water, and the second hydrophobic, repulsed by it. Quite naturally, these molecules tend to assemble in a double layer (placing the two opposing extremities opposite to each other), formed by the molecules lining up and finally adopting the form of a sphere to protect the hydrophobic extremities from water. Like soap bubbles, these lipid spheres are semi-permeable and imprison the many chemical components trapped during its formation. They do however actively channel in and out the most appropriate chemicals for maintaining themselves.

In assimilating living organisms to autopoietic systems, Varela et al. (1974) were the first to insist that this membrane should be endogenously produced by the elements and the reactions making up the network (for example lipids would come from the reactions of the network themselves) and would in return promotes the emergence and self-maintenance of the network. The membrane can help with the appearance of the reactive and growing network by the frontiers that it sets up, the concentration of certain molecules trapped in it or by acting as a catalyst to some of the reactions due to its geometry or its make-up. Basically, autopoiesis requires a cogeneration of the membrane and of the reactive network which it "walls up". The network presents a double closure – one chemical, linked to the cycling chain of its reactions and another physical, due to the frontiers produced by the membrane. In the cellular automata model of Varela (Varela et al., 1974; Mc Mullin and Varela, 1997) illustrated in figure 4, there are three types of particles capable of moving around a two-dimensional surface; "substrates", "catalysts" and "links". The working and updating rules of this cellular automata go as follows:

- If two substrates are near a catalyst, they disappear to create one single link where one of the two was located.
- If two links are near each other they link up and attach themselves to each other. Once attached these links become immobile.
- Each link is only allowed to attach itself to two other links at the most. This allows the links to form chains and to be able to make up a closed membrane.
- The substrates can diffuse through the links and their attachments, while the catalysts and the other links cannot. We can therefore understand how the process of the cogeneration comes about. The membranes shut in the catalysts and the links, which in turn support the membrane by being essential to its formation and regeneration once locally destroyed.

- The reactions creating the links are reversible, as the links can recreate the two original substrates (and thus cause the membrane to deteriorate), but at a lower speed. When this happens, the attachment between the links also disappears.

Continuous updating and execution of these rules produces minimal versions of reactive systems, physically closed and confined by means of a membrane, which is itself produced by the reactive system. For Varela and the others following him, this turns to be an essential stage in the road to life. Running the software, many difficulties are encountered such as the simple attainment of a closed cell on account of the many more possibilities for the membranes to unfold in a straight way. As a matter of fact, the original simulations departed from an already in place minimal cell, the rules then consisting essentially in stabilizing the cell despite the intermittent disappearance of the membrane parts. And even this turns out to be far from obvious as reported in (Mc Mullin and Varela, 1997), exploiting this difficulty for urging the different researchers interested in programming minimal life to respect good programming practices such as OO patterns, to render their code available, understandable and as well commented as possible. Only software simulations can interconnect the physical compartment played by the membrane with the generating metabolism, and further show how far from obvious it is for these two systems to mutually sustain each other.

Figure 4: Simulation by means of a cellular automata of the autopoietic model originally proposed by Varela. The minimal cell can easily be seen together with the catalysts and the substrates that it encapsulates.

The whole, interactive "metabolism and membrane" prefigures a minimal elementary cell, which already seems capable both of maintaining itself and detaching itself from its environment and from cells similar to it. It is at this stage of our successive conceptual additions on the way to establishing a better and more exact characterisation of life, that the definition given by Luigi Luisi (Luisi, 2002) takes on its full meaning (restating the idea of autopoeisis in more biological terms). "*Life is a system which can be self-maintaining by using external energy and nutritional sources to the production of its internal constituents. This system is spatially circumscribed by a semi-permeable membrane of its composition.*" In the food steps of Varela, considering life impossible without a way for individuation and compartmentalisation, the constitution of the membrane by simple self-organisation or self-assembly processes of bipolar molecules (hydrophilic and hydrophobic) has become a very popular field of artificial life in its own right. It is indeed rather simple to reproduce this phenomenon in software (like illustrated in figure 5). You need water molecules that just randomly move, in blue in the figure. You need two kind of sub-molecules (call then A and B) which when meet form, though the only authorized additive chemical reaction, an A-B molecule (A is hydrophobic and B hydrophilic) whose two poles are connected by a small string. You need also to adjust the degree of repulsion between A and water, between B and water, the strength of the string of the A-B molecule and the random component (akin to the thermal noise) to add on each of the intermolecular forces. Nevertheless, the final outcome turns out to be rather robust. The bi-layer of B-A/A-B molecules will very naturally and spontaneously form such as for real cells.

Figure 5: Simulation of a minimal cell based on A-B (A is hydrophobic and B hydrophilic) and water molecules. All molecules move in reaction of repulsive forces of different intensity and thermal agitation. The pink dots are the A, the grey dots are the B that do connect to give A-B (represented in red and black) by a simple chemical reaction. The blue dots are the water molecules.

Again as for the Varela's minimal cell, the closure turns out to be quite delicate to obtain. One very simple way to obtain it is to locate the source of A sub-molecules (the pink dots in the figure) in a singular point, so that the closed membrane will simply surround that source, the circular shape being the local minimal of the mechanical energy connecting all A-B together. Like in Varela's model, and somewhat paradoxically, the source needs to be circumscribed by the membrane for that same membrane to close on itself. However, in contrast with this autopoietic model, once in place the membrane cannot deteriorate and thus no further internal chemistry is required to endogenously produced what would be needed to fix it. Ultimately, this membrane should exhibit some selective channelling in and channelling out (akin, for some authors (Luisi, 2002), to a very primitive form of cognition) providing its internal metabolism with the right nutrients and the right evacuating way out so as to facilitate the cell self-maintaining. These two software models raise interesting questions for the biologists like: how are the molecular parts of the membrane generated (endogenously or exogenously) and is this cogeneration of the membrane and the internal metabolism the signature of minimal life ?

Self-replication of this elementary cell

Self-replication, or the ability of a system to produce a copy of itself on its own, is one of the essential characteristics which has most intrigued and impassioned disciples of artificial life, beginning with the second father of computing: John Von Neumann, after Turing. Biology, and in particular this faculty of self-replication, fascinated Von Neumann to the extent that he devoted the essential part of his ending scientific life to it. For if we want to compare a cell to a computer and a genome to a code, we need to explain how the computer itself was able to be created out of this code. Let us follow the reasoning of this genius step by step, as it is the perfect illustration of an "artificial life" type of approach: no material instantiation but just pure functions or rules. To a sequence of purely functional questions and showing an almost complete ignorance of biological materiality, his reasoning led to a logical solution, the content of which retraces astonishingly closely those lessons we have since learned about the way biology functions. Von Neumann departs from the principle that a universal constructor C must exist, which, based on the plan of some kind of machine P_M (P the plan, M the machine), must be capable of constructing the machine M^P. This idea may be simply translated by $C(P_M) = M^P$. The question of self-replication with is then raised is "Is this universal constructor capable of constructing itself?" In order to do so, it must, following the example of other construction products, have a plan of what it wants to construct; in this specific case, it is the constructor's plan P_C. The problem is then expressed as follows; can $C(P_C)$ give $C(P_C)$ in order for there to be a perfect replication of the original? Von Neumann therefore realised that the question at issue is that of the fate of the construction plan, because if the constructor constructs itself, it has to add the plan itself to the product of the construction. Von Neumann proposed then allotting two tasks to the universal constructor; first, that of constructing the machine according to the given plan and thus adding the original plan to this construction. The constructor's new formula then becomes: $C(P_M) = M^P(P_M)$. If the constructor applies itself to its own plan, this time the replication will be perfect: $C^P(P_C) = C^P(P_C)$.

The fascinating aspect of Von Neumann's solution is that it anticipated the two essential functions that, as we have since discovered, are the main attributions of the protein tools constituting the cell; constructing and maintaining this cell, but also duplicating the code in order for this construction to be able to prolong itself for further generations. Starting with

DNA, the whole protein machinery first of all builds the cell then, by an additional procedure, duplicates this same DNA. Von Neumann did not stop at duplication, because, at the same time, he imagined how this same machinery could evolve and become gradually more complex as a result of random mutations taking place while the plan recopies itself. Von Neumann gave also a cellular automata solution of the problem in which each cell of the automata possessed 5 neighbours and 29 states, and around 200 000 cells were necessary for the phenomenon of self-replication to take place. Many years later, Chris Langton, (Langton 1984, 1989) the organizer of the first conference on artificial life in 1989 proposed an extremely simplified version of this (8 states, but 219 rules remain necessary), although it still follows the pattern mapped out by Von Neumann. This automaton, shown in figure 6 incessantly reproduces a little motif shaped as a loop.

Figure 6: The Langton's self-replicating cellular automata

For many biologists as opposed to Varela, Luisi, Ganti, Maynard-Smith, life is not simply indissociable from but also essentially reducible to this capacity for self-replication. Nevertheless, they still need to explain how life can actually reproduce without an entire pre-existent metabolic chemical machinery. Departing from the elementary cell introduced in the preceding chapter, and in the interest of an unbroken narrative, let us imagine a simpler scenario leading to self-replication. The closed circuit of chemical reactions could be destabilized by some kind of disturbance, causing a growth in concentration of some of its constituents, including those involved in the formation of membranes. This would also be the case provided all the reactions of the metabolism turn out to be autocatalytic entailing the concentration exponential growth of all its molecular elements (including again the membrane constituents). The membrane and the elements which it captures begin to grow (as illustrated in figure 7) until they reach the fatal point where the balance is upset. This is followed by the production of a new cell produced by and from the old one. When the new one comes, it quickly grows fast enough to catch up with the "generator" and "nursing" cell, as a chemical network is capable of some degree of self-regeneration due to its intrinsic stability; each molecule looks around for another that it can couple up to. This reconstitutes the natural chain reaction of the whole. The new membrane and the new chemical network reconstitute on their own by helping each other. Again, obtaining such duplication is far from obvious since, any cell being intrinsically stable, only a thermal but quite unnatural agitation would do the job. Alife hackers ought to find some more convincing mechanisms including the genetic template in the years to come.

Figure 7: The elementary minimal cell of figure 5 in a process of self-replication induced by the growing and the division of the chemical metabolic network together with the membrane individuating it. A lot of random thermal noise is here indispensable to destabilize the initial cell.

Rather than this elementary form of chemical self-replication coupled to the physical self-replication induced by the growth and division of the membrane, life has opted for a more sophisticated physico-chemical version of it, more promising for the evolution to come: self-replication by the interposing of an "information template". Each element of the template can only couple itself with one complementary element. The new coming elements will as a whole naturally reconstitute the template they were attracted to, causing then the replication of the entire template. In biology, it is the extraordinarily emblematic double helix of DNA that acts as a template, shouldering the major role in the history of life – that of the first known replicator. Our elementary cell must now be internally equipped with this information template. Since the 1950s, Timor Ganti, (Ganti 2003), a largely misunderstood researcher, but nevertheless a key precursor of artificial life, proposed a first minimum mathematical system,

named "chemoton", represented in the figure 8. This is the first abstract computational proto-cell that we know, construed by Ganti as the original ancestor of living organisms. It possesses three autocatalytic chemically linked sub-systems; a metabolic network, a membrane and an information template responsible for scheduling and regulating self-replication. They all three grow exponentially until to reproduce and essentially depend on each other for their existence and their stability. The metabolism feeds the membrane and the template, the membrane concentrates the metabolites, the template mechanism dictates the reproduction of the whole. The triadic ensemble is indeed capable of a whole synchronous self-replication and tries to computationally answer questions about the three sub-systems and their interdependency, such as "how does the self-replication of the template automatically accompanies the self-replication of the whole". This complex software object, the "chemoton", has also become these recent years the topic of many software developments and experimentations and is emblematic of artificial life at its best.

Figure 8: The schematic representation of Ganti's chemoton. One can easily see the three autocatalytic subsystems: the metabolism, the membrane and the information template, chemically coupled.

Genetic coding and evolution by mutation, recombination and selection

The information template introduced in the last chapter can in no way be described as "information" by chance, given that, apart from its self-replication, each letter constituting it contributes to the code of a functional component essential to the cell and designed on the basis of that code – a protein. As soon as he hears anyone talking about code, the software specialist has, quite legitimately, to put his head in thorough the window, because it is to him and him alone that we in fact owe the metaphor of the genetic code. Since Darwin and thereafter throughout all evolutionary science, we have a good idea of what the last chapter of the history of life is. Doubtless what has stimulated the most developments in "artificial life" (primarily from the point of view of engineering), is the fact that the genetic code can evolve through mutation and sexual crossing between the old machines, evolving so as to produce new machines that are more and more efficient. Over the past twenty years, many of those developments into artificial life have been eager to show how beneficial this idea is for the research and the automated discovery of sophisticated solutions to complex problems. As illustrated in figure 9, this research can take place though a succession of mutations and recombinations operating at the level of the code, with the best solutions proposed being preserved in the next generation in order to be used for a new cycle of these same operations. The brute force of the computer is used to its full effect.

Figure 9: Illustration of the genetic algorithms: the recurrent iterated sequence of: selection, mutation and recombination to easily lead to an interesting solution of a complex optimisation problem.

These are the same genetic algorithms which Dawkins used in his Darwinian crusade, when he developed his biomorphs. It should also be stressed that another element in Dawkins' program is that when it is finally evaluated, the phenotype is not directly obtained from the genotype, as would be the case for classical optimization in a real or combinatory space. In his work, the biomorphs are the product of a recursive sophisticated program which is carried out starting from a given genotype to give a phenotype. A great "semantic distance" is maintained between these genotypes and phenotypes, which reflect the long process of cell construction from the genetic code and the need for a sophisticated metabolism building the machine out of the code. In brief, this constant program, able to interpret the evolving genotype, is much more important for the complexity of the final outcome than the genotype

itself. Similarly, a very prized derivative of these algorithms is genetic programming, (Koza, 1992) where the individuals now to be optimized are software codes.

Conclusions

Parallelism, functional emergence and adaptability are the conditions necessary to allow these new biologically inspired artefacts to emerge, to "face the world". We are jumping straight into the robotics branch of artificial life (Brooks, 1990). The interfacing with the real world required by these robots needs a parallel information reception mechanism, because the environment subjects it to a constant bombardment of stimuli. They have to learn to organize and master this avalanche falling on their perceptions. They have to learn to build their own concepts, fed and stimulated by this environment, and which, in return allows them to master it. The conceptual high level cognitive processes are born out of motor-sensory interactions and serve to support them. Cognitive systems extend at new levels what the minimal cell in the primitive soup does, with a flow of matter and energy crossing straight through, maintaining itself by selectively integrating this influx to form a closed reactor network and the membrane enclosing it. Like the simplest membrane, they do not passively receive a predetermined world, but integrate it in a way which is adapted to their structure and their maintenance in the world. Biological systems initially function in a closed way before opening themselves to the outside, and what they do with this "outside" is first and foremost a result of the concern of stabilizing themselves.

My conclusions are addressed to the three partners: the biologist, the engineer and the philosopher. To the first, the outcomes of artificial life consist in bringing out what the computer and the biology share intimately: an elementary way of working at the ultimate lowest level, but which by the brute force of parallelism and incessantly repeated iterations, can make unknown and sophisticated phenomena to emerge at higher levels. These processes give rise to many different "ways of being", which will be sorted by a further phase of selection aimed at a higher "well-being" or viability of organism. I have elsewhere (Bersini, 2004) defended the view that only biology, since it includes in its interpretation a mechanical observer, not human, and the selection as much natural to which these collective self-organized phenomena are submitted, can attribute to them the ontological status of "emergence". The qualitative aspect of these simulations can give them new roles in the vast scientific register : use it for education, illustrate biological principles which are already understood, open up possible experiences of thought, play and replay multiple biological scenarios very quickly, titillate the imagination by on-screen representations, call into question some of the ambiguously-interpreted but commonly-accepted facts and, when detailed at most, be able to predict experimental measurements.

The second partner, the engineer, is vigorously encouraged to use the computer for what it is best at doing – this infinite possibility of trial and error. He must exclude from the loop what he cannot take in, the time for the computer to propose this immense range of solutions to the problem facing him. Then, like the child playing at "hot and cold", he will be able to guide the computer as he wishes, because he remains the only one who will know the nature of the problem and be capable of appreciating the quality of the solutions. It is a perfect synergy, where both participants complement each other ideally. While the engineer must bow to the computer in terms of calculating power, this is compensated for by his judgement. Genetic algorithms, ant colonies, neural networks, reinforced learning have enriched the engineer's toolbox. It is rare for new problems to be completely created piece by piece and we are still

left with the usual data processing, with situations requiring either optimization or regulation – but these new algorithms coming from artificial life have the singular advantage of adding simplicity to performance and leaving the computer by its calculating power softens the torment undergone by the engineer in his inferential progression to better solutions.

Finally, for the philosopher, at each attempt at a definition of life, artificial life makes a real attempt to achieve a computerized version in conformity with this definition. For the sceptic, unhappy with this computerized "lining", the question now becomes how to refine his definition, to complete it or to renounce the possibility that there is no definition which cannot be computerized. The other possibility, doubtless more logical but more difficult for many philosophers to accept would be that life poses no problem for a computer snapshot since it is computational at its roots. The beneficial effect of such an attitude is to help de-sanctify the idea of life in its most primitive form, when it is the privilege of the most elementary organisms, and not, as in more evolved organisms, when it takes root and becomes indistinguishable from manifestations of consciousness; in substance this means to keep life and consciousness of life clearly separate. By referring a philosophical article which has become famous in the artificial intelligence community (Nagel, 1974), if a computer cannot know what it is like to be a bat, living like one could be much more within its reach.

Bibliography

Barabasi, L-A. (2002). *Linked. The New Science of Networks*. Perseus Books Group.

Bersini, H. (2004). Whatever emerges should be intrinsically useful. In *Proceedings of Artificial Life 9. pp. 226-231. MIT Press.*

Brooks, R. (1990). Elephants don't play chess. In *Designing Autonomous Agents*, MIT Press. P. Maes (ed.)

Dawkins, R. (1986) *The Blind Watchmaker*. New York: W. W. Norton & Company, Inc. ISBN 0-393-31570-3

De Duve, C. (2002) *Life Evolving: Molecules, Mind, and Meaning*. Oxford University Press, USA

Dyson, F. (1999) *Origins of life*. Cambridge University Press; 2 edition (September 28, 1999)

Fontana, W. (1992) Algorithmic Chemistry. In *Artificial Life II*: A Proceedings Volume in the SFI Studies in the Sciences of Complexity (C.G. Langton, J.D. Farmer, S. Rasmussen, C. Taylor, eds.), vol. 10. Addison-Wesley, Reading, Mass

Ganti, T. (2003) *The Principles of Life* – Oxford University Press.

Goldberg, D.E. (1989). *Genetic Algorithms in Search, Optimization, and Machine Learning* - Addison-Wesley Professional; 1 edition (January 11, 1989)

Kauffman, S. (1993): *The Origins of Order: Self-Organization and Selection in Evolution* – Oxford University Press

Kauffman, S. (1995). *At home in the universe. The search for the laws of self-organisation and complexity*. New-York. Oxford University Press.

Koza, J. (1992). *Genetic Programming*. Cambridge. MIT Press.

Langton, C.G. (1984). Self-Reproduction in Cellular Automata. *Physica D*, 10, pp. 135-144.

Langton, C.G. (ed.) (1989). *Artificial Life I*, Addison Wesley.

T.Lenaerts and H. Bersini. (2009) A synthon approach to artificial chemistry. *Artificial Life journal.* Winter 2009, Vol. 15, No. 1: 89–103.

Lovelock, J. (2000). *Gaia: A new look at life on earth*. Oxford University Press.

Luisi, P.L. (2002). Some open questions about the origin of life – In : *Fundamentals of Life,* Elsevier Paris, ISBN 2-84299-303-9, 287-301.

Maynard Smith, J. and E. Szathmary. (1999). *The origins of life: from the birth of life to the origin of language*. Oxford University Press.

McMullin, B. and F.R. Varela. (1994). Rediscovering computational autopoiesis. In Husband, P. and Harvey, I. (eds.) . *Proceedings of the 4th European Conference on Artificial Life*, pp.38. Cambridge, MA: MIT Press.

Meinhardt, H. (1998). *The Algorithmic Beauty of Sea Shells*. 2nd enlarged edition Springer, Heidelberg, New York.

Nagel, T. (1974). "What is it Like to be a Bat?" *Philosophical Review* 83. pp. 435–50. Reprinted in his *Mortal Questions*. New York: Cambridge University Press pp. 165–180.

Plasson, R., Kondepudi, D. K., Bersini, H., Commeyras, A. and Asakura, K. (2007). Emergence of homochirality in far-from-equilibrium systems: mechanisms and role in prebiotic chemistry. *Chirality*, 19:589–600.

Shapiro, R. (2007) "A Simpler Origin for Life", *Scientific American*, 296, June, 46-53

Turing, A. M. (1952), 'The chemical basis of morphogenesis', *Phil. Trans. R. Soc. London B* 237: 37-72; also in *The Collected Works of A. M. Turing: Morphogenesis,* ed. P. T. Saunders, Amsterdam: North-Holland (1992)

Varela, F.R., Maturana, H.R. and Uribe, R. (1974). Autopoiesis: The organisation of living systems, its characterization and a model. *BioSystems*, 5:197.
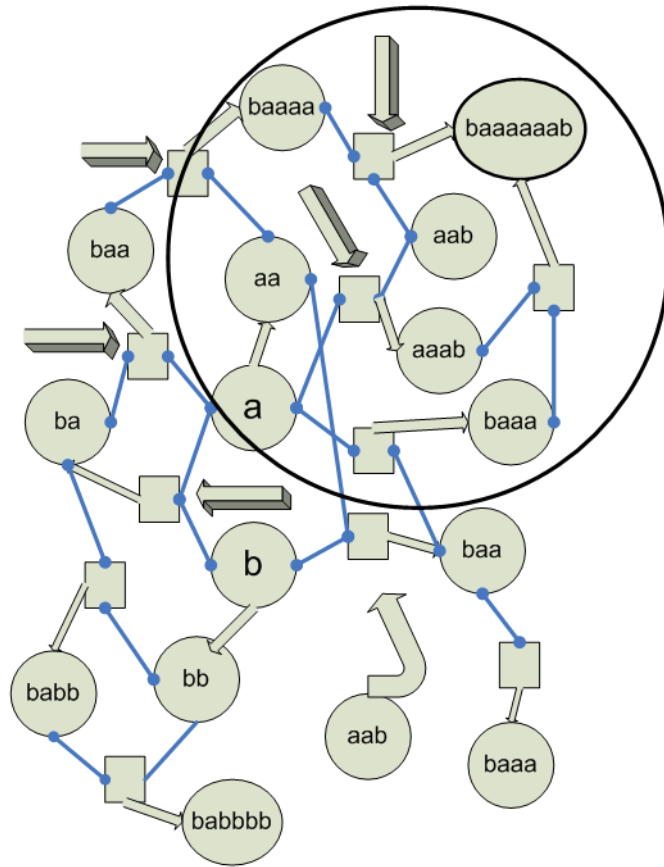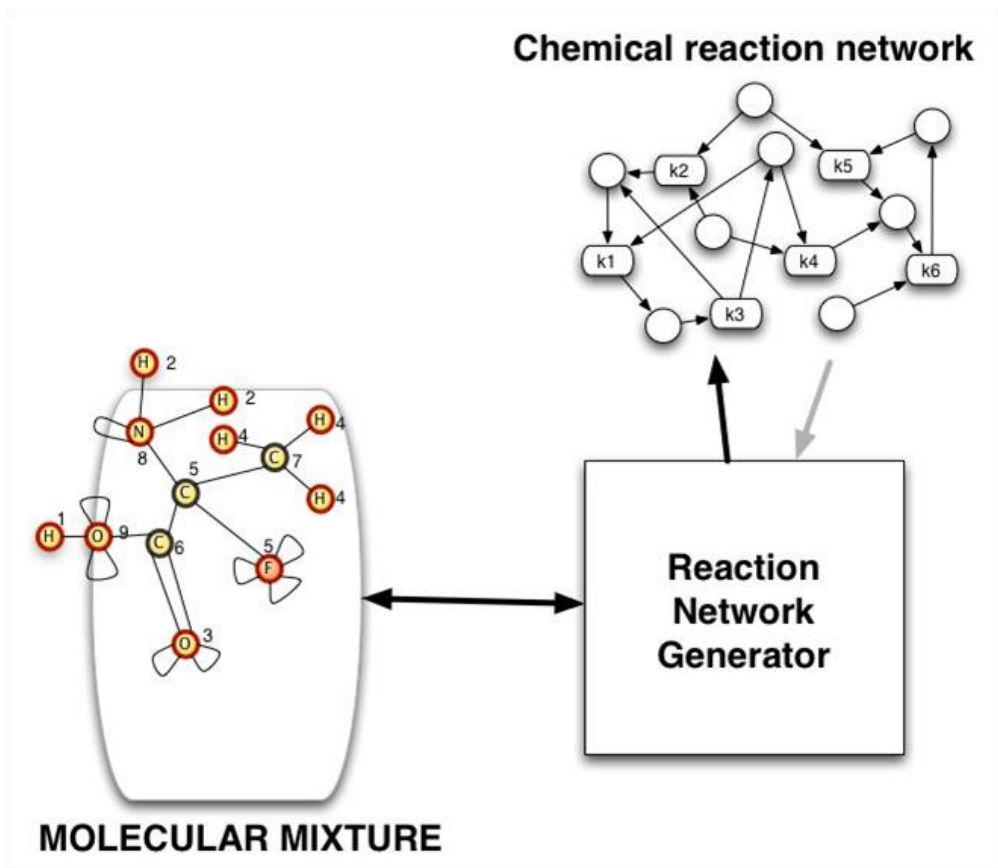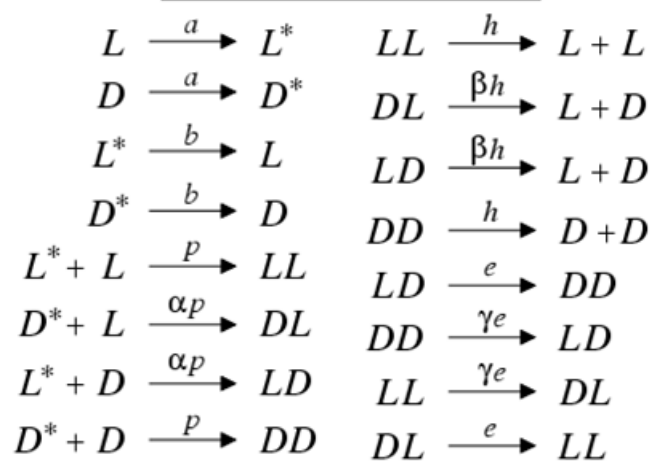
Figure 1

Figure 2

Chemical reactions

$$L \xrightarrow{a} L^*$$
$$D \xrightarrow{a} D^*$$
$$L^* \xrightarrow{b} L$$
$$D^* \xrightarrow{b} D$$
$$L^* + L \xrightarrow{p} LL$$
$$D^* + L \xrightarrow{\alpha p} DL$$
$$L^* + D \xrightarrow{\alpha p} LD$$
$$D^* + D \xrightarrow{p} DD$$

$$LL \xrightarrow{h} L + L$$
$$DL \xrightarrow{\beta h} L + D$$
$$LD \xrightarrow{\beta h} L + D$$
$$DD \xrightarrow{h} D + D$$
$$LD \xrightarrow{e} DD$$
$$DD \xrightarrow{\gamma e} LD$$
$$LL \xrightarrow{\gamma e} DL$$
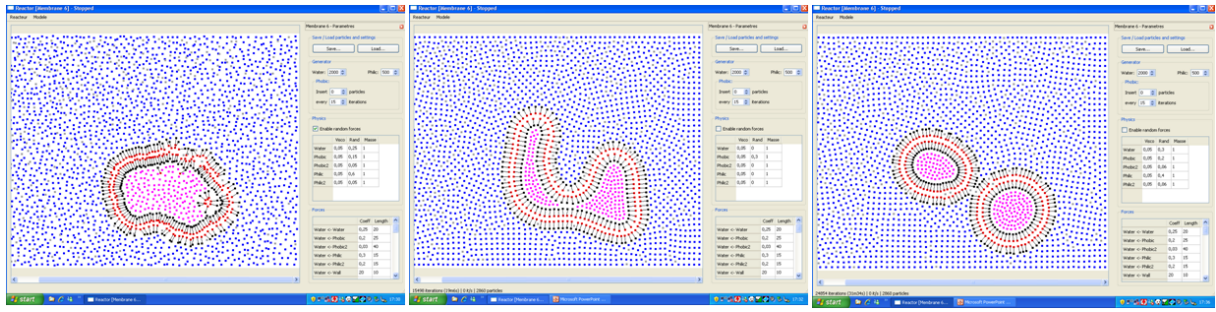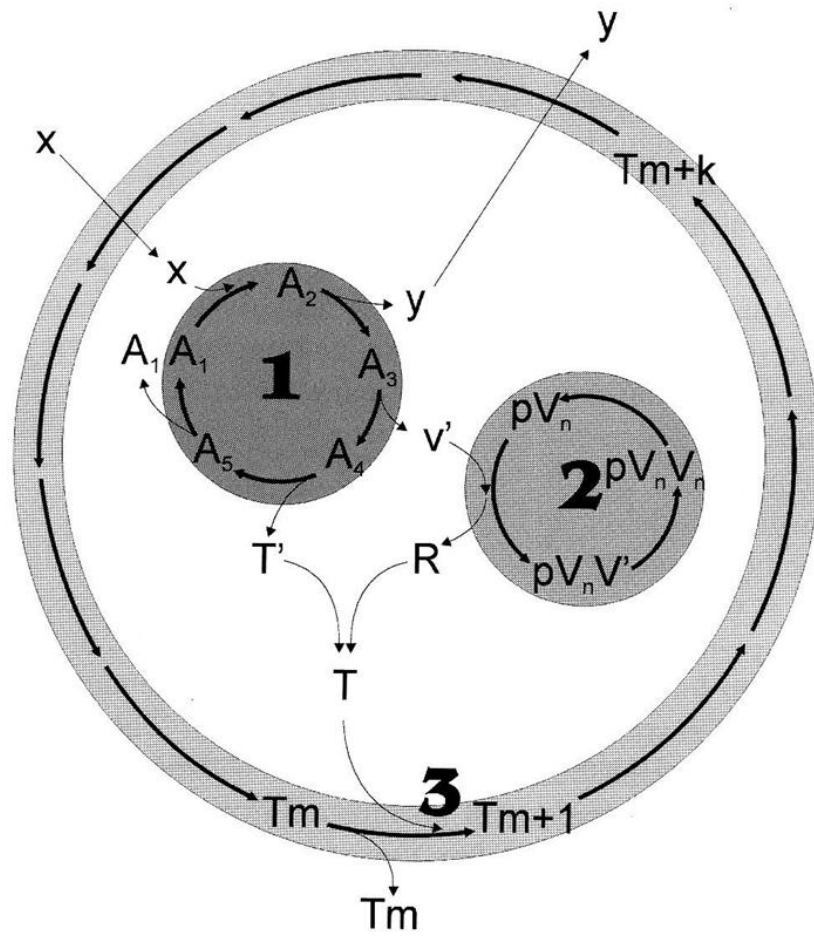$$DL \xrightarrow{e} LL$$

Reaction network
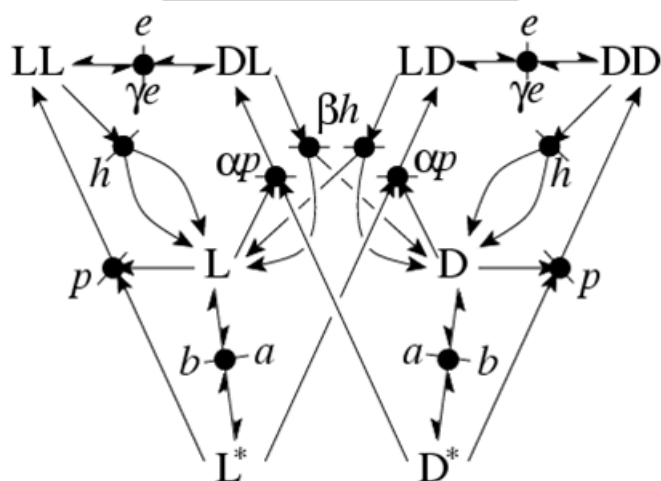
Figure 3
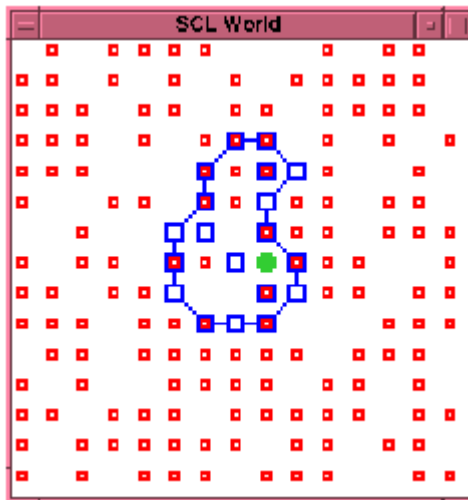
Figure 4
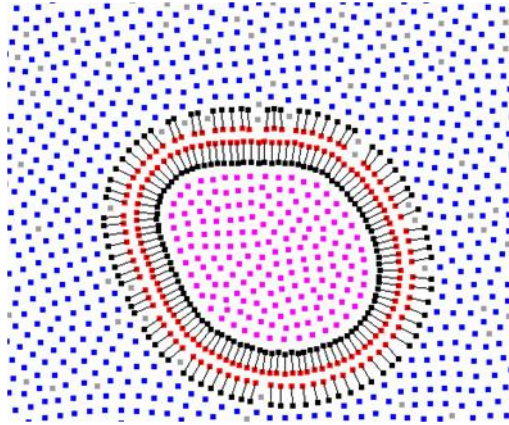
Figure 5
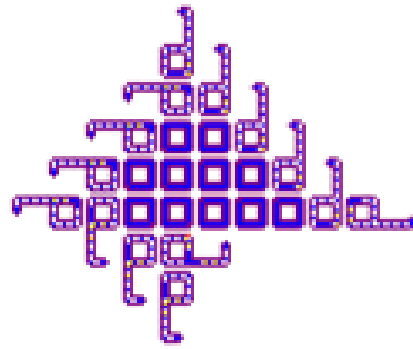
Figure 6

Figure 7

Figure 8

Figure 9