J Mol Evol (1982) 18:344–350

Journal of Molecular Evolution

A Model for the Origin of Life

Freeman J. Dyson

Institute for Advanced Study, Princeton, New Jersey 08540, USA

Summary. A simple statistical model is constructed, describing the transition from disorder to order in a population of mutually catalytic molecules undergoing random mutations. The consequences of the model are calculated, and its possible relevance to the problem of the origin of life is discussed. The main conclusion of the analysis is that the model allows populations of several thousand molecular units to make the transition from disorder to order with reasonable probability.

Key words: Origin of life – Statistical model

I. Assumptions

The elucidation of the origin of life depends primarily on the work of experimental chemists (Miller and Orgel 1974; Lohrmann et al. 1980; Biebricher et al. 1981). Only a detailed study of reaction rates and dissociation rates can identify plausible pathways of prebiotic evolution. The purpose of the present note is to describe a simple abstract model of the transition from disorder to order in prebiotic structures. The model is not intended to be a theory of the origin of life. It provides only an empty mathematical framework within which questions about the origin of life can be posed with some degree of precision. The model will have served its purpose if it helps us to ask questions which geologists and chemists may be able to answer.

The model is in essence only an elementary exercise in population biology, following the ideas of Fisher (1930), Wright (1931) and Kimura (1970). The assumptions of the model are the following (1–9).

1. Molecular evolution occurs in small isolated populations which we call "islands." An island might be a colloidal droplet or a solid particle with molecules adsorbed on its surface. Each island exchanges molecular components slowly with the surrounding medium, which serves as a source of chemical free energy for reactions within the island.

2. Evolution occurs by random genetic drift only. Natural selection and Darwinian evolution belong to a later stage of development, when the island populations begin to grow and to compete with one another for nutrients.

3. Each island contains a fixed number N of molecular units (monomers) of various species. Some monomers may be free, while others are combined into polymers in an initially random way.

4. The population of polymers changes by discrete mutations, one monomer at a time being added, subtracted, or substituted in a polymer.

5. The multidimensional random walk of polymer mutations is mapped onto a one-dimensional random walk by counting only the numbers of monomers which are "active" and "inactive." A monomer is active if it happens to be correctly placed as part of a structure catalyzing the synthesis of other catalytic structures. Otherwise it is inactive.

6. Each of the N monomers in an island mutates with equal probability (1/N).

7. When a mutation occurs in an island with k monomers active, the probability that the mutated unit be active is $\phi(k/N)$, where $\phi(x)$ is a function describing the autocatalytic capability of the whole assemblage of active monomers.

8. The function $\phi(x)$ is monotonically increasing on the interval $0 \leqslant x \leqslant 1$.

9. The equation

$$\phi(x) = x \tag{1}$$

has three solutions, $x = \alpha, \beta, \gamma$, with

0022–2844/82/0018/0344/$01.40

$$0 < \alpha < \beta < \gamma < 1. \tag{2}$$

The crucial items in this list of assumptions are 7 and 9. Assumption 7 states that the effectiveness of active monomers in catalyzing the placement of other active monomers depends only on the total number of active monomers present and not on their detailed arrangement. This assumption is analogous to the "mean-field approximation" in the physical theory of ferromagnetism. It drastically simplifies the description of molecular populations by reducing all autocatalytic tendencies to a single parameter.

Assumption 9 states that there are three values of $x = (k/N)$, such that an island population with k active monomers is in statistical equilibrium. The condition (1) states that the number of active monomers is on the average unchanged by mutations. Conditions (1) and (2) together imply

$$\phi'(\alpha) < 1, \phi'(\beta) > 1, \phi'(\gamma) < 1, \tag{3}$$

which means that the equilibrium states $x = \alpha, \gamma$ are stable while the state $x = \beta$ is unstable. Each island, starting with a random population of monomers, will rapidly approach the "disordered state" $x = \alpha$ and remain for a long time executing small statistical fluctuations in the neighborhood of the disordered state. There is only a very small probability that the population may occasionally suffer a large statistical fluctuation which takes it over the unstable saddle-point $x = \beta$ to the "ordered state" $x = \gamma$. It may then fluctuate around the ordered state for a long time before it jumps back over the saddle to the disordered state.

The idea underlying our model is that the population of an island in the ordered state is in some sense "alive." The jump over the saddle from ordered to disordered state is "death." The jump upward from disordered to ordered state is the crucial event in the "origin of life." In the following Section II we work out the quantitative behavior of the model and calculate the probabilities of transitions between ordered and disordered states. Then in Section III we discuss a list of questions which the model suggests for further investigation; these questions are mostly concerned with the implications of the model for the nature of early biological evolution.

II. Consequences

The exact equations describing the evolution of populations in the model are

$$P_j(k+1) - P_j(k) = \psi_j(k) - \psi_{j-1}(k), \tag{4}$$

$$\psi_j(k) = f((j+1)/N)\, P_{j+1}(k) - g(j/N)\, P_j(k), \tag{5}$$

$$f(x) = x(1 - \phi(x)), \; g(x) = (1 - x)\, \phi(x), \tag{6}$$

where $P_j(k)$ is the probability for finding j active monomers in the population after k mutations have occurred. Eq. (4) holds for $j = 0,1, \ldots, N$, with the boundary conditions

$$\psi_N(k) = \psi_{-1}(k) = 0. \tag{7}$$

The unique stationary solution of Eq. (4) is obtained by setting

$$P_j(k) = P_j \tag{8}$$

independent of k. Then Eq. (4), (7) imply

$$\psi_j(k) = 0, \tag{9}$$

and Eq. (2) gives the solution

$$P_j = P_0 \exp(-W(j)), \tag{10}$$

$$W(j) = \Sigma_{\ell=1}^{j} \left[\log(\ell/(N+1-\ell)) - \log(\phi((\ell-1)/N)\,(1-\phi(\ell/N)))\right]. \tag{11}$$

When N is large we may use the continuum approximation

$$W(j) = NU(j/N) + 1/2 \log(2\pi N u(j/N)), \qquad u(x) = [(x-x^2)\,\phi(x)\,(1-\phi(x))/(\phi(0)\,(1-\phi(0)))], \tag{12}$$

$$U(x) = \int_0^x [\log(y/(1-y)) - \log(\phi(y)/(1-\phi(y)))]\, dy. \tag{13}$$

The potential U(x) has minima at $x = \alpha, \gamma$ and a maximum at $x = \beta$. The stationary distribution P_j will be concentrated around the two minima. The distribution around the disordered minimum at $j = N\alpha$ will be approximately Gaussian with variance

$$V(\alpha) = (N/U''(\alpha)) = N((\alpha-\alpha^2)/(1-\phi'(\alpha))), \tag{14}$$

and similarly for the ordered minimum at $j = N\gamma$. The ratio of ordered to disordered populations in the stationary solution will be

$$[V(\gamma)\, u(\alpha)/V(\alpha)\; u(\gamma)]^{1/2} \exp[N(U(\alpha) - U(\gamma))]\,. \tag{15}$$

But the crucial question for the origin of life is not the abundance of ordered populations in the stationary solution but the rate of transitions from disorder to order.

To calculate the transition-rate, we consider an artificial steady-flow situation in which all populations arriving at the ordered state $j = N\gamma$ are immediately removed and replaced by disordered populations at $j = 0$. Eq. (4) then becomes

$$\psi_j - \psi_{j-1} = \epsilon, j = N\gamma,$$
$$\psi_j - \psi_{j-1} = -\epsilon, j = 0, \quad (16)$$
$$\psi_j - \psi_{j-1} = 0, \text{otherwise},$$

where ϵ is the transition-rate, the probability per mutation that a population crosses the barrier from disorder to order. Eqs. (5), (16) are to be solved with the boundary condition

$$P_j = 0, \quad j \geqslant N\gamma \quad . \quad (17)$$

The steady-flow solution is

$$P_j = \epsilon \sum_{\ell=j}^{N\gamma-1} (g(\ell/N))^{-1} \exp[W(\ell) - W(j)]. \quad (18).$$

Since the total probability is normalized to unity, Eq. (18) implies

$$\epsilon^{-1} = \Sigma\Sigma_{\ell \geqslant j} (g(\ell/N))^{-1} \exp[W(\ell) - W(j)]. \quad (19)$$

When N is large, Eq. (19) becomes

$$\epsilon^{-1} = N^2 \int_0^{\gamma} dx \int_x^{\gamma} dy\, (g(y))^{-1} (u(y)/u(x))^{1/2}$$
$$\exp[N(U(y) - U(x))]. \quad (20)$$

The main contribution to the double integral (20) comes from x near α and y near β. Using a Gaussian approximation for the integrand near α and β, we find

$$\epsilon = (2\pi N)^{-1} [(1-\phi'(\alpha))(\phi'(\beta)-1)((\alpha-\alpha^2)/(\beta-\beta^2))]^{1/2}$$
$$\exp[N(U(\alpha) - U(\beta))] \quad . \quad (21)$$

The rate of transition depends exponentially on the potential difference $(U(\beta) - U(\alpha))$ between the saddle-point β and the disordered minimum α. A similar calculation gives the rate of transition in the reverse direction from order to disorder,

$$\eta = (2\pi N)^{-1} [(1-\phi'(\gamma))(\phi'(\beta)-1)((\gamma-\gamma^2)/(\beta-\beta^2))]^{1/2}$$
$$\exp[N(U(\gamma) - U(\beta)] \quad . \quad (22)$$

The ratio of the rates (η/ϵ) is the inverse of the population-ratio (15), in accordance with the principle of detailed balance.

Up to this point, the model is general and abstract and says nothing about the detailed chemical mechanisms by which population-changes occur. All the details are hidden in the function $\phi(x)$ which relates the catalytic activity of a newly-placed monomer to the catalytic activity of its parent population. To make the model concrete, a particular form of $\phi(x)$ must be chosen. After $\phi(x)$ is chosen, it will be easy to calculate numerically from Eq. (21) and (22) the rates of transitions between order and disorder. Ideally, the choice of $\phi(x)$ should be based on a complete theory of prebiotic chemistry. Since no such theory exists, the choice must be made arbitrarily.

First a short digression on the application of the model to present-day organisms. Kirkwood (1980) has discussed a similar model in connection with a theory of aging of cells. The curve $y = \phi(x)$ appropriate to modern cells has an unsymmetrical S-shape, crossing the line y = x at three unevenly spaced points, for example

$$\alpha = 0.05, \quad \beta = 0.999, \quad \gamma = 0.9999 \quad . \quad (23)$$

The behavior of the model is then qualitatively correct. The cell has two stable equilibrium states, the live state with an error-rate of 10^{-4}, and the dead state with an error-rate of 0.95, and an unstable equilibrium state with an error-rate of 10^{-3}. If a sudden injury produces an error-rate less than 10^{-3}, the cell will almost certainly recover and return to the live state. If an injury produces an error-rate greater than 10^{-3}, the cell will almost certainly die. The model correctly predicts that the death of a live cell is statistically possible whereas the resurrection of a dead cell is essentially impossible. The extreme asymmetry of the curve $y = \phi(x)$, with the crossing-points β and γ squeezed together close to x = 1, is a reflection of the extreme precision and fine-tuning of the modern metabolic apparatus. We may conjecture that the modern highly asymmetric $\phi(x)$ evolved gradually out of a primitive $\phi(x)$ which was less fine-tuned and less asymmetric.

Since we know little about prebiotic conditions, our choice of a primitive $\phi(x)$ is designed to be simple rather than realistic. Realistic details may be added later as knowledge of prebiotic chemistry increases. The primitive $\phi(x)$ is chosen to depend on two parameters a and b, specifying respectively the diversity of the population of monomers and the precision of the polymerizing catalysts. We assume that the reacting monomers are divided into (a + 1) equally abundant species, so that each site in a catalyst may be occupied either by one active monomer or by one of a inactive monomers. Thus

$$\phi(0) = (1 + a)^{-1} \quad (24)$$

is the probability that any particular site will be correctly occupied in the absence of autocatalysis. We assume that every catalyst which is itself correctly constituted can discriminate active from inactive monomers by a factor b. Thus

$$\phi(1) = (1 + (a/b))^{-1} \quad (25)$$

is the chance that a newly placed monomer will be active in a population which has all its previously placed monomers active. The form of $\phi(x)$ for intermediate values of x is suggested by thermodynamics. Each perfect catalyst lowers the activation energy required for correct placement of a monomer by

$$d = kT \log b \quad , \tag{26}$$

where k is Boltzmann's constant and T is the absolute temperature. We assume that, in a population with the fraction x of the monomers active, each catalyst lowers the activation energy for correct placement by (xd). The form of $\phi(x)$ is thereby determined to be

$$\phi(x) = [1 + ab^{-x}]^{-1} = [1 + \exp(A - Bx)]^{-1} \quad , \tag{27}$$

with

$$A = \log a \quad , \quad B = \log b \quad . \tag{28}$$

The potential (13) corresponding to this choice takes the simple form

$$U(x) = x \log x + (1-x) \log (1-x) + Ax - \frac{1}{2}Bx^2 \quad . \tag{29}$$

The crucial assumption, that every imperfect catalyst produces an energy-lowering proportional to x, is an approximation consistent with our Assumption 7 in Section I. The approximation would obviously be false for an elaborate catalytic apparatus such as a modern ribosome. For a population containing a variety of primitive catalysts of relatively simple structure, the approximation should be roughly correct in some average sense.

Assumption 9 of Section I imposes restrictions on the allowed values of the parameters a, b. For the curve $y = \phi(x)$ to cross the line $y = x$ three times, it is necessary (but not sufficient) that the slope at the point of inflection be greater than unity. So when $\phi(x)$ has the form (27), a necessary condition for the existence of stable ordered and disordered equilibrium populations is

$$B > 4 \quad , \quad b > e^4 = 54.6 \quad . \tag{30}$$

This means that the catalysts must have discrimination-factors of the order of 100. They need not be as fine-tuned as modern enzymes which usually have discrimination-factors between 10^3 and 10^4. Another necessary condition for a triple-crossing is

$$A > 2 \quad , \quad a > e^2 = 7.4 \quad , \tag{31}$$

which follows from the requirement that the curve has slope exceeding unity at the unstable crossing-point β. Eq. (31) means that there must not be too few species of monomer. It may be significant that no model satisfying Eq. (3) can be constructed with $a = 3$, which would describe a pure nucleic-acid system built out of 4 types of nucleotide. Systems built out of 10 or 20 species of amino-acid, or mixed systems built out of amino-acids and nucleotides, can be accommodated within our model without difficulty.

The fact that the model fails with less than 9 species of monomer may seem paradoxical. The model can of course be defined for any value of a. What happens for $a \leqslant 7$ is that the model does not allow an order-disorder transition; there is either a disordered equilibrium state (if the catalysts are weak) or an ordered equilibrium state (if the catalysts are strong) but no possibility of order and disorder coexisting. When $a \leqslant 7$, the model says that cells will either be incapable of living or incapable of dying.

The precise range of values of the parameters (A,B) which allow an order-disorder transition is given by

$$A_- \leqslant A \leqslant A_+ \quad , \tag{32}$$

where

$$A_- = 1 + \exp(-\theta) + \theta \quad , \quad A_+ = 1 + \exp(\theta) - \theta \quad , \tag{33}$$

and θ is defined by

$$B = A_- + A_+ = 2 + 2 \cosh \theta \quad . \tag{34}$$

The conditions (32)–(34) define a wedge-shaped region in the (A,B)-plane extending upward and to the right from the cusp at $A = 2$, $B = 4$. We shall study three representative sequences of models, one sequence which we call "symmetric" having

$$A = \frac{1}{2}B \quad , \quad b = a^2 \quad , \tag{35}$$

one sequence called "marginally alive" having

$$A = A_+ \quad , \tag{36}$$

and the third sequence called "marginally immortal" having

$$A = A_- \quad . \tag{37}$$

The symmetric models have the curve $y = \phi(x)$ symmetrical about its point of inflection at $x = y = 1/2$, so that the three solutions of Eq. (1) are

$$\alpha, \beta = \frac{1}{2} \quad , \quad \gamma = 1 - \alpha \quad , \tag{38}$$

and the potential U(x) has equal minima at α and γ. While there is no physical or chemical reason to expect Eq. (35) to hold exactly, it is reasonable to expect the

efficiency of catalysis (specified by b) to increase with the number of monomer species (specified by a). For example, the symmetric model a = 10, b = 100 would describe proto-enzymes of modest performance built out of a restricted set of 11 types of amino-acid, while the model a = 19, b = 361 would describe more capable enzymes built out of a full modern complement of 20 amino-acids. The marginally alive and immortal models are as unsymmetrical as possible consistent with Eq. (32). They have the weakest and the strongest catalysts allowing the existence of an order-disorder transition with a given number of monomer species. They have the curve $y = \phi(x)$ touching the line $y = x$, so that the three solutions of Eq. (1) are for the marginally alive models

$$\alpha, \beta = \gamma = [1 + \exp(-\theta)]^{-1} \quad , \tag{39}$$

and for the marginally immortal models

$$\alpha = \beta = [1 + \exp(\theta)]^{-1} \quad , \quad \gamma \quad . \tag{40}$$

These models are on the borderline between triple-crossing and single-crossing. In the marginally alive models, the potential U(x) has a point of inflection rather than a minimum at γ, so that the ordered equilibrium state is marginally unstable. If the catalysts were infinitesimally stronger there would be a true disorder-order transition; if the catalysts were infinitesimally weaker there would be no ordered state at all. Similarly, in the marginally immortal models, U(x) has a point of inflection at α, and the disordered state is marginally unstable. If the catalysts were infinitesimally weaker there would be a true disorder-order transition; if the catalysts were infinitesimally stronger there would be no disordered state. The symmetric, marginally alive and marginally immortal sequences cover the range of interesting models and have the advantage of being easy to calculate analytically. Intermediate cases could be computed numerically if required.

In the symmetric models, it is convenient to express everything in terms of a parameter q defined by

$$\alpha = (q + 1)^{-1} \quad , \quad \gamma = q(q + 1)^{-1} \quad , \quad q > 1 \quad . \tag{41}$$

Eq. (1) with $x = \alpha$ gives

$$A = ((q + 1)/(q - 1)) \log q \quad , \tag{42}$$

and the coefficient of (-N) in the exponent of Eq. (21) is

$$\Delta = U(\beta) - U(\alpha) = \log(\tfrac{1}{2}(1 + q)) - \\ - ((3q + 1)/(4q + 4)) \log q. \tag{43}$$

The mean time for a population to make the transition from disorder to order is

$$t = F\tau \exp(\Delta N) \quad , \tag{44}$$

where τ is the average time between mutations at a given site, Δ is defined by Eq. (43), and F is an unimportant numerical factor depending on q.

In the marginally alive models, it is convenient to express everything in terms of a positive parameter ω, so that

$$A = \omega^2 p^{-1} + \log(p/r) \quad , \tag{45}$$

$$B = \omega^2 (p^{-1} + r^{-1}) \quad , \tag{46}$$

$$p = \exp(\omega) - 1 - \omega \quad , \quad r = \exp(-\omega) - 1 + \omega \quad . \tag{47}$$

The solutions of Eq. (1) are then

$$\alpha = r(p + r)^{-1} \quad , \quad \beta = \gamma = r\omega^{-1}(\omega - r)^{-1} \quad . \tag{48}$$

The lifetime formula (44) holds with

$$\Delta = \tfrac{1}{2}(s - 1 - \log s) \quad , \tag{49}$$

$$s = \omega^2 (p + r)^{-1} \quad . \tag{50}$$

In the marginally immortal models, Eq. (46), (47) still hold, but now

$$A = \omega^2 r^{-1} + \log(r/p) \quad , \tag{51}$$

$$\alpha = \beta = p\omega^{-1}(\omega + p)^{-1} \quad , \quad \gamma = p(p + r)^{-1} \quad , \tag{52}$$

and Δ is zero by definition since $\alpha = \beta$. The vanishing of Δ means that the transition from disorder to order becomes infinitely rapid as the parameters of the model tend to the limit $A = A_{\_}$.

We have no way to guess the number of island populations that may have existed in the remote past, or the duration of their existence. If we make the conservative assumption that in some suitably favorable environment 10^{10} islands existed for 10^5 mutation-times, then a substantial number of them could have made the transition from disorder to order according to Eq. (44), provided that

$$N < N_c = 30\,\Delta^{-1} \quad . \tag{53}$$

In any case, Eq. (53) gives a rough idea of the maximum population which could be expected to make the order-disorder jump with reasonable probability. Table 1 gives values of $a, b, \alpha, \beta, \gamma, \Delta, N_c$ for an assortment of symmetric, marginally alive and marginally immortal models.

The main conclusion to be drawn from Table 1 is that the critical population-sizes can be large for what appear to be reasonable choices of the input parameters a and b. As a representative and not extreme case, we may take the symmetric model with parameters a = 10 (11 species

Table 1. Parameters of representative models

a	b	α	β	γ	Δ	N_c
8	62.9	0.32	0.59	0.59	0.002695	11131
8	64	0.33	0.50	0.67	0.001129[a]	26566
8	65.7	0.39	0.39	0.70	0	∞
10	89.4	0.19	0.67	0.67	0.0298	1006
10	100.0	0.20	0.50	0.80	0.0145	2070
10	128.0	0.29	0.29	0.87	0	∞
19	219.3	0.07	0.75	0.75	0.1906	157
19	361	0.08	0.50	0.92	0.1051	285
19	3195	0.14	0.14	0.99	0	∞

In each model, (1+a) is the number of monomer species and b is the discrimination factor of the catalysts. For each value of a, the model with $\beta = \gamma$ is marginally alive, the model with $\beta = 1/2$ is symmetric, and the model with $\beta = \alpha$ is marginally immortal. The fraction of monomers active is α in the disordered state, γ in the ordered state. N_c is the maximum size of population for which a disorder-order transition occurs with reasonable probability.

a As noted by H.C. Longuet-Higgins, this value of Δ is equal to (log 3 -(19/12) log 2), which is known to musicians as the fractional difference in pitch between a true fifth and an equi-tempered fifth. Its smallness is the main reason for the possibility of a harmonious equi-tempered scale

of monomer) and b = 100 (catalysts of moderate specificity). This model will have 20% of the monomers active in the disordered phase and 80% active in the ordered phase. It allows populations of up to 2000 monomers to make the transition from disorder to order without requiring the occurrence of a miracle.

III. Questions

This section consists of a list of questions with a brief discussion of possible answers.

1. Were the first living creatures composed of polypeptides or nucleic acids or a mixture of the two?

For a discussion of chemical evidence bearing on this question see Miller and Orgel (1974), especially chapter 12. For recent studies of models of the origin of life including nucleic acid replication from the beginning, see Niesert, Harnasch and Bresch (1980, 1981), or Kuhn and Waser (1981). For a theory not including nucleic acids at the beginning, see Weiss (1981). If the model of this paper has anything to do with the origin of life, then the first living creatures contained a population of a few thousand monomers assembled into mutually catalytic structures. It is difficult to imagine that a few thousand loosely organized monomers with a fraction of the order of 20% incorrectly placed could comprise anything resembling the modern genetic apparatus. It is easier to think of the earliest creatures as composed of amino-acids and other miscellaneous chemicals, polymerized into proto-enzymes with a mainly polypeptide structure. In other words, the model seems to imply that enzymes came first, genes second.

2. At what stage did random genetic drift give way to natural selection?

The model does not allow natural selection to operate, because it does not allow the island populations to grow or to reproduce. So long as there is no birth and death of islands, there can be no natural selection. However, if an island population has once reached the ordered state as defined in the model, there is a good chance that it will pass into a new phase of evolution by systematically assimilating fresh monomers from its environment. An island which increases its population N by assimilation will quickly become stabilized against reversion to the disordered state according to Eq. (22). It can then continue to grow until some physical disturbance causes it to divide. If it divides into two, there is a good chance that both daughter populations contain a sufficient assortment of catalysts to remain in the ordered state. The processes of growth and division can continue until the islands begin to exhaust the supply of nutrient monomers. When the nutrients are in short supply, some islands will lose their substance and die. From that point on, evolution will be driven by natural selection.

3. Does the model contradict the Central Dogma of molecular biology?

The Central Dogma (Crick 1957) states that genetic information can pass from nucleic acid to nucleic acid or to protein but cannot pass from protein to nucleic acid or to protein. The dogma is true for all contemporary organisms. The model implies that it was probably untrue for the earliest organisms. According to the model, the first organisms probably passed genetic information to their offspring in the form of enzymes, i.e. primitive proteins. There is no logical reason why a population of enzymes mutually catalyzing each other's synthesis should not serve as a carrier of genetic information.

4. How did nucleic acids originate?

If it is true that the first organisms contained no nucleic acids, then it is likely that they learned first to utilize nucleotide-related molecules such as ATP as energy-carriers. Efficient enzymatic machinery for synthesizing and using ATP would have evolved. Cells would then contain substantial populations of nucleotide-related molecules. If in one of these cells there should arise by chance an enzyme similar to the Q_β replicase in the experiment of Biebricher, Eigen and Luce (1981), then the nucleotides in the cell could organize themselves into RNA as they did in the experiment. The proliferation of RNA in the cell would begin as a parasitic infection and develop later into a symbiosis.

5. How did the modern genetic apparatus evolve?

After RNA was established as a normal constituent of cells, a genetic apparatus might have evolved by a sequence of steps such as the following. (a) Non-specific binding of RNA to free amino-acids, activating them for

easier polymerization. (b) Specific binding of RNA to catalytic sites to give them structural precision. (c) RNA bound to amino-acids becomes transfer RNA. (d) RNA bound to catalytic sites becomes ribosomal RNA. (e) Catalytic sites evolve from special-purpose to general-purpose by using transfer RNA instead of amino-acids for recognition. (f) Recognition unit splits off from ribosomal RNA and becomes messenger RNA. (g) Ribosomal structure becomes unique as the genetic code takes over the function of recognition. This is only one of many possible pathways which might have led to the evolution of the genetic code. The essential point is that all such pathways appear to be long and tortuous.

6. How late was the latest common ancestor of all living species?

The model asserts that cells came before enzymes, enzymes before genes. The geological record tells us that cells existed very early, as long as 3.5 eons ago. But there is no evidence that the earliest cells which are preserved as microfossils contained a modern genetic apparatus. The evolution described in the discussion of questions 4 and 5 may have taken eons to complete. We know from the universality of the genetic code that the latest common ancestor came after the end of that evolution. The pace of evolution may have accelerated after the genetic code was established, allowing the development from ancestral prokaryote to eukaryotic cells and multicellular organisms to be completed in less time than it took to go from primitive cell to ancestral prokaryote. It is therefore possible that the latest common ancestor came late in the history of life, perhaps as late as two-thirds of the way from the beginning.

7. Does there exist a concrete realization of the model, for example a population of a few thousand amino-acids forming an association of polypeptides which can catalyze each other's synthesis with 80% efficiency? Can such an association maintain itself in homeostatic equilibrium?

These are the crucial questions which only experiment can answer.

Acknowledgements. I am indebted to Leslie Orgel for starting me thinking about the origin of life, to Manfred Eigen, Hyman Hartman and John Bahcall for helpful discussions, and to Martin Rees for inviting me to the conference on "Matter into Life" in September 1981 at Cambridge, where the ideas contained in this paper took shape.

References

Biebricher CK, Eigen M, Luce R (1981) Kinetic Analysis of Template-instructed and de novo RNA Synthesis by Q_β Replicase. J Mol Biol 148:391–410

Bresch C, Niesert U, Harnasch D (1980) Hypercycles, Parasites and Packages. J Theor Biol 85:399–405

Crick FHC (1957) On Protein Synthesis. Symp Soc Exp Biol 12:138–163

Fisher RA (1930) The Genetical Theory of Natural Selection, (New York, Dover Publications reprint, 1958) Chapter 4

Kimura M (1970) Stochastic Processes in Population Genetics, In: Kojima KI (ed) Mathematical Topics in Population Genetics. Springer, Berlin Heidelberg New York, p 178–209

Kirkwood TBL (1980) Error Propagation in Intracellular Transfer. J Theor Biol 82:363–382

Kuhn H, Waser J (1981) Molecular Self-Organization and the Origin of Life. Angew Chem Int Ed Engl 20:500–520

Lohrmann R, Bridson PK, Orgel LE (1980) Efficient Metal-Ion Catalyzed Template-directed Oligonucleotide Synthesis. Science 208:1464–1465

Miller SM, Orgel LE (1974) The Origins of Life on the Earth. Prentice-Hall, Englewood Cliffs NJ

Niesert U, Harnasch D, Bresch C (1981) Origin of Life between Scylla and Charybdis. J Mol Evol 17 :348–353

Weiss A (1981) Replication and Evolution in Inorganic Systems. Angew Chem Int Ed Engl 20:850–860

Wright S (1931) Evolution in Mendelian Populations. Genetics 10:97–159

Received December 14, 1981 / Revised April 30, 1982