A Statistical Chemistry Approach to the Origin of Life

Daniel Segré *and* ***Doron Lancet****, Department of Molecular Genetics and the Genome Center, The Weizmann Institute of Science, Rehovot 76100, Israel*

We revisit some theoretical models dealing with the chemical emergence of life-like properties in prebiotic systems. Special emphasis is given to models involving random assemblies of mutually catalytic organic molecules, as opposed to scenarios in which individual molecular species are endowed with the capacity of self-replication. We highlight here the challenge of tracing the very first steps of biogenesis, when self-replication, mutation, selection and evolution may have been hardly recognizable. The models we discuss share the assumption that a large repertoire of relatively simple organic compounds could spontaneously form prebiotically, and the notion that a statistical approach, independent of detailed molecular properties, can uncover some general principles underlying biogenic processes.

Fundamental models, put forward by Dyson and Kauffman, describe very early scenarios, whose statistical nature is reflected in the possibility of characterizing many random, mutually catalytic interactions with relatively few parameters. Further theoretical considerations indicate that mutually catalytic assemblies might also entail a primitive information transfer system, exclusively based on idiosyncratic chemical compositions, a situation described here as the inheritance of a "compositional genome". Amphiphilic molecules, due to their peculiar attributes, are suggested to potentially embody many of the properties necessary for these systems to emerge spontaneously, hinting to the possibility of an exclusively lipid-based origin of life.

We stress that modern trends in molecular complementarity, combinatorial chemistry and enzyme mimetics represent a source of conceptual and experimental information that can help extend previous models. This is exemplified here by the Graded Autocatalysis Replication Domain (GARD) model we developed, based on a statistical distribution of catalytic activities. A further extension of this model, the Amphiphile-GARD, aims at a more realistic and testable theoretical description of some scenarios for early prebiotic evolution.

Address correspondence and reprint requests to: Doron Lancet, Department of Molecular Genetics and the Genome Center, The Weizmann Institute of Science, Rehovot 76100, Israel.
E-mail: Doron.Lancet@weizmann.ac.il. Received for publication March 17, 1998; revised November 24, 1998.

Many investigators have studied the formation of organic molecules under prebiotic conditions. While the fine details and the actual molecular mechanisms remain somewhat elusive, the generation of a large diversity of organic molecules under simulated prebiotic conditions constitutes a widely accepted notion[1-4]. In what follows, we cover scenarios that assume that such molecules were present in a large number of types and high enough concentration, at least in certain locales.

Most origin of life scenarios invoke a "continuity principle" in one form or another[5-7]. Some researchers propose that present-day RNA might have been the seed of biogenesis[8-10], or that proteins with a similar chemistry to that found in living cells today were a prerequisite for early catalysis[11, 12]. Less stringent continuity models suggest that biopolymers are essential, but allow other monomer chemistries[13, 14]. Obviously, a weaker continuity principle helps avoid major mechanistic omissions, but increases the spectrum of pathways that need to be analyzed. The statistical chemistry view of the origin of life, as presented below, allows one to safely use the least stringent continuity principle, namely that *any* organic molecules could initiate life[15]. The severe loss of specific information thus sustained, is overcome through the use of statistical tools that quantitate molecular interactions despite the lack of a detailed structural definition[16-21]. At the extreme, one could question the necessity of organic compounds, invoking a "mineral origin". However, as delineated below, an "organic origin" has strong advantages in the realms of diversity generation and soft matter dynamics.

In many treatises of the origin of life[22-29] individual molecules are described that can generate their own copies. In this case, the replicated information is typically carried in a sequence of monomers strung covalently to form a polymer. The view that such molecules seeded life is often referred to as "genome first". An alternative scenario[5, 15, 18, 20, 30, 31] states that early information could be carried in the composition of molecular sets, propagated by mechanisms that involved the catalytic replication† of the entire molecular assembly. This "metabolism first" view is considered by some to be more appropriate for the very early steps of chemical evolution[32], when the high fidelity catalysts needed for templating and polymerization were not yet present. We show that the tools of statistical chemistry are useful for analyzing the properties of such composition-based chemical information systems‡.

Homeostasis and Self-replication in Mutually Catalytic Assemblies

In the first half of this century, Oparin established the "metabolism first" scenario, stressing the primary role of cell-like coacervate droplets[33]. He did not

† Some texts use the word "reproduction" when dealing with complex assemblies, and reserve "replication" to refer to the copying process of DNA-like molecules. Here we do not make this distinction, and use replication throughout.

‡ The present review is not meant to be comprehensive, and complementary overviews can be found elsewhere[77-79]. The theoretical models covered here represent examples within a promising but still largely unexplored domain.

refer to DNA/RNA-like molecules, because he and his contemporaries had little idea of the fundamental role they played. The concept of homeostasis, and the idea of self-replication as a collective property of molecular assemblies were very clearly stated[33]:

> ...the stationary drop of a coacervate, or any other open system, may be preserved as a whole for a certain time while changing continually in regard to both its composition and the network of processes taking place within it, always assuming that these changes do not disturb its dynamic stability. ... This constant repetition of connected reactions, co-ordinated in a single network, also led to the emergence of a property characteristic of living things, that of self-reproduction. This may be taken as the origin of life.

Based on the considerably deeper awareness of molecular mechanisms in modern biochemistry, Dyson[34] and Kauffman[35] subsequently set out to devise quantitative models for homeostasis and potential self-replication in molecular assemblies without genomes. Following Oparin's teachings, they conjectured that:
1) Collections of molecules may exist in a state that entails a high degree of mutual catalysis among their components. 2) If fed with energy and nutrient molecules from the environment, such assemblies might expand, split and give rise to catalysis-rich progeny. 3) A rudimentary process of selection might be initiated, based on the diversity of the ensuing catalytic compositions.

Dyson expresses the importance of the time-dependent process that leads to homeostasis[31], while Kauffman champions the role of increasing diversity in leading to self-replication[15]. We emphasize here that these concepts are highly related. As analyzed below, both Dyson's and Kauffman's models describe possible transitions between a poorly organized catalytic network and a well connected system of catalytic interactions, capable of increasing the global rate of synthesis of the constituent molecules from externally available "food" material.

We propose to use the general term RHEA* models (Replicative-Homeostatic Early Assemblies) to denote these and other paradigms for very early evolution, describing how molecular assemblies became endowed with rudimentary life-like properties, en-route to the first living cells (Figure 1).

Dyson's model

Dyson[31, 34] assumes that molecules enclosed in a bounded microenvironment are composed of monomers that exist in two states: catalytically active and catalytically inactive. An autocatalytic phenomenon is assumed, whereby active molecules will convert others into a similarly privileged state (Figure 1). In other, simpler models, such behavior may result in an exponential explosion of the amount of the active molecules, even when starting with one active copy[29, 36]. Yet, because Dyson's model includes a backward reaction, in which an active species may be inactivated, exponential growth does not occur, and steady states are reached under certain conditions.

Dyson shows that given the probability 1/a of specific catalytic activation (a is

* The Greek goddess Rhea, born to Gaea (Earth) and Uranus (Heaven), mother of Zeus and Hera, and ancestor to all Olympian gods, has analogously played an intermediate role between the physical, inanimate creation and a world teeming with life.

roughly the number of monomer types) and if one assumes a single mean-field catalytic enhancement factor b, then the efficacy of an existing molecular population in promoting the formation of new catalysts within it may be computed as a function of x, the fraction of active catalysts (Figure 1b). For some combinations of the parameters a and b, there are three steady states. An assembly can move through single substitution mutations from the disordered steady state (α) to the ordered one (γ), by passing through the unstable steady state (β). This process leads to the emergence of an organized metabolism. The average time for this transition to occur can be expressed as a function of a, b, and of the total number of monomers.

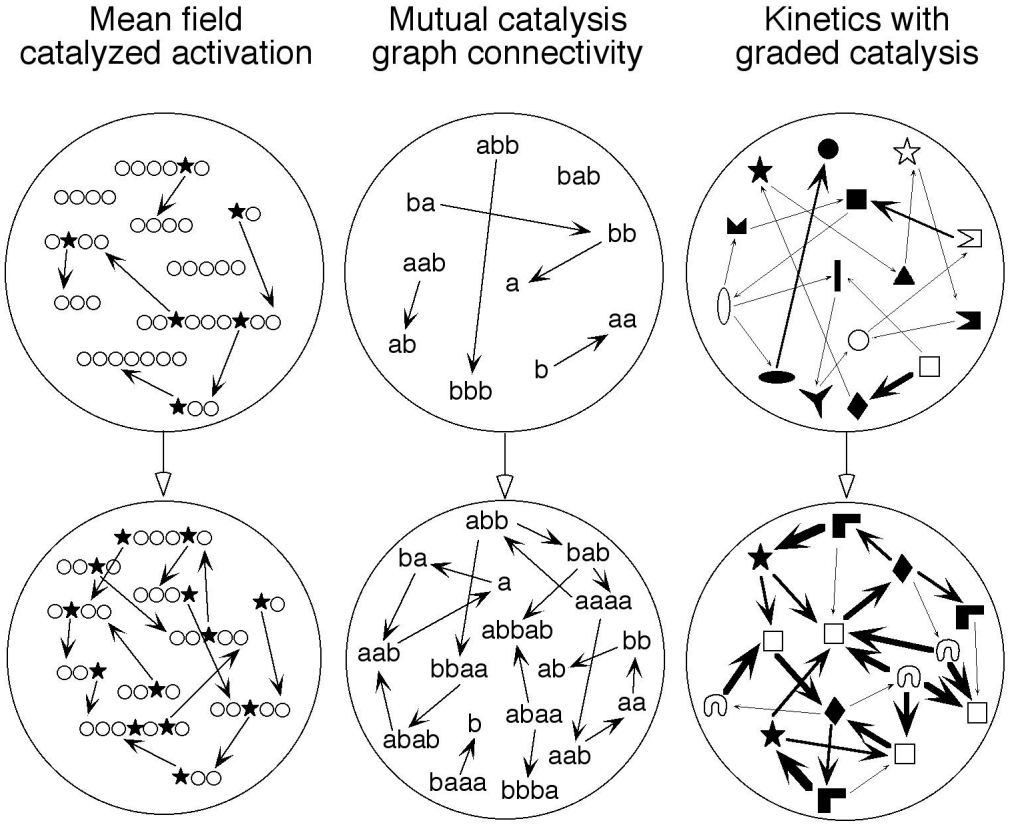

Figure 1. A comparative scheme for three classes of Replicative-Homeostatic Early Assemblies (RHEA) models. Left column: Dyson's mean field statistical model[31, 34]; Center column: Kauffman's graph theory model for mutual catalysis; Right column: Explicit chemical kinetics models for graded mutual catalysis, exemplified by the autocatalytic metabolism model of Bagley et al.[38, 39] and our Graded Autocatalysis Replication Domain (GARD) model[18, 19, 20]. **A,** The chemistry of RHEA models, showing a transition of enclosed molecular assemblies from a simpler, more poorly connected state (top row) to a state with a denser network of mutual catalysis events, endowed with homeostatic/replicative properties. In Dyson's model (left column), monomers may be catalytically active (star) or inactive (circle). Arrows represent events of catalytic activation by the "starred" species. In the two other columns, arrows depict catalysis exerted by one species on the synthesis of another, and thickness represents the catalytic power. In Kauffman's model (center column), the chemical species are exemplified by oligomers of two monomer types (a,b), while the shapes in the graded catalysis models (right column) represent arbitrary molecular species.

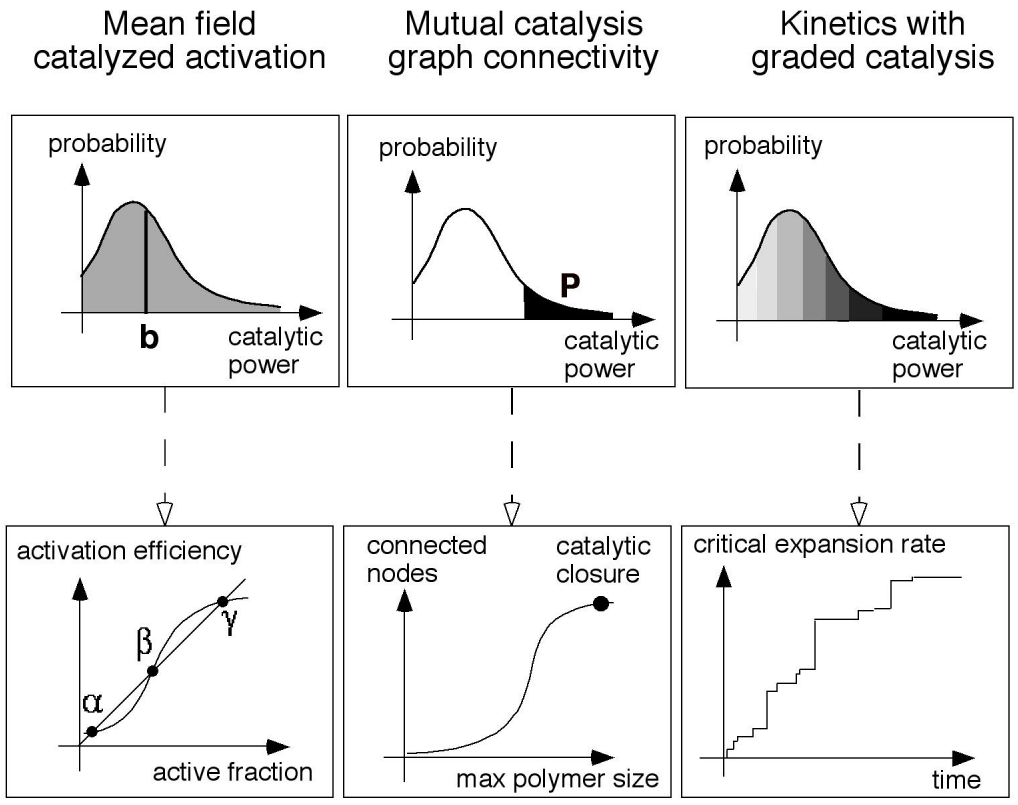

Figure 1. *(continued)***B.** The formalisms of RHEA models, showing the statistical description of the catalytic powers (top row) and the predicted transition from randomness to the more organized states (bottom row). The curves in the top row represent an ideal probability distribution of catalytic enhancement factors, as would be predicted to occur among randomly-picked organic compounds (see Figure 3). In Dyson's mean field approach (left column), a constant-probability single-valued catalytic enhancement b represents the average of all catalytic factors in the distribution. Kauffman's model (middle column) utilizes an above-threshold catalytic potency with a constant-probability P, corresponding to the area of the dark region under the curve. In the graded catalysis models (right column), the entire distribution of catalytic powers is employed**.** The transition curves in the bottom row relate to different formal approaches for the three model classes. In Dyson's model, the activation efficiency $\phi(x)$ of the existing population of catalysts in promoting the formation of a new catalyst is plotted against the fraction x of active monomers, yielding an S-shaped curve. The three points of intersection with the diagonal represent the steady states: α, the system's stable disordered state, which may undergo a transition to β, the unstable middle point en-route to γ, the stable organized state. In Kauffman's model a phase transition to a catalytically-closed state occurs as chemical diversity increases, as a result of augmented polymer size. For the graded catalysis models, a curve is shown, as predicted by a GARD simulation[18, 20], in which assemblies with better interconnected catalytic networks (manifesting a higher critical expansion rate) emerge as a consequence of compositional mutations. Different degrees of homeostatic preservation can arise in a punctuated equilibrium fashion[20, 80], corresponding to different speeds of replication for the molecular assemblies.

Dyson's model constitutes a clear example of a statistical approach where molecules have no specific identity, or structure, but only a simple functional property (catalytic power). The mean field assumption underlying the model is equivalent to postulating that the behavior of a mixture of catalytic chemicals can be approximated by averaging the catalytic factors of all the individual molecular species within the set.

Kauffman's paradigm

A different strategy is at the core of Kauffman's model for autocatalytic sets[15, 35]. He demonstrates that a set of random molecules (e.g. biopolymers) can be thought as self-replicating in its entirety, provided that it entails sufficient diversity of molecular species (Figure 1A).

In this embodiment of mutual catalytic sets, monomers of different kinds can form polymers with a length M or shorter. It is shown that as one considers increasing values of M, the number of possible polymer types goes up as an exponent of M, but the number of synthesis and cleavage reactions grows even faster. The ratio of reaction count to polymer types thus increases linearly with M. The statistical nature of Kauffman's analysis is reflected in a constant probability parameter P, which describes the chance that a molecular species in the set will catalyze any reaction leading to the formation of another species (Figure 1B). For M large enough, each species in the set will have the last step for its synthesis catalyzed by at least one other molecular species. The set thus emerging, described as "catalytically closed", may resemble in its behavior an individual autocatalyst.

Based on a graph theory derivation, Kauffman argues that any sufficiently complex set of molecules may attain the emergent collective property of catalytic closure. It is then possible to compute the likelihood of catalytic closure with increasing M, as a function of the probability of catalysis P. As explored also by other authors[37, 38, 39] the mutually catalytic set represents a good candidate for a primordial unit on which natural selection could act.

A Compositional Genome

An important question is whether a rudimentary genetic "memory" may emerge, based on the simplified statistical rules of mutually catalytic networks. Many are tempted to regard RNA and DNA as the major devices for biological information storage and transfer[22, 40]. Nevertheless, as previously pointed out[15, 30, 31], even present-day living cells bequeath a considerable amount of biochemical information in their metabolic composition. Such a simpler system of propagating information could predate nucleic acids. Perhaps, paraphrasing Occam's razor, "*Nuclei Acida non sunt multiplicanda praeter necessitate*" (Nucleic acids should not be multiplied unnecessarily).

The basic idea of a "compositional genome" is that in a mixture of relatively simple chemicals, the array of relevant concentrations may be viewed as a vehicle for information storage. This has been stated as"Information is stored not in a stable inert structure, such as template-replicating DNA, but in the self-consistent web of transformations"[15, 41]; and "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"[31]. Analogous ideas have been elaborated in less formal, but chemically more defined contexts[42, 43].

A detailed analysis of information transfer without biopolymers in prebiotic molecular assemblies has been put forward by Morowitz[5, 30, 44]:"*Memory can also exist without specific macromolecules, but may be instantiated in chemical networks, catalytic loops and reflexive autocatalysis"*[30]. The concept explored is that in very early life, the "memory" and the "operating system" were embodied unitedly, and that the separation of informational biopolymers from metabolic catalysts only occurred later in evolution (Figure 2). Analogous distributed memory

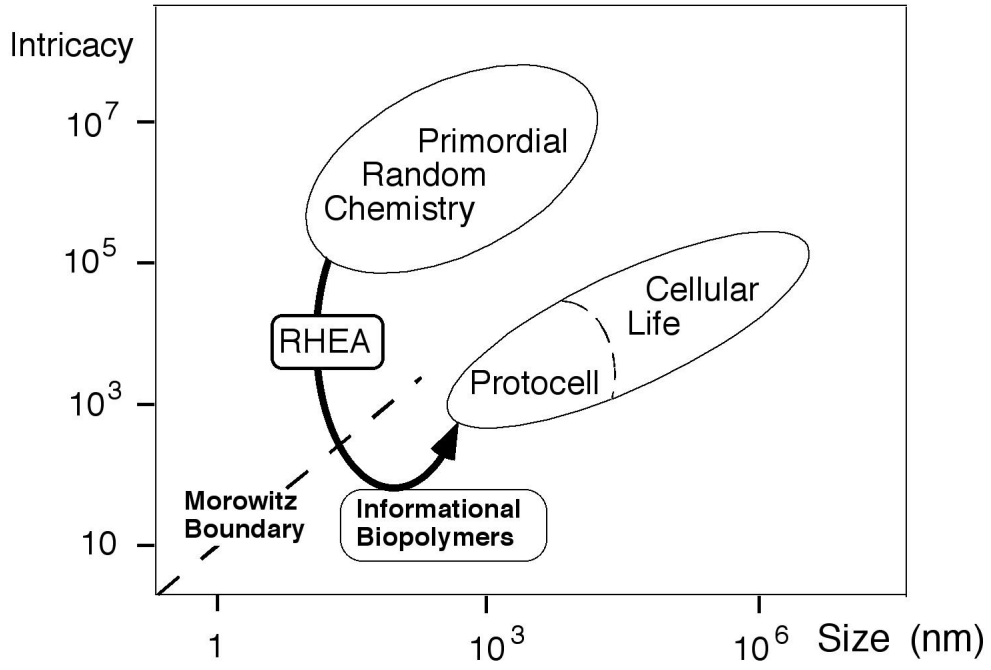

Figure 2. A compositional space diagram illustrating conjectural rough relationships between assembly composition and size in early prebiotic events. Intricacy[47], the number of different types of molecular species in an assembly, is plotted against assembly size. Early prebiotic chemistry is assumed to have generated many millions of different organic molecules in the size range of a few hundred Dalton. These may have formed small assemblies (e.g. micelles) in the 0.1-10 μm range. Such primordial assemblies might lie in a region of the diagram (upper oval) where intricacy is related to assembly size as dictated by a multinomial distribution statistics[21]. A hypothetical path is shown (thick arrow) from prebiotic random chemistry to more biased, life-like assemblies, capable of transmitting compositional information. It is proposed that RHEA models, i.e. mutual catalysis-based,"metabolism first" scenarios, provide a likely path for such assemblies to cross the Morowitz boundary (broken line, representing a probability=0.5 for successful division), by selecting subsets of molecules capable of efficient homeostasis (see text). Below this boundary, those assemblies that grow and split can propagate their compositional information with some fidelity. This consequent decreased intricacy, potentially associated with only minor changes of assembly size, means the generation of assemblies with a relatively small number of types of low molecular weight chemical species ("monomers"), thus paving the way for informational biopolymers. Only then may intricacy rise again, as the complex attributes leading to cellular life begin to emerge. Protocells may have been characterized by an augmented size, and the appearance of structural complexity, with only modest increase of intricacy. This trend is clearly manifested in present day cells which may be up to 1mm in size, contain trillions of molecules, but may have only a few ten thousands different kinds of molecular species, including proteins.

devices may be observed in many complex systems, e.g. spin glasses and associative neural networks[15, 45].

As an example, consider N=128 nucleotides. Orderly strung in a polymer, they encode $\log_2(4^N)=2N=256$ bits of information, whereas if placed in a compositional "bag" they only contain about $\log_2(N^3/3!)\approx 18$ bits. However, while replicating a biopolymer requires a specialized mechanism of templating and catalyzed covalent bond formation, non-covalent assemblies may undergo replication in a much simpler way: growth and splitting (see "The amphiphile advantage" below).

One of the critical questions to be asked[44] is how a molecular assembly could ensure its compositional inheritance. Morowitz shows how constraints on the transmissibility of compositional information are related to the size of the assembly and the diversity of its molecular species[5, 44, 46]. It is demonstrated that if an assembly contains m molecular types (the complexity of the assembly or its "intricacy"[47]), and if each is present with an average copy number 2r per assembly, then P_d, the probability that all molecular types are represented in the progeny in at least one copy, is given by $P_d = (1 - e^{-r})^m$. Assuming that all the components are catalytically essential, P_d represents the likelyhood of a successful functional division. In the above example, m=4 and 2r=32, and the probability of faithful division P_d is very near to 1 (only one in a million chance for a "wrong" division). However, if the assembly were much smaller, say with 2r=4, then the splitting success probability would go down to about $P_d = 0.55$. For a given average copy number, it is thus possible to define a "Morowitz boundary", corresponding to the assembly size that yields $P_d = 0.5$ (Figure 2).

The ideas of "compositional genome" and "Morowitz boundary" may serve as a basis for modelling early evolutionary processes with random chemistry tools.

Primordial Random Chemistry

Combinatorial or random chemistry has been defined as a "set of techniques for creating a multiplicity of compounds and then testing them for activity"[48]. Such compound repertoires, exemplified by phage display libraries [49, 50] are generated by random synthesis or by optimal design[51], and represent haphazard "cloud" configurations in chemical "shape space"[52, 53]. Besides their pharmaceutical importance for drug design, combinatorial chemistry tools have been used for addressing important questions about biocatalysis and self-replication in proteins[54] and nucleic acids[55].

A statistical chemistry approach would view the random emergence of diverse organic molecules, and their assemblies, as a natural consequence of undirected prebiotic organosynthesis[56]. It would then ask how life-like processes could emerge within such primordial random assortments (Figure 3), rather than study their specific molecular content (e.g. for α-amino acids or nucleotide bases). The RHEA models basically suggest that the transition from random chemistry to self-replicating entities would occur because of intrinsic statistical factors, e.g. the probability for mutual catalysis among randomly selected counterparts within a mixture. In this view, two chance-selected compounds that

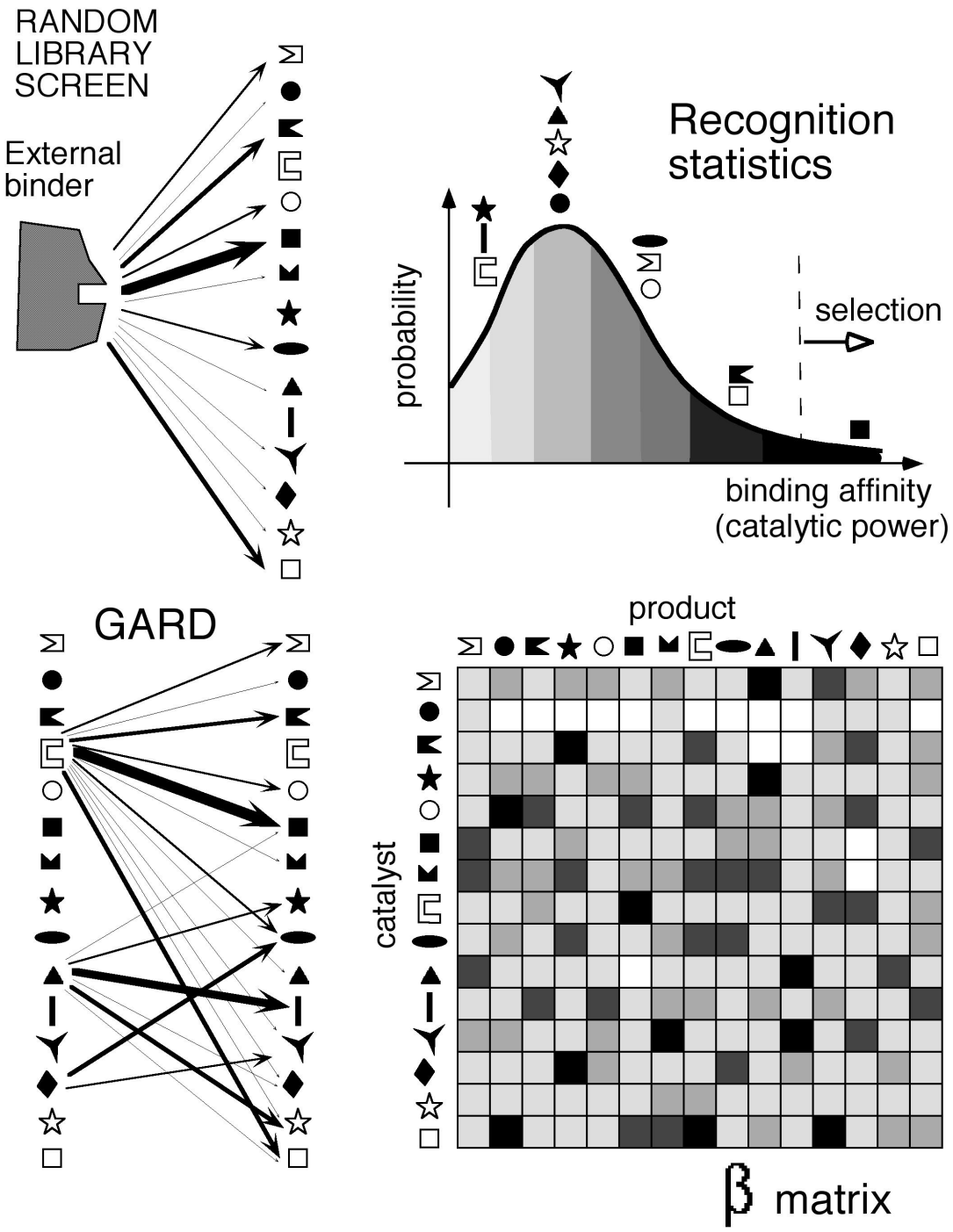

Figure 3. The analogy of random chemistry principles between current experimental systems and early primordial scenarios. Nowadays, random library screening (e.g. with a phage display library) is used for searching a maximal affinity ligand (dark square) for highly evolved receptors or antibodies (external binder, top left). Statistical models for the affinity distribution that governs such selection (top right) allow one to quantitate this process, with the highest affinity ligand being at the far right end of the distribution. A similar statistical distribution may have prevailed when both binders and ligands (or catalysts, substrates and products) were indistinguishable members of randomly formed mixture of prebiotic organics, as embodied in the GARD model (bottom left). The self-replication capacity of a set of molecules is related to a specific β matrix (bottom right), encompassing all the cross- and auto-catalysis enhancements within the set. In the right side figures, darker shade denotes higher catalytic potency.

manifest mutual catalysis obey the same chemical principles as those revealed when a high affinity peptide emerges from a phage display library. The main difference is that in the traditional combinatorial chemistry studies one of the counterparts is usually a highly evolved biological macromolecule, introduced by the experimenter, while prebiotic scenarios assume that both reactants are derived from the same combinatorial collection (Figure 3). A crucial step in the establishment of quantitative RHEA models, based on statistical/random chemistry concepts, is the definition of a metric for random catalytic interactions. As pointed out[15], such fortuitous, weak catalytic events may be thought to resemble those manifested in cases of enzyme mimetic organic molecules. An analogous problem of defining a metric of interactions is at the base of molecular recognition within random receptor repertoires such as in the immune and the olfactory pathways. Studies of such systems led to the delineation of the Receptor Affinity Distribution (RAD) model[57-60].

Later such concepts were extended to describe substrate-catalyst interactions, using a matrix of graded catalytic potencies[16, 18, 19] (β matrix, Figure 3). This forms an important cornerstone of the GARD model described below, and is analogous to the use of string matching rules to describe mutually catalytic interactions[16, 38]. Dyson's catalytic potency b with a probability 1/a, as well as Kauffman's catalytic events with a probability P, constitute special cases of the β matrix. The use of graded catalysis matrices may alleviate normalization problems that arise when employing one constant probability value for catalysis[36].

Quantitative Kinetics of Mutual Catalysis

The power of the statistical approach to mutually catalytic sets is greatly enhanced when combined with a quantitative kinetic analysis. This is essential for a rigorous demonstration that such sets may undergo self-replication, and for answering how very large numbers of molecular species may interact despite their extreme dilution(cf.ref.[24]). A computational approach realistic enough for faithfully simulating interactions among molecules in a prebiotic environment should thus best involve equations of chemical kinetics, as well as a statistical account of catalytic events. In this, quantitative RHEA models complement λ-calculus methods[61] and other Artificial Life approaches[62, 63].

The GARD model

The Graded Autocatalysis Replication Domain (GARD) model[17-21] provides a framework for a thermodynamic and kinetic analysis of mutually catalytic assemblies combined with statistical tools. It assumes a finite enclosure (e.g. an amphiphile vesicle), containing the catalytic set members, and absorbing energy-rich chemical precursors ("foodstuff"[15]) from the external environment. GARD's chemical and physical rules are computationally implemented by numerical solution of differential equations[19] or by Monte Carlo simulation procedures[21].

An important principle incorporated in GARD is the equivalence of homeostasis and replication[31, 46]. Simply formulated, if an enclosed, externally-fed catalytic set undergoes expansion, and if its catalytic network is effective, then its idiosyncratic internal composition will tend to be homeostatically preserved, despite the continuous growth of volume. If the compartment subsequently undergoes splitting due to physical forces, the "progeny" may constitute compositional similes of the original, and replication may ensue, subject to the Morowitz boundary formula.

The GARD formalism does not describe a specific molecular scenario, and may therefore apply equally well different kinds of chemistries (e.g. peptides, nucleotides or amphiphiles) and reaction schemes (isomerizations[18], covalent dimerization[19] or non covalent associations[21]). In GARD, every species A_j may in principle catalyze every possible reaction in which another species A_i is formed/decomposed, with a catalytic potency $\beta_{ij} \geq 0$. The β matrix of catalysis thus defined is populated with values derived from a catalytic potency distribution[16, 18]. Large values are very unlikely, but the long tail of the distribution predicts finite probabilities for high catalytic potencies. GARD is formally always catalytically closed[15], but it is possible to define within it realms the *degree* of catalytic closure for every conceivable assembly. This is embodied in a measure of homeostatic preservation, λ_c, defined as the highest expansion rate that still allows an above-threshold maintenance of the original composition[19].

A feature of the GARD model is its potential ability to provide quantitative predictions for primordial processes, based on measurable random chemistry data. Thus, for example, it is possible to relate the probability that an assembly with a given homeostatic preservation capacity will appear on earth, based on the planetary volumes and the geological time window available for its spontaneous formation[18]. GARD also allows to carry out chemically-realistic prebiotic evolutionary simulations, using the critical expansion value λ_c as a fitness parameter. In this way, highly improbable "self-organized" assemblies may be shown to arise by a combination of random composition seeding and gradual mutation-like small compositional changes[18, 20].

The Amphiphile Advantage: Non Covalent Assemblies

The idea that "*the simplest protocell which fulfills the principle of continuity is a bilayer vesicle made from a single species or a mixture of small amphiphiles*" [46], has been pursued by several researchers, both experimentally and theoretically[4, 21, 64, 65, 66]. This line of investigation is supported, among others, by evidence for the presence in meteorites of amphiphiles capable of forming micelles and vesicles[4, 67]. Present-day life is still based largely on the interactions of lipid-like molecules among themselves, as well as with other substances.

A large proportion of the organic molecules that potentially formed on primitive earth would have been rather water insoluble. Many such molecules could be amphiphilic, with one aspect containing polar heteroatoms (oxygen, nitrogen, sulfur, phosphorous) and the other being more hydrophobic. In the following section, we entertain the possibility that amphiphile assemblies could be early life precursors, and not just make up the vesicular enclosures that contain other biomolecules. In this context, amphiphiles possess the following unique advantages:

1. An ability to undergo spontaneous aggregation, overcoming the problem of dilution. This could allow molecules of different kinds to come to close proximity, so that mutual catalytic effects may be manifested

2. "Soft matter" properties, which are intermediate between those of aqueous solutions and solid minerals, and allow free internal diffusion and exchange[64]

3. A limit on aggregate size, which is crucial for the statistical generation of early diversity under conditions of high chemical diversity, whereby not all molecular types are present within a given assembly (Figure 2)

4. A facile capacity to recruit and lose components, leading to an ability to undergo compositional mutations, flexibly grow in size, and most importantly, to undergo spontaneous, physically-dictated splitting, generating primitive progeny[68]

5. A capability to undergo a fusion reaction, the reverse of splitting[69], entailing a potential for compositional symbiosis

6. Surface properties such as absorption, rudimentary molecular recognition and surface catalysis[66, 70], which may depend on the diversity and juxtaposition of polar "head-groups" [71], not unlike modern day proteins, which also possess a hydrophobic core and a diverse chemically active exterior

7. A propensity to form more elaborate structures, such as vesicles with an aqueous interior, a decisive step in leading to early protocells[46, 64, 72]

8. In the vesicular embodiment, a capacity of acting as an energy transducers, via electric potentials and concentration gradients[5, 73], possibly coupled to photochemistry.[46, 64]

A "Lipid combinatorics" origin

It is possible to combine GARD's detailed kinetic and statistical formalisms, with the advantages exhibited by amphiphile assemblies (Figure 4). The Amphiphile-GARD model[21] describes spontaneously aggregating molecules that have the capacity of catalyzing reactions by which other amphiphiles join/leave the assemblies or form/decompose within them. Such catalytic events at the surface of lipid structures are actually prevalent in present day cells[74].

In Amphiphile-GARD, individual amphipathic molecules form mixed assemblies with a very high degree of internal diversity and inter-assembly variation, comparable to the one observed in small subsets drawn from a random library of bio-oligomers. If different amphiphile assemblies contain different configurations of mutual catalysis, as governed by the GARD β matrix, then catalytic networks with different efficacy will result. Along the GARD formalism, some would maintain their idiosyncratic assembly compositions better than others. This, together with a facile capacity to gain and lose molecular constitutents, conforms with the basic definition of metabolism[31], and provides a mechanism for compositional mutations. A similar process, of "random access" chemical changes, is much more difficult to envisage in a primordial covalent polymer.

A significant feature of the Amphiphile-GARD concept is a natural and self consistent quantitation of assembly growth. In earlier GARD embodiments[19], as well as in other models[15, 31], the coupling between the potential growth of a vesicle-like enclosure and the internal catalytic activity is less clearly defined. In Amphiphile-GARD it is assumed that the same molecules that form the assembly, a micelle-like structure, are also responsible for the mutually catalytic functions. In this way, growth of assemblies becomes a kinetic consequence of successful mutual catalysis.

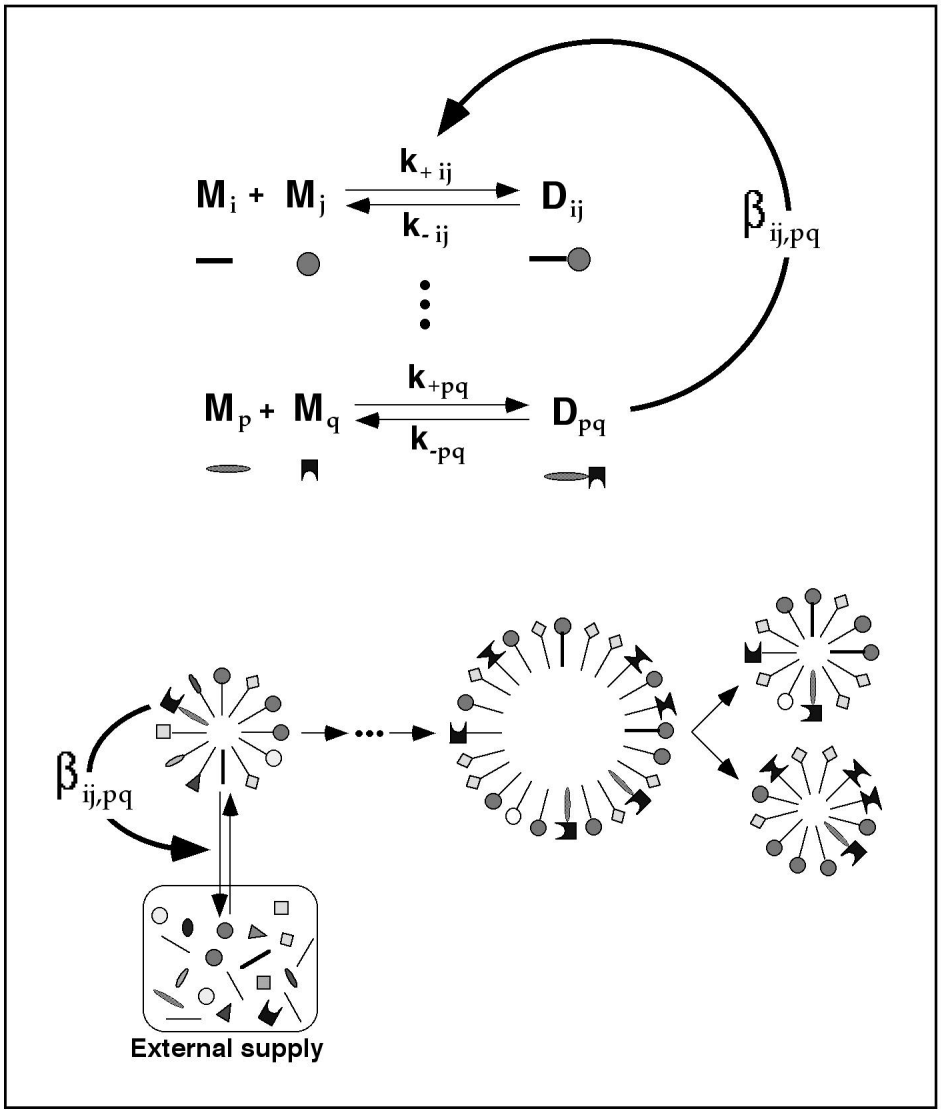

Figure 4. A schematic depiction of the GARD model. In the embodiment shown, amphiphilic dimers are catalytically formed from polar heads (geometrical shapes) and apolar tails (sticks). Dimers are formed in a reversible bimolecular reactions with defined forward and reverse kinetic constants[19, 81]. Each dimer may act as a catalyst for any of the dimerization reactions, with rate enhancement factors governed by a β matrix (Figure 3). These amphiphile dimers form micelles-like assemblies (bottom), within which the mutually catalytic reactions may favorably take place. Catalyzed joining and chemical modification may lead to assembly growth, potentially accompanied by physically-regulated splitting processes. These processes may be rigorously modelled by computers simulations[21] (see text).

The results of numerical simulations, based on the Amphiphile-GARD model[21], implemented with a stochastic single molecule numerical simulation approach[75] demonstrated the spontaneous emergence of compositionally biased, catalytically superior assemblies out of a random "soup" of amphiphiles. These catalyzed assemblies, unlike the randomly-formed ones that lack significant mutual catalysis, tend to lie below the Morowitz Boundary (Figure 2), hence are candidates for a more successful self-replication upon splitting.

Future Prospects

We have described here some models for the behavior of random collections of mutually catalytic chemicals. These Replicative-Homeostatic Early Assemblies (RHEA) models, based on the statistics of large molecular repertoires, may be regarded as capable of filling the gap between the hypothetical primordial soup and the relatively advanced prebiotic stages, at which information-carrying biopolymers began to appear.

It seems that time is ripe for an enhanced quantitative exploration of such concepts, including the modern tools of molecular recognition and molecular dynamics. A two-fold development may be envisaged:

1. Theoretical models aided by computer simulations should allow one to perform increasingly realistic *in silico* experiments for testing the possible consequences of different hypothetical prebiotic evolution scenarios, and for determining their typical time and probability scales. This should be facilitated by advances in ultrafast parallel computers.

2. Experimental tests of the capacity of assemblies of mutually catalytic species to self-replicate and undergo selection and evolution could complement the theoretical simulations and help in improving them. For this, the mainstream studies of self-replication of individual molecular types[24, 26, 27, 28, 66, 76] should be merged with random chemistry experiments with heterogeneous molecular assemblies.

The statistical chemistry approach may open the way to new avenues of studying the origin of life. The profound knowledge of simple free-living organisms and their metabolic pathways, as brought about by the world Genome Project, could serve to guide extensions of the simplest RHEA models. These future models could incorporate an increasing complexity, including the emergence of biopolymers and perhaps even a primitive genetic code, thereby leading to a better understanding of how life originated.

Acknowledgements

We thank Dr. David Deamer, Dr. Amit Meller, Tzachi Pilpel, Shai Rosenwald, and Gustavo Glusman for their critical comments, and Dr. Nathan Sharon for the catalysis.

References

(1) Hargreaves, W.R., Mulvihill, S. Deamer, D.W. *Nature* **1977**, *266,* 78-80.
(2) Miller, S.L. *Science* **1953**, *117,* 528-529.
(3) Miller, S.L. *Chemtracts, this volume* **1998**,
(4) Deamer, D.W. *Chemtracts, this volume* **1998**,
(5) Morowitz, H.J. *Beginnings of cellular life*; Yale University Press: London, 1992.
(6) Lahav, N. S., N. *Origins Life Evol. Biosphere* **1997**, *27,* 377-395.
(7) Burgess, A.W. *J. Theor. Biol.* **1982**, *96,* 21-38.
(8) Gilbert, W. *Nature* **1986**, *319,* 618.
(9) Joyce, G.F. *The New Biologist* **1991**, *3,* 399-407.
(10) Eigen, M. Winkler-Oswatitsch, R. *Steps towards life*; Oxford University Press: Oxford, 1992.
(11) Fox, S.W. Harada, K. *Science* **1958**, *128,* 1214.

(12) Pappelis, A. Fox, S.W. In *Chemical Evolution: Physics of the Origin and Evolution of Life*, Chela-Flores, J. Raulin, F., ed. Kluwer: 1996; pp 157-165.
(13) de Graaf, R.M., Visscher, J. Schwartz, A.W. *Nature* **1995**, *378,* 474-477.
(14) Shapiro, R. *Origins Life Evol. Biosphere* **1984**, *14 (1-4),* 565-570.
(15) Kauffman, S.A. *The origins of order - Self-organization and selection in evolution*; Oxford University Press: Oxford, 1993.
(16) Lancet, D., Kedem, O. Pilpel, Y. *Ber Bunsenges. Phys. Chem.* **1994**, *98,* 1166-1169.
(17) Lancet, D., Glusman, G., Segré, D., Kedem, O. Pilpel, Y. *Origins of Life and Evolution of the Biosphere* **1996**, *26,* 270-271.
(18) Segré, D., Pilpel, Y., Glusman, G. Lancet, D. In *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, C.B., Bowyer, S. Werthimer, D., ed. Editrice Compositori: Bologna, 1997; pp 469-476.
(19) Segré, D., Lancet, D., Kedem, O. Pilpel, Y. *Origins Life Evol. Biosphere* **1998**, *28,* 501-514
(20) Segré, D., Pilpel, Y. Lancet, D. *Physica A* **1998**, *249,* 558-564.
(21) Segré, D. Lancet, D. In *Exobiology: matter, energy and information in the origin and evolution of life in the universe*, Chela-Flores, J. Raulin, F., ed. Kluwer, 123-131.
(22) Eigen, M. Schuster, P. *The Hypercycle*; Springer Verlag: Berlin, 1979.
(23) Maynard Smith, J. Szathmary, E. *The Major Transitions in Evolution*; W.H. Freeman & Co.: New York, NY, 1995.
(24) Orgel, L.E. *Nature* **1992**, *358,* 203-209.
(25) Küppers, B. *Molecular theory of evolution*; Springer-Verlag: Berlin-Heidelberg, 1983.
(26) Rebek, J. *Scientific American* **1994**, 48-55.
(27) Lee, D.H., Granja, J.R., Martinez, J.A., Severin, K. Ghadiri, M.R. *Nature* **1996**, *382,* 525-528.
(28) Li, T. Nicolaou, K.C. *Nature.* **1994**, *369,* 218-221.
(29) Szathmáry, E. Maynard Smith, J. *J. Theor. Biol.* **1997**, *187,* 555-571.
(30) Morowitz, H.J. *A theory of biochemical organization, metabolic pathways and evolution*; **1996**, Santa Fe Institute, 96-04-014.
(31) Dyson, F. *Origins of Life*; Cambridge University Press: Cambridge, 1985.
(32) Zubay, G. *Chemtracts-Biochem. and Mol. Biol.* **1994**, *5,* 179-198.
(33) Oparin, A.I. *The origin of life on the earth*; Oliver and Boyd: London, 1957.
(34) Dyson, F.J. *J. Mol. Evol.* **1982**, *18,* 344-350.
(35) Kauffman, S.A. *J. Theor. Biol.* **1986**, *119,* 1-24.
(36) Lifson, S. *J. Molec. Evol.* **1996**, *44,* 1-8.
(37) Stadler, P.F., Fontana, W. Miller, J.H. **1993**, *63,* 378-392.
(38) Bagley, R.J. Farmer, J.D. In *Artificial Life II, SFI Studies in the Sciences of Complexity*, Langton, C.G., Taylor, C., Farmer, J.D. Rasmussen, S., ed. Addison-Wesley: 1991; pp 93-140.
(39) Bagley, R.J., Farmer, J.D. Fontana, W. In *Artificial Life II, SFI Studies in the Sciences of Complexity*, Langton, C.G., Taylor, C., Farmer, J.D. Rasmussen, S., ed. Addison-Wesley: 1991; pp 141-158.
(40) Yockey, H.P. *Information theory and molecular biology*; Cambridge University Press: 1992.
(41) Kauffman, S. *At Home in the Universe*; Oxford University Press: 1995.
(42) Wachtershauser, G. *Proc. Natl. Acad. Sci. USA* **1990**, *87,* 200-204.
(43) Huber, C. Wächtershäuser *Science* **1997**, *276,* 245.
(44) Morowitz, H.J. In *Progress in Theoretical Biology*, Snell, F.M., ed. Academic Press: 1967; pp 35-58.
(45) Stein, D.L. Anderson, P.W. *Proc. Natl. Acad. USA* **1984**, *81,* 1751-1753.
(46) Morowitz, H.J., Heinz, B. Deamer, D.W. *Origins of Life and Evolution of the Biosphere* **1988**, *18,* 281-287.
(47) Smith, T.F. Morowitz, H.J. *J. Mol. Evol.* **1982**, *18,* 265-282.
(48) Borman, S. *Chem. & Eng. News* **1997**, 43-62.
(49) Scott, J.K. Smith, G.P. *Science* **1990**, *249,* 386-390.
(50) Collins, J. In *Annual reports in combinatorial chemistry and molecular diversity*, W.H.

Moos, M.R.P., A. D. Ellington and B. K. Kay, ed. Leiden: 1997; pp 210-262.
(51) Hassan, M., Bielawski, J.P., Hempel, J.C. Waldman, M. *Mol. Divers.* **1996**, *2(1-2),* 64-74.
(52) De Boer, R.J., Segel, L.A. Perelson, A.S. *J. Theor. Biol.* **1992**, *155 (3),* 295-333.
(53) Holm, L. Sander, C. *Science* **1996**, *273,* 595-602.
(54) Moore, J.C., Jin, H., Kuchner, O. Arnold, F.H. *J. Mol. Biol.* **1997**, *272,* 336-347.
(55) Gold, L., Brown, D., He, Y., Shtatland, T., Singer, B.S. Wu, Y. *Proc. Natl. Acad. Sci. USA* **1997**, *94,* 59-64.
(56) Schwartz, A.W. *Chemistry and Biology* **1996**, *3,* 515-518.
(57) Lancet, D., Horovitz, A. Katchalski-Katzir, E. In *Perspectives in supramolecular chemistry*, Behr, J.-P., ed. J. Wiley New-York: 1994
(58) Lancet, D., Sadovsky, E. Seidemann, E. *Proc. Natl. Acad. USA.* **1993**, *90,* 3715-3719.
(59) Levitan, B. In *Annual Reports in Combinatorial Chemistry and Molecular Diversity*, Moos, W.H., Pavia, M.R., Ellington, A.D. Kay, B.K., ed. ESCOM, The Netherlands: 1997; pp 95-152.
(60) Rosenwald, S. Lancet, D. *(unpublished)* **1998**,
(61) Fontana, W. Buss, L.W. *Proc. Natl. Acad. Sci. USA.* **1994**, *91,* 757-761.
(62) Moreno, F., Moran, A., Chacon, J.J. Merelo, A. *Advances in Artificial Life*; ed. Springer: Granada, Spain, 1995.
(63) Langton, C.G., Taylor, C., Farmer, J.D. Rasmussen, S. *Artificial life II*; ed. Addison-Wesley: 1992.
(64) Deamer, D.W. *Microbiol. Molecular Biology Reviews* **1997**, *61(2),* 239-261.
(65) Ourisson, G. Nakatani, Y. *Chemistry & Biology* **1994**, *1,* 11-23.
(66) Bachmann, P., Luisi, P. Lang, J. *Nature* **1992**, *357,* 57-59.
(67) Deamer, D.W. *Origins Life Evol. Biosphere* **1989**, *19,* 21-38.
(68) Lipowsky, R. *Nature* **1991**, *349,* 475-481.
(69) Blumenthal, R. Dimitrov, D.S. In *Handbook of physiology*, Hoffman, J.F. Jamieson, J.C., ed. Oxford University Press: New York, 1997; pp 563-603.
(70) Cuccovia, I.M., Quina, F.H. Chaimovich, H. *Tetrahedron* **1982**, *38,* 917-920.
(71) Lipowsky, R. *Curr. Opin. Struct. Biol.* **1995**, *5(4),* 531-540.
(72) Luisi, P.L., Walde, P. Oberholzer, T. *Ber. Bunsenges. Phys. Chem.* **1994**, *98,* 1160-1165.
(73) Pohorille, A. Wilson, M.A. *Origins Life Evol. Biosphere* **1995**, *25,* 21-46.
(74) Zachowski, A., Henry, J.P. Devaux, P.F. *Nature* **1989**, *340,* 75-76.
(75) Bartholomay, A.F. *Biochemistry* **1962**, *1,* 223-230.
(76) Sievers, D. Von-Kiedrowski, G. *Nature.* **1994**, *369,* 221-224.
(77) Lazcano, A. Miller, S.L. *Cell* **1996**, *85,* 793-798.
(78) Deamer, D.W. Fleischaker, G.R. *Origins of life - The central concepts*; Jones & Bartlett: Boston, 1994.
(79) Lahav, N. *Biogenesys: theories of life's origin*. Oxford University press, New-York, 1999.
(80) Bak, P. Paczuski, M. *Proc. Natl. Acad. USA.* **1995**, *92,* 6689-6696.
(81) Goldbeter, A. Nicolis, G. *Biophysik* **1972**, *8,* 212-223.