Bulletin of Mathematical Biology, Vol. 41, pp. 803 812
Pergamon Press Ltd. 1979. Printed in Great Britain

0007-4985/79/1201-0803 $02.00/0

SOME REMARKS ON THE DYNAMICS OF MOLECULAR SELF-ORGANIZATION

BERND KÜPPERS
Max-Planck-Institut für biophysikalische Chemie,
D-3400 Göttingen,
W. Germany

The dynamics of a self-organizing molecular system is described in terms of its normal modes. Each normal mode is associated with a certain eigenvalue, the average of which reflects most directly the overall process of self-organization. For the temporal change of this quantity a maximum principle holds.

1. Introduction. Within the last few years the molecular theory of evolution (Eigen, 1971) has been developed to a considerable extent. Even though it will never be possible to reconstruct the precise historical route of evolution, one can however reveal the general principles by which evolutionary processes are governed. In the following a model system is to be described that enables us to make some general conclusions about the dynamics of molecular self-organization.

2. Selective Self-Reproduction of Biological Macromolecules. In order to derive the basic equations that describe selection phenomena at the molecular level, we consider the model system depicted in Figure 1. A rigorous justification for the choice of this particular model can be found elsewhere (Eigen, 1971; Küppers, 1979).

Each macromolecule within the box can be associated with a population variable x_i (number of macromolecules of sort i per unit volume). For simplicity we assume all polymeric sequences to have a uniform "length" of ν monomeric digits. Thus if there are l classes of monomers, the running index i defines one out of l^ν alternative sequences (species). In order to make the evolutionary problem non-trivial we require that the total population number of the system is very small compared to the number of all possible species. As a consequence of this condition the non-instructed formation of any individual macromolecule is completely negligible.

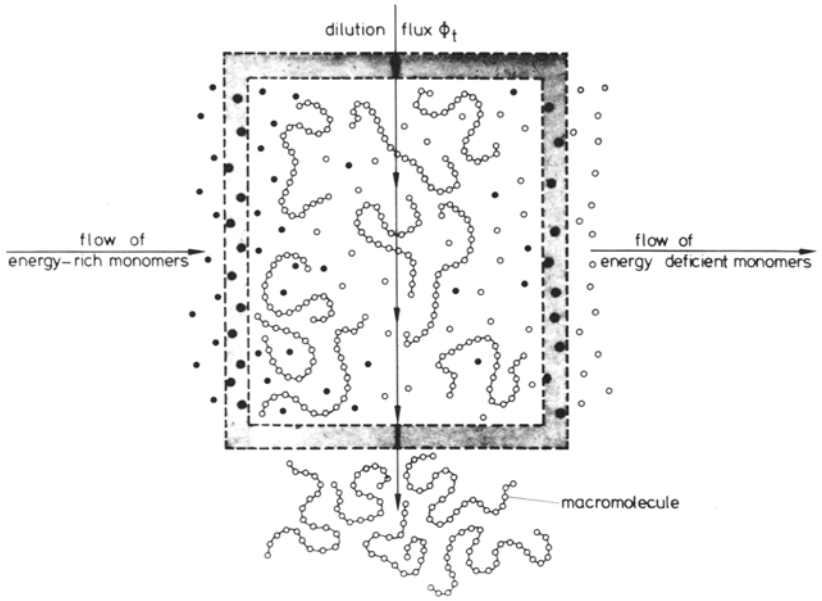

Figure 1. Model of a self-organizing molecular system (Eigen, 1971; Küppers, 1979). The box contains biological macromolecules (i.e. nucleic acids), which are steadily assembled from energy-rich monomers and decomposed into energy-deficient monomers. The formation of the macromolecular species is *self-instructed*, i.e. any sequence of monomers will instruct its own reproduction. If any error occurs in the reproduction, the error copy will be further reproduced. Furthermore the box is assumed to have semipermeable walls through which energy-rich and energy-deficient monomers can flow in and out. By a dilution flux ϕ_t the overall population density of the macromolecular species can be regulated

Then—within the framework of our model—the temporal change in x_i is determined by the following equation (Eigen, 1971) ($\dot{x} = \mathrm{d}x/\mathrm{d}t$, $t =$ time):

$$\dot{x}_i = W_i x_i + \sum_{k \neq i} \varphi_{ik} x_k - \phi_i, \quad i = 1, \ldots, N. \tag{1}$$

The different terms on the right hand side specify the essential properties of the system: $W_i x_i$ refers to the self-instructed formation of the species i. $\varphi_{ik} x_k$ results from an imprecise copying of the species k, which leads to the species i (back flow from a mutant k to its master copy i). ϕ_i represents a flow term, by which the concentration of the species i can be externally regulated. Moreover we assume that each species contributes to the total flow ϕ_t in proportion to its presence, i.e.

$$\phi_i = f x_i. \tag{2}$$

The particular form of (1) has been chosen in order to express formally the kinetics of *inherent* autocatalysis. It does not necessarily imply first-order reaction behavior. The rate parameter W_i is certainly a function of the monomer concentration and, moreover, may depend on the population variables x_k. By controlling its flow rate the monomeric digits can be buffered—a condition which will be chosen for most evolutionary experiments (Küppers, 1979). Throughout this paper we will assume the rate parameters W_i as well as φ_{ik} to be constant in time and *non-negative*.

One can specify W_i further in purely phenomenological terms, i.e.

$$W_i = A_i Q_i - D_i, \tag{3}$$

where A_i is a rate constant which describes how fast synthesis is directed by the template i, regardless of whether the reproduction leads to precise copies or mutants. With Q_i we introduce a measure for the fraction of precise copies, so that in the extreme cases we have $Q_i = 0$ (i.e. all copies of species i are incorrect) and $Q_i = 1$ (i.e. all copies of species i are correct). D_i finally represents an individual decomposition term.

So much for the structure of (1). For a more detailed discussion reference is made to the extensive literature (Eigen, 1971; Eigen and Schuster, 1977, 1978). To make the system selective we have to force its macromolecular species into real competition for survival. From an experimental point of view (Küppers, 1979), there is one straightforward procedure to impose a selection pressure, that is the constraint of constant overall population density

$$\sum_i x_i = n = \text{const.} \tag{4}$$

It is obvious that this condition may lead to selection since any increase in the concentration of one species must necessarily be compensated by a decrease in the other species. In order to maintain the constraint (4) the total flow ϕ_t has to be adjusted so as to compensate for the excess overall production

$$\phi_t = \sum_i \phi_i = \sum_i (A_i - D_i) x_i = \sum_i E_i x_i, \tag{5}$$

where we call $E_i = A_i - D_i$ the (excess) productivity of the template i. From (2) it follows that

$$\phi_t = \sum_i f x_i = fn. \tag{6}$$

Combining (5) and (6) we have

$$f \equiv \bar{E}(t) = \frac{1}{n} \sum_i E_i x_i \tag{7}$$

and (1) now becomes

$$\dot{x}_i = (W_i - \bar{E}) x_i + \sum_{k \neq i} \varphi_{ik} x_k, \quad i = 1, \ldots, N. \tag{8}$$

This is the basic equation that describes the selective reproduction of biological macromolecules under the constraint of constant overall population density (Eigen, 1971). There are two types of couplings among the macromolecular species: (a) all species are forced into mutual competition by the coupling factor $\bar{E}$; and (b) between closely related species a (weak) cooperation is provided by the exchange factor $\sum_{k \neq i} \varphi_{ik} x_k$.

For the following discussion it will turn out to be proper to reorganize the system (Eigen and Schuster, 1977, 1978): instead of subdividing the total population into N species we define a new set of N quasi-species, for which the population variables y_i are linear combinations of the original variables x_i, such that

$$x_i = \sum_j a_{ij} y_j, \quad j = 1, \ldots, N. \tag{9}$$

The coefficients a_{ij} should be chosen so that in the absence of any population control (i.e. $f(t) \equiv 0$) equations (1) are transformed into

$$\dot{y}_i = \lambda_i y_i, \tag{10}$$

where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of the matrix

$$M = (m_{ik}) = \begin{pmatrix} W_1 & \varphi_{12} & \cdots & \varphi_{1N} \\ \varphi_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \varphi_{N1} & \cdots & \cdots & W_N \end{pmatrix}. \tag{11}$$

These eigenvalues are attributed to certain "normal modes" of reaction, represented by the y_i. For a detailed physical interpretation reference is made to the extensive discussion by Eigen (1971) and Eigen and Schuster (1977, 1978).

We now take into account a population control of the kind expressed by

(4). Applying the same affine transformation to (8) we obtain

$$\dot{y}_i = (\lambda_i - \bar{E}) y_i \tag{12}$$

with the condition

$$\sum_i x_i = \sum_i \sum_j a_{ij} y_j = \sum_j \left\{ \sum_i a_{ij} \right\} y_j = \text{const.} \tag{13}$$

According to (12) we can normalize each eigenvector y_j by the constant $\sum_i a_{ij}$. This accounts for the fact that in the course of reorganization of the system the total number of organized monomeric digits and hence the overall population density must be conserved:

$$\sum_i x_i = \sum_j y_j = n. \tag{14}$$

We now add all equations (12) and obtain with relation to (14),

$$\dot{n} = (\bar{\lambda} - \bar{E}) n. \tag{15}$$

From the conservation condition we have $\dot{n} = 0$ and thus

$$\bar{E} \equiv \bar{\lambda} = \frac{1}{n} \sum_j \lambda_j y_j, \tag{16}$$

i.e. $\bar{E}$ remains invariant in the transformation and can be expressed now as the average of the λ_i's. Thus the selective self-reproduction of a quasi-species population is described by the following set of differential equations:

$$\dot{y}_i = (\lambda_i - \bar{\lambda}) y_i, \quad i = 1, \ldots, N. \tag{17}$$

These equations have already been obtained in a slightly different procedure by other authors (Thompson and McBride, 1974; Jones *et al.*, 1976). The solutions of (17) show the following characteristics: all equations are coupled via the average eigenvalue $\bar{\lambda}$, which provides a sliding and self-adjusting threshold reflecting the self-organization of the system. Only those quasi-species (characterized by an eigenvalue λ_i and a population variable y_i) will grow, whose λ_i-value is above the threshold $\bar{\lambda}$. As a consequence of their growth they shift the threshold steadily to higher values until an optimum of $\bar{\lambda}$ is reached which matches the maximum

eigenvalue of the distribution,

$$\lim_{t\to\infty}\bar{\lambda}(t)=\lambda_{\max}. \tag{18}$$

By then all quasi-species but one—i.e. the one associated with the maximum eigenvalue $\lambda_{\max}$—will have died out. Their population variables have become zero. This means selection.

The term "quasi species" has a real physical meaning. It accounts for the fact that a certain selected nucleotide sequence (master copy) in general does not exist alone, but rather is always surrounded by a whole set of probable and viable mutants. This is true even for the state of selection equilibrium. Thus the quasi species represents the ensemble of information carriers, consisting of a master copy and a comet tail of mutants, the relative frequency of which is determined by the probability of their template directed formation. A more detailed discussion of the concept of quasi-species can be found elsewhere (Eigen and Schuster, 1977, 1978). From the above representation, however, it is already obvious that the overall process of molecular self-organization is essentially reflected by the temporal behavior of the threshold function $\bar{\lambda}$.

3. A Maximum Principle for Molecular Self-Organization. Before we start to formulate a maximum principle for the temporal change of $\bar{\lambda}$, we notice some other interesting properties of this function. Let us first consider the time derivatives of $\bar{\lambda}$:

$$\dot{\bar{\lambda}}=\frac{1}{n}\sum_i \lambda_i\dot{y}_i=\frac{1}{n}\sum_i \lambda_i(\lambda_i-\bar{\lambda})y_i. \tag{19}$$

By rearranging, it follows

$$\dot{\bar{\lambda}}=\sum_i(\lambda_i-\bar{\lambda})^2\frac{y_i}{n}=\overline{\lambda^2}-\bar{\lambda}^2\geqq 0, \quad \text{at constant } \lambda_i \text{ and } n. \tag{20}$$

This relation bears a strong analogy to Fisher's Fundamental Theorem of population biology (Crow and Kimura, 1970): i.e. the effectiveness of selection in changing the composition of a population depends on the variance in the reproduction rate spectrum of all species present. Thus, outside of the selection equilibrium we have $\dot{\bar{\lambda}}>0$, whereas the equilibrium is characterized by $\dot{\bar{\lambda}}=0$. For the temporal change in the variance V it immediately follows that:

$$\dot{V}=\sum_i(\lambda_i-\bar{\lambda})^2\frac{\dot{y}_i}{n}-2\sum_i(\lambda_i-\bar{\lambda})\dot{\bar{\lambda}}\frac{y_i}{n}=\sum_i(\lambda_i-\bar{\lambda})^3\frac{y_i}{n}, \tag{21}$$

i.e. the temporal change in the variance V is equal to the third moment about the mean of the distribution.

We are now prepared to formulate the following theorem about the rate of molecular self-organization (Maximum Principle):

THEOREM. *The process of molecular self-organization proceeds so as to maximize the rate of increase in the average eigenvalue of a quasi-species population.*

Let us express this theorem more explictly: during the process of selection the "motion" of the concentration variables y_i are such that the variation of the line integral

$$I=\int_1^2 \mathrm{d}\bar{\lambda} \tag{22}$$

is—for fixed t_1 and t_2—equal to zero, i.e.

$$\delta I=\delta\int_{t_1}^{t_2}\dot{\bar{\lambda}}\mathrm{d}t=0. \tag{23}$$

To demonstrate this principle we shall show that the maximization of $\dot{\bar{\lambda}}$ leads to the selection equations (17). The procedure is well known from classical mechanics where Newton's laws of motion can be deduced from Hamilton's integral principle.

In order to maximize $\dot{\bar{\lambda}}$ we make use of the standard method of Lagrange multipliers. Two constraints have been taken into account. The first constraint results from the conservation condition (14), i.e.

$$\sum_i \delta y_i=0. \tag{24}$$

The second condition is obtained by the requirement that the variance is (momentarily) constant. A combination from (17) and (20) yields

$$\frac{1}{n}\sum_i \frac{1}{y_i}\left(\frac{\delta y_i}{\delta t}\right)^2=\overline{\lambda^2}-\bar{\lambda}^2. \tag{25}$$

For the variation $\delta\bar{\lambda}$ one obtains from (16)

$$\delta\bar{\lambda}=\frac{1}{n}\sum_i \lambda_i\delta y_i. \tag{26}$$

We now follow the usual procedure in the calculus of variations. We define a function ψ by

$$\psi = \delta\bar{\lambda} + \xi_1 \left\{ \sum_i \delta y_i \right\} + \xi_2 \left\{ \overline{\lambda^2}(\delta t)^2 - \bar{\lambda}^2(\delta t)^2 - \frac{1}{n} \sum_i \frac{1}{y_i} (\delta y_i)^2 \right\}, \tag{27}$$

where ξ_1 and ξ_2 are for the time being undetermined constants (Lagrange multipliers). The requirement $\partial\psi/\partial(\delta y_i) = 0$ then yields N equations of the type

$$\frac{\lambda_i}{n} + \xi_1 + \frac{2\xi_2 \delta y_i}{n y_i} = 0. \tag{28}$$

We multiply the equation by y_i and add all N equations

$$\frac{1}{n} \sum_i \lambda_i y_i + \xi_1 \sum_i y_i + \frac{2\xi_2}{n} \sum_i \delta y_i = 0. \tag{29}$$

From the condition (24) we know that $\sum_i \delta y_i = 0$. So we obtain

$$\bar{\lambda} + \xi_1 n = 0, \tag{30}$$

and determine from this equation the first multiplier

$$\xi_1 = -\bar{\lambda}/n \tag{31}$$

Substituting ξ_1 into (28) yields

$$\delta y_i = -\frac{(\lambda_i - \bar{\lambda}) y_i}{2\xi_2} \tag{32}$$

and with this relation it follows from (25)

$$\overline{\lambda^2} - \bar{\lambda}^2 - \frac{1}{n} \sum_i \frac{(\lambda_i - \bar{\lambda})^2 y_i}{4\xi_2^2 (\delta t)^2} = 0. \tag{33}$$

Solving this equation for ξ_2 we first obtain

$$\xi_2 = \pm \frac{1}{2\delta t}, \tag{34}$$

and then from (28) our desired selection equation (cf. eq. (17))

$$\frac{\delta y_i}{\delta t} = (\lambda_i - \bar{\lambda}) y_i \tag{35}$$

4. Discussion. The quantity which most directly reflects the self-organization of a molecular system is $\bar{\lambda}$, the average eigenvalue of the associated quasi-species population. Three essential statements about the properties of $\bar{\lambda}$ can be made (the λ_i's are assumed to be constant in time):

1. At each moment the temporal change of $\bar{\lambda}$ is given by the variance V and therefore is greater than (or equal to) zero;

$$\dot{\bar{\lambda}} \equiv V = \overline{\lambda^2} - \bar{\lambda}^2 \geqq 0.$$

2. The temporal change of the variance V is equal to the third moment about the mean of the distribution;

$$\dot{V} = \frac{1}{n} \sum_i (\lambda_i - \bar{\lambda})^3 y_i.$$

3. A maximum principle holds for $\dot{\bar{\lambda}}$:

$$\delta \int_1^2 \mathrm{d}\bar{\lambda} = 0.$$

i.e. with a given amount of variance V in a short time interval the normal mode concentrations change in such a way, that the increase of $\bar{\lambda}$ is maximised.

The last statement is illustrated in Figure 2 for the case of a population with only two normal modes of self-organization. The average eigenvalue associated with each point in the (y_1, y_2)-plane is located on a two dimensional surface. Now let us assume the system to be in the state (y_1', y_2') at time t. Any virtual displacement $(\delta y_1, \delta y_2)$ then must lie on the ellipse given by (26) in order to be compatible with the constraints imposed. The Maximum Principle states that among the various possible infinitesimal changes in the y-variables the path actually taken by the system is the one which leads to the largest increase in the ordinate, i.e. in the $\bar{\lambda}$-value.

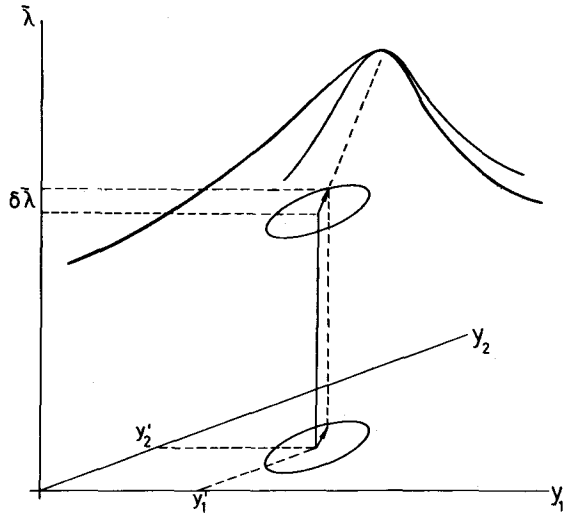

Figure 2. Two dimensional representation of the $\bar{\lambda}$-surface. In order to be compatible with the constraints any virtual displacement $(\delta y_1, \delta y_2)$ must lie on an ellipse with half axes $a_1 = \sqrt{y_1 n V}\,\delta t$ and $a_2 = \sqrt{y_2 n V}\,\delta t$. For details see text. In particular reference is also made to population genetics theory (Crow and Kimura, 1970)

I would like to thank Prof. M. Eigen and Prof. I. Epstein for many helpful discussions.

LITERATURE

Crow, J. F. and M. Kimura. 1970. *An Introduction to Population Genetics Theory.* New York: Harper & Row.

Eigen, M. 1971. "Self-Organisation of Matter and the Evolution of Biological Macromolecules." *Naturwissenschaften,* **58,** 465–523.

——— and P. Schuster. 1977 and 1978. "The Hypercycle—A Principle of Natural Self-Organization." Part A: "Emergence of the Hypercycle." *Naturwissenschaften,* **64,** 541–565. Part B: "The Abstract Hypercycle." *Naturwissenschaften,* **65,** 7–41. Part C: "The Realistic Hypercycle." *Naturwissenschaften,* **65,** 341–369.

Jones, B. I., R. H. Enns and S. S. Rangnekar. 1976. "On the Theory of Selection of Coupled Macromolecular Systems." *Bull. Math. Biol.,* **38,** 15–28.

Küppers, B. 1979. "Towards an Experimental Analysis of Molecular Self-Organization and Precellular Darwinian Evolution." *Naturwissenschaften,* (in press).

Thompson, C. J. and J. L. McBride. 1974. "On Eigen's Theory of Self-Organization of Matter and the Evolution of Biological Macromolecules." *Math. Biosci.,* **21,** 127–142.

RECEIVED 7-12-78
REVISED 11-8-78