J Mol Evol (1981) 17:348–353

Journal of Molecular Evolution

Origin of Life Between Scylla and Charybdis

U. Niesert[1], D. Harnasch[1], and C. Bresch[2]

[1] Zentrallabor für Mutagenitätsprüfung der Deutschen Forschungsgemeinschaft, Breisacher Strasse 33, D–7800 Freiburg, Federal Republic of Germany

[2] Institut für Biologie III der Universität Freiburg, Schänzlestrasse 1, D–7800 Freiburg, Federal Republic of Germany

Summary. The "package model" discussed here is concerned with the preservation of genetic information by primordial compartments. Each viable "package" encloses a complete set of unlinked genes in varying numbers of copies. Due to stochastic distribution and error-prone replication two potent perils endanger the informational integrity of packages: fluctuation and mutation.

A computer simulation was used to quantify the effects of fluctuation, mutation, and package death by accident. Assuming reasonable rates for these parameters it is suggested that life started out with compartments containing not more than 3 different genes.

Key words: Origin of life – Primitive compartments – Hypercycle – Mutation rate – Computer simulation

Introduction

The understanding of the origin of life has recently been deepened substantially by Eigen and Schuster (1977, 1978a, b). They argue convincingly that at spontaneous error rates of about 10^{-2} per nucleotide (Pörschke 1977) the first faithfully reproducing molecules should be single-stranded RNAs not exceeding 50 to 100 nucleotides. Stability against hydrolysis (Usher and McHale 1976) and replicability of the template (Lohrmann and Orgel 1979) might favor RNAs that display a loop-and-stem structure similar to that of modern tRNAs. Some of them are supposed to participate in ribosome-free translation of an appropriate messenger as suggested by Crick et al. (1976). Reproducible translation would be greatly enhanced by a periodic frame in both, messengers and tRNAs. Eigen and Schuster suggest this to be a GNC-frame in all RNAs, permitting a single RNA ancestor. The identity of the GNC-frame with its complementary counterpart facilitates internal folding and hence yields high single-strand stabilities, an advantage in thermodynamic selection. The GNC-frame can be favored against its CNG-competitor, because its four codons GGC, GCC, GAC, GUC are translated into glycine, alanine, aspartic acid, and valine, i.e. into those four amino acids that arise most abundantly by chemical evolution (Miller and Orgel 1974; Oro et al. 1971).

Furthermore, Eigen and Schuster have elaborated the earlier abstract concept of the hypercycle (Eigen 1971) as a model for co-evolution of several spatially unlinked RNAs by cyclic functional coupling: Each member of the hypercycle, a pair of RNA + and – strands, has to encode a polypeptide supporting specifically the replication of the next member and to display an appropriate target for the "replicase" encoded by the preceeding unit. In addition, at least some members, serving as tRNAs have to participate in the general translation mechanism.

Though we are fascinated by the fundamental suggestions on the physico-chemical nature of primordial nucleic acids we have pronounced reservations about the role of a hypercyclic organization at the origin of life (Bresch et al. 1980). These doubts about the evolutionary potential of the hypercycle are based on the indispensable requirement of triple function for its RNA-members while only one feature, namely the target quality is positively selected. This results in the general weakness in the ability of hypercycles to overcome uncooperative mutants that constantly arise from erroneous replication of the members of the hypercycle.

Abbreviations: A = adenine, C = cytosine, G = guanine, U = uracil, N = any one of these 4 bases

Offprint requests to: Ursula Niesert

0022–2844/81/0017/0348/$01.20

In this paper we will discuss an alternative "package" model, which is consistent with Eigen and Schuster's assumptions relative to primordial RNA properties, but circumvents the threats inherent in the hypercycle. A simple version of this genetic model was computer simulated to demonstrate its viability in principle.

Package Model

Packages are conceived as small compartments in the primordial broth, enclosing a few RNA molecules and a few polypeptides. Copying of RNA molecules is aided by an enzymatic activity of a polypeptide, the "replicase". Primitive "translation" is achieved by cooperation of RNAs with tRNA-character and an RNA with messenger function to yield the replicase. All RNAs in a package compete for the standard replicase. But if all RNAs are derived from a common ancestor, they all will display identical targets. Thus, no systematic preferences occur among the RNAs. Occasional fission of the package is due to unspecified physical forces. When a package is split its contents are randomly distributed to the arising daughter packages.

In this model *Darwinian selection* prevails: The unit of selection is the entire package. Being enclosed in packages, the genes[1] can evolve to better functions, since mutational improvements will enhance the reproductive capacity of the package and its descendants. Moreover, an increase in information content of the package can occur if a mutated RNA displays a new function.

A package is threatened by several dangers:

- Inherent in the model is the "fluctuation abyss". As a consequence of the stochastic nature of RNA replication and package fission some daughter packages will be deprived entirely of one gene or the other. Such packages have irreversibly lost their reproductive capacity.

Less transparent in their implication are two main types of harmful mutations:

- "Parasites" are defined as mutated genes which have an impaired function, but an unchanged replication rate.
- "Lethals" are "dominant" mutants that can destroy the entire offspring. For example, mutants that cause highly error-prone replication or translation lead to an error catastrophe. Similarly, mutants with an increased target affinity towards the replicase would, by their selfish behaviour of constantly retaining the replicase, drive the lineage down the fluctuation abyss.

The terms "parasite rate" and "lethal rate" refer to the corresponding mutation probabilities per RNA replication.

A crucial parameter of the package model is the average *N*umber *O*f *R*eplicated *M*olecules (NORM) between two package fissions which is influenced by a multitude of environmental and internal package factors. Obviously at too small NORMs package families are approaching the fluctuation abyss. To avoid this, it seems necessary to steer towards higher NORMs. But resourceful Odysseus will immediately realize the menace in this decision, namely, the higher the NORM the larger the risk to the package of falling victim to deleterious mutations of the parasite or lethal type ("mutational reefs").

Thus, packages can be viable only within a limited NORM-range. Increase in the number of different genes to be preserved results in a narrowing of this range of viability. Furthermore, high mutation rates will drastically cut down this range as will high "accident" rates, i.e. rates of package death due to unspecific causes (e.g. heat). Conditions under which packages still prove viable were determined by a computer simulation.

Simulation

A simplified version of the package model was analyzed by a computer simulation written in FORTRAN IV and run on the Univac 1100 of the University of Freiburg.

A population of packages is stored as a matrix in which each package is represented by the numbers of copies of the different genes. Within a generation cycle all packages are "processed" (including fission). To avoid a population overflow, only the most "promising" (see below) daughters are returned to the parental store.

Processing of an individual package consists in the following steps:

1. Consecutive replications according to an assumed constant NORM-value (variation of NORM in a package population, though certainly nearer to reality, has not been taken into account in this simple simulation). The following cycle of replication is repeated NORM times: The molecule to be copied is chosen randomly from the pool present in the package. If a functioning RNA molecule is chosen, the quality of its copy (error-free, parasitic, or lethal) is determined stochastically according to the respective copying (in)fidelities. A parasite chosen, however, is always copied as a parasite, or even mutated into a lethal. Then, the original molecule and the new copy are returned to the pool before the next cycle of replication starts.

[1] A gene in this context is considered any information, coded as an RNA molecule, contributing an essential function to the package.

2. Fission: After replication is completed, the molecules are randomly distributed between two daughter packages.
3. The viability of both daughter packages is checked. Any package that lacks one of the genes or contains a lethal mutant will be dismissed. Furthermore, it is stochastically determined whether a package will suffer an accident before its next fission.

Exponentially growing package populations as resulting from the model would soon surpass any computer capacity. This problem can be avoided by randomly discarding surplus packages. We, however, preferred to discard those supernumerary packages with a low chance of leaving viable descendents. This prognosis depends essentially on three properties: "package balance", i.e. degree of equipartition of the copies among the different genes, "parasitic load", i.e. number of parasites carried, and "package size", i.e. total number of gene copies.

Attaching different importance to these three aspects, one can define arbitrarily a variety of "prospective values". The formula used which we considered suitable is discussed in a short appendix.

Results

To determine the range of viability between the fluctuation abyss and the mutational reefs, the computer simulation takes the following into account:

- as fundamental quantity: the number of participating genes,
- as leading variable: the NORM-value,
- "costarring": lethal rate, parasite rate, and accident rate.

Starting from whatever conditions, the average number of RNA molecules per package will, after a few generations, approach values close to NORM. Hence, we initially equipped the ancestral package with about NORM copies evenly distributed among the different genes. The bonus of starting with optimal packages is worn out after about 10 generations, and thus does not distort the results essentially. To assure statistical significance 25 identical packages are used to start with. Any run is repeated twice, totalling to 75 packages tested as ancestors. Finally, a set of conditions used is classified "hospitable", if packages are still present after 1000 generations in each of the three parallel trials.

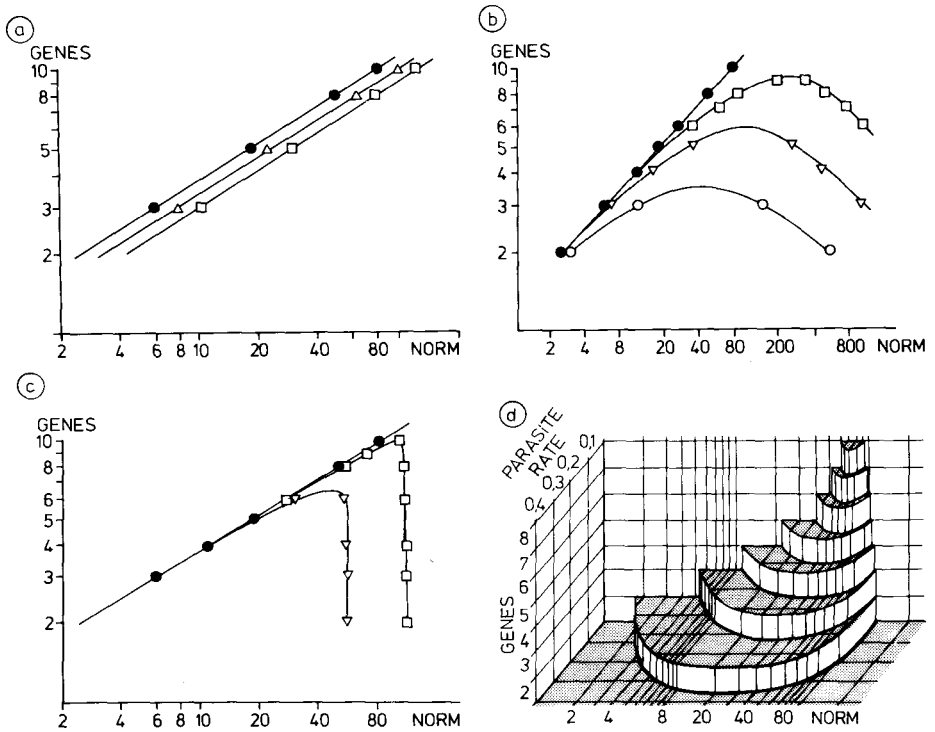

Fig. 1 a-d. Limits to the realm of viability. The plots illustrate the relation between the two crucial variables: the number of different genes in the set to be conserved and NORM, i.e. the *N*umber *O*f *R*eplicated *M*olecules in a package between two fissions. Each curve indicates the maximum number of genes that can be conserved at a given NORM and thus represents a limit to the realm of viability. The standard curve (– • –in plots a-c) represents the model-inherent "fluctuation abyss" (see text) without any further restraints. a) Accidents (– △ – 0.1, – □ – 0.2 per package) cause a parallel shift of the fluctuation abyss. b) Mutations of the parasite type (– □ – 0.05, – △ – 0.1, – ○ – 0.2 per replicated molecule) limit the realm of viability with respect to high NORMs and – even more drastically – with respect to gene number. c) Mutations of the lethal type (– □ – 0.005, – △ – 0.01, per replicated molecule) entail a sharp boundary at NORMs where almost 50% of the packages have acquired one such mutant. d) Here, the plot scheme is extended by a third important quantity: the rate of (parasite) mutations that is maximally tolerated at defined NORM and gene number. The other two parameters which can be assessed more easily (cf. a and c) were kept constant (accident rate = 0.05, lethal rate = 0.005)

Reduction in the range of viability as a consequence of accident rate, parasite rate, or lethal rate was first analyzed separately: The minimal NORM-values to evade the fluctuation abyss are shown in Fig. 1a for different accident rates. The existence of an upper limit in high NORM-values appears with the introduction of parasites (Fig. 1b). This is even more pronounced with lethal mutations leading to a sharp boundary (independent of the gene-number assumed) presented for two lethal rates in Fig. 1c.

Obviously, a multitude of *combinations* of different parameter values can be tested. In this approach we confined ourselves to assuming a constant accident rate of 1 in 20 packages and a constant lethal rate of 1 in 200 replicated molecules. Under these conditions Fig. 1d demonstrates the maximally tolerated parasite rate as a function of NORM for different numbers of participating genes. Data relevant for comparison with estimates on spontaneous replication fidelities (Pörschke 1977) are extracted from Fig. 1d and represented in Table 1.

Table 1. Maximally tolerated infidelities for different numbers of genes at an accident rate of 0.05 per package and a constant lethal rate of 0.005 per replication

Number of genes	Maximally tolerated parasite rate per RNA replication	In the NORM-range of	Corresponding infidelities per nucleotide in RNAs of	
			50 N [%]	100 N [%]
2	0.35	10–40	0.83	0.42
3	0.2	20–80	0.44	0.22
4	0.13	30–80	0.26	0.13
5	0.09	40–80	0.18	0.09

Discussion

As the origin-of-life problem bears important consequences on the general understanding of cosmos and mankind, there have been many attempts to investigate this question. Miller's approach of synthesizing amino acids under prebiotic conditions (1953) opened a new era. Since then, the geological, physical, and chemical requirements for the beginning of life have been discussed with diligence (Horowitz et al. 1970; Oparin 1976; Kaplan 1978; Margulis et al. 1979; Woese 1979). The data of Oparin (1953) and Fox (1965) have, furthermore cast some light on the "bio"chemistry of primitive compartments.

The package model is quite remote from such detailed biochemical efforts. It is only supplementing them by contributing the *genetic* aspect of early life. The model is to assess the problem of how to conserve the genetic information in spite of package-inherent disturbances (namely the stochastic character of fission and replication) and the impairing influence of accidents and mutations.

To explore the realm of viability between fluctuation abyss and mutational reefs the computer simulation requires a decision on several technical quantities. Their possible influences on the results obtained shall be discussed now:

1. *Parental storage size.* Obviously, an unlimited propagation of packages is incompatible with a computer simulation. For too small a storage size no discrimination is achieved between hospitable and inhospitable conditions as even a small series of unfortunate events will sooner or later eliminate any lineage. In a series of test runs the storage size of 25 packages proved a satisfactory compromise between calculation time and statistical significance. This limitation of storage results in a reduction of the realm of viability.
2. *Prospective value.* Limited computer capacity requires supernumerary daughter packages to be discarded in every generation according to a certainly arbitrary *prospective value.* Whatever formula one uses the reduction mechanism will never artificially expand the realm of viability beyond its "true" size. Inadequate formulae, though, will reduce it more drastically than suitable ones.
3. *Number of generations.* Any limitation in this technical quantity will apparently enlarge the realm of viability. To be classified hospitable a set of parameter values was requested to allow viable packages for 1000 generations in each of the three trials. This might seem exaggerated. Smaller generation numbers, however, would increase artificially the realm of viability even further. We wanted to keep this increase as small as justifiable with respect to computer time.
4. *Variance of NORM.* Certainly the number of molecules replicated within a fission cycle would vary in different cycles and among different packages. We arbitrarily used a fixed NORM-value for all packages in a computer run, although any substantial variation in NORM would reduce the realm of viability.

In brief: while the limitation in storage size and the prospective value lead to an underestimate of the realm of viability, the limitation in the number of the generations tested and the lack of variance in NORM tend to increase it. In spite of the distortions by these simplifications and approximations, it seems possible to recognize certain restrictions in the conditions hospitable to the origin of life.

The multitude of parameters involved required stringently limiting certain parameters to plausible constant values. We therefore confined our calculation to an accident rate of 0.05. If this value seems

too low one should take into consideration that at different locations of the young planet different values may prevail and that those areas with low accident rates will be the interesting ones. Furthermore, we assumed lethal mutations to occur in 1 out of 200 RNA replications. If 1 out of 5 replications will show any type of mutation, then the ratio of mutations leading to a loss of function to mutations with lethal consequences would be 40:1. Both assumptions seem acceptable to us.

Enzyme-free replication of RNA works with an inherent mutation rate per nucleotide of about 2%, this value being estimated from base pairing stabilities (Pörschke 1977; Eigen and Schuster 1977, 1978a, b). If one agrees with the assumption that the participation of an early replicase would not increase the fidelity more than tenfold, then two conclusions can be drawn: 1. The origin of life could not have involved more than 3 genes. 2. The average number of RNA molecules replicated within one fission cycle must not have been smaller than 20 or larger than 80.

The reader will certainly have realized that the simplified model used for the computer simulation has neglected entirely two essential features involved in the beginning of life: the complementarit of RNA strands and the mechanism of translation. Accounting for these two aspects will obviously result in even more restricted conditions. Without a detailed simulation of the more complicated model it remains a pure guess whether these further restrictions could still be accomodated by an acceptable set of parameter values.

The simplified version presented here considers the mechanism of *translation* to operate in sufficient quantity and quality, i.e. to have no effect on the survival prospects of the package lineage. The influence of the *complementarity* could even operate in two directions: on one hand the alternation between + and – strands in the course of replication would halve the number of functionally efficient molecules and thus require a higher NORM-value for viability of the lineage. On the other hand complementarity could permit a reduction in the minimal gene number if both, + and – strand of a tRNA gene would serve as tRNAs for two different amino acids (N.B. if one strand can form a clover-leaf structure, the complementary strand can do so as well).

However, relief from these restrictions will – perhaps very quickly – be achieved by fortunate mutations, that certainly will occur and improve the prospects for viability of the lineage. In particular, evolution of the replicase gene will be accomplished in two ways: Replicases with *higher affinity towards the standard target* will reduce the rate of lethal target mutations. Replicases of *increased fidelity* will reduce the probabilities for *all* dangerous mutations and permit higher NORM-values and even the elongation of genes. All in all this improvement opens the possibility for an increase in the number of different participating genes: tRNA genes with + and – strands performing as tRNAs for two different amino acids may expand to two non-overlapping tRNA genes by divergent evolution of + and – strands; additional structural genes could be used to improve the efficiency of the translation process; and finally, additional tRNA genes would render accessible advantages of a more elaborate code. But before such progress can occur life has to find its way to pass through the narrows between fluctuation abyss and mutational reefs.

Acknowledgements. This work was supported by the Deutsche Forschungsgemeinschaft. The Albert-Ludwigs-Universität kindly provided the computer facilities.

Appendix

The Prospective Value Used

With

k being the number of different genes assumed,
n_i the number of copies of the i^{th} gene, and
n_p the number of parasites in a package, with

$$N = \sum_{i=1}^{k} n_i \quad \text{, and}$$

$T = N + n_p$ the number of RNA molecules carrying the standard target, we can introduce:

$$\frac{n_1}{N} \cdot \frac{n_2}{N} \cdots \frac{n_k}{N} = \prod_{i=1}^{k} \frac{n_i}{N}$$ the "balance factor",

measuring the degree of equipartition of the copy numbers among the different genes,

k^k a constant in each computer run to normalize the balance factor for any number of different genes assumed,

$\left(\frac{N}{T}\right)^k$ the "parasitic load factor",
accounting for the fraction of parasites in a package. The exponent k was introduced, since the same fraction of parasites becomes the more dangerous the larger the number of different genes sharing in the functional copies,

$\log_2 (N - k + 2)$ the "surplus factor",
which reflects the package size – or more precisely: the number of gene copies surpassing the minimal requirement for viability of one copy per gene. This factor was constructed to become 1, if there is just one copy per gene in the package. The use of the logarithm decreases the weight of more and more surplus copies added. The base 2 for the logarithm is completely arbitrary as is finally – let us stress again – the entire formula for the prospective value we used:

$$V = k^k \cdot \left(\prod_{i=1}^{k} \frac{n_i}{N}\right) \cdot \left(\frac{N}{T}\right)^k \cdot \log_2 (N - k + 2) .$$

References

Bresch C, Niesert U, Harnasch D (1980) J Theor Biol 85:399–405

Crick FHC, Brenner S, Klug A, Pieczenik G (1976) Orig Life 7:389–397

Eigen M (1971) Naturwissenschaften 58:465–523

Eigen M, Schuster P (1977) Naturwissenschaften 64:541–565

Eigen M, Schuster P (1978a) Naturwissenschaften 65:7–41

Eigen M, Schuster P (1978b) Naturwissenschaften 65:341–369

Fox SW (1965) Nature 205:328–340

Horowitz NH, Drake FD, Miller SL, Orgel LE, Sagan C (1970) The origins of life. In: Handler P (ed) Biology and the Future of Man. Oxford University Press, London, pp 163–201

Kaplan RW (1978) Der Ursprung des Lebens. Biogenetik, ein Forschungsgebiet heutiger Naturwissenschaft, 2nd ed. Georg Thieme Verlag, Stuttgart

Lohrmann R, Orgel LE (1979) J Mol Evol 14:243–250

Margulis L, Mazur P, Barghoorn ES, Halvorson HO, Jukes TH, Kaplan IR (1979) J Mol Evol 14:223–232

Miller SL (1953) Science 117:528–529

Miller SL, Orgel LE (1974) The origins of life on earth. Prentice Hall New Jersey

Oparin A (1953) The origin of life (republication of 1938 ed) Dover Publ, New York

Oparin A (1976) Orig Life 7:3–8

Oro J, Gilbert J, Lichtenstein H, Wikstrom S, Flory DA (1971) Nature 230:105–106

Pörschke D (1977) Elementary steps of base recognition and helix-coil transitions in nucleic acids. In: Pecht I, Rigler R (eds) Mol Biol Biochem Biophys 24:191–218

Usher DA, McHale AH (1976) Proc Natl Acad Sci USA 73: 1149–1153

Woese CR (1979) J Mol Evol 13:95–101

Received September 4, 1980/Revised February 18, 1981