Sequence evolution

Carsten Kemena

July 22, 2007

Abstract

This composition explains parts of the model introduced by Eigen and Schuster in 1977. It explains the evolutionary process of molecules and is applicable on the development of the DNA. Thus it illustrates an important step in the development of life.

1 Introduction

All living organisms have one thing in common, the usage of DNA as memory to save all information needed to control every single process within a cell. The DNA consists in every organism of four different bases and in addition also every codon codes for the same amino acid in almost every species. Exceptions are only known for mitochondriums, chloroplasts and some specific single-celled eucaryots. It is due to this circumstance that the DNA has to be developed at a very early point of evolution even before the development of cells. This process can be modelled with differential equations as described in the model of Eigen and Schuster [1].

2 Molecular Evolution

To model the evolution of molecules it is important to differentiate between groups of molecules. For this purpose the term 'molecular species' is now introduced. A molecular species is the entirety of all molecules of the same type. Even though 'molecular species' can be used for every type of molecules, it is primarily associated with DNA in this text. For example two DNA molecules with the same base sequence belong to the same species, whereas two DNA molecules with different base sequences belong to different (molecular) species.

The following functions model the change of concentrations of molecular species and therefore the evolutionary process. To do this correctly we need to know which characteristics are common for evolutionary behaviour and therefore need to be parameters of the functions. Three requirements have to be fulfilled for evolution on the level of molecules as well as on the level of living organisms:

1. Metabolism
2. Self-reproduction
3. Mutability

The term metabolism relates to the formation and degeneration of molecules. Both processes have to be spontaneous and independent from each other. This condition cannot be fulfilled in equilibrated systems, because in that case the degradation of one molecule would be connected with the formation of a new one. Therefore a process is needed turning energy-rich material into energy-deficient material (including some intermediate states e.g. DNA molecules). The selection process during evolution is effective on the intermediate states because only these molecules have a function.

In the case that the formation process only consists of spontaneous accumulation of single components, the information content of a resulting DNA sequence would be pure chance. No evolutionary process could take place because the information of a DNA sequence would be lost during the degeneration process. That is the reason why self-reproduction is needed. Self-replicating molecules allow the degeneration of single molecules without information loss if other molecules of the same species, results of the self-replication, will still exist.

Mutation is a physical result of the replication process and cannot be avoided. This fact has two aspects. The positive one is that without mutations it would be impossible to gain new information and hence no adaption could take place in case of a change in the environment. But if the mutation rate is too high the already gained information in a DNA sequence could be lost and the resulting molecule would not function properly.

This three parameters can be modelled with differential equations. This equations are explained in the next section.

3 Modelling of the evolutionary process

First the metabolism is modelled. The formation of new molecules depends on the self-replication rate k. A molecule X needs some raw material then it can replicate itself with rate k. The degeneration of a molecule to some end product is depending on a rate k'.

$$A + X \xrightarrow{k} 2X, \quad X \xrightarrow{k'} F \tag{1}$$

Usually we not only have one molecular species but several. Each one has its own self-replication and degeneration rate.

$$\begin{aligned} A + X_1 &\xrightarrow{k_1} 2X_1, \quad X_1 \xrightarrow{k'_1} F \\ A + X_2 &\xrightarrow{k_2} 2X_2, \quad X_2 \xrightarrow{k'_2} F \\ &\vdots \\ A + X_n &\xrightarrow{k_n} 2X_n, \quad X_n \xrightarrow{k'_n} F \end{aligned} \tag{2}$$

Each process described in equation 2 is now changed into a differential equation. The change in concentration of a species i depends on the proportion of the two rates k_i and k'_i. n_i denotes the current concentration.

$$\frac{dn_i}{dt} = k_i n_i - k'_i n_i = (k_i - k'_i) n_i \tag{3}$$

The solution of this differential equation is the following:

$$n_i(t) = n_{i,0} e^{(k_i - k_i')t} \tag{4}$$

The start concentration $n_{i,0}$ has no influence on the increasing or decreasing of the function value. If $k_i < k_i'$ holds, the concentration of a species decreases, if $k_i > k_i'$ the concentration increases exponentially. The curve progression is shown in Figure 1. In the case of $k_i = k_i'$ the concentration remains on the level of the start concentration.

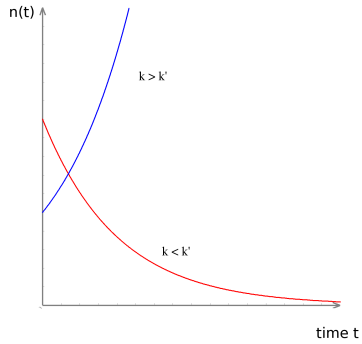

Figure 1: The behaviour of $n(t) = n_0 e^{(k-k')t}$ depends on the replication and degeneration rate in proportion to each other.

No kind of limitations (like limited raw material) are considered until now. Hence, the concentration of each species with $k > k'$ will grow infinitely. To make the model more realistic the selection process is added to differential equation 3.

There are two possibilities to model the selection process. The first method is to restrict the mass of available raw material. In this case molecules cannot replicate as often as the replication rate permits. The second possibility is the one Eigen and Schuster [1] choose. A limit to the total number of macromolecules is set. To achieve this goal a permanent dilution is assumed. This can be modelled by adding an additional term to the differential equation designed until now.

$$\frac{dn_i}{dt} = (k_i - k_i^{'})n_i - \varphi n_i \tag{5}$$

φ denotes the rate of dilution and is for each species the same. Setting a limit on the total number is the same as setting the change to zero. Thus the correct value of φ can be calculated by setting the sum over all differential equations to zero.

$$\begin{aligned} 0 = \sum_i \frac{dn_i}{dt} &= \sum_i (k_i - k_i^{'})n_i - \varphi \sum_i n_i \\ \Rightarrow \varphi = \frac{\sum_i (k_i - k_i^{'})n_i}{\sum_i n_i} &= \frac{\sum_i E_i n_i}{\sum_i n_i} = \bar{E} \end{aligned} \tag{6}$$

The selection value E of a species denotes the excess productivity of a single species. The average excess productivity is denoted by $\bar{E}$, the selection threshold. Differential equation 5 can now be rewritten into the following equation:

$$\frac{dn_i}{dt} = (E_i - \bar{E})n_i \tag{7}$$

The increase and decrease of the concentrations of a species i now depends on the proportion of E_i to $\bar{E}$. In the case of $E_i > \bar{E}$ the concentration of a species increases, if $E_i < \bar{E}$ the concentration decreases. This behaviour is illustrated in Figure 2. Each of the three species has the same start concentration but different selection values.

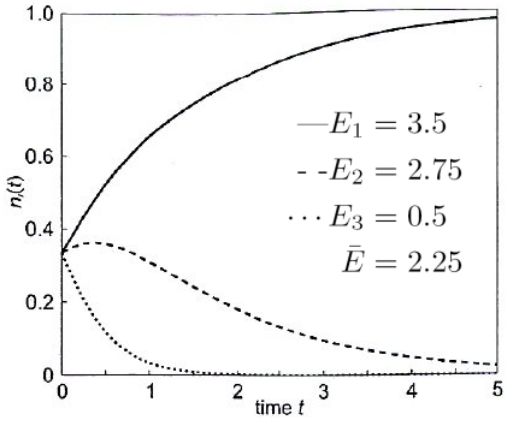

Figure 2: Development of concentrations under selection pressure. Each of the three species begin with the same start concentration but behave differently depending of the selection value. Species one is the only surviving one. (source: [2])

At the beginning the concentration of species 1 and 2 increases whereas the concentration of species 3 steadily decreases. But after a short time the concentration of species 2 will decrease, too. This is caused by the increase of the selection threshold (see Figure 3). As can be seen in equation 6 the value of the selection threshold depends on the concentration and the excess productivity of each species. That is why the threshold increases when the concentration of species 1 increases and the concentration of species 3 decreases. Therefore the mean excess productivity of the system growths. At the end of this process the selection threshold reaches the highest selection value of all species. Hence only the species belonging to this selection value will survive the selection process.

The only remaining step is to include the mutation process into the model. Instead of the replication rate k we have a rate of error-free replication F. F is a combination of the usual replication rate k and the percentage of error-free replication Q. Therefore $F = Qk$.

An erroneous replication will result in a sequence of another species. Taking this into consideration results in the addition of $\sum_j m_{ij} n_j$ to differential equation 5. m_{ij} denotes the rate of mutation from a species j to a species i.

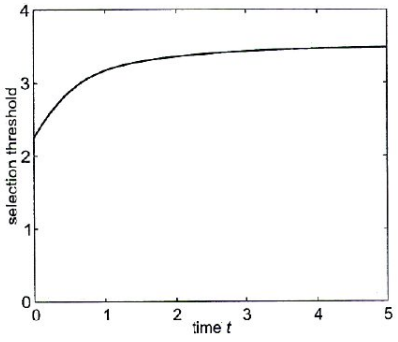

Figure 3: Development of the selection threshold in Figure 2. The increase of species 1 with simultaneous decrease of species 3 and later species 2 results in an increase of the selection threshold. (source: [2])

$$\frac{dn_i}{dt} = F_i n_i - k_i^{'} n_i + \sum_j m_{ij} n_j - \varphi n_i \tag{8}$$

This differential equation leads to a new selection value W_i of a species. It is calculated as following: $W_i = Q_i k_i - k_i'$. The concentration of species with a selection value above $\bar{E}$ will increase, other species will die out.

Mutations will occur even after one molecular species is selected. So instead of one single species there always exist a number of closely related sequences. The species with the highest selection value is called master sequence. The master sequence in combination with the cloud of closely related sequences is called quasi species. This quasi species is the goal of selection. The result is that the selection process ends with several different sequences instead of only one even if these sequences are related very closely.

The average of this distribution, the master sequence, is often called the wild type. Figure 4 shows the sequence distribution. Although the master sequence is the species with the highest population, the species occurs only with a small percentage in the whole distribution. The master sequence does not stay the same during the evolutionary process. If the environmental conditions changes another molecular species can become the master sequence, if it is better adapted to the changed conditions.

4 Summary & Outlook

The model presented includes the basic properties of evolution. It can explain how the evolution of DNA sequences happened but it cannot explain how many different molecular species can exist side by side. The only chance that several very different quasi species exist in this model would be that they occur in different geographic regions.

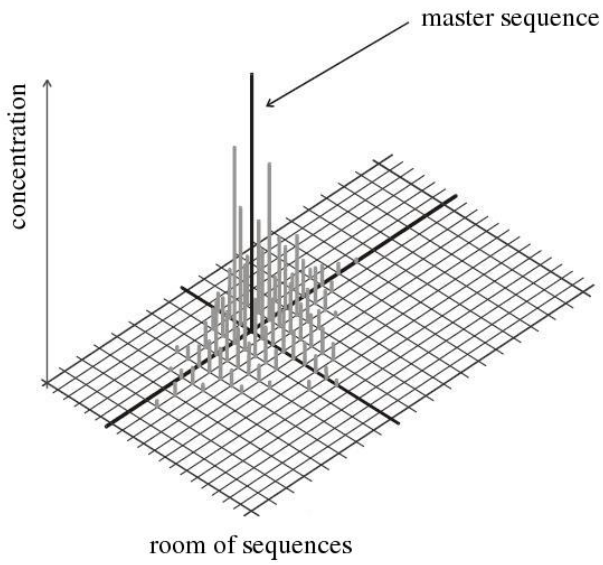

Figure 4: Distribution of a quasi species. The master sequence has the highest concentration but many other closely related species exist alongside. (source: [3])

But in reality there exists a wide range of quasi sequences alongside. This is explained in the second part of the model developed by Eigen and Schuster [1]. For this purpose the concept of cooperation is needed. The model developed from this is named hypercycle. It allows several different sequences to coexist in the same environment if they cooperate.

References

[1] Manfred Eigen and Peter Schuster. The hypercycle: a principle of natural self-organisation, part a. *Naturwissenschaften*, 64(11):541–565, 1977.

[2] Competition and selection in biological systems. Lecture 'Modellierung biologischer Systeme', HU, 2000.

[3] Peter Schuster. Ursprung des lebens und evolution von molekülen. Lecture 'Evolution und Schöpfung', Universität Wien, 1998.