Central European Science Journals
w w w . c e s j . c o m

Central European Journal of Physics
Central European Science Journals
CEJP 3 (2003) 516–555

An Introduction to Physical Theory of Molecular Evolution

Bartolo Luque*

Departamento Matemática Aplicada y Estadística,
Escuela Superior de Ingenieros Aeronáuticos,
Universidad Politécnica de Madrid,
Plaza Cardenal Cisneros 3, Madrid 28040, Spain.

Received 5 March 2003; revised 10 July

Abstract: This work is a tutorial in Molecular Evolution from the point of view of Physics. We discuss Eigen's model, a link between evolutionary theory and physics. We will begin by assuming the existence of (macro)molecules or replicators with the template property, that is, the capacity to self-replicate. According to this assumption, information will be randomly generated and destroyed by mutations in the code (i.e., errors in the copying process) and new bits of information will be fixed (made stable) by the existence of an external pressure on the system (i.e.,selection), and the ability of the molecules to replicate themselves. Our aim is to build a model in order to describe molecular evolution from as general a standpoint as possible. As we will see, even very simple models from the theoretical point of view will have surprisingly deep consequences.

Keywords: Molecular Evolution, Quasi-species, Hypercycles, Biological Physics; Adaptation and Self-Organizing Systems; Soft Condensed Matter; Statistical Mechanics
PACS (2000): 87.10.+e, 87.23.-n, 89.75.-k

1 The birth of information

The question of how life arose on Earth is one of the most engaging intellectual challenges in science today [1, 2]. During recent years we have been able to reconstruct a tentative phylogenetic tree starting from a common ancestor by putting together the data provided both by Paleontology and Molecular Biology [3]. The particular historical order in which different branches have appeared and disappeared from this tree is, however, still subject to much debate. Furthermore, since the concrete history of life on this planet is probably

* E-mail: bartolo@dmae.upm.es

full of accidental situations, hardly reproducible by their own nature, it is likely that this will continue to be the state of affairs for a long time indeed. Unless we are willing to accept Earth's uniqueness in the Universe, there must be a set of general principles that make life a viable phenomenon [4]. It is about these general principles that we should concern ourselves, and it is in this spirit that we will try to tackle the problems put forward in this introduction.

The late nineteenth century Biology, developed in the midst of the industrial revolution, put forward a vision of living systems in which their internal mechanisms and energy transformations were the key elements. However, the telecommunications dominated society of today sees life as a dynamical state of matter organized by and around information. Energy transformations are not enough. Today we know that the genetic code stores information, and we are learning to decode it. The letters of this alphabet are the chemical bases that form the DNA, codons may be thought of as words, operons as paragraphs. Thus, we can rephrase the old question about the origin of life as: What is the origin of the information stored in living systems, and why is it stable? [5]

A few preliminary conclusions can already be established by examining the physical and informational constraints that (pre)living systems need to achieve:

- First, one should observe that all fluctuations are absorbed, eliminated, in a system in thermodynamic equilibrium. The spontaneous generation of order and information in this picture seems a very difficult (if not impossible) feat to accomplish. Our first conclusion therefore is that we will need to model processes that take place *out of equilibrium* [6].
- As the second piece of the puzzle, classical Thermodynamics tells us about the unstoppable increase of entropy that takes place in isolated systems, about information loss. Any way out of this second constraint needs thus to provide an energy flux, an *open system*. In principle, this energy flux could be used either to prevent the degradation of the information carrying units or to fabricate new replicas of them (in a timescale shorter than their lifetime, obviously). In the prebiotic stage, however, these units (prebiotic molecules) will be small and easy to build. Therefore, in our context the latter route will tend to be the most efficient way to prevent the information loss associated with entropy increase.
- The third and central component of the picture is the need for *evolution*. An immediate reflection about what evolution means in the biological sense leads to the conclusion that we must have two different mechanisms acting at once in an evolving system. First, the information contained in it must be modifiable, and must be able to change. Secondly, a particular direction of change must be selected by an external pressure or force exerted on the system, i.e., there must be a *selection* mechanism at work. We want to emphasize that evolution does occur in every population of entities with three characteristics: reproduction, heredity and mutation [2].

In this introduction we will assume the existence of (macro)molecules or replicators with the template property, that is, the capacity to self-replicate. On the basis of this assumption, information will be randomly generated and destroyed by mutations in the

code i.e., errors in the copying process, and new bits of information will be made stable due to the existence of an external pressure on the system i.e., selection, and the ability of the molecules to replicate themselves. The available experimental evidence points towards RNA chains as the components of this ancient prebiotic world on Earth: the so-called "RNA world" scenario [7]. There are at least three main questions in the RNA world scenario: first the synthesis of ancient nucleotides, second the emergence of the first self-replicating molecules, and third the error-catastrophe [8]. In this introduction we shall focus our attention on the last point. What is of real importance to us, however, is that until the work of Manfred Eigen [9] in the 70's there had been no convincing model for the evolution of complexity of this early replicators. Eigen has created a link between evolutionary theory and physics. This area is well-known today as the Physical Theory of Molecular Evolution and is the main focus of this work. Our aim will be to tackle the subject from a theoretical point of view. Thus, we aim to build a model which can describe molecular evolution from a very general framework. As we shall see, even simple theoretical models have surprisingly deep consequences.

A few words before we start: this work is not an extensive review and can be regarded as a friendly introduction to Molecular Evolution to those who like to be initiated into the field. Many interesting issues are just mentioned and can be complemented with updated bibliography. Complex details have been removed for the sake of simplicity and clarity. Three excellent reviews which go in depth are: "Biological evolution and statistical physics" by Barbara Drossel [10], "Introduction to the statistical theory of Darwinian evolution" by Luca Peliti [11] and "Molecular replicator dynamics" by B. M. R. Stadler and P. F. Stadler [12].

2 Experimental evolution?

There are two well-known criticisms to the Darwinian evolutionary theory: the selection concept implies a tautology, and that it cannot be checked experimentally. However, it is already possible today to observe evolutionary processes in short time scales: the evolution of the *influenza* or RNA viruses [13] are well known examples.

One of the first well-known example was the phago Q-β, a RNA virus that attacks bacteria. It stores its own information in RNA form and uses inverse transcriptasa to make its complementary copy. In the 60's, Haruna and Spiegelman cut off the phago Q-β's catalyst, the Q-β replicasa. This enzyme is the one responsible for the RNA copy. If we combine together in a test tube Q-β replicasa along with monomers like ATP, GTP, UTP or CTP, the viral RNA no longer needs hosts for its self-reproduction. In this way it is easy to define the fitness of a molecule by simply calculating its replication and decomposition rate. Thus, the tautology is avoided. Note that this is obviously not a natural environment for a virus. The capsule construction instructions stored in the RNA have become unnecessary since the now reproduction capsid is not needed any more, for example. As in the replicating process some mistakes may occur, the next generation of copies could conceivably obtain some reproduction benefits in this new environment

due to mutations. Thus, for example, the now accessory capsid instructions could be detached from the chain and dismissed, in order to shorten the length of the chain and increase its reproduction rate.

A experiment (see figure 1) with the same design as the one just described was carried out by Spigelman and collaborators [14]. A Q-β phage was inserted in the first test tube. After a few hours, part of the population of the tube was extracted and subsequently inserted in a new test tube. The end result after seventy repetitions of this process was that the chains had reached a stable configuration in which they had lost their infective capabilities. The RNA virus had only 17% its original length, and was reproducing at a rate 15 times faster than its initial replication rate. Nowadays, molecular or viral evolution experiments similar to the one described here yield spectacular results, and are providing us with a wealth of information on evolution [13].

3 The Eigen's model

The model that we are going to introduce now follows the scheme suggested by Eigen and Schuster [15]. We consider a system (a flux reactor as depicted in figure 2(a) formed by s molecular species: $I_1, I_2, \ldots, I_s$ with capacity for self-replication. Rich energetic monomers μ^* are introduced by valve 1 and the products of molecule degradation (μ) are eliminated by means of valve 2. Let $x_i(t) \geq 0$ be the concentration of the molecular species I_i in the system. We designate the *replication rate* of species I_i by A_i and its *degradation rate* by D_i. The process of molecular degradation is a direct consequence of molecular interactions, be they interaction of molecules with radiation or collisions amongst molecules, for example. We will however simplify the book-keeping of all these processes by assuming that molecules degrade at a constant rate.

3.1 Simple replication

If we assume simple replication and degradation in the system, we can write the kinetic reactions as:

- Replication: $\mu^* + I_i \xrightarrow{A_i} 2I_i$.
- Degradation: $I_i \xrightarrow{D_i} \mu$.

If by valve 1 we maintain μ^* constant in this model the self-replicating units or replicators are totally independent. There are no interactions between different species, and the population growth of a given species does not affect the growth of another.

We will thus have a continuous dynamical equation for the concentrations of each species:

$$\frac{dx_i}{dt} = (A_i - D_i)x_i(t) = E_i x_i(t), \qquad i = 1, \ldots, s. \tag{1}$$

Where we define $E_i = A_i - D_i$ as the *productivity* of species I_i. The solutions to this equation are very simple:

$$x_i(t) = x_i(0) \exp(E_i t), \qquad i = 1, \ldots, s. \tag{2}$$

The productivity acts obviously as the exponential growth rate for the species, so that if $E_i < 0$ species I_i will disappear and if $E_i > 0$ its population will grow unboundedly.

In figure 2(b) we see this happening for four species with $E_1 = 0.8$, $E_2 = 0.2$, $E_3 = -0.1$ and $E_4 = -0.5$. All the species initially have identical concentration $x_i(0) = 0.25$. We can see how the two with positive productivity (I_1 and I_2) grow without limit, while the two with negative productivity (I_3 and I_4) disappear exponentially. Thus, with this model we obtain segregation of species into two classes: only those species with $E_i > 0$ survive, while the rest disappear. This however does not correspond to selection in the biological sense. If we want a selection mechanism we need to introduce competition.

3.2 Replication and Selection

A straightforward manner to introduce competition in our model is to limit the resources and/or space available. Equivalently, we may keep the total molecular population constant (in the literature this constraint is referred to as *constant population* or CP). Although it is evident that in the prebiotic world such a constraint does not exist, we would like to introduce a term that mimics natural selection and see what lessons can be learned from it. The CP condition can expressed as:

$$\sum_{i=1}^{s} x_i(t) = 1, \tag{3}$$

where we have set the constant equal to one merely for convenience. This constraint implies that there will be no temporal variation in the total molecular population inside our reactor:

$$\frac{d}{dt}\Big[\sum_{i=1}^{s} x_i(t)\Big] = 0. \tag{4}$$

Physically, this can be arranged through an evacuation flux $\Phi(t)$, a non-chemical rate of change modulated by the valve 3 in picture 2(a). Introducing this constraint in our equations yields:

$$\frac{dx_i}{dt} = (E_i - \Phi(t))x_i(t), \qquad i = 1, \ldots, s \tag{5}$$

where we assume that the evacuation flux of a given species is proportional to its population size. If we sum all the equations and use (3) and (4), we obtain:

$$\sum_{i=1}^{s} \frac{dx_i}{dt} = \sum_{i=1}^{s} (E_i - \Phi(t))x_i(t) = 0, \tag{6}$$

and thus we have an expression for the temporal dependence of the flux:

$$\Phi(t) = \sum_{i=1}^{s} E_i x_i(t) = \langle E(t) \rangle \,. \tag{7}$$

The simple meaning of this expression is that if we want to keep the population constant, at any instant we have to evacuate molecules at a rate equal to the *average productivity*

at that instant $\langle E(t)\rangle$. Thus, the outgoing flux will be equal and opposite to the rate at which the population would change in a system without restrictions.

The equations (5) can be written:

$$\frac{dx_i}{dt} = (E_i - \langle E(t)\rangle)x_i(t), \qquad i = 1, \ldots, s \tag{8}$$

These equations are now non-linear because $\langle E(t)\rangle$ depends on all the concentrations as we can see in (7). The solutions are not difficult to compute analytically, but before we do that it will be useful to give some qualitative considerations without explicit calculus.

Let us assume that initially we have several species at equal concentrations, with an initial mean productivity $\langle E(t=0)\rangle$. Obviously, some of the species initially present will have a productivity below the mean. What equation 8 tells us is that the concentrations of these molecules will subsequently start to decrease, while the concentrations of species with productivity above the mean will grow. As time goes on however, this process will in turn make the mean productivity grow. But as the mean productivity grows, some species that initially had productivities above the mean will fall below it, and their concentrations will start to decrease, and so on. In this manner the average productivity will keep growing, and different species will keep disappearing, until $\langle E(t)\rangle$ reaches a value equal to the highest productivity present in the system. At this point, only the species corresponding to this productivity will have a non-zero concentration, and we can say that this species has been selected, hence the *selective equilibrium* is reached.

This fact is illustrated in figures 3(a) and 3(b). In figure 3(a) we start with four species with positive productivities and identical initial concentrations. Initially, the mean productivity is $\langle E(t=0)\rangle = 0.45$. We can see that species with productivity below the mean productivity ($E_1 = 0.1$ and $E_2 = 0.3$) immediately start to decrease their concentration. Meanwhile, the concentrations of those species with productivity above the initial mean ($E_3 = 0.6$ and $E_4 = 0.8$) start to increase in time. As an outcome, the mean productivity increases too until it grows above $E_3 = 0.6$, at which point the concentration of I_3 also starts to decrease. Finally, we reach a state in which all the population is made up from the species with the highest productivity, I_4, with $E_4 = 0.8$. At this point the average productivity is of course stabilized at $\langle E(t > 30)\rangle = E_4 = 0.8$.

In figure 3(b) we perform the same experiment, but with an initial condition in which I_4 is not present. We can see that the selective equilibrium is reached when species I_3 saturates the population and the mean productivity stabilizes at 0.6. To break up this state at $t = 25$ we introduce an infinitesimal concentration of I_4, and we see how this species displaces I_3 from the system until a new selective equilibrium at $\langle E(t > 30)\rangle = 0.8$ has been reached.

It is not difficult to compute explicitly the solutions of equations 8. First, notice that the steady state of the k-th equation, $dx_k/dt = 0$, corresponds to either $x_k = 0$ or to $E_k = \overline{E}$, where $\overline{E} = \langle E(t \to \infty)\rangle$ is the asymptotic value of $\langle E(t)\rangle$. If we consider a non-degenerate system, that is, $E_i \neq E_j$ for all i, j, only one species can have productivity $E_i = \overline{E}$. Thus, for the total system, the trivial fixed points defined by $\Omega_0^* \equiv (0, 0, ..., 0, ..., 0)$ and the set $\Omega_k^* \equiv \{(0, 0, ..., x_k^* = 1, ..., 0); i = 1, \ldots, s\}$ provide the $s + 1$ global steady

states for a system of s species. It would therefore seem that coexistence is not possible.

We can study the evolution of the system using classical stability analysis. We perturb the state of the system taking it slightly away from the fixed point $\Omega_k^* \equiv (0, 0, ..., x_k^* = 1, ..., 0)$ and study the stability of the system against the perturbation. For this we need to compute the Jacobi matrix of the system. Its elements are:

$$L_{ij}(t) = \frac{\partial}{\partial x_j}\Big[\Big(E_i - \langle E(t)\rangle\Big)x_i(t)\Big] = \frac{\partial}{\partial x_j}\Big[\Big(E_i - \sum_{r=1}^{s} E_r x_r(t)\Big)x_i(t)\Big]$$
$$= \begin{cases} E_i - \sum_{r=1(r\neq i)}^{s} E_r x_r(t) - 2E_i x_i(t) & \text{if } i = j \\ -E_j x_i(t) & \text{if } i \neq j \end{cases}. \tag{9}$$

If we have as solution $\Omega_k^* \equiv (0, 0, ..., x_k^* = 1, ..., 0)$, then $x_r = 0$ for all $r \neq k$ and we obtain

$$L_{ij}(\Omega_k^*) = \begin{cases} E_i - E_k & \text{if } k \neq i = j \\ -E_k & \text{if } k = i = j \\ -E_j & \text{if } k = i \neq j \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

or, in matrix form,

$$\mathbf{L}(\Omega_k^*) = \begin{pmatrix} E_1 - E_k & 0 & \vdots & 0 & \vdots & 0 \\ 0 & E_2 - E_k & \vdots & 0 & \vdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & 0 \\ -E_1 & -E_2 & \vdots & -E_k & \vdots & -E_s \\ \vdots & \vdots & \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & E_s - E_k \end{pmatrix} \tag{11}$$

We can compute the characteristic polynomial through the determinant

$$|\mathbf{L}(\Omega_k^*) - \lambda\mathbf{I}| = 0. \tag{12}$$

Therefore,

$$P(\lambda) = (E_k - \lambda)\prod_{r=1\neq k}^{s}(E_r - E_k - \lambda) = 0, \tag{13}$$

whence the associated eigenvalues will be $\lambda_r = E_r - E_k$ for $j \neq k$ and $\lambda_k = -E_k$.

The fixed point generally is locally stable if $Re(\lambda) < 0$. Naturally, λ here is real. Evidently this will be the case if and only if the point x_k^* represents the concentration of a species such that $E_k > E_r$ for all $r \neq k$. Thus, only the best replicator survives.

Now we have authentic selection at work, but alas, no evolution. Once a selective equilibrium has been reached the system remains in that state indefinitely (indeed, that is why it is called equilibrium!). The only way to introduce a change seems to be by hand, as we did in the experiment of figure 3(b). The next question is thus: How can we introduce novelty in the system in an endogenous way? The answer is: easily, if only we remember that we can learn from our mistakes...

3.3 Replication, selection and mutation

If we want evolution we need mutation, we need errors in the copying process. Suppose that the *replication accuracy* of species I_i is $Q_{ii} \in [0,1]$. Let us use this to represent the probability of producing a perfect copy. Thus, if $Q_{ii} = 1$ every copy produced by species I_i will be identical to the original. If $Q_{ii} \neq 1$ however there will be a non-zero probability, $1 - Q_{ii}$, for the copying process to have an error. In this case, the copying of a molecule of type I_i will produce a molecule of another species, say I_j. Let us therefore define Q_{ij} as the probability that a molecule I_i produces, as the result of an erroneous replication process, a copy belonging to species I_j. Thus, we will have

$$1 - Q_{ii} = \sum_{j=1 \neq i}^{s} Q_{ij}. \tag{14}$$

The kinetic reactions will now be:

- Replication without error: $\mu^* + I_i \xrightarrow{A_i Q_{ii}} 2I_i$.
- Replication with error: $\mu^* + I_i \xrightarrow{A_i Q_{ij}} I_i + I_{j \neq i}$.
- Degradation: $I_i \xrightarrow{D_i} \mu$.

It is not difficult to incorporate the effect of the mutation into our equations (5):

$$\frac{dx_i}{dt} = (A_i Q_{ii} - D_i)x_i(t) + \sum_{j \neq i} A_j Q_{ij} x_j(t) - \Phi(t)x_i(t), \qquad i = 1, \ldots, s \tag{15}$$

The first term of the equation represents the increase in the population as a result of the production of correct replicas, the second represents the apparition of copies of I_i as a consequence of defective replication of other species, while the third is the flux term used to implement the CP constraint.

We can again find a value for the flux by adding all the equations and taking into account that now:

$$\sum_{i=1}^{s} \sum_{j \neq i}^{s} A_j Q_{ij} x_j(t) = \sum_{i=1}^{s} A_i (1 - Q_{ii}) x_i(t), \qquad i = 1, \ldots, s. \tag{16}$$

As was to be expected, again the mean productivity equals the flux. Using this, equation (15) can be rewritten in a more convenient form as

$$\frac{dx_i}{dt} = \sum_{i=1}^{s} W_{ij} x_j(t) - \langle E(t) \rangle x_i(t), \qquad i = 1, \ldots, s. \tag{17}$$

The coefficients $W_{ii} = A_i Q_{ii} - D_i$ are the so-called *selective values*, while $W_{ij} = A_j Q_{ij}$ for $i \neq j$ are known as the *mutation values*. Obviously, if $Q_{ii} = 1$ then $Q_{ij} = 0$ for all $j \neq i$ and the equations revert back to system 8. In fact if $Q_{ij} \approx 0$ we can neglect the mutational term. Then the solution to our equations will contain a single species that will dominate the selective equilibrium, as was seen before. We call this selective winner

the *master sequence*. In general, however, Q_{ii} will not be strictly equal to 1. Then the master sequence will always be accompanied by a number of mutants. The set formed by this cloud of mutants together with the master sequence is known as the *quasispecies* [16]. Figure 4 graphically illustrates the concept.

It is not difficult though to solve the equations without neglecting the mutational term [17]. In matrix form, equations 17 take the form:

$$\frac{d\mathbf{X}}{dt} = (\mathbf{W} - \langle E(t)\rangle)\mathbf{X}. \tag{18}$$

To solve this system it suffices to perform an orthogonal change of variables to take $\mathbf{W}$ to a diagonal form. If $\mathbf{O}$ is a matrix such that $\mathbf{OWO^{-1}} = \lambda\mathbf{I}$ and $\mathbf{U} = \mathbf{OX}$, equation 18 transforms to:

$$\frac{d\mathbf{U}}{dt} = (\lambda\mathbf{I} - \langle E(t)\rangle)\mathbf{U}, \tag{19}$$

or, equivalently,

$$\frac{du_i}{dt} = (\lambda_i - \langle E(t)\rangle)u_i(t), \qquad i = 1, \ldots, s. \tag{20}$$

These equations are formally equivalent to 8. Now, the eigenvalues λ_i act as the productivity E_i and the u_i act as concentrations. Furthermore, it is easy to show that:

$$\langle E(t)\rangle = \sum_{i=1}^{s} \lambda_i u_i. \tag{21}$$

That is: the mean productivity is invariant under the orthogonal transformation. Solution of (20) is obviously identical, step-by-step, to the preceding case. Therefore, we know that only one u_i will be selected (i.e., the one with the highest λ_i). Now, however, u_i does not represent a single species but a linear combination of them, a quasi-species. Thus, a quasi-species is a dynamically stable distribution of species. It will be the quasi-species, rather than the single species, which will dominate the selective equilibrium.

We illustrate, following reference [18], the quasi-species concept in figure 5(a), where we show the results of a simulation for a $\mathbf{W}$ given by

$$\mathbf{W} = \begin{pmatrix} 1 & 0.001 & 0.001 & 0.01 \\ 0.1 & 2 & 0.01 & 0.01 \\ 0.001 & 0.1 & 3 & 0.01 \\ 0.001 & 0.001 & 0.001 & 4 \end{pmatrix}$$

We have chosen selective values $W_{ii} = i$ purely for convenience so that I_4 has the highest selective value. We start our simulation with all the population formed by species I_1. Subsequently, species I_2, I_3 and I_4 appear by mutation. As we can see, when all the mutational values are small the selective equilibrium state is dominated by the master sequence and the presence of mutants is very small.

If, however, we choose higher values for the W_{ij}, such as for example:

$$\mathbf{W} = \begin{pmatrix} 1 & 0.001 & 0.001 & 1 \\ 0.1 & 2 & 0.01 & 1 \\ 0.001 & 0.1 & 3 & 1 \\ 0.001 & .001 & 0.001 & 4 \end{pmatrix},$$

the net result is again a coexistence of all the species, but now the relative concentrations of mutants is of the same order as that of the master sequence. Furthermore, for these high mutational values the majority of the population in the selective equilibrium state is in the form of mutants.

4 Tierra: silicon quasispecies

The entities taking part in the type of dynamics described in the previous sections do not necessarily have to belong to the realm of chemistry. Any object able to self-replicate, mutate, and degrade, may indeed follow the same evolutive equations [19]. A particularly interesting example is provided by computer programs. If one makes a computer code susceptible to mutations, (i.e., one puts into it some errors) in general the result will be a non working code. T. Ray [20] solved this problem, known as *brittleness*, designing a closed code in assembly language made up of five bit instructions. He wrote his code in a way that a mutation in any of these bits led to other instruction with sense in the code. For example, the "codon" 11111 in Tierra means "divide" in assembly language. A single mutation may take this codon to 01111, the "inc" instruction. In addition, the use of templates (patterns of instructions) for addressing purposes rather than absolute addresses could avoid the brittleness problem. A codon's sequence is a creature in Tierra. Each creature has a CPU time (it has some resources). CPU time is limited (this is the equivalent to the CP condition).

In the previous equations we have not taken into account the effect of the finite size of the replicant population (for advances in the analytical treatment of finite population in the quasi-species model see [21] and the problem of Muller's Ratchet [22]). In general, only a tiny fraction of all the possible species will be represented in the population at any time. Therefore, to begin with, the population structure will stabilize around a quasi-species distribution formed by some of the species that were initially present. Obviously though this quasi-species will in general not be the one with the highest productivity, and thus mutation will in time bring on a master sequence with a fitness higher than the dominant one. This higher fitness mutant will subsequently oust the dominant quasi-species and generate a new metastable quasi-species with higher λ than the initial one, and so on, until we reach the quasi-species with the maximum λ available. Note that it is an optimization process, a search for better solutions. In this sense the quasi-species form an example of *genetic algorithm* [23].

In figure 6, a typical Tierra simulation with mutation rate $q = 0.5 \times 10^{-8}$ per bit is shown [24]. The vertical axis represents the most common fitness of the existing creatures, the most popular genotype. Productivity is associated with fitness. Time is represented on the horizontal axis in units of 10^9 instructions executed by the CPU. As it is shown in the plot what we get is a *punctuated* picture of evolution: stasis periods, epochs when a master sequence and its cohort of mutants dominate, are interrupted by very abrupt, discontinuous changes. A new mutant with higher fitness appears in the system destabilizing the dominant quasi-species and taking the system to a new metastable

selective equilibrium. Classically mutations correspond to statistical fluctuations, needed as a continuous source of evolutionary novelty. The object of selection, the target, is the 'wild type' or master sequence. In the Eigen model the quasi-species is what is selected.

Let us briefly comment on some important characteristics of this evolutionary paths that generally appear in many different systems [19]. In Tierra, it is clear that a separation of time scales exists: the one for stasis being much greater than the jump or nucleation that acts as relaxation time. This is reminiscent of a Self-Organized Criticality (SOC) [25]. In figure 7 we represent the power spectrum of the figure 8. The dashed line is a fit [24] to $P(f) \sim f^{-\beta}$ with $\beta = 2.0 \pm 0.05$. As we know, this kind of noise is typical of SOC, but its presence is not a sufficient condition to warrant the presence of SOC behavior [26].

In the plot 8 the integrated distribution of the times of domination for each quasi-species, $N(t)$, is shown

$$M(\tau) = \frac{1}{\tau} \int_{\tau}^{\infty} N(t) dt \tag{22}$$

The integrated distribution is distributed with the same exponent as $N(t)$ but it is more significant statistically, especially for a small number of measurements. The plotted results have been obtained from 50 simulations under identical initial conditions [24]. In all, 512 times were measured. Fitting the data yields

$$M(\tau) = C\tau^{-\alpha} \exp\left(\frac{-\tau}{T}\right). \tag{23}$$

with $\alpha = 1.1 \pm 0.05$ and a cut-off parameter $T = 450 \pm 50$. We find stasis periods and power laws exponents indicative of SOC. The extraordinary difference is that now we are dealing with computer creatures or molecules, rather than with living animals that inhabited the World, the planet Tierra.

5 Modeling quasi-species in a computer: the error catastrophe

Up until now we have not made any reference to the chemical nature of our self-replicative molecules. If we take ribozimes for instance (RNA with self-catalytic capacity which in fact have been proposed as candidates for the prebiotic stages on Earth [27]), we would be speaking of chains of monomers. A given molecule of length N can be described as a sequence of variables $(s_1, s_2, ..., s_N)$, where the s_i can take four values in order to describe the four types of monomers that can occupy the locus i of the sequence. (Thus for instance $s_1 = -2$ would imply that the first monomer is citosine C, $s_1 = -1$ that it is urine U, $s_1 = 1$ guanine G, and $s_1 = 2$ adenine, to choose but one particular value assignation). A molecule formed by N bases can take 4^N different configurations, and this configurations will form the sequence space. Each possible molecule will have a productivity associated to it: its fitness. The geometrical representation of the multidimensional space of all the possible sequences, with their associated fitness represented as 'heights' is the so-called *fitness landscape*. An idea first introduced by Sewall Wright and later extended by several authors [28].

Following [29], in order to simulate the quasi-species model we will make a simplifying approximation of our molecules as follows: each molecule will be a sequence of $N = 50$ monomers, and each monomer will belong to only two possible types, purines (either G or A indistinctly) an pyrimidines (C or U). Thus the s_i will take only two values (-1, +1), and the whole chain is made up of a sequence of plus and minus ones $(1, 1, -1, ...)$. There are then 2^{50} possible configurations in the sequence space. In this space we will define a distance between sequences: the Hamming distance.

We start with a population of sequences where all the strings are 1111...1111, i.e. only plus ones, and they are the best-adapted type (that is, the one with higher fitness). This master sequence replicates itself ten times faster than the rest, with all the other sequences replicating at the same fixed rate, of say 1 for example. The fitness landscape is then very simple: a peak with height 10 for the master sequence and a platform with identical value 1 for the rest of the sequences. In the literature this fitness landscape is known as the single peak landscape.

In this picture, point mutation naturally corresponds to the substitution of a +1 for a -1 (or vice versa) at locus i as the chain replicates. As the system evolves in time the chains replicate and point-mutate, moving away from the master copy in the space of sequences, but lower replication rates for most mutants will tend to keep the copies close to the master sequence. We define the error rate for the copying of individual symbols as $1 - q$. That is, the probability that a -1 changes to 1 (or a 1 to -1) in the copying process of every molecule is $1 - q$. Thus, the probability for a string I_i to produce a mutant offspring I_j at a Hamming distance d away will be

$$w_{ij} = \binom{N}{d} q^{N-d}(1-q)^d \tag{24}$$

We will designate the master sequence concentration at time t by $x_0(t)$, the combined concentrations of all the single-error mutants by $x_1(t)$, and in general, the combined concentration of mutants at a distance d by $x_d(t)$. In this notation the CP condition takes the form $\sum_{d=0}^{50} x_d(t) = 1$.

We start the simulations with $x_0(t) = 1$, i.e., all the initial strings belong to the master sequence, and let the system evolve to the quasi-species equilibrium. Figure 9 represents the distribution of equilibrium concentrations as a function of the error rate as it ranges from $1 - q = 0$ (error free) to 0.06. Only in the error free case we are led to an all-or-none selection. As the error rate increases, the concentration of the master sequence decreases rapidly. Nevertheless the quasi-species remains stable, there is a *consensus sequence*, i.e., it is possible to recuperate the information statistically about the master sequence. This is so until we reach a value of $(1 - q) = 0.046$. This value, known as the *error threshold*, marks a sharp transition. Beyond it, the distribution of chains suddenly changes and all the information about the original master sequence is lost. The process clearly seems to resemble a phase transition.

This can bee seen more clearly in the logarithmic plot (figure 9 down). Below the error threshold we have a non-trivial distribution, the quasi-species, which as we can see, can be regarded as the master sequence with a comet like tail of mutants accompanying

it, while beyond the error threshold we have a uniform distribution, a random population. The biological interpretation of this sharp dependence of the model on the mutational rate is amazing.

RNA viruses offer an opportunity for exploring long term evolution under controlled conditions [30]. Indeed, let us suppose that our chains represent RNA viruses genomes. We catalog new viral entities via their consensus sequence, or "wild type", (master sequence in our jargon). For a population of viruses in general, the consensus sequence is extracted via statistical methods (population sampling). We know, however, that the probability of finding a virus representative of the consensus sequence is negligible. According to our model, this fact is indicates that viruses have a high mutational rate. But it cannot be too high a value because the consensus sequence would altogether disappear, and the virus for instance would loose its infective capabilities. The values of the mutational rates thus indicate that many RNA viruses find themselves just at the edge of the error catastrophe. What selective advantage do they find craning their necks out into the abyss? Viruses are identified and eliminated by immune systems. If the virus shows itself in many different variations the immune system will be in deep trouble to recognize and eliminate them all, and the virus will get to escape its prosecutors. Viruses therefore place themselves exactly at a point (value of its mutational rate) where the mutant diversity is biggest while the biological identity is still preserved. For additional information about the selective benefit of higher mutation rate see [31].

Now that we know this natural strategy though we can play dirty with the virus. If we push a virus a little further, if we increase its mutational rate externally, then it will fall beyond the error catastrophe, losing its infective capacity (and indeed its identity). This idea has been successfully applied by E. Domingo et al. [32] to the aftosa virus in vitro. An amazing success of molecular evolutionary theory.

6 From living systems to magnets: the Ising connection

It is a remarkable fact that there is a well-known system in statistical physics that seems to closely resemble the type of dynamics that we have just described. Indeed, a 1-dimensional Ising chain of length N can be described in terms of N s_i's, $(s_1, s_2, ..., s_N)$, with $s_i = \pm 1$. As the chain evolves in time, nearby spins tend to align in the same direction to achieve the least energetic configuration, whereas thermal fluctuations will tend to make spins flip (a close analogy to point mutations). In 1985-6, Leuthäusser [33] took this conceptual analogy between quasi-species and Ising model a step forward, demonstrating that the similarity between the two systems could be put in a rigorous form. He thus opened a door to the possibility of applying the well established framework of equilibrium phase transitions to understand the error threshold and the existence of quasi-species [34] (Mappings with other statistical systems have been proposed. For example with direct polymers in random media [35]).

The basic idea is rather simple, as it is often the case. Indeed, the constant population constraint does not fundamentally alter the dynamics of the system, amounting to a mere

(time dependent) normalization of the concentrations. Thus we may as well do away with it, which is easily accomplished by the change of variables

$$y_i(t) = x_i(t) e^{\int_0^t dt' \sum_{j=1}^{s} A_j x_j(t')} \tag{25}$$

We will identify every possible chain configuration with a different species. Thus, for chains with ν sites, the matrix $\mathbf{W}$ will now be of order 2^ν ($2^\nu \times 2^\nu$). In these variables the equation of motion for the system simply reads, in matrix form,

$$\frac{d\mathbf{y}}{dt} = \mathbf{W}\mathbf{y}. \tag{26}$$

This type of equation is of course very well known in Physics. $\mathbf{W}$ is analogous to the negative of the Hamiltonian in Euclidean Quantum Mechanics (possibly up to a factor). The elements of $\mathbf{W}$ are all real and greater than or equal to zero, since they represent mutation rates. In practice, although $\mathbf{W}$ need not be symmetric, the W_{ij} are chosen so that all the eigenvalues of $\mathbf{W}$ are real (inclusion of an imaginary part in the eigenvalues will merely produce damped or amplified oscillations). If the highest eigenvalue λ_0 is unique and isolated then the evolution operator $e^{\mathbf{W}t}$ will project any initial population into a final state given by the eigenvector of λ_0 (i.e., the ground state of the equivalent quantum theory). That is,

$$\mathbf{y}(t \to \infty) = e^{\mathbf{W}t}\mathbf{y}(0) \to \mathbf{v_0}, \tag{27}$$

where $\mathbf{v_0}$ is the eigenvector corresponding to λ_0. Therefore, by observing the equivalence with Quantum Mechanics we have re-derived our main, previous result. Although this conclusion could have been reached by mere inspection of equation 26, the real advantage of the analogy with Euclidean Quantum Mechanics lies in exploiting the deep connection of the latter with Statistical Mechanics, where a wealth of techniques have been developed in order to explore the structure of the ground state and the equilibrium properties of systems with many (possibly infinite) degrees of freedom.

The fundamental tool used to describe a system in Statistical Physics is the partition function

$$Z = tr\ e^{-\beta \mathbf{H}} \tag{28}$$

in which β is the inverse temperature (in natural units), $\mathbf{H}$ the Hamiltonian of the system, and the trace is taken over all the possible states of the system. Z provides a means to determine the structure of the lowest energy (analogous to the highest fitness) state of the system. In particular, as we saw in the chapter devoted to the Ising model, in the case of lattice models with finite range interactions (for example nearest neighbors) it is possible to write the partition function with periodic boundary conditions in a particularly simple way in terms of the *transfer matrix*,

$$Z = tr\ \mathbf{T}^N, \tag{29}$$

in which N is a measure of the size of the system and $\mathbf{T}$ is a much simpler matrix than the original $e^{-\beta \mathbf{H}}$, typically involving only the interaction of an element with its nearest neighbors.

Returning to our problem, let us think in terms of a replication process driven by discrete time steps. Accordingly, let us assume that the population changes discretely from one generation to the next, the time interval needed to go from one generation to the next being Δt. Thus $\mathbf{y}(0)$ is the first generation population, $\mathbf{y}(\Delta t)$ the second, and so on. Figure 10 depicts the time-ordered lay out of successive generations starting from a single molecule, the up and down arrows at each point identifying the type of monomer present at each locus of the chain. The evolution operator that advances the initial chain forward in time n generations will then evidently be

$$e^{\mathbf{W}n\Delta t} = (e^{\mathbf{W}\Delta t})^n, \tag{30}$$

or,

$$(e^{\mathbf{W}})^n, \tag{31}$$

by setting $\Delta t = 1$. Thus, our time evolution operator can also be recast in the form of a simpler transfer matrix up to a power related to the system size. Note however that the involved size scale is now the *time* scale. This implies that we will be concerned only about the *surface* properties of the system described by this transfer matrix (that is, the properties of the system at time t) and not about the behavior of the bulk (i.e., the previous generations).

One could argue that although going from eqn. (28) to (29) involves a considerable simplification, the step from (30) to (31) does not, since after all in both cases we have to deal with $e^{\mathbf{W}}$. The trick lies however in realizing (as Leuthäusser did) that $\mathbf{W}$ is in fact already formally identical to the transfer matrix for the 2-dimensional Ising model. Indeed, it is easy to convince oneself that for chains of size ν we will have:

$$W_{ij} = A_i q^{\nu - d_{ij}} (1-q)^{d_{ij}}, \tag{32}$$

where A_i is the replication rate of chain i, q is the probability that a particular locus of the chain is copied correctly into the next generation and d_{ij} is the Hamming distance between chains i and j (that is, the number of elements of the chain that have to change to go from i to j. The Hamming distance however can be written as

$$d_{ij} = \frac{1}{2}\Big(\nu - \sum_{l=1}^{\nu} s_l^{(i)} s_l^{(j)}\Big) \tag{33}$$

(the reader can check this by using a simple example with $\nu = 4$ for instance). Substituting this into the expression for W_{ij} and performing a little rearrangement then yields

$$W_{ij} = A_i [q(1-q)]^{\nu/2} \left(\frac{1-q}{q}\right)^{-\frac{1}{2}\sum_{l=1}^{\nu} s_l^{(i)} s_l^{(j)}}, \tag{34}$$

or,

$$W_{ij} = A_i [q(1-q)]^{\nu/2} \exp\left(-K \sum_{l=1}^{\nu} s_l^{(i)} s_l^{(j)}\right), \tag{35}$$

where

$$K = \frac{1}{2} \ln \left(\frac{1-q}{q} \right). \tag{36}$$

It suffices now to compare this expression with the transfer matrix of the 2-d Ising model for rows of length ν

$$T_{ij} = \exp\left(-\beta E(\sigma^{(i)}, \sigma^{(i)})\right) \exp\left(-\beta J \sum_{l=1}^{\nu} \sigma_l^{(i)} \sigma_l^{(j)}\right). \tag{37}$$

In this expression $\sigma^{(i)}$ and $\sigma^{(j)}$ label the configurations i and j (out of the 2^ν possible configurations) of two neighboring spin rows, $E(\sigma^{(i)}, \sigma^{(i)}))$ is the interaction energy between spins of one row, and J is the strength of the ferromagnetic coupling. Comparison of these two expressions shows that K includes both the ferromagnetic case ($K < 0$ with $1/2 \leq q \leq 1$ being equivalent to $J < 0$) and the antiferromagnetic one ($K > 0$ with $0 \leq q \leq 1/2$ equivalent to $J > 0$. Since q is always close to unity however, we will be dealing with the first case only. Also, the transition from a binary to a quaternary system also leads to an exact solution, although this does not alter the basic conclusions presented here.

It is of utmost physical importance to observe the equivalence between temperature and mutation rate. Indeed,

$$\frac{1}{T} \sim \left| \ln \frac{1-q}{q} \right|. \tag{38}$$

In the thermodynamic limit, the 2-d Ising model is known to go through a phase transition at a critical temperature T_c that separates an ergodic phase at $T > T_c$, in which any finite region of available phase space has a non-vanishing probability to be explored, from a phase of broken ergodicity at $T < T_c$, in which the phase space is broken into disjoint sets. In the latter case phase space decomposes into a set of disjoint sets: if the system is initially in one of those sets, it remains there at a later time. The precise correspondence between expressions 35 and 37, and the equivalence between T and q immediately allows us to recognize the error catastrophe: the ergodic phase corresponds to the disappearance of the master copy and the spreading of the system over all the space of sequences, whereas the non-ergodic phase corresponds to the existence of the quasi-species. The whole formulation developed to study equilibrium phase transitions, spin glasses, etc, becomes thus available to us in order to study how evolution proceeds in more complex and realistic fitness landscapes [36]. In [37] have given examples of error catastrophes in mutation-replication models which are not phase transitions. In [37]a the author argues that in the models for which the error catastrophe is a true phase transition, this is due to the "singularity" of the potential being used, namely, that there is one fit master sequence and all others are equally unfit. In his more realistic models there is a "smooth transition" from the master sequence to the most unfit sequences. The author realized that one does not need to enforce CP at all, since it is only the relative proportions of the different sequences that matter, and not the total number of individuals in the population. Thus one can use linear algebra to study quasi-species. In this way the already known

results can be re-derived as well as some new ones, having to do with the appearance and time evolution of a quasi-species after infection and the relation between the fitness of the sequences and their relative abundances. In fact, the study of complex landscapes actually involves new aspects of the quasi-species model under active investigation. A concrete example is the dynamic fitness landscape: situations in which the replication or/and decay rates of the replicators change over time [38]. Another is neutral evolution. It is believed that many mutations of a molecule are evolutionarily neutral in the sense that they do not alter the fitness to perform the function for which it has been selected [39]. Despite the long history of the concept, many aspects of neutral evolution are still not well understood, and there are many interesting investigations in this line [40].

7 The informational crisis

Returning to equations (17), we know that if $Q_{ii} = 1$ there will be no evolution. It thus seems intuitively evident that a decrease of Q_{ii} will increase the novelty in the system and the 'velocity' of its evolutionary process. But as we saw, Q_{ii} can not decrease below the error threshold without compromising the stability of the quasi-species. We can make this fact explicit by taking a polymer chain of ν nucleotides (a master sequence) and considering its reproduction as an easy *branching processes* [41]. Let q be the copy faithfulness per nucleotide. Then our chain will be copied with exact fidelity with probability q^ν ($q^\nu = Q_{ii}$), and a copy will have at least one "bad" nucleotide with probability $(1-q^\nu)$. If we define the survival probability of a chain to hydrolysis in one generation as w, then $(1-w)$ will be the probability of disappearing via hydrolysis. Note that in one generation a molecule will have 0 offsprings with probability $(1-w)$ (the molecule degrades by hydrolysis), 1 with probability $w(1-q^\nu)$ (the molecule survives but its only offspring is a mutation, therefore only one copy of itself remains at the end of the process), and have 2 descendants with probability wq^ν (case of survival and faithful copy). The population of a given "correct" polynucleotide evolves following a branching processes of mean:

$$m = 0(1-w) + 1w(1-q^\nu) + 2wq^\nu = w(1+q^\nu) \tag{39}$$

The extinction is secure if: $w(1+q^\nu) \leq 1$, i.e., if

$$q^\nu \leq \frac{1-w}{w}. \tag{40}$$

Thus, for an indefinite survival of the molecule we must have

$$\nu < \frac{1}{\log q} \log \frac{1-w}{w}, \tag{41}$$

if $q \approx 1$ then we can approximate $\log q \approx q-1$ to get

$$\nu < \frac{1}{1-q} \log \frac{w}{1-w}.$$

This condition imposes a limit to the size ν of the molecule in order to avoid extinction, that is, a limit on the amount of information that can be stored in the master sequence.

Let us study this in the framework of our differential equations. We will try to determine the concentration $x_m(t)$ of the master sequence. If we suppose that the contribution to the master sequence concentration coming from mutations from other species is negligible, then from the equation (17) we get:

$$\frac{dx_m}{dt} \approx (W_{mm} - \langle E(t)\rangle)x_m(t). \tag{42}$$

Thus in the selective equilibrium, $\frac{dx_m}{dt} = 0$, we will have

$$\langle E(t \to \infty)\rangle = W_{mm}. \tag{43}$$

Let us define, following [42], the mean productivity of all the mutants except the master sequence as:

$$\langle E_{j\neq m}(t)\rangle = \frac{\sum_{j\neq m}^{s} E_j(t)x_j(t)}{\sum_{j\neq m}^{s} x_j(t)} = \frac{\sum_{j=1}^{s} E_j(t)x_j(t) - E_m(t)x_m(t)}{1 - x_m(t)}. \tag{44}$$

Then the master sequence concentration can be expressed as

$$x_m(t) = \frac{\langle E(t)\rangle - \langle E_{j\neq m}(t)\rangle}{E_m(t) - \langle E_{j\neq m}(t)\rangle}, \tag{45}$$

and in the selective equilibrium ($t \to \infty$)

$$x_m^* = \frac{W_{mm} - \langle E_{j\neq m}(\infty)\rangle}{E_m(\infty) - \langle E_{j\neq m}(\infty)\rangle}. \tag{46}$$

We see that, as Q_{ii} decreases the mutant cloud becomes more and more important: its mean productivity $\langle E_{j\neq i}(t)\rangle$ grows until it reaches the selective value of the master sequence W_{ii}. At this moment the master sequence is doomed to extinction. The information stored in it is lost. The error catastrophe has taken place. The value Q_{ii} for which the error catastrophe takes place thus corresponds of course to the error threshold, Q_c. Equating expression 46 to zero, we get

$$Q_c = \frac{D_m + \langle E_{j\neq m}\rangle}{A_m} = \frac{1}{\Theta} \tag{47}$$

where we have defined Θ, the *superiority* of the master sequence as the inverse of Q_c. Since as previously stated, $Q_{ii} = q^\nu$, at the threshold

$$\nu_c = -\frac{\log \Theta}{\log q} \approx \frac{\log \Theta}{1 - q} \tag{48}$$

Thus, in order to increase the length of the master sequence without getting into the error catastrophe either we should increase the superiority (increasing the replication rate A_m, decreasing the mean productivity of the mutants $\langle E_{j\neq m}\rangle$ and/or decreasing the degradation rate D_m) or we increase the fidelity q.

In the next table [43] we show the maximum possible sizes of different types of chains as a function of a variety of values of q and Θ. Together with them, the approximate

Quality factor	Superiority	Maximum size	Example
q	**Θ**	$\nu_{\max}$	
0.95	2	14	tRNA precursors
	20	60	($\nu \approx 80$)
	200	106	
0.9995	2	1386	RNA of phage Qβ
	20	5991	($\nu \approx 4500$)
	200	10597	
0.999995	2	0.7×10^6	Escherichia coli
	20	3.0×10^6	($\nu \approx 4 \times 10^6$)
	200	5.3×10^6	
0.99999995	2	0.7×10^9	Homo sapiens
	20	3.0×10^9	($\nu \approx 3 \times 10^9$)
	200	5.3×10^9	

Table 1 Maximum possibles sizes of different examples of chains as a function of a variety of values of quality factor q and superiority Θ.

sizes of tRNA, RNA of the phage Qβ, and DNA chains of Eschericia coli and of human beings are shown. The early replicators necessarily have a small size. The examples chosen show an increase in the complexity of the replicating entities as the sizes of their corresponding chains grow. *Grosso modo*, one could say that longer chains store more information. Note how if one does not want to fall prey to the error catastrophe, one should enormously increase the quality factor. If we assume the average point mutation rate to be approximately constant for all molecules, the only way to increase the quality factor is through code correctors, enzymes. However, a DNA of extraordinary complexity (and length!) is needed in order to be able to reproduce and have the enzymes at our disposal, given their prodigious specialties. Whence we find ourselves at a Gordian knot, a chicken-egg type of dilemma: What came first, the corrector enzymes or the DNA capable of making them? It seems to be impossible to increase the quality factor without long sequences, but it seems impossible to replicate long sequences without increasing the quality factor. A *cul de sac* for the increase of complexity in our model? An insuperable crisis of information?

8 The hypercycle

In figure 11(a) we see a group of different quasi-species. As we know, in our model quasi-species compete amongst them for selective equilibrium. Only one quasi-species win, and

only its information is kept. The information carried by the rest of quasi-species is lost. However, if we want to solve the informational crisis we will need the coexistence of many quasi-species. As was proposed by M. Eigen and P. Schuster [44], cooperation can be the way out to leave the information crisis behind us: mutual catalysis between quasi-species is a possible solution. As we shall now see, higher order interactions between quasi-species make coexistence possible. The model, know as *the hypercycle* [45], is formed by a set of catalytic reactions where some quasi-species catalyze the self-replication of other quasi-species.

The hypercycle model applies to any catalytic chain or cycle in a broad sense (i.e., chemical reactions, ecological interactions, etc...). In a general form, we can write the hypercycle equation as [46]

$$\frac{dq_i}{dt} = \Gamma_i(\mathbf{q}(t), \mathbf{k}) - \Phi(t) q_i(t), \qquad i = 1, \ldots, n \tag{49}$$

where $\mathbf{q}(t) = (q_i(t), \ldots, q_n(t))$ is a vector of temporal concentrations of n quasi-species, $\mathbf{k} = (k_i, \ldots, k_n)$ is a vector of kinetic constants and $\Phi(t)$ is a flux term to account for the CP constraint. Using the fact that the total concentration satisfies $\sum_{i=1}^{n} q_i(t) = 1$,, adding up all the equations and setting the sum of the derivatives equal to zero we obtain a value for the flux:

$$\Phi(t) = \sum_{i=1}^{n} \Gamma_i(\mathbf{q}(t), \mathbf{k}),$$

and the equations can be rewritten as

$$\frac{dq_i}{dt} = \Gamma_i(\mathbf{q}(t), \mathbf{k}) - \Big(\sum_{j=1}^{n} \Gamma_j(\mathbf{q}(t), \mathbf{k})\Big) q_i(t) \qquad i = 1, \ldots, n. \tag{50}$$

The typical ansatz for Γ_i is:

$$\Gamma_i = k_i + \sum_{j=1}^{n} k_{ij} q_j(t) + \sum_{j=1}^{n} \sum_{l=1}^{n} k_{ijl} q_j(t) q_l(t) + \ldots \tag{51}$$

where the contributions of order zero, first, second, etc. appear explicitly.

If we couple the quasi-species as in figure 11(b) in acyclical catalytic chains, we obtain:

$$\begin{aligned} \Gamma_1(t) &= k_1 q_1(t) \\ \Gamma_i(t) &= k_i q_i(t) q_{i-1}(t), \qquad i = 2, \ldots, n \end{aligned} \tag{52}$$

The net effect of this type of coupling is that the growth of each quasi-species is promoted by its predecessor in the chain. It can be shown [46] that this coupling usually does not result in the coexistence of all the quasi-species in the chain (and indeed it is not difficult to convince oneself that this must be the case). This is also the case for lineal chains with branches.

The next simple structure that we can study is a cyclic chain (this the hypercycle proper), that is, a structure like the one depicted in figure 11(c). Mathematically,

$$\begin{aligned} \Gamma_1(t) &= k_1 q_1(t) q_n(t) \\ \Gamma_i(t) &= k_i q_i(t) q_{i-1}(t), \qquad i = 2, \ldots, n \end{aligned} \tag{53}$$

If any k_i is equal to zero the hypercycle becomes an acyclical chain, and we go back to the previous case (52). The cyclical chain that we have just introduced is the so-called elemental hypercycle. Contrary to the acyclical chains above, in this case the coexistence of species is possible since there is no net benefactor in the cycle: all the elements contribute and they all benefit.

Let us consider some dynamical properties of elemental hypercycles. The basic equations are:

$$\frac{dq_1}{dt} = k_1 q_1(t) q_n(t) - q_1(t)\Big(k_1 q_1(t) q_n(t) + \sum_{r=2}^{n} k_r q_r(t) q_{r-1}(t))\Big)$$
$$\frac{dq_i}{dt} = k_i q_i(t) q_{i-1}(t) - q_i(t)\Big(k_1 q_1(t) q_n(t) + \sum_{r=2}^{n} k_r q_r(t) q_{r-1}(t))\Big), \quad i = 2, \ldots, n \quad (54)$$

In figure 12 we show the numerical solutions of these equations. The left hand side figures represent relative concentrations of the quasi-species in time, while the figures on the right represent the phase portrait for $q_1(t)$ and $q_2(t)$. For all these cases we have taken all the coupling constants to be $k_i = 1.0$. In (a) and (b) we have the equations 54 with $n = 2$ and initial conditions $\mathbf{q}(t = 0) = (0.99, 0.01)$. As we see the concentrations of the two quasi-species tend to an equilibrium value. This is depicted in the $q_1(t) - q_2(t)$ plane as a attractive focus.

In (c) and (d) we take $n = 3$ with initial conditions $\mathbf{q}(t = 0) = (0.98, 0.01, 0.01)$. We observe damped oscillations that quickly tend to a focus equilibrium in a spiral way.

In (e) and (f), the hypercycle is formed by $n = 4$ quasi-species. We can see that after a transient state with damping, in the final state the concentrations oscillate periodically (as we run through the cycle). In the $q_1(t) - q_2(t)$ plane we can see this as a spiral approximation to a limit cycle.

In figure 13(a) and 12(b) we take $n = 5$ and initial conditions $\mathbf{q}(t = 0)$ =(0.96, 0.01, 0.01, 0.01, 0.01). We see how the dynamics is dominated again by periodic oscillations, while in the $q_1(t) - q_2(t)$ figure we see that the limiting cycle is reached rapidly.

For values $n \geq 5$ the existence of a limiting cycle seems to hold, although this cannot be shown analytically. In (c) and (d) we take $n = 10$ with initial conditions $\mathbf{q}(t = 0) = (0.91, 0.01, ..., 0.01)$ and again we obtain cyclic behavior.

As the figures show, for $n = 2$ and $n = 3$ the concentrations tend to a constant value, whereas for $n \geq 4$ in general we will have oscillations and coexistence is always possible for all the quasi-species (something which is not always the case for $n < 4$). In (e) and (f) we take $n = 3$ with initial conditions $\mathbf{q}(t = 0) = (0.98, 0.01, 0.01)$ and asymmetric coupling constants $\mathbf{k} = (1, 1, 0.1)$. We see that the results are similar to the ones depicted above.

As the reader may imagine the elemental hypercycles are the simplest system that can avoid the information crisis. The catalytic terms can be more sophisticated and the order of the ansatz for Γ can however be higher. These combinatorial possibilities open up a true bestiarium of systems in which dynamical behaviors of all types appear, including chaos.

The hypercycle solution to the information crisis, however, has two main handicaps [18]. The first one is that selection of hypercycles operates in a 'once and for all' basis. If a mutant hypercycle with selective advantage appears it will not displace the presently dominating cycle, in contrast to the quasi-species model (remember Tierra for example). Why this difference between the metastability of hypercycles and quasi-species? With no CP conditions the quasi-species grows exponentially, thus the population becomes infinite in infinite time. But it is easy to see [46] that for elemental hypercycles without CP the increase of the population is hyperbolic. Thus an infinite population is reached in a finite time. This fact is a hard blow for the model. The price for eliminating the limited size constraint imposed by the error threshold is the incapacity for evolution! The second handicap is the instability of hypercycles against parasites. A parasite in the model is a branch in the elemental hypercycle. It is a quasi-species that benefits from catalysis but which does not contribute to the cycle. It seems that the hypercycle does not support parasites, it is rendered unstable by them.

A variety of solutions have been proposed to overcome these problems. We can have evolution within the hypercycle if the hypercycle itself acts as a core for growth, incorporating more and more quasi-species. But, the most interesting prospects may lie in the incorporation of spatial dependence to the model. Preliminary results show that, in this way, stability of hypercycles against parasites can be achieved [47]. However, the spatially explicit model is stable against parasites only if the number of coupled quasi-species is greater than 4. But the spatially explicit system is shown to be resistant against only the so-called "selfish parasites", whereas the "short-cut" parasites can invade it, decreasing the number of coexisting quasi-species in the hypercycle. The error threshold was a very serious problem for the Eigen model, which justified the introduction of the hypercycle concept, because at the time the general understanding was that only proteins can behave as a catalyst. A protein synthesis requires cooperation of roughly 100 genes, and the minimal organism with catalysis was considered quite complicated. The current view is, that RNA molecules can store information and play the role of enzymes, so the minimal organism requires only a single gene, which is able to reproduce itself. In this sense, the hypothesis of RNA catalysts reformulates the problem completely. See [48] and references therein, for an up-to-date discussion. Anyway, at this moment this is an active area of research [49].

However, the hypercycle is not the only possible resolution of the error-catastrophe problem, there are numerous alternative hypotheses. First of all there are other experimentally tested mechanisms (the product inhibition) which could maintain the coexistence of different replicators [50], and non-perfect mixing can have the same effect in open chaotic flows [51]. The other possibility is that co-evolution of replicator fidelity and template function can lead to the overcoming error catastrophe [48, 52].

9 A final remark

In 1994 E. Schrödinger wrote a little book called: "What Is Life?" [53] This essay posed basic and precise questions: How is the information to build and maintain an organism stored, read and executed at a molecular level? How is it copied and transmitted from generation to generation? These questions determined research in the field in the next decades. But subtly Schrödinger laid aside questions difficult to answer in the near future. To be specific, he did not speak about metabolism. Without really saying it, he replaced the complex problem of the origin of Life by the more simple question of how information is replicated. As mentioned before in the second paragraph of this work: "we can rephrase the old question about the origin of life as: What is the origin of the information stored in living systems, and why is it stable?" Sixty years later, metabolism is still aside so we ignore how thousands of molecules are organized within a cell. Today we find ourselves in the post-genomic era having sequenced several organisms, including a human being. Metabolism is our next big theoretical problem to understand the origin of life.

Acknowledgments

We would like to thank Antonio Ferrera, Fernando Ballesteros, Jordi Bascompte, Aida Agea, Ricardo García-Pelayo and two anonymous referees for their valuable contributions and opinions. This work has been supported by CICYT BFM2002-01812.

References

[1] (a) F. Crick: *Life Itself: Its Nature and Origin*, Simon and Schuster, Nueva York, 1981. (b) A.G. Cairns-Smith: *Seven Clues to the Origin of Life*, Cambridge University Press, Cambridge, 1985.

[2] M. Smith and E.Szathmáry: *The Major Transitions in Evolution*, Oxford University Press, New York, 1995.

[3] (a) D.R. Brooks and D.A. McLennan: *Phylogeny, Ecology, and Behavior*, University of Chicago Press, Chicago, 1991. (b) P.H. Harvey and M.D. Pagel: *The Comparative Method in Evolutionary Biology*, Oxford University Press, Oxford and New York, 1991.

[4] (a) J. Monod: *El azar y la necesidad*, Tusquets, Barcelona, 1981. (b) R. Shapiro: *Origins: A Skeptic's Guide to the Creation of Life on Earth*, Summit Books, Nueva York, 1986. (c) C. de Duve: *Vital Dust: Life as a Cosmic Imperative*, Basic Books, Nueva York, 1995. (d) L. Smolin: *The Life of the Cosmos*, Cambridge University Press, Cambridge, 1997.

[5] (a) B.O. Küpers: *Information and the Origin of Life*, Mass.: MIT Press, Cambridge, 1990. (b) H. Yockey: *Information Theory and Molecular Biology*, Cambridge University Press, Cambridge, 1992. (c) P. Davies: *The fifth miracle: The search for the origin and meaning of life*, Simon and Schuster, New York, 1999.

[6] Y. Prigogine and I. Stengers: *Order Out of Chaos*, Heinemann, London, 1984.

[7] (a) W. Gilbert: "The RNA World", *Nature*, Vol. 319, (1986), pp. 618. (b) G.F. Joyce: "RNA evolution and the origins of life", *Nature*, Vol. 338, (1989), pp. 217–224. (c) C.K.

Biebricher, and W. C. Gardiner: "Molecular evolution of RNA in vitro", *Biophys. Chem*, Vol. 66, (1997), pp. 179. (d) C. O. Wilke: "Selection for Fitness vs. Selection for Robustness in RNA Secondary Structure Folding", *Evolution*, Vol. 55, (2001), pp. 2412-2420.

[8] G. Joyce: "The antiquity of RNA-based evolution", *Nature*, Vol. 418, (2002), pp. 214–221.

[9] M. Eigen: *Steps toward Life*, Oxford University Press, Oxford, 1985.

[10] B. Drossel: "Biological Evolution and Statistical Physics", *Advances in Physics*, Vol. 50, (2001), pp. 209–295.

[11] L. Peliti: *Introduction to the statistical theory of Darwinian evolution*, Lectures at the Summer College on Frustrated System, Trieste, August 1997.

[12] B.M.R. Stadler and P.F.Stadler: "Molecular replicator dynamics", *Santa Fe working paper*, (2000), http://www.santafe.edu/sfi/publications/wpabstract/2002090409.

[13] (a) J.A.G.M. de Visser: "The fate of microbial mutators", *Microbiology*, Vol. 148, (2002), pp. 1247–1252. (b) J.E. LeClerc et al.: "Hypermutable Strains Among E. coli and Salmonella Pathogens", *Science*, Vol. 274, (1996), pp. 1208-1211. (c) P.D. Sniegowski, P. Gerrish, R. E. Lenski: "Evolution of high mutation rates in experimental populations of E. coli", *Nature*, Vol. 387, (1997), pp. 703705. (d) A. Giraud et al.: "Costs and Benefits of High Mutation Rates: Adaptive Evolution of Bacteria in the Mouse Gut", *Science*, Vol. 001, (2001), pp. 2606-2608.

[14] S. Spiegelman: "An approach to the experimental analysis of precellular evolution", *Quart. Rev. Biophys*, Vol. 4, (1997), pp. 213-253.

[15] (a) M. Eigen: "Selforganization of matter and the evolution of biological macromolecules", *Naturwissenshaften*, Vol. 58, (1971), pp. 465–523. (b) C.K. Biebricher, M. Eigen, W.C. Gardiner: "Kinetics of RNA replication competition and selection among self-replicating RNA species", *Biochemistry*, Vol. 24, (1983), pp. 6550–6560. (c) M. Eigen, J. McCaskill, P. Schuster: "Molecular quasispecies", *Phys. Chem*, Vol. 92, (1988), pp. 6881–6891. (d) M. Eigen, J. McCaskill, P. Schuster: "The molecular quasispecies", *Adv. Chem. Phys*, Vol. 75, (1989), pp. 149–263. (e) M. Eigen: "Viral quasispecies", *Scientif. Am*, Vol. 269, (1993), pp. 32–39.

[16] M.A. Nowak: "What is a quasispecies?", *Trends in Ecol. Evol*, Vol. 7, (1991), pp. 118.

[17] A.V. Ratner et al.: *Molecular Evolution*, Springer Verlag, Berlin, 1996.

[18] G.W. Rowe: *Theoretical Models in Biology*, Oxford University Press, New York, 1997.

[19] (a) C. Adami: *Introduction to Artificial Life*, TELOS Springer Verlag, Santa Clara, 1998. (b) C.O. Wilke et al.: "Evolution of Digital Organisms at High Mutation Rate Leads To Survial of the Flattest", *Nature*, Vol. 412, (2001), pp. 331-333.

[20] T.S. Ray: *An approach to the synthesis of life*, In: Artificial Life II: A Proccedings Volume in the Santa Fe Institute the Sciences of Complexity, Vol. 10, Addison-Wesley, Redwood City, 1991, pp. 371–408.

[21] (a) M. Nowak and P. Schuster: "Error Thresholds of Replication in Finite Populations", *J. Theor. Biol*, Vol. 137, (1989), pp. 375395. (b) Y-C Zhang: "Quasispecies Evolution of Finite Populations", *Phys. Rev. E*, Vol. 55, (1997), pp. R3817–R3819. (c) T. Wiehe: "Model dependency of error thresholds: the role of the fitness finctions and contrasts between the finite and infinite sites models", *Genet. Res. Camb*, Vol. 69, (1997), pp. 127-136.

[22] (a) L. Chao: "Fitness of RNA virus decreased by Muller's ratchet", *Nature*, Vol. 348, (1990), pp. 454. (b) G.P. Wagner and P. Krall: "What is the difference between models of error thresholds and Muller's ratchet?", *J. Math. Biol*, Vol. 32, (1993), pp. 33-44. (c) W. Gabriel and R. Burger: "Fixation of Clonal Lineages under Muller's Ratchet", *Evolution*, Vol. 54, (2000), pp. 1116-1125.

[23] M. Mitchell: *An Introduction to Genetic Algorithms*, Mass: MIT Press, Cambridge, 1996.

[24] C. Adami: "Self-organized criticality in living systems", *Phys. Lett. A*, Vol. 203, (1995), pp. 29–32.

[25] H.J. Jensen: *Self-Organized Criticality: Emergent Complex Behavior in Physical and Biological Systems*, Cambridge University Press, New York, 2000.

[26] M. E. Newman et al.: "Self-organized criticality in living systems- Comment", *Psys. Lett. A*, Vol. 228,(1997), pp. 202.

[27] J.R. Lorsch and J.W. Szostak: "*In vitro* evolution of new ribozymes with polynucleotide kinase activity", *Nature*, Vol. 371, (1994), pp. 31–36.

[28] S. Kauffman: *Origins of Order*, Oxford University Press, Oxford, 1990.

[29] (a) J. Swetina and P. Schuster: "Self-Replication with error - a model for polynucleotide replication", *Biophys. Chem*, Vol. 16, (1982), pp. 329-345.
(b) P. Schuster, J. Swetina: "Stationary mutant distributions and evolutionary optimization", *Bulletin of Mathematical Biology*, Vol. 50, (1988), pp. 329–345.

[30] R.V. Solé et al: "Red Queen dynamics, competition and critical points in a model of RNA virus quasispecies", *J. theor. Biol*, Vol. 198, (1999), pp. 47–59.

[31] (a) C.L. Burch and L. Chao: "Evolvability of RNA virus is determined by its mutation neighborhood", *Nature*, Vol. 406, (2000), pp. 625–628. (b) D.C. Krakauer, J.B. Plotkin: "Redundancy, antiredundancy, and robustness of genomes", *Proc. Nat. Acad. Sci*, Vol. 99, (2002), pp. 1405–1409.

[32] (a) E. Domingo and J.J. Holland: *The evolutionary biology of viruses*, S. Morse(ed.) Raven Press, New York, 1994. (b) E. Domingo and J.J. Holland: "RNA virus mutations and fitness for survival", *Annu. Rev. Microbiol*, Vol. 51, (1997), pp. 151. (c) E. Domingo, C.K. Biebricher, M. Eigen, J.J. Holland: *Quasispecies and RNA virus evolution: Principles and consequences*, Landes Bioscience, Georgetown, 2001. (d) S. Crotty, C.E. Cameron, R. Andino: "RNA virus error catastrophe: direct molecular test using ribavirin", *Proc. Nat. Acad. Sci*, Vol. 98, (2001), pp. 6895-6900.

[33] (a) I. Leuthaüsser: "Statistical mechanics of Eigen's evolution model", *J. Stat. Phys*, Vol. 48, (1987), pp. 343. (b) I. Leuthaüsser: "An exact correspondence between Eigen evolution model and two-dimensional ising system", *J. Chem. Phys*, Vol. 84, (1986), pp. 1884–1885.

[34] (a) P. Tarazona: "Error threshold for molecular quasispecies as phase transitions: From simple landscapes to spin-glass models", *Phys. Rev. A*, Vol. 45, (1992), pp. 6038–6050. (b) D. Alves and J.F. Fontanari: "Population genetics approach to the quasispecies model", *Phys. Rev. E*, Vol. 54, (1996), pp. 4048. (c) E. Baake, M. Baake, H. Wagner: "Ising quantum chain is equivalent to a model of biological evolution", *Phys. Rev. Lett*, Vol. 78, (1997), pp. 559–562. (d) E. Baake, M. Baake, H. Wagner: "Quantum mechanics versus classical propability in biological evolution", *Phys. Rev. E*, Vol. 57, (1998), pp. 1191-1192.

[35] (a) T. Halpin-Healy and Y.C. Zhang: "Kinetic Roughening, Stochastic Growth, Directed Polymers and All That", *Phys. Rep*, Vol. 254, (1995), pp. 215. (b) A. Galluccio, R. Graber, Y.C. Zhang: "Diffusion on a hypercubic lattice with pinning potential: exact results for the error-catastrophe problem in biological evolution", *J. Phys. A*, Vol. 29, (1996), pp. L249. (c) S. Galluccio: "Exact solution of the quasispecies model in a sharply-peaked landscape", *Phys. Rev. E*, Vol. 56, (1997), pp. 4526.

[36] (a) P.W. Anderson: *Proc. Natl. Acad. Sci. USA*, Vol. 80, (1983), pp. 3386–3390. (b) L. Demetrius: "Statistical mechanics and population biology", *J. Stat. Phys*, Vol. 30, (1983), pp. 709-753. (c) C. Amitrano, L. Peliti, M. Saber: "Population dynamics in a spin-glass model of chemical evolution", *J. Mol. Evol*, Vol. 29, (1989), pp. 513. (d) D. Alves and J.F. Fontanari: "Error threshold in the evolution of diploid organisms", *J. Phys. A Math. Gen*, Vol. 30, (1997), pp. 2601-2607. (e) P.R.A. Campos and J.F. Fontanari: "Finite-size scaling of the quasispecies model", *Phys. Rev. E*, Vol. 58, (1998), pp. 2664-2667. (f) P.R.A. Campos and J.F. Fontanari: "Finite-size scaling of the error threshold transition", *J. Phys. A*, Vol. 32, (1999), pp. L1-L7. (g) J. Hermisson, H. Wagner, M. Baake: "Four-state quantum chain as a model for sequence evolution", *J. Stat. Phys*, Vol. 102, (2001), pp. 315-343. (h) E. Baake and W. Gabriel: "Biological evolution through mutation, selection, and drift: An introductory review", *Ann. Rev. Comp. Phys. VII, World Scientific*, Vol. 9, (2000), pp. 203–264.

[37] (a) R. García-Pelayo: "A linear algebra model for quasispecies", *Physica A*, Vol. 309, (2002), pp. 131–156. (b) T. Wiehe: "Model dependency of error thresholds: the role of fitness functions and contrasts between the finite and infinite sites models", *Genet. Res. Camb*, Vol. 69, (1997), pp. 127.

[38] (a) S. Franz and L. Peliti: "Error threshold in simple landscapes", *J. Phys. A*, Vol. 30, (1997), pp. 4481. (b) M. Nilsson and N. Snoad: "Optimal mutation rates on dynamic landscapes", *Phys. Rev. Lett*, Vol. 84, (2000), pp. 191-194. (c) C.O. Wilke, C. Ronnewinkel, T. Martinetz: "Dynamic fitness landscapes in molecular evolution", *Phys. Rep*, Vol. 349, (2001), pp. 395–446.

[39] (a) M. Kimura: "The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations", *Genetics*, Vol. 61, (1969), pp. 893. (b) M. Kimura: *The Neutral Theory of Molecular Evolution*, Cambridge University Press, Cambridge, 1983. (c) M. Kimura, *Population Genetics, Molecular Evolution, and the Neutral Theory: Selected Papers*, The University of Chicago Press, Chicago, 1994.

[40] (a) C. Reidys, P. Stadler, P. Schuster: "Replication and mutation on neutral networks", *Bull. Math. Biol*, Vol. 59, (1997), pp. 339. (b) M.E.J. Newman and R. Engelhardt: "Effects of selective neutrality on the evolution of molecular species", *Proc. R. Soc. London B*, Vol. 265, (1998), pp. 1333–1338.

[41] (a) J. Hofbauer and K. Sigmund: *The Theory of Evolution and Dynamical Systems*, Cambridge University Press, Cambridge, 1988. (b) L. Demetrius, P. Schuster, K. Sigmund: "Polynucleotide evolution and branching processes", *Bull. Math. Biol.* Vol. 47, (1985), pp. 239-262.

[42] F. Montero and F. Morán: *Biofísica*, Eudema, Madrid, 1992.

[43] F. Montero, C.E. Sanz, M. A. Andrade: *Evolucin Prebiótica*, Eudema, Madrid, 1993.

[44] M. Eigen and P. Schuster: *The hypercycle: a principle of natural self-organization*, Springer-Verlag, Berlin, 1979.

[45] (a) J.M. Smith: " Hypercycles and the origin of life", *Nature*, Vol. 280, (1979), pp. 445–446. (b) R.M. May: "Hypercycles spring to life", *Nature*, Vol. 353, (1991), pp. 607.

[46] (a) B.O. Küpers: *Molecular Theory of Evolution*, Springer Verlag, Berlin, 1988. (b) R. Burger: *The Mathematical Theory of Selection, Recombination, and Mutation*, Wiley, Chichester, 2001.

[47] (a) M. Boerlijst and P.Hogeweg: "Spiral wave structure in pre-biotic evolution: hypercycles stable against parasites", *Physica D*, Vol. 48, (1991), pp. 17. (b) M.B. Cronhjort: "Hypercycles versus parasites in the origin of life: model dependence in spatial hypercycle dynamics", *Origins Life Ecol. Biosphere*, Vol. 25, (1993), pp. 227. (c) M.B. Cronhjort and C. Blomberg: "Hypercycles versus parasites in a two-dimensional partial differential equations", *J. theor. Biol*, Vol. 169, (1994), pp. 34.

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

[49] D.H. Lee, K. Severin, Y. Yokobayashi, M.R. Ghadiri: "Emergence of Symbiosis in Peptide Self-Replication through a Hypercyclic Network", *Nature*, Vol. 390, (1997), pp. 591-594.

[50] (a) E. Szathmáry: "The Integration of the Earliest Genetic Information", *Trends in Ecology and Evolution*, Vol. 4, (1989), pp. 200–204. (b) D. Sievert and G. von Kiedrowski: "Self-replication of Complementary Nucleotide Based Oligomers", *Nature*, Vol. 369, (1994), pp. 221–224. (c) I. Scheruing, and E. Szathmáry: "Survival of Replicators with Parabolic Growth Tendency and Exponential Decay", *J. theor. Biol*, Vol. 212, (2001), pp. 99–105. (d) B.M.R. Stadler, P.F. Stadler, P.R. Wills: "Evolution in Systems of Ligation Based Replicators", *Z. Phys. Chem*, Vol. 216, (2002), pp. 21–34.

[51] (a) G. Károlyi et al.: "Chaotic Flow: the Physics of Species Coexistence", *Proc. Nat. Acad. Sci*, Vol. 97, (2000), 13661–13665. (b) I. Scheuring et al.: "Competing Populations in Flows with Chaotic Mixing", *Theor. Popul. Biol*, Vol. 63, (2003), pp. 77–90.

[52] (a) A. Poole, D. Jeffares, D. Penny: "Early Evolution: Prokaryotes, the New Kids on the Block", *BioEssays*, Vol. 21, (1999), pp. 880–889.

[53] E. Schrödinger: *What Is Life?*, Cambridge University Press, Cambridge, 1944.

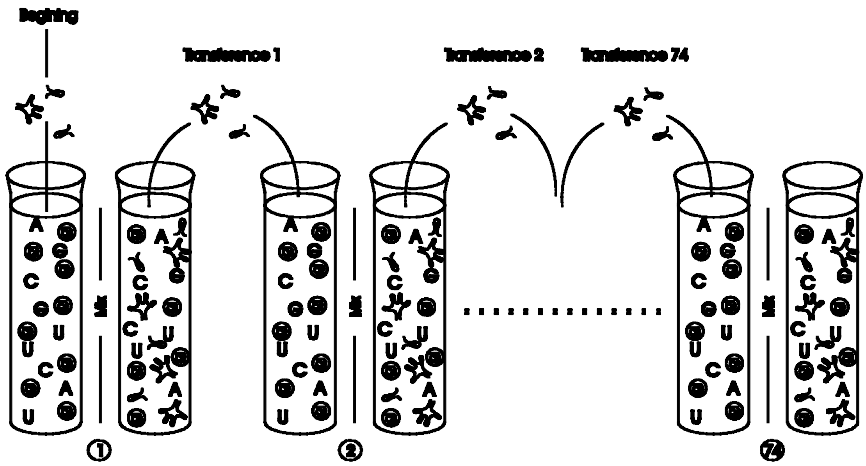
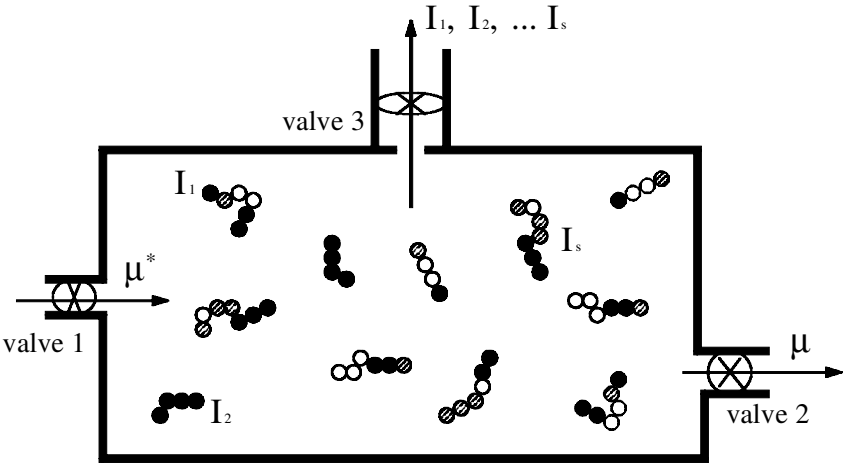
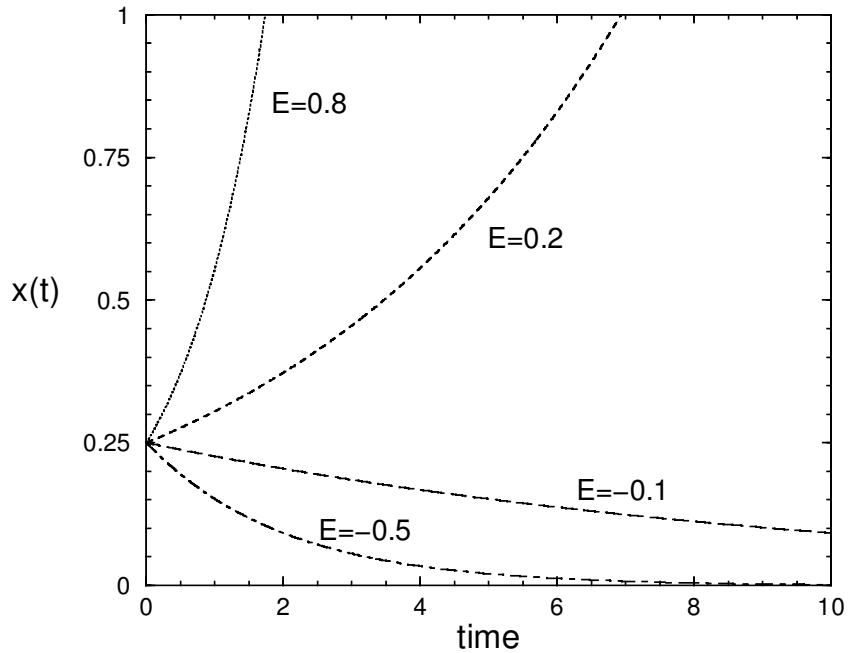


Fig. 1 Experiment of evolution in vitro of phago Q-β. A tube collection with Q-β*replicasa* along with monomers like ATP, GTP, UTP or CTP is assembled. We begin inserting phago Q-β in the first tube. After a few hours we transfer an aliquot of the first tube into a second tube. And repeat successively to order of seventeen twice. The result is a RNA virus that has a length of only 17% its original one, and was reproducing at a rate 15 times faster.

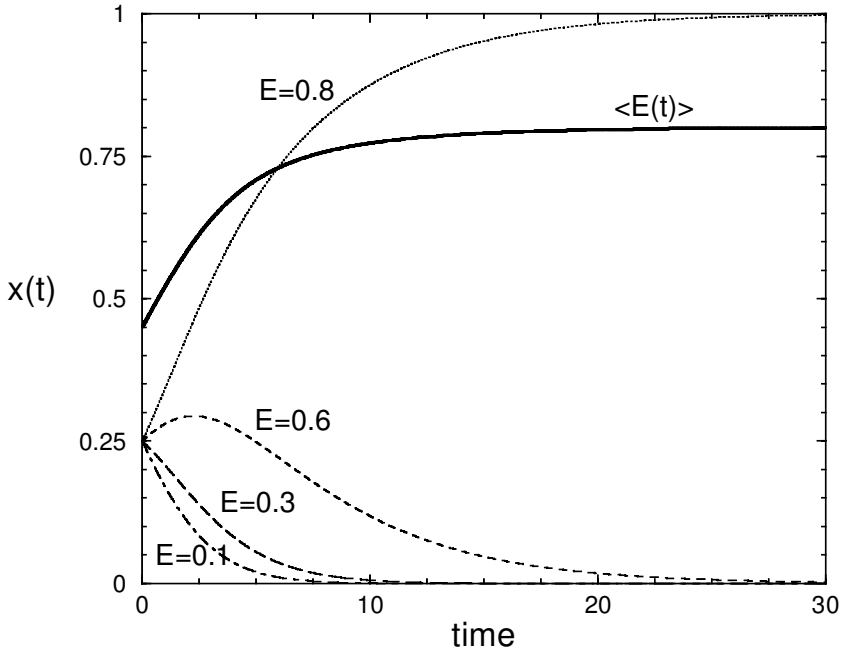


(a)

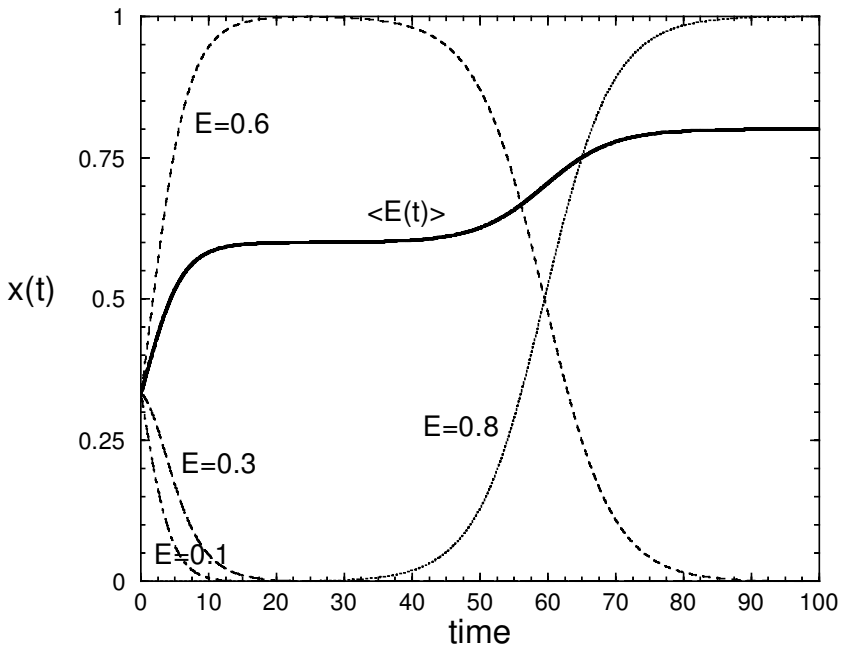


(b)

Fig. 2 (a): Schematic flux reactor for molecular evolution. The self-replicating molecules are represented as ball chains. Rich energetic monomers are introduced by valve 1 and residual products are eliminated by valve 2. We can maintain the molecular population constant by valve 3. (b): Segregation process between survival and extinct molecular species. In the model without restrictions only the molecules with positive productivity survive and grow without limit.



(a)



(b)

Fig. 3 (a): Simulation for four species with constant population conditions. The best replicator dominates the system. The selective equilibrium is reached in a few steps. (b): We repeat the same numerical experiment as in figure (a) but without the best replicator. The selective equilibrium is reached for the molecular species with productivity $E = 0.6$, the "fittest" of the three initial species. At time $t = 25$ we introduce an infinitesimal amount of a species with $E = 0.8$. A new process of selection takes place and a new selective equilibrium is reached.

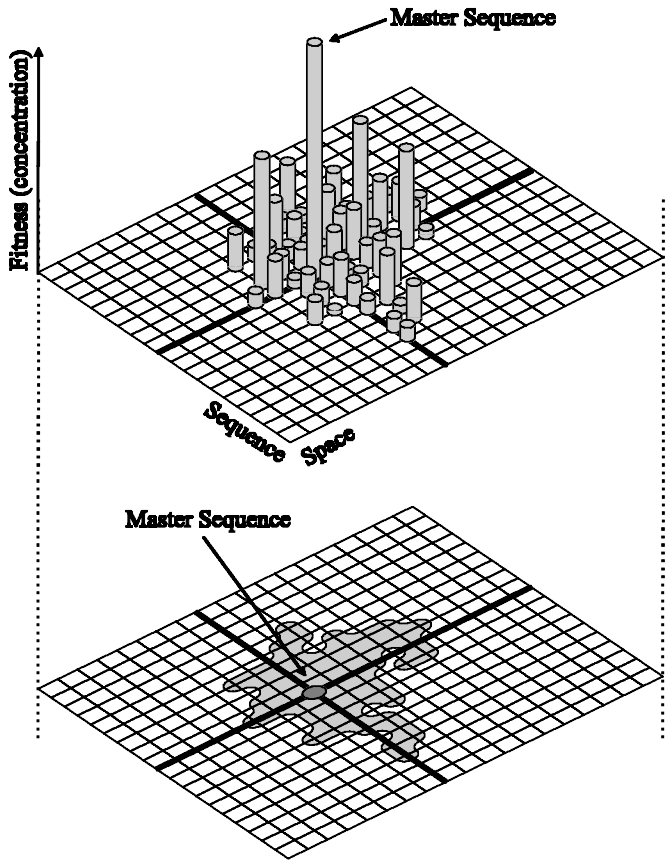
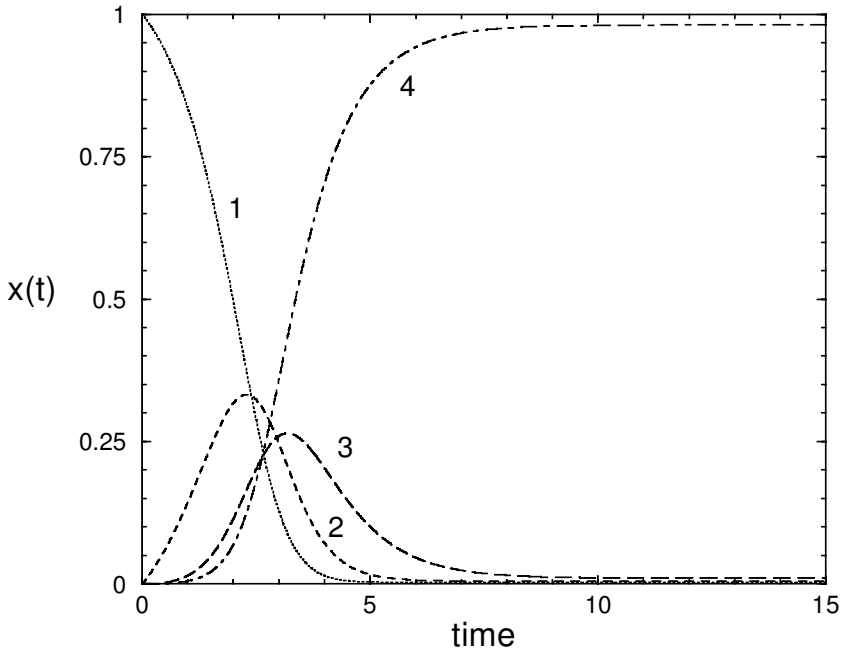
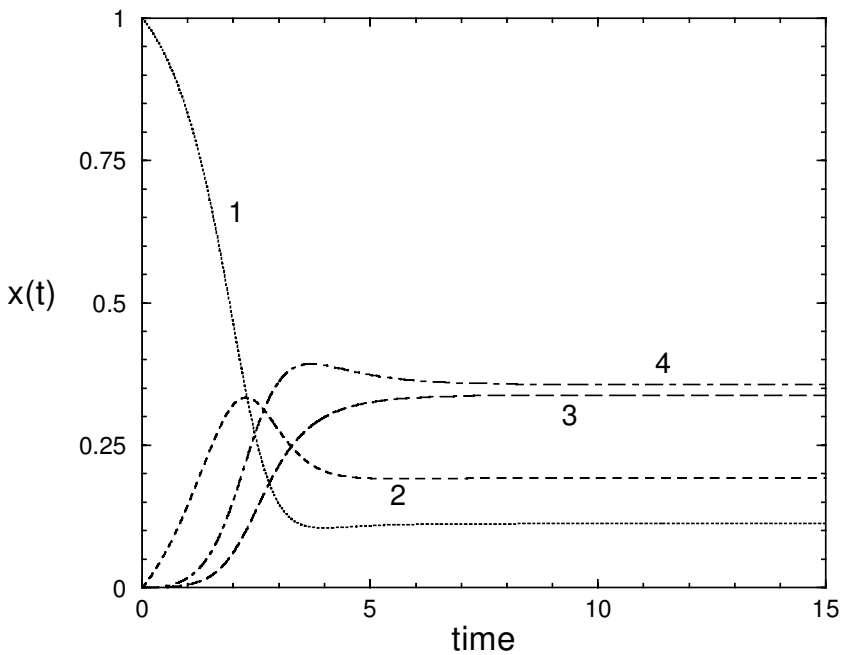


Fig. 4 Schematic representation of a quasi-species. The grid represents the space of sequences. The vertical lines show the fitness or relative concentration of the corresponding sequence in the quasi-species distribution. The master sequence in this case is the most represented. The cloud around the master sequence in the sequence space indicates the region of the sequence space occupied by the quasi-species.



(a)



(b)

Fig. 5 (a): Simulation of the quasi-species model with four different species. The mutational values W_{ij} are small in this case. Thus the system behaves as in the previous model without mutation and CP constraint: when the selective equilibrium is reached, the master sequence dominates the population. The mutant cloud is minimally represented. (b): We repeat the simulation of figure 5(a) but now the mutational matrix has non-negligible elements. We see now that the relative concentrations of the different species (master sequence and mutants) are of equal order.

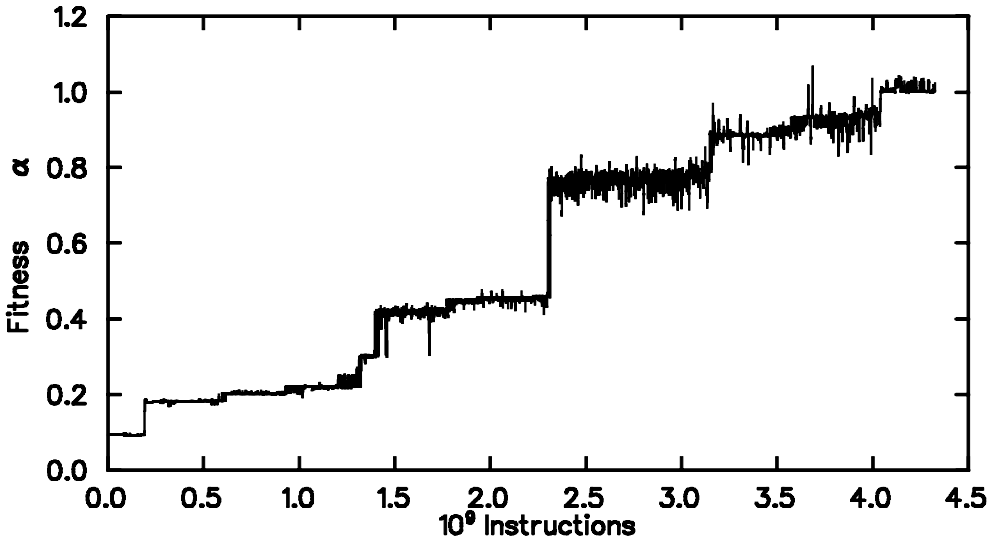


Fig. 6 Path evolution of a simulation in Tierra with a mutational rate of $q = 0.5 \times 10^{-8}$ per bit. The time in horizontal axis is in units of 10^9 executed instructions. Shown on the vertical axis is the most common fitness of the existing creatures, the most popular genotype (From [24]).

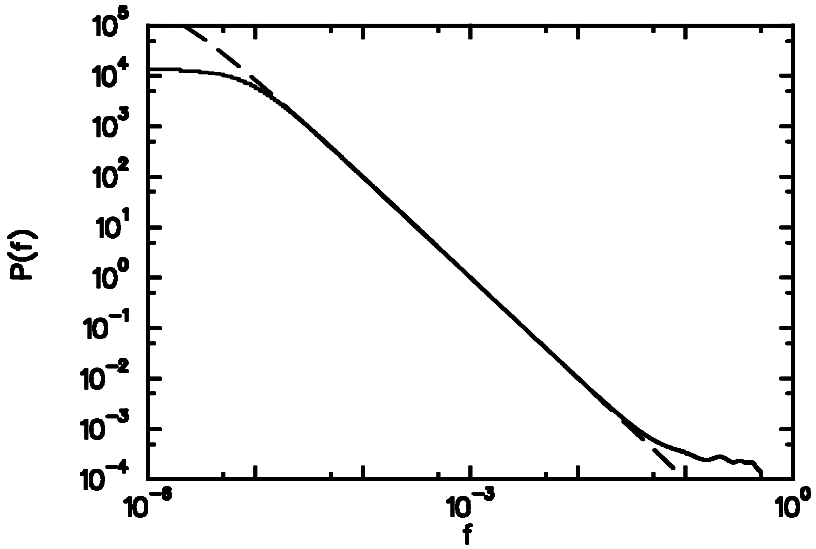


Fig. 7 Power spectrum of figure 6 (From [24]).

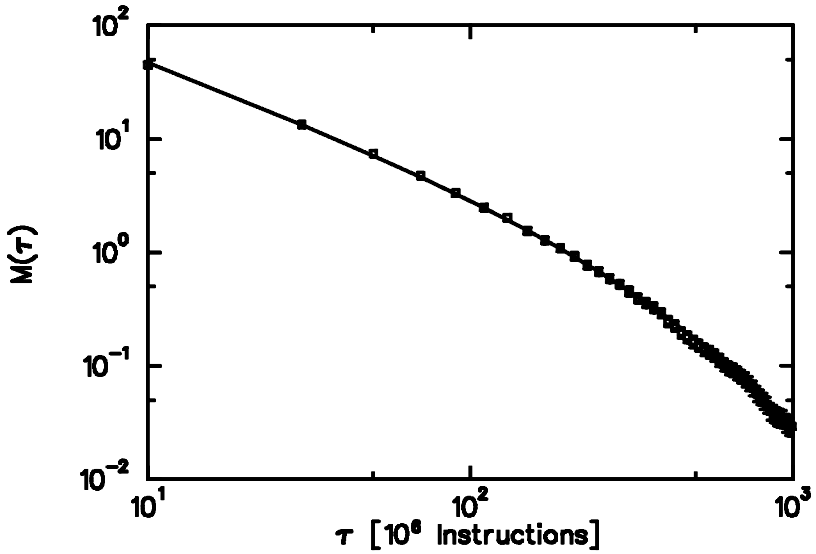


Fig. 8 Integrated distribution of dominant times $N(t)$ of cuasispecies.(From [24])

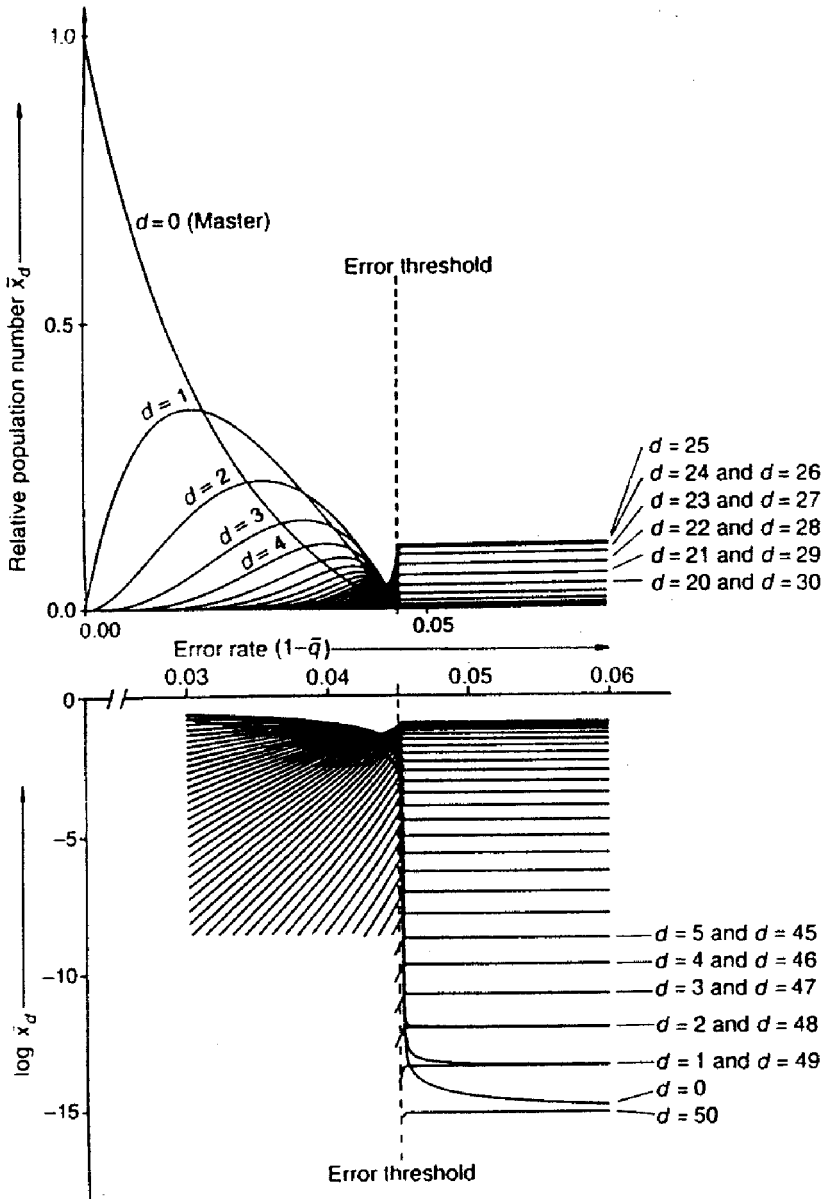


Fig. 9 Above: variation of relative populations x_d (equilibrium concentrations) as a function of the error rate $1 - q$. The error threshold $1 - q = 0.046$ is marked with a vertical dashed line. Beyond it the distribution of relative populations becomes a random distribution. This is the error catastrophe. Below: The same graph but with a logarithm scale in the vertical axis to clearly appreciate the sharp drop of the master sequence to zero concentration (From [29]).

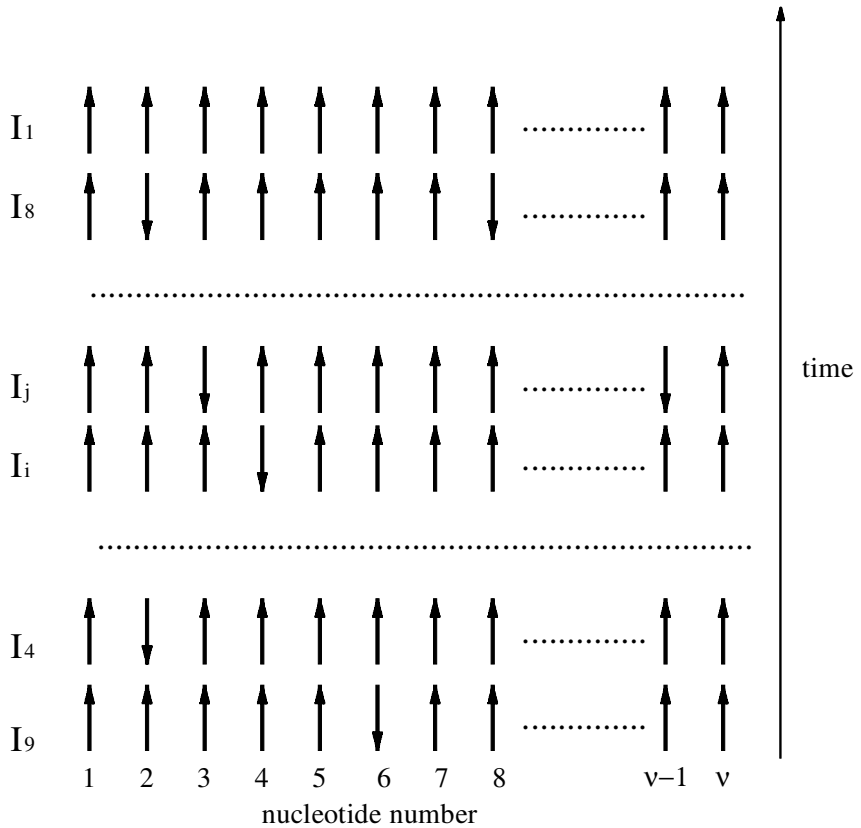


Fig. 10 Successive reproductions in time of a initial spin chain I_9 subject to spin flips (mutations).

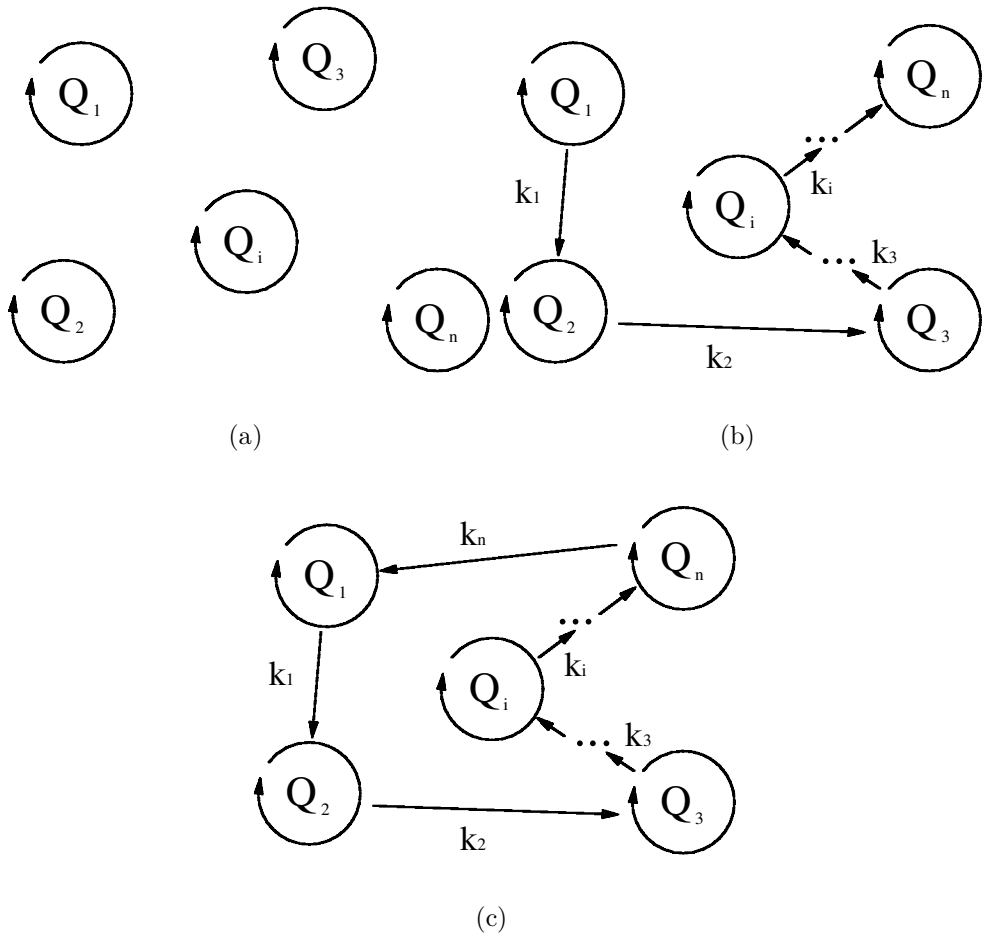
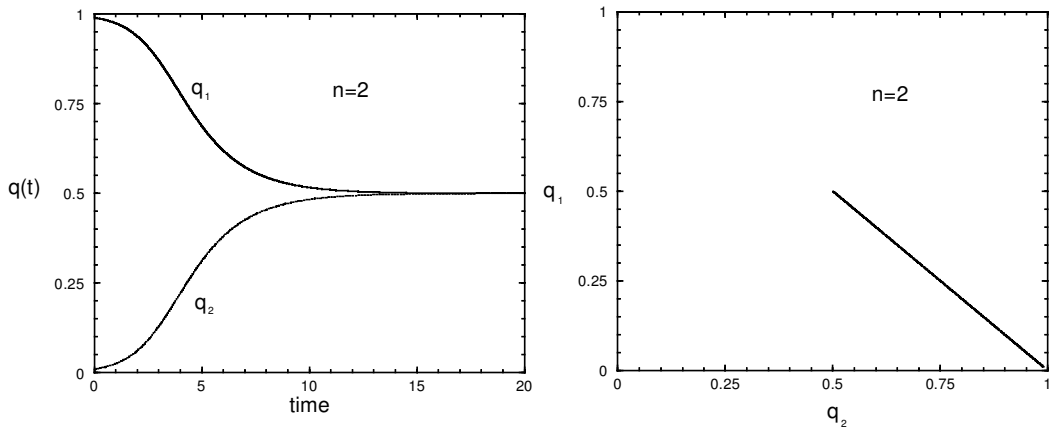
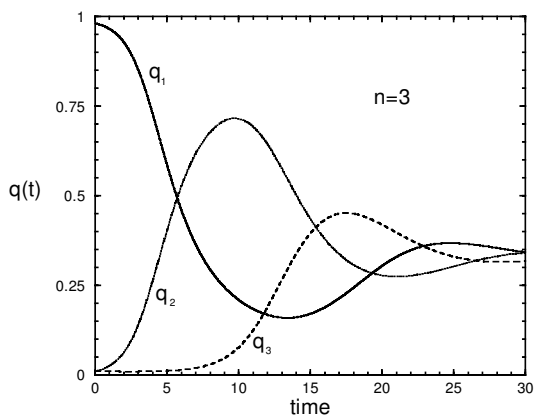


Fig. 11 (a): A set of independent quasi-species. (b): An acyclic coupled quasi-species by mutual catalysis. (c): A cyclic chain of coupled quasi-species by mutual catalysis, more aptly, a hypercycle.

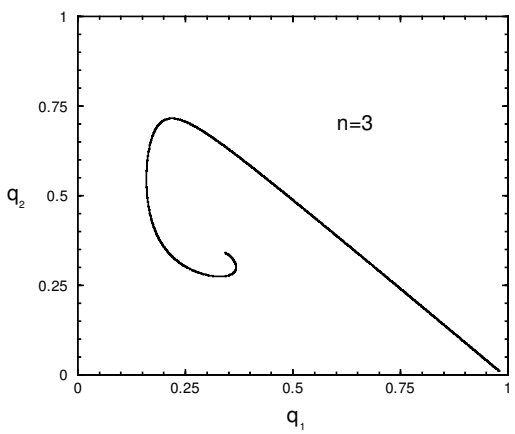


(a)

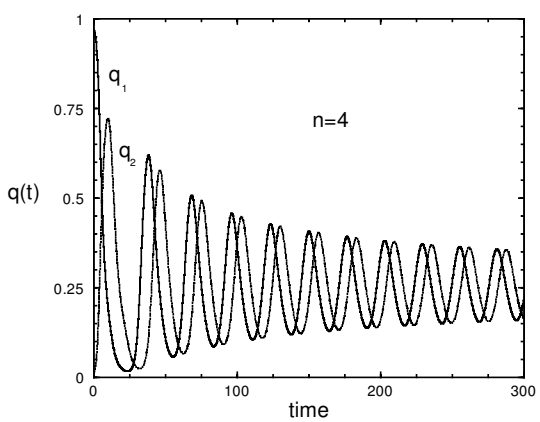
(b)



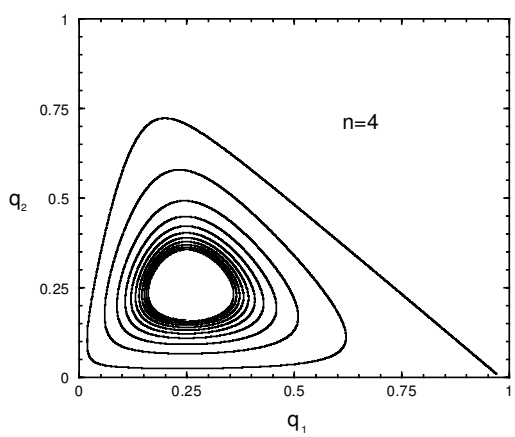
(c)



(d)

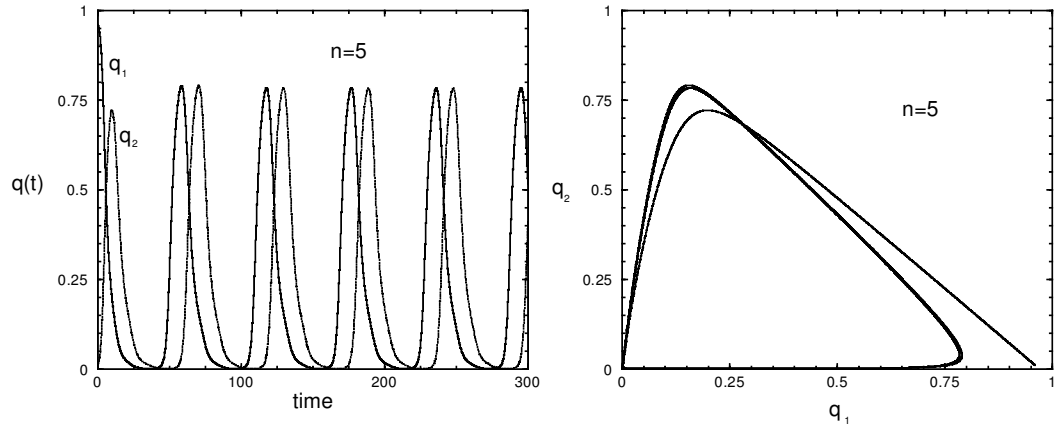


(e)

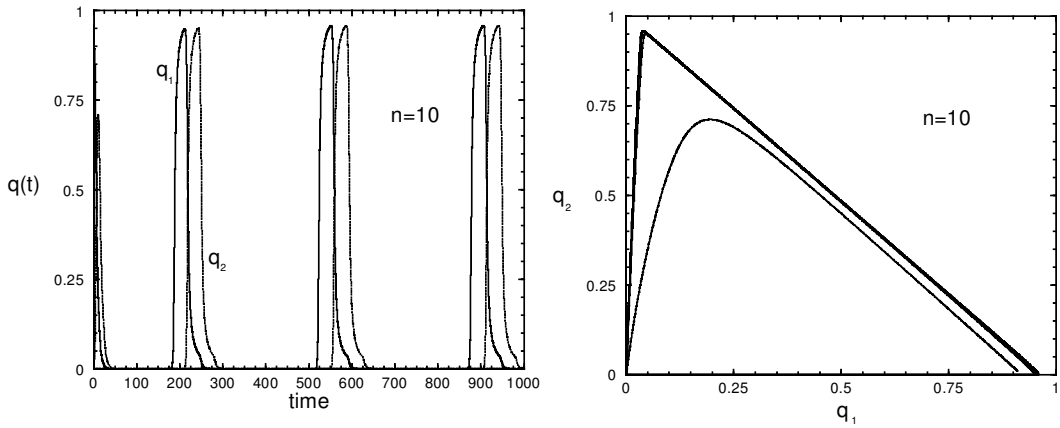


(f)

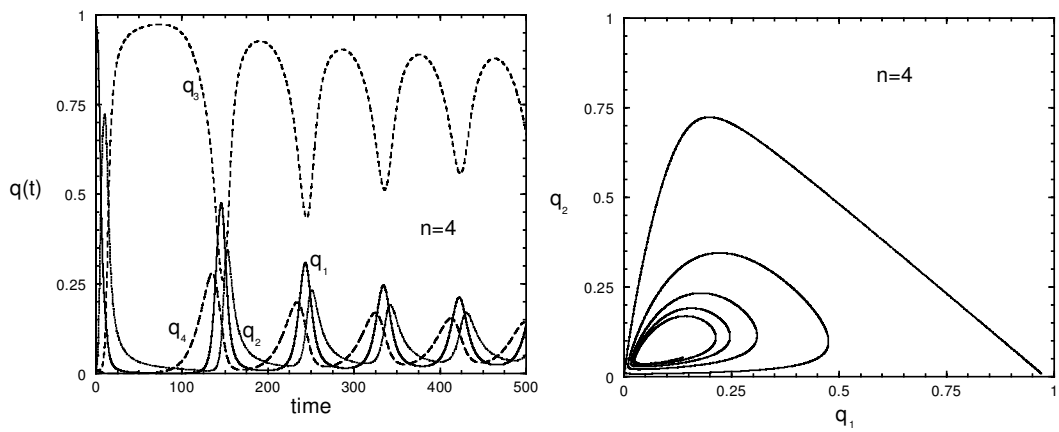
Fig. 12 Dynamical behavior of the elemental hypercycles. Concentrations in time of a hypercycle with: (a): $n = 2$ quasi-species, (c): $n = 3$ and (d): $n = 4$. Phase portrait, plane $q_1(t) - q_2(t)$, for hypercycle with: (b): $n = 2$, a direct focus, (d): $n = 3$ a focus in spiral and (f): $n = 4$, a limit cycle.



(a) (b)



(c) (d)



(e) (f)

Fig. 13 The same numerical experiments as in figure 12 but with $n = 5$ in (a), (b) and $n = 10$ in (c), (d). In both cases the behavior is the limit cycle. In the last case (e), (f) we represent a hypercycle with $n = 4$ but now with non-symmetric constants coupled. Again the behavior is the limit cyclic.