MUTUALLY CATALYTIC AMPHIPHILES: SIMULATED CHEMICAL EVOLUTION AND IMPLICATIONS TO EXOBIOLOGY*

DANIEL SEGRÉ and DORON LANCET[1]
Department of Molecular Genetics, and [1]Head, Genome Center
The Weizmann Institute of Science, Rehovot 76100, Israel

Abstract. A description of the emergence of life should delineate a chemically rigorous gradual transition from random collections of simple organic molecules to spatially confined assemblies displaying rudimentary self-reproduction capacity. It has been suggested that large sets of mutually catalytic molecules, and not self-replicating information-carrying biopolymers, could have been the precursors of life. We present here a stochastic model in which the mutually catalytic molecules are spontaneously aggregating amphiphiles. When such amphiphiles exert on each other random catalytic effects, biased molecular compositions emerge, that are endowed with replication-like properties. This approach may have important consequences to the understanding of very early chemical evolution. It could also guide a search for extraterrestrial forms of very primitive life.

1. Introduction

Since the discovery of DNA and its important role in living systems, models for the origin of life have been mostly aimed at understanding properties such as evolution of self-replicating biopolymers [1, 2], mechanisms for coding [3, 4], and template-mediated oligomerization [5]. The successful results obtained in these fields still leave many unanswered questions about the very early steps of chemical evolution. It has been suggested [6] that, for understanding these stages of biogenesis, classical concepts, such as Oparin's coacervate droplets [7], should be reconsidered in the light of modern biological knowledge and with the aid of today's computer power. The mean-field model that Dyson himself suggested in 1982 [8], and Kauffman's results on sets of mutually catalytic polymers [9] are among the main theoretical models for analyzing prebiotic systems, in which the interactions among different kinds of molecules lead to the emergence of replication-like properties. One of the most important aspects of such models is that ensemble replication can happen even if none of its constituent molecules by themselves are endowed with the capacity to generate their own copies or to undergo autocatalysis. In a similar spirit, some authors [10, 11, 12, 13] have emphasized the potential importance of amphiphiles at the first stages of chemical evolution. It was proposed that spontaneous aggregation of such compounds to form micelles or vesicles in a prebiotic environment could have ensured high local concentrations, and the subsequent enhancement of catalyzed chemical reactions among the constituent or enclosed chemicals. Our previously described Graded Autocatalysis Replication Domain (GARD) model [14, 15, 16, 17] which analyzes the explicit chemical kinetics of mutually catalytic sets, is extended here to exploring the emergence of self-reproducing molecular assemblies

composed of amphiphiles. A further elaboration of the previous model is the use of a stochastic simulation algorithm instead of solving sets of differential kinetic equations. The scenario we propose, based on former teachings [6, 7, 9, 10, 11], can be summarized in the following phases:

(i) Spontaneous formation of assemblies, by aggregation of molecular species (eg: amphiphiles) due to physicochemical interactions (eg: hydrophobic effects). The high local concentrations thus attained allow more effective catalysis.

(ii) Emergence of poorly organized metabolic processes, in the form of quantifiable mutually catalytic interactions among members of the assemblies. Assemblies, which may appear by chance, whose molecular compositions are more successful in catalytically recruiting additional monomeric species from the environment, have a selective advantage.

(iii) Growing and splitting of the assemblies initiate an evolution-like mechanism which leads to increased abundance of more efficient catalytic assemblies. The assemblies, when splitting, propagate their unique composition to daughter assemblies [10, 18], creating a primordial system for storing and transmitting information. A quantitative description of this scenario may help understand important aspects of very early self-organization, potentially responsible for the emergence of life on Earth and elsewhere in the universe.

## 2. Chemical dynamics of molecular assemblies: a Monte Carlo approach

We analyze possible dynamic patterns in what we would like to call the "primordial random chemistry era". For this, we developed a method for depicting the chemical kinetics of self-assembling molecular sets by realistic computer simulations. These are based on standard mass action chemical laws, and on stochastic dynamical rules [19]. This approach, that we named SMACC (Single Molecule Algorithms for Complex Chemistry) [16] is aimed at reproducing the global kinetics of chemical reactions, without loosing the capacity to resolve individual molecules. The computer simulations whose results are presented here, analyze the behavior of sets of amphiphiles subjected to spontaneous aggregation and to mutually catalytic interactions. Amphiphiles are assumed to include diverse chemical configurations sharing the general property of having a hydrophobic region ("tail") and a hydrophilic, charged or polar, region ("head"). As in other origin of life scenarios [20] it is assumed that primordial chemistry would provide an ample supply of such molecules [11, 21]. The crucial qualitative aspects of the model are: a) Amphiphilic monomers may form mixed assemblies (e.g. mixed micelles) in an energetically favorable reaction; b) The assemblies are dynamic [22], having "soft matter" properties, and may gain or lose individual monomers by insertion/removal reactions; c) The monomer addition/removal reactions may be catalyzed by configurations of neighboring monomers; d) When assemblies reach a critical size they undergo splitting to two smaller assemblies. While a variety of critical sizes and splitting pathways are conceivable, we consider here a simple case whereby all assemblies behave identically and both split progeny are identical in size. These constraints may be easily relieved in future analyses.

We consider the existence of a N_G global number of amphiphile types (A_i) whose concentrations in the surrounding aqueous medium are buffered by a constant external supply. For simplicity, all A_i concentrations are assumed to be equal.
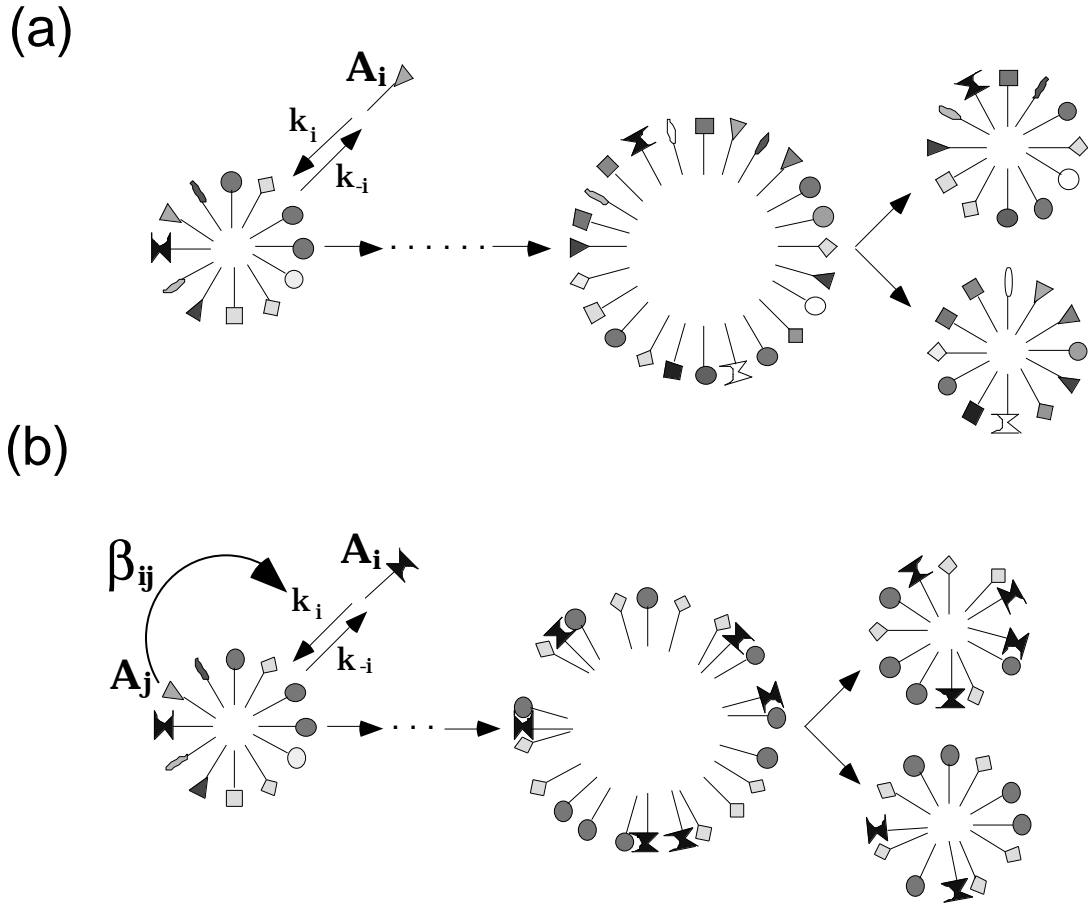
Figure 1. A schematic depiction of the reactions assumed to be responsible for growing and splitting of heterogeneous micelles. Amphiphiles are represented as having hydrophobic tails (sticks) and a variety of hydrophilic heads (geometrical shapes). In future simulations, tails could also be assumed to come in a large variety of types. Mutually catalytic interactions are symbolized by a geometrical complementarity between some neighboring species. Two scenarios are shown: **(a)** In absence of catalysis all the molecular species are equally likely to join/leave a micelle. Consequently, upon growth, if the number of possible species is much larger than the amount of molecules composing an assembly, no bias in the composition is expected to develop. When a micelle reaches a critical size (twice the initial one) its constituents molecules are randomly distributed into two new daughter assemblies. None of these new assemblies will probably resemble in the composition previously existed micelles. **(b)** When catalysis is exerted on joining/leaving of amphiphiles (arrow β_{ij}), species enhancing each other's reactions will tend to prevail after some time. When the composition becomes biased, and few species are present in large count, splitting generates assemblies whose composition resembles the one of the parent micelle.

The reaction by which a monomer A_i joins an assembly M_N composed of N monomers (n_1 of kind A_1, n_2 of kind A_2, etc.) can be represented formally as follows [23]:

$$M_N(n_1,n_2,...,n_i,...,n_{N_G}) + A_i \xrightarrow{\ k_i\ } M_{N+1}(n_1,n_2,...,n_i+1,...,n_{N_G}) \qquad (1)$$
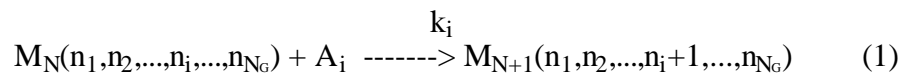
Species A_i has a probability of insertion proportional to a basal forward rate k_i. Similarly, the reverse reactions are described by

$$M_{N+1}(n_1,n_2,...,n_i+1,...,n_{N_G}) \xrightarrow{k_{-i}} M_N(n_1,n_2,...,n_i,...,n_{N_G}) + A_i \qquad (2)$$
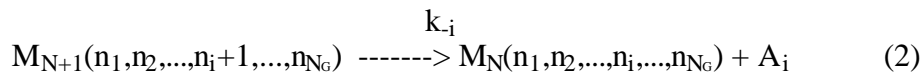
governed by a probability proportional to the backward basal rate k_{-i}.

It is postulated here as an additional simplification that, in absence of catalysis, the average time for a monomer to join the assembly does not depend on the specific monomer being inserted. In other words, that all k_i (and similarly all k_{-i}) are independent of i. In the presence of catalysis, it is assumed that the rates of the assembly growth/decline reactions are regulated by the kinetic terms $k_i(1+<n_j\beta_{ij}>_j)$ and $k_{-i}(1+<n_j\beta_{ij}>_j)$ for the forward and reverse reactions, respectively. The assembly average $<...>_j$ indicates the mean value of the catalysis exerted by the all molecules (A_j) present in the assembly at the time of insertion/deletion of the considered A_i. This averaging step is equivalent to an assumption of fast pre-equilibrium for the dynamic collisions between the incoming/outgoing A_i and all other members of the assembly. Similar mean-field assumptions have been used in other models [6].

The catalytic enhancement factors β_{ij} are the elements of an $N_G \times N_G$ matrix (the β matrix) with all the mutual catalysis values for the chemistry considered. The β matrix components are drawn from a probability distribution [15, 24, 25]. Based on these probabilities, a Monte Carlo simulation process determines the fates of the single molecules. The splitting event

$$M_{2N} \longrightarrow M'_N + M''_N \qquad (3)$$
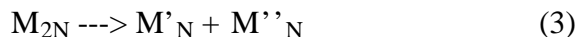
occurs whenever a micelle reaches twice the initial size.

Interesting effects can be seen only by simulating a large number of reactions. This makes it necessary to simplify the chemical rules in the described manner. More accurate mathematical models for the kinetics of amphiphiles assemblies [23] might be used in more advanced versions.

3. Random assemblies and the "vital bias".

The number of assemblies that can form by spontaneous aggregation of N molecules of N_G different kinds can provide a wide diversity of possible compositions. In particular, for $N_G \gg N$, a vast majority of the species will be present in the assembly in a single copy, and most species will not be present in the assembly at all (figure 2A). Spontaneously formed assemblies will practically never have strong compositional biases, i.e. many copies of relatively few species types. In contrast, a crucial characteristics of present day life is that it contains a relatively small number of molecular types, as compared to the astronomical number of organic compounds that are generally conceivable. A common tendency in origins of life research is to try to explain how these exact compounds present in today's cellular life could have formed or could have been selected from the many possible substances. This process is often seen as a prerequisite needed for life to emerge. A different mechanism might be envisaged, in which the selection of the "right" building blocks coincides with the process by which life emerges. The difference is subtle, but it has a very important implication: the latter view - unlike the former - just calls for a general

decrease in the number of chemical species, without requiring that very specific compounds (e.g. nucleotides or amino acids) will be enriched.

The involvement of non covalent assemblies in our model has significant consequences.While in the case of covalently linked biopolymers, changes in sequences seem to require laborious mechanisms involving sophisticated enzymatic catalysis, non covalent assemblies can easily change their composition by monomers joining/leaving. On the other hand it is often considered difficult to see how non-covalent assemblies may acquire the capacity of storing and propagating information, a property usually ascribed to biopolymers. However, as suggested by Morowitz [18], an assembly that grows and splits into two random halves, can propagate the specific composition to the daughter assemblies, provided that enough copies of each species are present before splitting. Precise relations between the size of the assembly and the number of species present in it can be deduced following this criterion [18]. In the results illustrated below, we show how random catalytic interactions within randomly seeded molecular assemblies could lead to the emergence of micelles capable of propagate their highly improbable biased compositions.
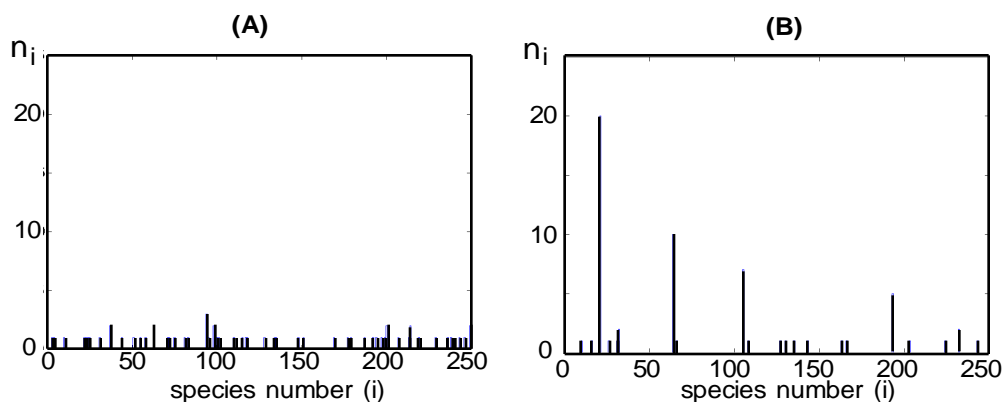
Figure 2. Histograms of the composition of two assemblies. For each species A_i, the number of individual molecules present in the assembly (n_i) is plotted. In general it is $\sum n_i = N$, with N between a minimal size (30) and the critical splitting size (60). The histograms are relative to micelles at their critical splitting size. **(A)** An assembly that underwent very little catalyzed joining/leaving reactions has no bias in its composition. This situation corresponds to an assembly of generation 1 (figure 3). **(B)** An assembly that underwent many steps of the catalyzed growing and splitting process (generation 15 in figure 3). The program for the Monte Carlo simulation of the reactions in an Amphiphilic GARD was implemented in Matlab on a DEC Alpha workstation.

4. Results of a computer experiment

Figure 2 shows some results from a computer experiment performed with an amphiphilic GARD as described above. After seeding randomly two micelles, we let them grow by mutually catalytic monomer joining reactions. The first that reaches the critical splitting size divides, and the two micelles born thereof start growing again in a similar way. The process is repeated cyclically, and biases in the compositions of the micelles are

monitored. For quantifying the degree of bias we use a measure of compositional information $I=-\log(P)$ where P is derived from the multinomial probability distribution

$$P=(1/N_G)^N \cdot N!/(n_1!n_2!...n_{N_G}!). \qquad (4)$$

The spontaneous increase of I (i.e. decreased probability of the assembly composition) as a function of generations is seen in figure 3. The maximal increase is roughly 20 standard deviations away from the mean of the distribution of randomly generated amphiphile assemblies (figure 3, inset).

These changes constitute a demonstration of the capacity of amphiphile assemblies to undergo an evolution-like change towards less probable structures. In the present model, progeny assemblies will practically never have the exact same composition of the original one.
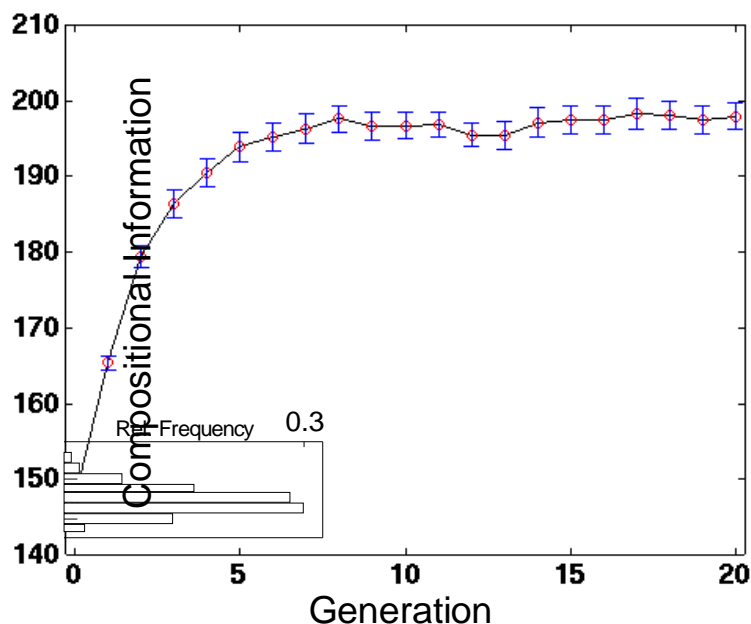
Figure 3. The curves shown are averages of 100 independent simulations, with bars indicating the standard deviation of the mean. Assemblies of size N=30 are randomly seeded at t=0 from an amphiphile chemistry of size $N_G=250$. In each generation two parallel assemblies undergo monomer addition/loss reactions governed by the kinetics described in Fig. 1. When one of the two assemblies reaches the critical splitting size N=60, its randomly sorted halves start the cycle again. This process is repeated for 20 generations. This amounts to following a trajectory on a tree by considering the faster-growing branch at each generation. For each generation, the value of the negative logarithm of the multinomial probability (I, see equation 4) is plotted. In the inset the distribution of I for spontaneously formed micelles is shown.

The fidelity of replication may actually be computed as a distance between "parent" and progeny using the metric

$$D=\sum |n_i'-n_i|/(n_i'+n_i), \qquad (5)$$

where n_i', n_i are the counts for species A_i in the two generations considered. The similarity

(defined as 1/D) becomes stronger for the higher generations.

The very high correlation coefficient (0.99) between the curves for the growth of I and of 1/D throughout the generations, indicates that when a more idiosyncratic composition appears via mutual catalysis, the assembly constitutes a more effective replicator. This may be intuitively understood by observing that as mutual catalysis results in an increased copy number of certain molecular species, the probability of equal copy number of such species in the two progeny assemblies increases. This kind of analysis seems therefore very appropriate for investigating in further detail the idea of compositional information [26].

5. Consequences of the amphiphile GARD for a Galactic Traveler in search for extraterrestrial life

One of the important future outcomes of origins of life studies should be the generation of guidelines for the search of life on other planets. Any experiment aimed at discovering signs of life would be biased by what we expect to find. This is particularly true in those cases where life has been "frozen" at very early stages of its emergence. The model described in the previous paragraphs allows one to delineate the following predictions on the nature of early life relics representing consecutive stages of prebiotic processes:

1) The occurrence of simple organic molecules of any kind, with no absolute requirement for specific molecular types, such as amino acids or nucleotides.

2) The occurrence of more complex organic molecules, up to several dozen carbon atoms, not necessarily oligomers of molecules known to constitute present day life.

3) The occurrence of amphiphilic molecules, i.e. those that have a polar part and a lipophilic part, without other constraints on chemical structure.

4) A high degree of diversity of the molecules described in (2) and (3): this is a necessary stage en route to the "narrowing" of molecular diversity that occurs later on.

5) The appearance of assemblies of molecules in the form of micelle-like structures.

6) Biased internal composition of such assemblies, whereby relatively few types are present. This criterion may apply to assemblies, but can also be discerned in disperse molecules if the assemblies did not survive.

7) A high degree of similarity in the internal composition of assemblies. If the assemblies did not survive, this may still manifest itself as macroscopic regions containing biased organic chemical composition.

8) Rudimentary mutually catalytic or photochemical properties of the chemicals present in assemblies or in spatial domains.

9) More advanced ultrastructure of assemblies, e.g. vesicles or higher structures.

Points 4 to 7 in this list have been explicitly modelled through the computer experiments illustrated in this work. Biased compositions were shown to be possible important hints for the presence of assemblies in which mutually catalytic interactions had become more and more efficient. Based on these results, one can speculate how an imaginary Galactic Traveler would plan a search for extraterrestrial "prebiotic fossils", i.e. relics of very early chemical evolution. For the Galactic Traveler, an unexpected chemical composition might sometimes justify a serious suspicion of past biogenic phenomena. Counter examples would be crystals or minerals, arrays of very few chemical species, which present an extremely biased chemical composition.
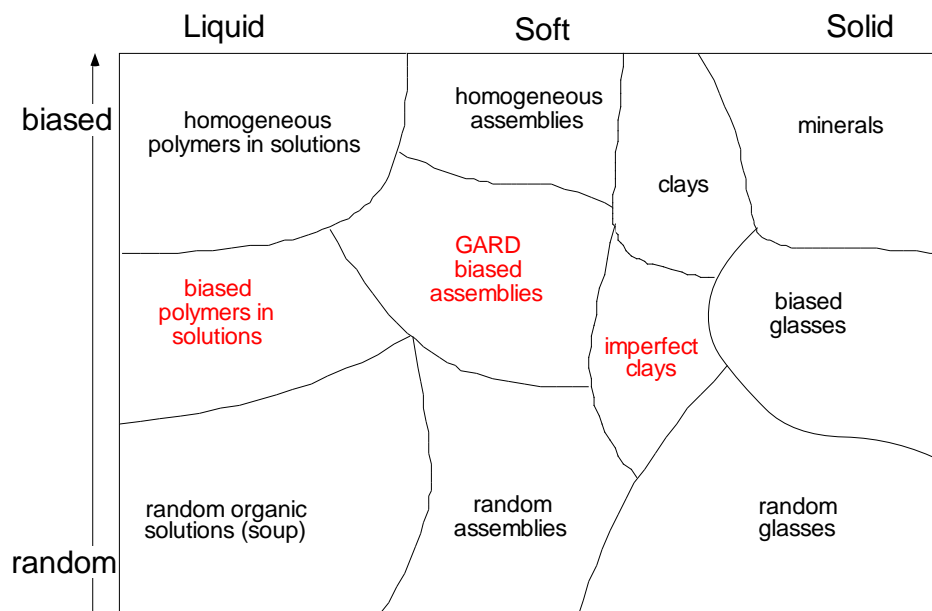
Figure 4. A diagram for a Galactic Traveler examining a sample of extraterrestrial material. The two attributes monitored here are the degree of compositional bias (vertical axis) and the compactness of the sample (horizontal axis). The idea that amphiphilic GARD-like assemblies could be good candidates for a primordial self-replicating system would suggest to survey with particular interest soft matter samples with some deviation from randomness in their composition. Such samples would be represented in the central region of the diagram.

These would usually not be considered as potential steps towards life. On the other hand minerals with imperfect crystal structures display a greater degree of chemical diversity and their life-like properties have been explored [27]. We suggest that an intermediate degree of chemical bias in soft matter (or what is left of it) would be indicative of very primitive prebiotic entities. Soft matter has the great advantage of affording the chemical dynamics needed for mutual catalysis and rudimentary replication, but at the same time allows effective chemical proximity which is not present in aqueous solutions. Thus the central region of the diagram (figure 4) is most likely to be interesting to a Galactic Traveler in search for the seeds of life.

6. Conclusions

A computer analysis of a model for the emergence of self-replicating units has been described, based on the idea that selective growth of micelle-like assemblies might result from random catalytic interactions among prebiotically formed amphiphiles of many different kinds. This idea can serve as a simple chemically realistic benchmark for performing *in silico* experiments that will help us understanding some aspects of early chemical evolution. As an alternative to scenarios focused mostly on present day life biopolymers, the model we suggest also indicates new possible viewpoints for exobiological explorations.

References

1. Eigen, M. and P. Schuster. "The Hypercycle." 1979 Springer Verlag. Berlin.

2. Orgel, L. E. "Molecular replication." Nature. **358**: 203-209, 1992.

3. Hartman, H. "Speculations on the origin of the genetic code." J. Mol. Evol. **40**: 541-544, 1995.

4. Nieselt-Struwe, K. and P. R. Wills. "The Emergence of Genetic Coding in Physical Systems." J. theor. Biol. **187**: 1-14, 1997.

5. Wright, M. C. and G. F. Joyce. "Continuous in Vitro Evolution of Catalytic Function." Science. **276**(25 april): 614-617, 1997.

6. Dyson, F. "Origins of Life." 1985 Cambridge University Press. Cambridge.

7. Oparin, A. I. "The origin of life." 1953 Dover Pub. N.Y.

8. Dyson, F. J. "A Model for the Origin of Life." Journal of Molecular Evolution. **18**: 344-350, 1982.

9. Kauffman, S. A. "The origins of order- Self-organization and selection in evolution." 1993 Oxford University Press. Oxford.

10. Morowitz, H. J., B. Heinz and D. W. Deamer. "The chemical logic of a minimum protocell." Origins of Life and Evolution of the Biosphere. **18**: 281-287, 1988.

11. Deamer, D. W. "Role of amphiphilic compounds in the evolution of membrane structure on the early earth." Origins of Life and Evolution of the Biosphere. **17**: 3-25, 1986.

12. Bachmann, P., P. Luisi and J. Lang. "Autocatalytic self-replicating micelles as models for prebiotic structures." Nature. **357**(7 May): 57-59, 1992.

13. Ourisson, G. and Y. Nakatani. "The terpenoid theory of the origin of cellular life: the evolution of terpenoids to choloesterol." Chemistry & Biology. **1**(September): 11-23, 1994.

14. Lancet, D., G. Glusman, D. Segré, O. Kedem and Y. Pilpel. "Self replication and chemical selection in primordial mutually catalytic sets." Origins of Life and Evolution of the Biosphere. **26**: 270-271, 1996.

15. Segré, D., Y. Pilpel, G. Glusman and D. Lancet. "Self-replication and Evolution in Primordial Mutually Catalytic Sets." Astronomical and Biochemical Origins and the Search for Life in the Universe, Proceedings of the 5th International Conference on Bioastronomy, IAU Colloquium N.161. Cosmovici, S.Bowyer and Werthimer ed. 1997 Editrice Compositori. Bologna.

16. Segré, D., Y. Pilpel and D. Lancet. "Mutual catalysis in sets of prebiotic organic molecules: evolution through computer simulated chemical kinetics." Physica A. **249**(1-4): 558-564, 1998.

17. Segré, D., D. Lancet, O. Kedem and Y. Pilpel. "Graded Autocatalysis Replication Domain (GARD): Kinetic Analysis of Self-Replication in Mutually Catalytic Sets." Origins of Life and Evolution of the Biosphere. **28** (in press): 1998.

18. Morowitz, H. J. "Biological Self-Replicating Systems." Progress in Theoretical Biology. Snell ed. 1967 Academic Press.

19. Bartholomay, A. F. "A stochastic approach to statistical kinetics with application to enzyme kinetics." Biochemistry. **1**(2): 223-230, 1962.

20. Miller, S. L. "A production of amino acids under possible primitive earth conditions." Science. **117**: 528-529, 1953.

21. Deamer, D. W. and G. R. Fleischaker. "Origins of life-The central concepts." 1994 Jones & Bartlett. Boston.

22. Lipowsky, R. "The conformation of membranes." **349**(7 February, 1991): 475-481, 1991.

23. Chizmadzhew, Y. A., M. Maestro and F. Mavelli. "A simplified kinetic model for an autopoietic synthesis of micelles." Chem. Phys. Lett. **226**: 56-62, 1994.

24. Lancet, D., O. Kedem and Y. Pilpel. "Emergence of order in small autocatalytic sets maintained far from equilibrium:application of receptor affinity distribution (RAD) model." Ber Bunsenges. Phys. Chem. **98**(9): 1166-1169, 1994.

25. Lancet, D., E. Sadovsky and E. Seidemann. "Probability model for molecular recognition in biological receptor repertoires: Significance to the olfactory system." Proc. Natl. Acad. USA. **90**: 3715-3719, 1993.

26. Morowitz, H. J. "Beginnings of Cellular Life." 1992 Yale University Press. London.

27. Cairns-Smith, A. G. "Genetic Takeover and the Mineral Origins of Life." 1982 The Cambridge University Press. Cambridge, U.K.