Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 8253–8256, December 1987
Biochemistry

Selection by differential molecular survival: A possible mechanism of early chemical evolution

(origin of life/catalysts/metabolism)

CHRISTIAN DE DUVE

The Rockefeller University, New York, NY 10021 and International Institute of Cellular and Molecular Pathology, 75, Avenue Hippocrate, B-1200 Brussels, Belgium

Contributed by Christian de Duve, August 3, 1987

ABSTRACT **A model is proposed to account for selective chemical evolution, progressing from a relatively simple initial set of abiotic synthetic phenomena up to the elaborately sophisticated processes that are almost certainly required to produce the complex molecules, such as replicatable RNA-like oligonucleotides, needed for a Darwinian form of selection to start operating. The model makes the following assumptions: (*i*) that a small number of micromolecular substances were present at high concentration; (*ii*) that a random assembly mechanism combined these molecules into a variety of multimeric compounds comprising a wide repertoire of rudimentary catalytic activities; and (*iii*) that a lytic system capable of breaking down the assembled products existed. The model assumes further that catalysts supplied with substrates were significantly protected against breakdown. It is shown that, by granting these assumptions, an increasingly complex network of metabolic pathways would progressively be established. At the same time, the catalysts concerned would accumulate selectively to become choice substrates for elongation and other modifications that could enhance their efficiency, as well as their survival. Chemical evolution would thus proceed by a dual process of metabolic extension and catalytic innovation. Such a process should be largely deterministic and predictable from initial conditions.**

Most theories of the origin of life start from the premise that primitive earth conditions were such that all the components necessary for some sort of Darwinian selection to start operating were provided. These theories take for granted that abiotic syntheses, driven by no more than the prevailing physicochemical conditions, proceeded all the way to the formation of authentic replicating information-conserving molecules capable of affecting their environment in a manner that reflects back on their own rate of replication. RNA-like polynucleotides and RNA-encoded polypeptides are the favorite candidates for the primitive genotype–phenotype couple (for instance, see ref. 1).

As pointed out by Shapiro (2) and, even more forcefully by Cairns-Smith (3), this belief, especially with regard to the synthesis of polynucleotides, credits the random operation of primitive chemical mechanisms with powers of discrimination that even the most inventive organic chemists have not been able to equal under the highly artificial and sophisticated conditions of the laboratory. It is accepted much less for its likelihood than for the lack of an alternative. A few workers refuse to be so resigned and are sufficiently impressed by the implausibility argument to go on searching for an alternative. Direct protein replication has been postulated by Dillon (4) and by Shapiro (2), but they offer no corroborative evidence, except for presumptive properties of the scrapie agent and other "prions" (5). However, the possibility that these infectious particles might be made of self-replicating proteins, not coded by nucleic acids, has not been borne out (6–8). A more radical proposal has been made and elaborated in great detail by Cairns-Smith (3) who postulates an initial phase governed by mineral genes, probably made of clay, during which the whole protein–nucleic acid apparatus developed progressively until it took over control, and the clay genes were discarded. A difficulty with this theory, in addition to its lack of empirical or experimental support, is that it does not convincingly explain what kind of replicative advantage clay genes might derive from the development of an increasingly complex organic machinery.

Both of these alternatives remain firmly rooted in a Darwinian selection mechanism operating by way of preferential replication. They differ from the conventional theory only by the nature of the postulated first genetic material, protein or clay, instead of RNA. The only recent theory that tries to evade a Darwinian mechanism altogether is that of Dyson (9), who has presented a model based on Kimura's theory of evolution by genetic drift (22). His model is attractive and yields plausible quantitative predictions, but it calls for a population of catalytic oligopeptides (or other comparable oligomers) busily reshuffling each other's structures—not a very realistic assumption.

In the present paper, it is pointed out that selection at the chemical level can operate by the preferential survival of useful molecules, as well as by their preferential replication, and a model is proposed to explain the development of an increasingly complex metabolic network from relatively simple abiotic precursors by such a mechanism.

Main Features of the Model

The model requires abiotic mechanisms to provide the following ingredients: (*i*) a number of small molecules comprising building blocks for the synthesizer mentioned below and one or more substrates for the catalytic activities of its products; (*ii*) a random synthesizer assembling building blocks into a variety of oligomeric or polymeric compounds, a number of which exhibit some sort of specific catalytic activity; and (*iii*) a lytic activity capable of breaking down the products of the synthesizer.

In such a system, assuming that lysis is the only reaction causing the disappearance of synthetic products and that it obeys first-order kinetics, the amount N_j of synthetic compound j present under steady-state conditions is given by:

$$N_j = \frac{R_j}{k_j}, \qquad [1]$$

in which R_j is its rate of synthesis and k_j its breakdown rate constant.

What Eq. 1 expresses is the obvious fact that the amount of a given synthetic product is determined both by its rate of synthesis and by its rate of breakdown. However, all prebiotic models so far have singled out enhanced synthesis

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

as the response to the selective criterion of usefulness (by way of a Darwinian feedback loop). It is proposed here that usefulness could be linked to decreased breakdown through the stabilization of catalytically active synthetic products by their substrates. This is a well-known property of enzymes. The extent to which it could characterize the more primitive catalysts considered here defines the main uncertainty of the model.

Granting this assumption, it is clear that those catalysts made by the synthesizer that find appropriate substrates in the environment will be selectively protected and will, therefore, accumulate preferentially in the system. But this is only a first step. Through their activities, they will create new products, which will stabilize new catalysts, which in turn will make new products, and so on. In this way, an increasingly complex network of prebiotic "metabolic" pathways will be laid down, and the necessary catalysts will emerge, with as an only additional requirement, besides the need for catalyst–substrate stabilization, a sufficiently rich repertoire of catalytic activities among the synthesizer products to make further development of the network possible.

As catalysts accumulate in the system, they become themselves favored substrates (by mass-action effect) for chemical modification, including elongation by the synthesizer. Improved or novel catalysts arising in this way will in turn be selected if they are supplied with substrates that stabilize them. The model thus allows a progressive exploration of the catalytic as well as of the metabolic landscape. It is assumed that this dual process of metabolic extension and catalytic innovation would progress, under the kind of selective pressure envisaged and in what could well be a mutually reinforcing fashion, up to the point from which other models start—where a nucleic acid–protein system could begin to evolve by Darwinian selection operating at the level of synthesis through favored replication.

Details of the Model

Small Molecules. The number and nature of the components of the initial "soup" are dictated mainly by what is needed by the synthesizer to produce enough catalytic diversity for evolution to start. One or more potential substrates are required as well, but these are likely to be found among the building blocks used by the synthesizer.

What these requirements amount to is difficult to estimate, but it is perfectly possible that the process could be primed with a small number of simple molecules of the kind that are readily produced in Miller-type experiments. Indeed, the synthesizer products would have to be fairly long multimeric molecules if they were to display the variety of catalytic activities demanded by the model. Therefore, diversity could be achieved by combinatorial variation rather than by multiplication of the types of building blocks. In the case of decamers for example, only eight interchangeable building blocks are needed to allow some 10^9 distinct combinations. This value may well set some sort of upper limit to the degree of diversity that can be tolerated if each multimer is to be present in the system at more than a vanishingly small concentration. In the case of 10^9 distinct varieties, a volume the size of a eukaryotic cell filled with a millimolar multimer solution (about 0.1% in the case of decapeptides) would, on average, contain no more than about five molecules of each kind. Greater variety could well be self-defeating.

Synthesizer. It is tempting to think of the synthetic products as oligo- or polypeptides. However, the fact that they are expected ultimately to evolve into proteins does not mean that they should necessarily start as pure peptides of the kind we know today. This is actually not very likely. For one thing, we would expect both D- and L-amino acids to contribute to their formation, as is still true for such bacterial products as cell-wall peptidoglycan and the antibiotics gramicidin S and tyrocidin. Furthermore, given a random and presumably unspecific assembly mechanism, other building blocks besides amino acids could well participate in the synthetic process to give products comparable to amino acid conjugates or to more complex compounds such as pantetheine. Most hybrid compounds of this type arise by the dehydrating assembly of a carboxylic acid with another molecule, a property also of peptides. This is a plausible description of the activity of our putative synthesizer, especially in view of the abundance of carboxylic acids among the products of experiments aimed at simulating prebiotic syntheses.

Assembly would have to proceed initially without catalysts or with only mineral catalysts, such as clay (3). As it evolved, however, the system could be expected to generate its own catalysts subject to substrate stabilization. Interestingly, this kind of selection would particularly favor iterative assembly (elongation) catalysts, whose products are also substrates and, therefore, can act as stabilizers.

Whatever the mechanism of the assembly process, it most likely required energy, especially if it took place in an aqueous medium, which seems probable. Heat, electric discharges, ultraviolet radiation, and, perhaps, thunder shock waves have for quite some time been choice candidates for the role of energy source in the view of workers who have been trying to mimic abiotic syntheses in the laboratory. However, only heat, in the somewhat unlikely conditions adopted by Fox (see ref. 10) for the synthesis of "proteinoids" from amino acid mixtures, has been shown to support dehydrating assembly reactions. The possibility of an early dependence on chemical activation, therefore, deserves consideration. Indeed, several authors, among them Lipmann (11), Folsome (12), and Morowitz (13), have proposed that life started with the development of a system allowing the conversion of radiant or electron-transfer energy into a chemically usable form.

Two choices, not necessarily mutually exclusive, are open in this respect. One relies on activation by pyrophosphates, either provided by the environment or manufactured by a coupled mechanism powered by light, electron transfer, or some ionic gradient. The other depends on thioesterification, which could be mediated by pyrophosphate or, perhaps, take place directly under favorable conditions. There is much to be said for the participation of thioesters in primitive activation mechanisms, a view already expressed by Lars Onsager [quoted by Lipmann (14)]. The thioester bond occupies a central place in biochemistry. It is at the root of substrate-level phosphorylation, presumably the oldest form of metabolic ATP regeneration. It is the precursor of virtually all acyl ester bonds and of many carbon-carbon bonds, as in fatty acids, Krebs-cycle intermediates (through citrate), porphyrins (through 5-aminolevulinate), and the vast terpenoid-steroid family (through mevalonate). Even more suggestive, it is also the precursor of some acyl-amino bonds found, for example, in amino acid conjugates and in those possible vestigial products, the stereochemically mixed peptides gramicidin S and tyrocidin. In the synthesis of all these compounds, the phosphate ester of pantetheine, which is itself an interesting hybrid that could also have an ancient history, serves as universal acyl carrier. It does so as such (fatty acids, peptides) or as part of coenzyme A, possibly a later invention developed when ATP came on the scene. The suggestion by Lipmann (11, 14) that pantetheine-dependent peptide synthesis may have preceded the RNA-dependent mechanism certainly deserves serious consideration. As intimated here, it could go back to prebiotic times.

If there is any truth to the above speculation, one would tend to look for a H_2S-rich medium as the site of abiotic assembly processes. Such an environment remains the

choice ecological niche of many microorganisms today, including a number of archaebacteria. It could well have been more widespread on a young planet subject to intense volcanic activity. Also pointing to a sulfide-rich cradle of early life are the molecular properties of the iron-sulfur protein ferredoxin, which Eck and Dayhoff (15) believe to be derived through iteration and mutation from a particularly remote ancestor, the tetrapeptide alanyl-aspartyl-seryl-glycine, not an unlikely product for our putative synthesizer.

Lytic Activity. Low pH, heat, radiation, and other physical factors could be agents of breakdown. However, the most obvious lytic system, and also the one most likely to be hampered by substrate binding, is catalytic in nature. The appearance of hydrolytic activities against peptide and other similar bonds is clearly to be expected in a system required by definition to produce a wide variety of catalysts.

Advantages of the Model

The proposed model presents a possible mechanism for what may be called the prereplicative phase of chemical evolution. Its main merit lies in the relative simplicity of the demands it makes on random abiotic chemistry. It shows how, starting from conditions that do not greatly exceed what has been achieved without too much contrivance in laboratory simulations of prebiotic syntheses, a primitive network of metabolic pathways catalyzed by protoenzymes could develop progressively up to the point where a Darwinian kind of evolution could begin to operate. The degree of sophistication needed for this point to be reached must have been appreciable and may well have included an ATP-based economy and many other complex molecules and mechanisms, according to some of the scenarios that have been invented to explain the origin of the genetic system (for example, see refs. 1, 3, 12, and 16–20). According to the model, all this intricacy was "discovered" in the course of a protracted coevolutionary process of metabolic and catalytic exploration, instead of arising by chance and somehow coming together amid a jumble of molecular "misfits," as is postulated by most other theories.

The model also offers a possible solution to the chirality problem, which many consider to be one of the most intractable problems connected with the origin of life. Starting, as mentioned, with heterogeneous peptides containing both D- and L-amino acids, the system could have evolved toward chiral homogeneity, because chirally homogeneous peptides bound their substrates more strongly, were more stable, or had some other property likely to favor their survival. The choice of one enantiomorphic form over another could have been accidental or dictated by substrate conformations.

The model has in common with other evolutionary theories that it relies on a random mechanism—the synthesizer—to produce the diversity on which selection operates. But it selects favorable "clones" by having them live longer, rather than reproduce faster. Survival is also a factor in Darwinian selection but only as a factor favoring reproduction. Another difference is that the model uses chance in a largely deterministic fashion. It can work only, at least at first, if the synthesizer produces all possible variants so that a steady supply of those selected is ensured. As previously pointed out, this also means that the system had to start with a restricted choice of small oligomers—a requirement that clearly fits an early prebiotic mechanism. Only later, after specific catalysts for catalyst elongation or modification appeared, could the system afford to explore a wider space of variability. An intriguing implication of this deterministic behavior is that the whole blueprint of the prebiotic metabolic network must have been contained in the initial conditions. It was all given at the start, though not as a minute part of a vast chemical hodgepodge, as must be assumed in conventional theories, but as something to be fashioned selectively by a progressive process of chemical evolution, which could well be largely predictable.

These are all attractive features, but it remains to be seen whether the model does not strain credibility in other ways. How plausible are its two basic assumptions that the products of a random assembly system will include a variety of catalysts and that these will be significantly protected against lysis by their substrates?

Plausibility of the Model

Generation of Catalysts. Simple calculations show that even the smallest enzyme could not possibly have arisen full-blown by chance, unless the same activity can be realized by a truly astronomical number of different amino acid combinations (21). This possibility cannot be rejected, but it seems more likely that enzymes had more modest beginnings and started, not as the finely honed finished products they are today, but as relatively short oligopeptides endowed with only rudimentary catalytic properties—just enough for evolution to start "tinkering." The model postulates just that and even offers a possible mechanism for pre-Darwinian selection. That random assembly could suffice to produce the first enzyme precursors is indicated by the finding by Fox (see ref. 10) that proteinoids obtained by heating amino acid mixtures display several weak catalytic activities.

Stabilization by Substrates. It is a well-known property of enzymes that they are often stabilized by their substrates. It is at least plausible that the rudimentary catalysts postulated by the model would be similarly protected, but whether this effect would be strong enough to affect the life span of the molecules significantly is more questionable. Taking the most favorable assumption that the catalyst–substrate complex is totally resistant to degradation, we find that the ratio of the steady-state amount N_j' of catalyst j in the presence of substrate to the amount N_j in the absence of substrate is given by:

$$\frac{N_j'}{N_j} = 1 + \frac{S_j}{K_j}, \qquad [2]$$

in which S_j is the substrate concentration and K_j the dissociation constant of the catalyst–substrate complex. For significant protection, we need substrate concentrations at least equal to, and preferably distinctly higher than, the dissociation constants of the complexes. Since K_j is not likely to be smaller than 10^{-3} M—it could well be much higher—we may take it that only substrates present at a concentration of the order of 10 mM or higher could conceivably participate in the postulated mechanism. This means that some concentrations would even have to reach or exceed molarity to allow for unfavorable equilibria.

This condition might seem so severe as to invalidate the whole model, except that the concentration problem plagues all theories of the origin of life. The thickness of the "primeval soup" is a recurrent topic of discussion and various phenomena—accumulation over aeons of time, evaporation, adsorption, confinement—have been evoked to take care of the concentration problem. In the present model at least, all that is needed is a high concentration of those few simple molecules required for priming the system. Other metabolites are expected to arise later by specific catalytic action and should build up to comparable levels of concentration within the limits set by reaction equilibria. This is a much less-stringent condition than to have all the necessary ingredients spontaneously reach concentrations that might allow the synthesis of something as complex as an oligonucleotide by a random assembly process. In any case, it is

hardly to be expected that a primitive prebiotic metabolism would operate immediately with the low levels of substrates and intermediates found in living cells.

It could also be pointed out that selection by survival is not an indispensable feature of the model. As long as all the catalysts needed are present, the development of a metabolic network can proceed as postulated. However, a parallel refinement of the catalysts cannot be explained without accumulation to promote it. At least not every catalyst need be protected for the model to function.

In conclusion, the proposed model does not seem to stress probability beyond the bounds of the possible. By most criteria, it does so distinctly less than do other prebiotic models. Its cornerstone is the assumption that precursors to a wide variety of enzymes will be found within the oligomeric products of the random assembly of simple prebiotic building blocks. With the growing availability and improvement of automated peptide synthesis, this assumption, as well as the assumption of stabilization by substrate, could possibly become subject to experimental testing.

Henri Beaufay and Miklos Müller kindly read the manuscript and made useful comments.

1. Eigen, M., Gardiner, W., Schuster, P. & Winkler-Ostwatitsch, R. (1981) *Sci. Am.* **244** (4), 88–118.
2. Shapiro, R. (1985) *Origins, A Skeptic's Guide to the Creation of Life on Earth* (Summit, New York).
3. Cairns-Smith, A. G. (1982) *Genetic Takeover and the Mineral Origins of Life* (Cambridge Univ. Press, Cambridge, U.K.).
4. Dillon, L. S. (1978) *The Genetic Mechanism and the Origin of Life* (Plenum, New York).
5. Prusiner, S. (1984) *Sci. Am.* **251** (4), 48–57.
6. Robertson, H. D., Branch, A. D. & Dahlberg, J. E. (1985) *Cell* **40**, 725–727.
7. Oesch, B., Westaway, D., Wälchli, M., McKinley, M. P., Kent, S. B. H., Aebersold, R., Barry, R. A., Tempst, P., Teplow, D. P., Hood, L. E., Prusiner, S. & Weissmann, C. (1985) *Cell* **40**, 735–746.
8. Chesebro, B., Race, R., Wehrly, K., Nishio, J., Bloom, M., Lechner, D., Bergström, S., Robbins, K., Mayer, L., Keith, J. M., Garon, C. & Haase, A. (1985) *Nature (London)* **315**, 331–333.
9. Dyson, F. (1985) *Origins of Life* (Cambridge Univ. Press, Cambridge, U.K.).
10. Fox, S. W. & Dose, K. (1972) *Molecular Evolution and the Origin of Life* (Freeman, San Francisco).
11. Lipmann, F. (1965) in *The Origins of Prebiological Systems and of their Molecular Matrices*, ed. Fox, S. (Academic, New York), pp. 259–280.
12. Folsome, C. E. (1979) *The Origin of Life, A Warm Little Pond* (Freeman, San Francisco).
13. Morowitz, H. J. (1987) *Cosmic Joy and Local Pain, Musings of a Mystic Scientist* (Scribner, New York).
14. Lipmann, F. (1984) *Annu. Rev. Biochem.* **53**, 1–33.
15. Eck, R. U. & Dayhoff, M. O. (1966) *Science* **152**, 363–366.
16. Crick, F. (1968) *J. Mol. Biol.* **38**, 367–379.
17. Orgel, L. (1973) *The Origins of Life, Molecules and Natural Selection* (Wiley, New York).
18. Eigen, M. & Schuster, P. (1977) *Naturwissenschaften* **64**, 541–565.
19. Eigen, M. & Schuster, P. (1978) *Naturwissenschaften* **65**, 341–369.
20. Ninio, J. (1979) *Approches Moléculaires de l'Evolution* (Masson, Paris).
21. de Duve, C. (1984) *A Guided Tour of the Living Cell* (Sci. Am. Books, New York), pp. 252–253.
22. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, U.K.).