Version: 6/1/03

THEORETICAL AND COMPUTATIONAL APPROACHES TO THE STUDY OF THE ORIGIN OF LIFE

D. SEGRÈ[1] **AND D. LANCET**[2]
[1]*Lipper Center for Computational Genetics and Department of Genetics Harvard Medical School, 200 Longwood Ave., Boston, MA 02115.*
[2]*Department of Molecular Genetics and The Crown Genome Center The Weizmann Institute of Science, 76100 Rehovot, Israel.*

1. Prebiotic synthesis and molecular diversity

1.1. THE SUBSTRATE OF EARLY EVOLUTION

Investigating the evolutionary path that led to present-day cells is a little like tracing a maze, where the entrance and exit are known, but much work remains to find the paths that connect them. For the origin of life, the "entrance" is the multifaceted environment of a water- and carbon-rich planet and the "exit" is a protocellular structure subject to Darwinian evolution.

Whether terrestrial life started on our planet or elsewhere in the Universe is a recurrent subject of debate, often nourished by proposals of Panspermia scenarios (Crick, 1981; McKay, 1997), and recently inspired by the finding of candidate relics of martian life forms (Kerr, 1996; McKay et al., 1996). An extraterrestrial origin cannot be absolutely discarded using the present knowledge, but there are two good reasons for setting aside this debate: (i) there is currently no solid evidence of extraterrestrial life, and (ii) moving the problem elsewhere still requires solving it. While exciting recent analyses have certified the presence of planetary systems other than the solar one (Lissauer et al., 2000), and future explorations might one day give us the opportunity to analyze frozen relics of extraterrestrial life, most origin of life scenarios assume that the cradle of life-as-we-know-it has been planet Earth (de Duve, 1995; Lahav, 1999; Oparin, 1957).

Although the details of prebiotic Earth conditions and of the nature of the earliest forms of life are probably lost forever, a lot is known about the general features primordial Earth. For example, a time window for the transition from inanimate to animate matter has been estimated using geophysical and astronomical methods. The lower bound of this window, 3.9 billion years ago, corresponds to the end of a period of intense meteorite bombardment (Chyba and Sagan, 1992), as testified, for example, by the analysis of the Moon's craters (Maher and Stevenson, 1988). For the upper bound we can rely on geological records, such as stromatolites and microscopic fossils, which provide evidence that 3.8 billion years ago life was already present on Earth (Mojzsis et al., 1996; Mojzsis et al., 1998; Schopf et al., 1965).

Since Oparin's first detailed description of how the emergence of life could be in principle explained on a molecular basis (Oparin, 1957), several experiments have been performed for evaluating potential paths towards the spontaneous emergence of a primordial cell. This has been mostly guided by the chemistry of present-day organisms, and in particular by the fundamental role of nucleic acids (Gesteland et al., 1999; Gilbert, 1986; Orgel, 1998) and proteins (Fox, 1991; Lee et al., 1996). This concept is inspired by an extreme of the Continuity Principle, stating that every stage in evolution is connected continuously to the previous ones (Morowitz, 1992; Orgel, 1968). It is assumed that very early life contained the same biopolymers as present day life (cf. Orgel, 1998).

At the other extreme, inorganic clay minerals were proposed to be the first substrate for life-like processes (Cairns-Smith, 1982). In the model proposed by Graham Cairns-Smith, imperfections in a growing crystal are the first form of inheritable information, and organic life is proposed to have arisen later, through a genetic takeover (Cairns-Smith, 1982). Clays are minerals made of micron-size particles, composed of complex crystals based on silicon, aluminum, iron and other inorganic compounds. The fluidity and the complexity of collections of these particles in water inspired the analogy with organic macromolecules. The clay mineral model, however, is considered a less likely predecessor of the cell, because it constitutes an extreme deviation form the Continuity Principle (Fry, 2000; Morowitz, 1992).

An intermediate approach may be envisaged, in which continuity is used to indicate that life started with carbon-based molecules, but without any assumption about the detailed structure of the primordial organic compounds. As opposed to the view invoking the initial abundant presence of present-day building blocks, this scenario envisages a selection process that gradually reduces the diversity of molecular building blocks, in analogy to the screening of large random libraries in random chemistry experiments (Segrè and Lancet, 1999).

1.2. SOURCES OF MOLECULAR DIVERSITY

The formation of the chemical elements necessary for the emergence of life started with the Big Bang (Macià et al., 1997; Orò, 2000; Sciama, 1971). Light nuclei, such as H, D, ^{3}He and ^{4}He, are believed to have formed during the early dense stages of the expanding Universe (Arnett, 1996). Additional synthesis of He and of heavier elements by thermonuclear burning followed the formation of stars and galaxies. Newly synthesized elements, from C to U, were ejected by exploding supernovae, and eventually incorporated into other stars or into planets (Arnett and Bazan, 1997; Trimble, 1997). The nuclear reaction responsible for the formation of the nuclide ^{12}C is the condensation of three alpha particles (Oberhummer et al., 2000). This reaction, as well as the formation of O in large amount, S, and small fractions of P (Macià et al., 1997) is thought to happen especially in carbon stars (Doty and Leung, 1998; Orò, 2000). Upon migrating to the circumstellar and cooler regions of the star, at temperatures of a few thousands degrees, the biogenic elements (H, C, O, N, S, P) produce more complex molecules (Doty and Leung, 1998), through ordinary chemical reactions. The newly formed compounds include diatomic and triatomic species such as C_2, CN, CO, CH, NH, OH and H_2O, as observed also in the atmosphere of the Sun (Orò, 2000).

Anders (Anders, 1989) and Chyba and Sagan (Chyba and Sagan, 1992) have found evidence that interplanetary dust particles (IDP) were the most abundant source of extra-terrestrial organic carbon in the late accretion phase of the early Earth. Comets and carbonaceous meteorites would contribute smaller amounts. The total delivery over a period of 100 million years of the late Hadean - early Archean eras is estimated to be on the order of 10^{16}-10^{18} kg (Chyba and Sagan, 1992). For comparison, the total organic carbon in the present biosphere is $6 \cdot 10^{14}$ kg, i.e. 10 to 1000 times smaller. Hence, organic material delivered as extraterrestrial infall was likely to have been a significant contribution to the organic inventory of the early Earth environment. In addition to extratrerrestrial delivery, several potential prebiotic reactions that could produce simple organic compounds on early Earth have been proposed (de Graaf et al., 1995; Huber and Wachtershauser, 1997; Miller and Urey, 1959; Schlesinger and Miller, 1983; Wachtershauser, 1988). Energy sources available to drive such reactions range from volcanoes and hydrothermal vents to lightning discharges, solar photochemistry and pyrite-dependent reduction. It should be stressed that in many cases, in addition to water-soluble organic compounds that are easily analyzed, insoluble residue (e.g. "tar" or "tholins" (Bernstein et al., 1999; McDonald et al., 1996) are reported in prebiotic synthesis simulations (Miller and Urey, 1959; Schlesinger and Miller, 1983). These are likely to contain sparingly soluble hydrophobic and amphiphilic compounds of relatively large molecular mass.

Why should one be interested in this large diversity of molecular shapes formed abiotically on early Earth or in outer space? Many scholars have been screening potential prebiotic mixtures in search for amino acids (Miller, 1953; Miyakawa et al., 2002), nucleosides (Ponnamperuma et al., 1963a; Robertson and Miller, 1995), and other precursors of present-day cellular life (McCollom et al., 1999; Ponnamperuma et al., 1963b; Reid and Orgel, 1967; Sagan, 1963). The success of such searches is sometimes mistakenly appraised as the solution to the problem of the origin of life. Indeed, the conclusion that a prebiotic mixture contains some of the building blocks of a present-day living cell constitutes a fundamental link between the inanimate and the animate worlds (Morowitz, 1992). Yet, a prominent alternative view is that one should take into account the whole primordial repertoire of organic compounds, and ask how its physical and chemical properties may lead to life-like characteristics (Fraser and Folsome, 1975; Lifson, 1996; Morowitz, 1992). This question, of how a “wild” prebiotic mixture (primordial soup) could self-organize to give rise to a self-reproducing and evolving system remains largely unanswered (Dyson, 1999).

Mathematical and computational models aimed at understanding these early stages of biological organization do not generally depend on a detailed knowledge of the composition of a primordial soup (Dyson, 1999; Kauffman, 1993). Rather, they make use of global measures, like the diversity of molecular species, or statistical parameters, such as the probability that a given molecules catalyzes the synthesis of any other molecule in a set. The question of prebiotic molecular diversity has been addressed explicitely by Morowitz (Morowitz, 1992), who reported the number of isomeric alkanes as a function of carbon number. The numbers fit an exponential law

$$N_G = Ke^{\alpha C} \qquad (1)$$

where N_G indicates the global number of different chemical species possible, C is the number of carbon atoms, and $K=6.7\cdot10^{-3}$ and $\alpha=0.915$ are parameters characteristic of the specific class of compounds. For 10 carbons the number of isomers is 75. A double number of carbons yields more than 360,000 compounds. It should be stressed that this calculation is relative to carbon and hydrogen atoms only. Hence, the above formula could be considered as a possible lower bound for molecular diversity. This kind of calculation, in addition to being useful for explicit mathematical modeling of prebiotic phenomena, should remind us of how narrow the chemical spectrum of present-day cellular life is with respect to the variegated universe of all possible chemical structures. Living cells typically contain less than thousand different low molecular weight "monomeric" compounds.

2. Chemical evolution

2.1. DARWINIAN EVOLUTION AT THE MOLECULAR LEVEL

The three classical elements of Darwin's theory, self-reproduction, mutation and selection, are widely accepted as the fundamental ingredients of evolution at the level of organisms. As stated by Li and Graur (Li and Graur, 1991), natural selection is the "differential reproduction of genetically distinct individuals or genotypes within a population". This definition could be extrapolated to early forms of life with primitive genomes, or even to freely floating lone genes capable of copying themselves (Eigen, 1971). But such definitions appear inapplicable to systems that do not possess genetic material. Any model that assumes early protocells that constitute ensembles of molecules without genetic machinery, would appear unable to explain how an early Darwinian process could take place. This puzzle is one facet of the Chicken-Egg paradox, summarized as follows: (i) Biopolymer-based self-reproducing systems must have evolved from simpler non-self-reproducing systems, (ii) Evolution can act only upon systems based on self-reproducing bio-polymers. Apparently, when dealing with primordial life, we should consider relinquishing one of these propositions. Relaxing proposition (i) implies that the first self-reproducing system emerged without evolving, i.e. that life originated in a one-time "lucky" jump. Such a view is sometimes accepted as an axiom of the inexplicability of life's origin (Yockey, 1992). Alternatively, self-organization processes have been suggested to potentially happen at the molecular level in absence of selective pressure (Horowitz, 1945), perhaps in analogy to random genetic drift (Dyson, 1999). Often the argument is presented that despite the extremely low probability for the chance sudden emergence of a self-reproducing system, the long available time and the huge set of possible geophysical settings could make such an event reasonably likely. An alternative view involves relaxing proposition (ii). This implies that a kind of natural selection may act on systems that do not contain self-reproducing biopolymers, perhaps on systems totally lacking genes and enzymes (Woolfson, 2000; Segrè et al., 2000). Is this possible? Could there be self-reproduction before genes? What would be the meaning of natural selection in a population of geneless protocellular systems?

2.2. INHERITANCE OF WHAT?

In modern living organisms, as the central dogma teaches us, biological information is contained in the genome (Crick, 1981). The flow of information is basically unidirectional, starting from DNA and ending with proteins. It is generally assumed that inheritance has to do with the copying of the message contained in the genome (Dawkins, 1996; Muller, 1966; Orgel, 1998). Indeed, this mechanism is so important, that the mutation of a single letter in this message might critically determine the fate of the cell receiving the modified sequence. When we talk about biological inheritance, however, we are not considering abstract systems that send and receive symbols like in Shannon's theory (Yockey, 2000). A DNA sequence alone has no meaning. Biological inheritance involves the self-reproduction of a complex molecular structure - the cell. The fact that one of its components, DNA, possesses an alphabet based string-like structure does not imply that information and inheritance should be ascribed only to sequences embodied by nucleic acid polymers. During cell division, lipids, proteins, whole organelles, cytosolic molecules and many other components are being inherited together with DNA. What is then biological information? What is inherited during a self-reproduction process?

Along the history of life, from an inanimate mixture of organic compounds to a sophisticated reproducing cell, there must have been intermediate stages involving proto-organisms capable of undergoing evolution without relying on the complex genetic machinery present today (Morowitz et al., 1988; Shapiro, 2000). It is conceivable that biopolymers were initially much shorter, and yet, an evolutionary process took place. It is necessary to envisage a simpler form of inheritance, which does not involve genes, coding and translation. This form of inheritance is the transmission of a peculiar chemical composition, as pioneered by Oparin (Oparin, 1957), and elaborated by many others (Dyson, 1982; Kauffman, 1986; Morowitz, 1967; Morowitz et al., 1988; New and Pohorille, 2000; Woolfson, 2000; Segrè et al., 2000). This is embodied in the "garbage bags" scenario (Dyson, 1999) where assortments of diverse simple molecules are capable of growing and splitting inaccurately. And interestingly, it is essentially true even today, where molecular compositions inherited during mitosis contain highly evolved, specialized and complex biopolymers, and the fate of each molecule upon cell division is carefully programmed. This compositional inheritance view is, however, still debated by many investigators. One of the reasons for this is the relatively small number of proposed formal analyses of such a process.

2.3. MOLECULAR SELF-REPLICATION

One of the fundamental characteristics of a living system is the capacity to make more of itself. The simplest chemical system bearing such attribute is an autocatalytic molecule, i.e. a molecule capable of accelerating the rate of its own production from precursors. (Lifson, 1996; Orgel, 1992; Szathmary and Maynard Smith, 1997). Autocatalytic molecules were experimentally designed in the laboratory, and it was shown that molecular self-replication in vitro is indeed possible. Examples involve modified nucleotides (Li and Nicolaou, 1994; Rebek, 1994; Tjivikua et al., 1990), as well as short peptides (Lee et al., 1996). One of the most attractive forms of autocatalyisis is the one involving biopolymers, as explored by Orgel’s through the template-directed synthesis of oligonucleo-

tide analogues (Orgel, 1992; Schwartz and Orgel, 1985; Sievers and von-Kiedrowski, 1994). Despite the increasing success in the design of autocatalytic molecules, it should be mentioned that no relevant example has been reported within any life form. Neither DNA nor RNA is truly autocatalytic, as they require the aid of specialized enzymes. How likely is it that nature could randomly hit upon a contrived self-replicating molecule similar to molecules whose synthesis presently requires the careful engineering of experienced organic chemists? This is a question that would require considerable future pursuit.

2.4. FORMAL ANALYSIS OF SELF-REPLICATING MOLECULES

Setting aside the plausibility of self-replicating molecules, it is legitimate to assume their existence and investigate their kinetic behavior. This approach brought to the initiation of a formal description of molecular systems that may undergo evolutionary processes. Such formalism is analogous to the evolutionary dynamics of more highly organized living systems. The most prominent mathematical model was developed by Manfred Eigen and Peter Schuster in the 70's (Eigen, 1971; Eigen and Schuster, 1979). This is analogous to the classical formalisms used in population biology for describing the increase and regulation of animal populations (Wilson and Bossert, 1971), based on a Malthusian growth law (Szathmary and Maynard Smith, 1997), as summarized by the logistic equation

$$\mathrm{d}N/\mathrm{d}t = aN\text{-}bN^2 \qquad (2)$$

In Eigen's scenario, a set of RNA polymers is enclosed in a flow reactor. Each such polymer is assumed to catalyze its own replication from constantly supplied activated monomers, and can therefore be referred to as a self-replicating information carrier, or replicator (Szathmary, 1999; Szathmary and Maynard Smith, 1997). The fidelity of a replicator is not always absolute: errors during copying may occur (Eigen and Schuster, 1979; Swetina and Schuster, 1982), whereby a polymer may catalyze the production of a variant polymer, rather than its own. Such replication errors are analogous to DNA sequence mutations.

In one of the possible formulations (Constant Population (CP)) the equation for the concentration of replicator *i* can be written as follows (Kuppers, 1983)

$$\frac{dx_i}{dt} = (W_i - E(t))x_i + \sum_{j \neq i} \psi_{ij} x_j \qquad (3)$$

where W_i is the autocatalytic rate enhancement (or selection value) of replicator *i*, *E(t)* is the average excess productivity, which keeps the overall concentration constant (in fulfillment of the CP constraint), and ψ_{ij} is the probability that replicator *i* will erroneously give rise to replicator *j*. An overall matrix of kinetic rates may be visualized as composed of the autocatalysis W_i terms (diagonal values) and the error rates ψ_{ij} (off-diagonal values). Eigen shows that a steady state is reached (Eigen, 1971; Eigen and Schuster, 1979; Kuppers, 1983), for which only a specific linear combinations of replicators, called a quasi-species, persists. In a transformed linear system of equations, the quasi-species corresponds to the eigenvector associated with the largest eigenvalue of the linear opera-

tor (Jain and Krishna, 1998). The quasi-species is hence associated with a distribution of RNA sequences with maximal replication rate. The larger the ψ_{ij} values with respect to W_i, the broader the distribution of different sequences contributing to the quasi-species.

2.5. ERROR CATASTROPHES

An important aspect of Eigen's analysis is the characterization of a threshold value for the rate of replication mistakes, which critically determines the information transfer property of the quasi-species (Alves and Fontanari, 1998; Eigen, 2000). Small error rates do not affect the replicating polymers in a dramatic way: the distribution of sequences at steady state is centered around a master polymer (i.e. the species with highest autocatalytic rate), whose information is faithfully propagated through the generations. When a critical value (the error threshold) is reached, however, the information of the master sequence is lost, and the mixture of self-replicating polymers becomes randomized (Swetina and Schuster, 1982). This occurs in a way analogous to a phase transition, referred to as the error catastrophe (Eigen, 2000). It should be mentioned that while a slow mutation rate prevents the loss of information, a fast mutation rate can increase the capacity of a system to explore the sequence space and optimize its fitness. A balance between these two opposite tendencies is what determines, according to Eigen, an optimal mutation rate (Eigen and Schuster, 1979). The error threshold analysis imposes a limit to the length of sequences that can be replicated with sufficient accuracy. Structures with a higher level of organization, the hypercycles, involving protein enzymes coded by the nucleic acid polymers, are shown to be necessary in order to afford longer polymers with the same accuracy of replication (Eigen, 2000).

2.6. RNA WORLD

An important origin of life scenario involving molecular replication is the RNA-world (Gesteland et al., 1999; Gilbert, 1986; Joyce, 1991). Following the discovery that RNA can perform catalyitc activities (Cech, 1986), it was proposed that the early evolution of life may have undergone through the following stages: (i) RNA polymers could act as primitive genes and enzymes (ribozymes, cf. Ellington et al., 1997; Lohse and Szostak, 1996), and propagate information through self-replication (Cech, 1986; Szostak, 1992), then (ii) RNA acquired the capability of synthesizing coded polypeptides (Lohse and Szostak, 1996), which would later specialize into becoming the major biocatalyst, and (iii) DNA took over for the role of genetic information carrier (Robertson and Miller, 1995). Ribozymes were invoked as possible indications of RNA's prebiotic role, since they persist today in tRNAs and within many cofactors, as well as in the active portion of the ribosome (Wimberly et al., 2000).

The interest in the potential role of RNA as an omnipotent biological molecule, generated major efforts in the last fifteen years (Gesteland et al., 1999), towards demonstrating whether indeed all the major cellular catalytic functions could be in principle performed by RNA. It was recently discovered, for example, that RNA may catalyze the synthesis of further nucleic acids building blocks (Unrau and Bartel, 1998). A comprehensive list of catalytic activities displayed by ribozymes, generated through *in vitro* RNA evolution (Ellington et al., 1997; Wright and Joyce, 1997), has been reported (Bartel and Unrau,

1999). In parallel to the interest in the historical path that has led to life, the RNA-world body of knowledge is often directed towards the attempt of engineering an artificial self-sustaining form of cellular life (Szostak et al., 2001). An engineered experimental realization of an RNA-based cell, however, would be unlikely to resemble the real protocells that once populated our planet. The spontaneous prebiotic emergence of a set of autocatalytic RNA polymers, in addition, encounters serious difficulties, mostly due to the implausibility of RNA synthesis and stability under prebiotic conditions (Bartel and Unrau, 1999; Niesert et al., 1981; Shapiro, 1984; Shapiro, 2000; Yarus, 1999).

2.7. ENSEMBLE REPRODUCTION

Despite the fact that the model of Eigen was constructed specifically for self-replicating nucleic acid polymers, the idea of describing molecular evolution with kinetic equations is quite general, and can be applied to cases of mutual catalysis rather than autocatalysis (Bagley and Farmer, 1991; Farmer et al., 1986; Segrè et al, 1998a; Segrè et al., 2000). The interpretation of variables and parameters in the two cases, though, is radically different. In Eigen's model the emphasis is on the autocatalytic replicating biopolymers, i.e. on the diagonal term of the matrix of interactions. The molecular interactions utilized in the compositional inheritance model (Bagley and Farmer, 1991; Segrè et al., 2000), on the other hand, do not require the ad hoc assumption of a self-replication capacity for individual molecules. The kinetic equations in this case are for catalyzed growth of entire molecular ensembles, due to the catalyzed aggregation and/or the formation of small organic compounds within non-covalent assemblies. The individual molecular species are not replicators, nor do they carry information. Both self-reproduction and inheritance are obtained as the emergent property of a collection of such molecules (Kauffman, 1986; Dyson, 1985). Interestingly, phenomena that resemble the quasi-species error threshold behavior can also observed for collective replication (Kauffman, 1993; Segrè et al., 2001a).

2.8. KAUFFMAN'S MODEL

This model (Kauffman, 1986) was one of the first attempts to formalize the idea that mutually catalytic metabolism-like networks of interacting molecules. It may lead to self-reproduction even if none of the individual molecules is a replicator (cf. McMullin, 1995; Oparin and Gladilin, 1980). Kauffman and colleagues demonstrated that a set of random polymers might undergo collective self-reproduction, provided that it entails sufficient diversity of molecular species (Farmer et al., 1986; Kauffman, 1986; Stadler et al., 1993). In Kauffman's words, this is an attempt to show that "self-reproduction and homeostasis, basic features of organisms, are natural collective expressions of polymer chemistry". According to this view, self-reproduction of a mixture of molecular species, exemplified by, but not restricted to oligomers, is caused by growth of the concentration of all the molecular species included. This is due to the mutual catalysis, or mutual rate acceleration, that each polymer exerts on the synthesis of others. Based on a graph theory derivation, Kauffman argues that any sufficiently complex set of molecules may attain the emergent collective property of catalytic closure, whereby every member of the set has at least one of the possible last steps in its formation catalyzed by some member of the set.

In the mathematical formulation of this model, monomers of different kinds can form polymers with length up to a maximal value M. A constant probability parameter P describes the chance that a molecular species in the set will catalyze any reaction leading to the formation of another species. 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 is shown then to increase linearly with M. When M is large enough, catalytic closure is realized and therefore each polymer in the set is efficiently synthesized. The outcome is that the whole set produces more of itself, despite the fact that none of the components of the set is assumed to be an autocatalyst. This concept of self-organizing metabolism-like network has been explored in different forms, including the study of systems of differential equations for the kinetics of polymers formation (Bagley and Farmer, 1991; Bagley et al., 1991) and the construction of an abstract artificial chemistry based on λ calculus (Fontana and Buss, 1994).

Kauffman's complexity-dependent closure reflects a widely diffused view that the emergence of life-like properties should resemble a phase transition, i.e. that the tuning of a fundamental parameter may determine in a critical way whether a system is in a state of randomness, or has undergone a sudden self-organization (Dyson, 1999). In the formalism of Eigen, this phase transition corresponds to the error threshold behavior described above. In Kauffman's model, the tuning parameter is the number of different mutually catalytic polymers, and the ensuing self-organization property is represented by catalyitc closure (Kauffman, 1986; Kauffman, 1993).

The self-reproduction behavior ensuing from a network of mutually catalytic interactions has been experimentally demonstrated using complementary nucleotide-based oligomers (Sievers and von-Kiedrowski, 1994), as well as an enhanced version of Ghadiri's self-replicating peptides (Kauffman, 1996; Lee et al., 1997; Lee et al., 1997). Both these experiments involve at most dual catalysis. More complex networks of mutually catalytic molecules are still missing experimental verification.

2.9. DYSON'S MODEL

While Kauffman champions the role of increasing diversity in leading to self-replication, Dyson's approach expresses the importance of a time-dependent process that leads to homeostasis. 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.

Dyson 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. Since Dyson's model includes a backward reaction, in which an active species may be inactivated, the exponential explosion typical of autocatalysis is prevented, and steady states are reached under certain conditions. Dyson shows that given the probability 1/a of specific catalytic activation (a is 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. 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.

The simplifications implied by Dyson's mean field approach are limiting the applicability of such model to reality. Nevertheless, the increasing available computational power may allow the implementation of simulations where explicit interactions between chemical species could be reproduced, as foreseen by Dyson himself. Dyson's point γ would then be substituted by a whole landscape of possible organized states, with different levels of complexity, stability and effectiveness in metabolizing external resources. The presence of multiple steady states may open also the way to a competition between different mutant assemblies, allowing evolutionary processes to occur. Interestingly, an experimental example of non-equilibrium bistability resembling Dyson's transition has been reported for mixed surfactant systems (Buhse et al., 1997)

It should be stressed that while in the first version of his model (Dyson, 1982; Dyson, 1985) Dyson depicted the monomers as linked into polymers that float in a vesicle-enclosed aqueous solution, in the 1999 edition (Dyson, 1999) the same membrane is assumed to enclose unbound monomers. Moreover, in the novel scenario, the catalyzed chemical reactions of activation and inactivation are identified with adsorption and desorption at sites of the vesicle rather than with changes in their position within polymers. This change of view reflects the concept of catalyzed amphiphilic monomers aggregation, which characterizes the Amphiphile-GARD model (Segrè and Lancet, 1998; Segrè and Lancet, 2000; Segrè et al., 2000).

2.10. THE MOROWITZ BOUNDARY

In models of collective self-reproduction, like the ones proposed by Dyson and Kauffman, the production of similar progeny is not evident as in the case of single molecule replication. If self-reproduction of a noncovalent assembly is to be obtained through a process of ongoing accretion or synthesis of molecular species through mutual catalysis, the generation of two or more entities out of a single one requires an additional step of physical separation. This has been classically described as a process of budding or splitting (Fox, 1976; Koch, 1985; Luisi et al., 1999; Morowitz, 1992; Morowitz et al., 1988). Under the assumption of randomly disposed molecules within an assembly, splitting may critically determine the fate of the progeny, i.e. whether the mixture of chemicals inherited by each daughter assembly may keep the catalytic network intact and guarantee continuing homeostasis. Statistical calculations may ascertain whether enough of each component is transferred to the progeny upon division. Morowitz shows how constraints on the transmissibility of information through direct inheritance of a molecular composition are related to the size of the assembly and the diversity of its molecular species (Morowitz, 1967). It is demonstrated that if an assembly contains m molecular types (the complexity of the assembly, or its intricacy (Smith and Morowitz, 1982)), 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. If, for example, m=4 and 2r=32, then 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 was 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 $P_d = 0.5$.

This concept may be crucial for determining specific ranges of parameters allowing the spontaneous emergence of protocellular structures capable of faithful collective self-reproduction (Segrè et al. 2000; Segrè et al., 2001a). In a description of membrane-enclosed assortments of genes, Szathmary proposed the "stochastic corrector model" for inheritance through protocell division, which is based on the same principle as above, but refers to a later evolutionary stage (Szathmary and Maynard Smith, 1997).

2.11. AUTOPOIESIS

An important demonstration that self-reproduction can be the property of a molecular assembly derives from the experiments done by Pier Luigi Luisi (Bachmann et al., 1992; Luisi et al., 1999). Inspired by the concept of autopoiesis (from the Greek 'auto' (self) and 'poiesis' (formation)), first proposed by Maturana and Varela (Luisi, 1997), Luisi demonstrated that micelles and vesicles can catalyze the production of their own constituents, and hence display a growth behavior akin to self-reproduction. Micelles and vesicles are self-assembled systems containing 10^3-10^6 amphiphilic monomers and forming thermodynamically stable physical structures through hydrophobic forces (Deamer, 1986; Morowitz et al., 1988; Tanford, 1978). The universal presence of structures like these in present day cells may indicate that lipid-like molecules played a major role during early evolution (Luisi et al., 1999).

In some of Luisi's experiments, long fatty acid chains were used as an amphiphile. A precursor of the surfactant (e.g. fatty acids anhydrides or esters) was initially deposited on a water solution, forming oil droplets floating on the surface. In a classical example, the hydrolysis of the ethylcaprylate precursors is responsible for the formation of caprylic acid vesicles (Bachmann et al., 1992). The hydrolysis is initially very slow, but its rate increases as vesicles are being formed. In other words, formed vesicles catalyze the production of further vesicle-forming components, in an autocatalytic reaction. Interestingly, it is the whole vesicle that acts as a catalyst rather than a single molecule, similarly to the cases of collective replication proposed by Kauffman and Dyson. Here too, single monomers are not replicators.

One of the drawbacks of Luisi's paradigm is the lack of diversity and therefore of evolution. Self-replicating micelles or vesicles would always look the same, unless different kinds of lipid monomers were introduced. Even if examples of mixed autocatalytic vesicles were designed (Bonaccio et al., 1994), the concept of compositional inheritance was not seriously taken into account, and the incorporation of self-replicating bio-polymers within self-reproducing vesicles was pursued as a paradigm for a minimal protocell (Oberholzer et al., 1995). In contrast, the recently proposed Lipid World hypothesis (Segrè et al., 2001) expands the autopoiesis concept to the realm of complex prebiotic

mixtures, and envisages lipid vesicles and micelles as possible carriers of compositional information.

2.12. A WORKING DEFINITION OF LIFE

Several formal definitions of the living state have been proposed (for a broad collection, see Lahav, 1999; Rizzotti, 1996), but no general agreement exists. This may be related to the fact that phenomenological descriptions of life's characteristics often refer to debated scenarios for the emergence of life. A good definition of life should describe a set that contains all the living systems we know, as well as any form of life possibly different from the one we know. This implies trying to distill, from our understanding of life, all those features that seem to be characteristic of the processes observed, without restrictions to specific molecular substrates. The latter may reflect "historical accidents" that happened along evolution. Without further reasoning on this matter, which has complex philosophical implications, one can formulate a working definition of life that is related to the problems under study. This will be helpful as a reference, since descriptions of life-like characteristics appear frequently in the models for prebiotic evolution presented here.

Life may be defined as an open chemical system far from thermodynamic equilibrium, whose linked reactions are organized in such a way that homeostasis and self-reproduction ensue. In more detail, the following features could be listed:

a) A living system is an open chemical system far from thermodynamic equilibrium. This involves the capacity of exchanging matter and energy with the surrounding environment. For a living system to emerge and perpetuate itself, an external energy source and an external energy sink must exist (Morowitz, 1979).

b) Each molecular species in the system (possibly an organic compound) is linked to other molecular species, forming a network of chemical transformations. The same molecular species may be involved in multiple functions within the reaction network and may be at the same time a substrate, a product and a catalyst for different reactions (Kauffman, 1986).

c) The network of chemical reactions and catalytic interactions displays a certain degree of organization. This may involve the presence of cycles or other ordered topologies (Morowitz, 1979).

d) From a functional perspective this organization implies two fundamental properties: the first is Homeostasis (Dyson, 1985), i.e. the capacity of the system to sustain itself, and maintain its internal order in spite of moderate fluctuations of environmental conditions. The second, intimately connected to the first one, is self-reproduction, i.e. the possibility to replenish molecular species that are diluted due to a growth in the total size and mass of the system. Upon a scission process, this maintenance of molecular concentrations upon growth ultimately leads to the duplication of the original system.

e) The self-reproduction process described above is, importantly, an imperfect one. The system is therefore prone to acquire mutation-like changes and may therefore evolve according to the Darwinian paradigm.

3. Random chemistry

What can be said about the interactions among organic molecules in a primordial mixture, without entering into details about their structures? Models for ensemble collective self-reproduction often utilize statistical assumptions, e.g. Kauffman's fixed probability P of catalysis between to molecules in a random set, or Dyson's mean field approximation for catalytic activity within a monomers' assembly. A deeper insight about the statistical properties of noncovalent interactions in large molecular ensembles re-quires a detour into the field of random and combinatorial chemistry.

3.1. PROBABILISTIC RECOGNITION

Many biological phenomena involve probabilistic recognition between randomly encountered molecules. This is especially manifest in biological repertoires, which have evolved to contain the molecular diversity necessary for binding any randomly encountered ligand with a functionally sufficient affinity. The immune repertoires, immunoglobulins and T-cell receptors, provide the most well known examples for systems displaying probability-based interactions. Other examples include the Multi Drug Resistance (MDR) proteins that underlie the cellular efflux of a large set of compounds (Bolhuis et al., 1997), the olfactory receptor repertoire, which recognizes diverse odorant molecules (Buck and Axel, 1991; Lancet, 1986; Lancet and Ben-Arie, 1993; Glusman et al., 2001), and biotransformation enzymes like cytochrome P450, which provide examples of the phenomenon of "probabilistic catalysis" (Cupp and Tracy, 1998). Probabilistic interactions are also found in catalytic antibodies (Janda et al., 1997; Schultz and Lerner, 1995).

Probability-based recognition is also at the core of the field of combinatorial chemistry. Here, ligand repertoires are used to find new binders for specific molecular targets (Collins, 1997; Hoogenboom, 1997; Lohse and Szostak, 1996; Lorsch and Szostak, 1994; Plunkett and Ellman, 1997; Scott and Smith, 1990). Within the realm of probabilistic recognition, the need for a formal depiction of the statistics that govern the interactions within ligand and receptor repertoires has brought to the concept of Affinity Distribution. This constitutes a frequency histogram for the affinities obtained when a single target is tested against numerous members within a repertoire. The existence of such an affinity distribution is widely recognized (Burnet, 1963; Inman, 1978; Kauvar et al., 1995; Lancet et al., 1993; Levitan, 1997; Levitan, 1998; Macken and Perelson, 1991; Richards, 1975; Vant-Hull et al., 1998). However, no general agreement exists on the specific functional shape of such distribution.

An appealing suggestion is that biological recognition between receptors and ligands obeys a simple, and very general, statistical law (Inman, 1978; Lancet et al., 1994a; Lancet et al., 1993; Rosenwald, 1998; Rosenwald et al., 2002). In other words, it is possible that a simple mathematical model could describe the affinity distribution for many different repertoire types, including receptor multi-gene families and combinatorial ligands. One of these statistical models, the Receptor Affinity Distribution (RAD) model (Lancet et al., 1993) suggests that the number of interactions L contributing to the energy of binding between a receptor and an arbitrary set of ligands is distributed binomially,

$$P(L) = \frac{B!}{L!(B-L)!}\left(\frac{1}{S}\right)^{L}\left(1-\frac{1}{S}\right)^{B-L} \qquad (4)$$

where B is the size of the binding site and S is the subsite diversity. The affinity of binding K is related to L through the following relation: $L = (RT/\alpha)\log(K)$ where α is the average energy contribution for a single subsite interaction, R is the gas constant and T is the absolute temperature. This distribution, in a Poisson approximation has recently been verified in reference to experimental data (Rosenwald et al., 2002).

3.2 QUANTITATIVE KINETICS OF MUTUAL CATALYSIS IN RANDOM ENSEMBLES

Besides their pharmaceutical importance for drug design, combinatorial chemistry tools have been used for addressing important questions about biocatalysis and self-replication in proteins (Moore et al., 1997) and nucleic acids (Gold et al., 1997). In the realm of origin of life, a statistical chemistry approach would view the random emergence of diverse organic molecules, and their assemblies, as a natural consequence of undirected prebiotic organosynthesis (Schwartz, 1996; Ehrenfreund and Charnley, 2000). It would then ask how life-like processes could emerge within such primordial random assortments (Cousins et al., 2000), rather than study their specific molecular content (e.g. for α-amino acids or nucleotide bases). One should not forget, however, that while in the traditional combinatorial chemistry studies one of the counterparts is usually a highly evolved biological macromolecule, introduced by the experimenter, prebiotic scenarios assume that both reactants are derived from the same combinatorial collection (Lancet et al., 1994b).

The importance of a probabilistic recognition approach to systems involving large numbers of random chemicals has been shown to extend beyond its initial scope. In previous models for prebiotic catalytic interactions (Chou et al., 1994; Dyson, 1982; Kauffman, 1986) the assumptions about unevolved molecular recognition processes was not based on explicit, biochemically rigorous results. This is where the RAD model can serve as a tool for exploring in a more accurate and quantitative way mutually catalytic sets (Segrè et al., 2001a; Shenhav et al., 2002).

The connection between RAD-governed binding affinity and catalytic activity in random collections of oligomers was explored previously (Lancet et al., 1994b). This expresses the fact that, since a fundamental step in catalytic reactions is the binding of the catalyst to the transition state, the extent of the catalytic enhancement can be assumed to distribute similarly to the binding affinity in a random repertoire.

The power of the statistical approach to mutually catalytic sets is greatly enhanced when combined with a quantitative kinetic analysis (Bagley and Farmer, 1991; Bagley et al., 1991; Wills 1997; Lancet 1994; Segrè et al., 2001a). This is essential for a rigorous demonstration that such sets may undergo self-replication. 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.

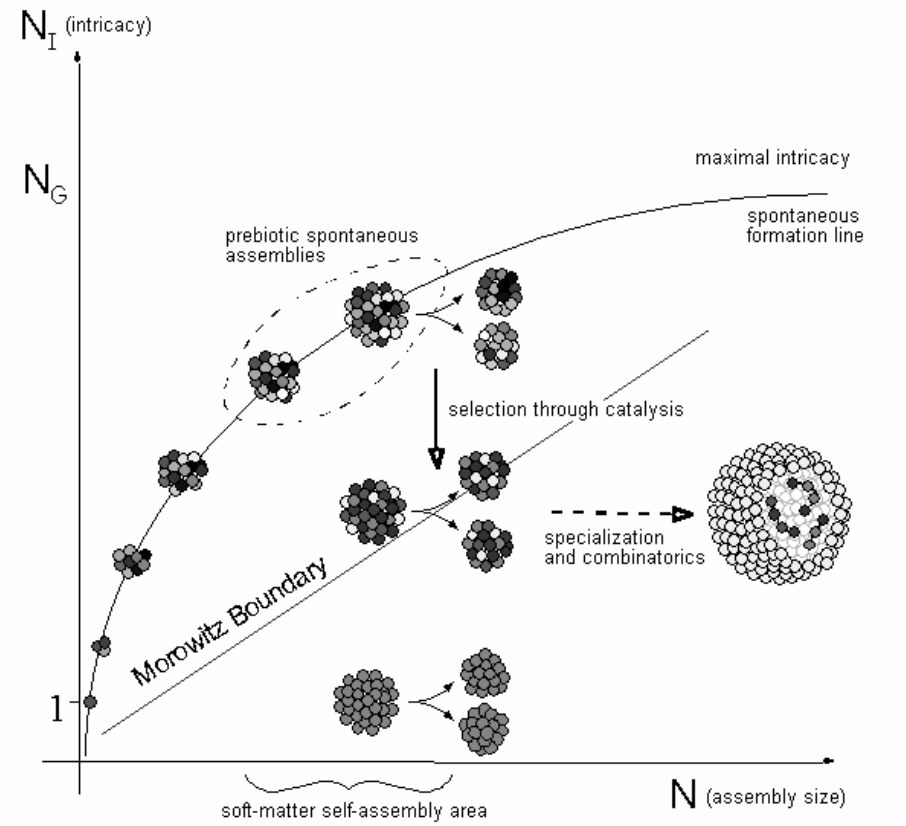

Figure 1 A schematic depiction of how assemblies with different Intricacy (number of different types of molecular species inside an assembly) and N (size, in total number of molecular components), can display different physical and self-organization properties relevant to the emergence of inheritance and self-replication (for more details, see Segrè and Lancet, 1999; Segrè and Lancet, 2000).

4. Accepting the twilight zone: the GARD model and the Lipid World

4.1. THE GARD "MANIFESTO"

Based on the aforementioned theoretical and experimental notions it is possible to delineate a set of guidelines for exploring theoretically the emergence of a primitive form of inheritance in the absence of biopolymers:

• Origin of life research can be viewed as the study of physico-chemical principles that were involved in primordial self-organization (Dyson, 1999; Kauffman, 1993; Oparin, 1957). The Graded Autocatalysis replication Domain (GARD) model (Segrè and Lancet, 1998; Segrè et al., 1998a) aims at distilling such principles and showing how life-like features could emerge based on testable assumptions. The exact details of the early his-

tory of life-as-we-know-it on Earth (Smith and Morowitz, 1982) are out of the scope of such approach.

• The central players of the model are molecules. At a minimal level of detail, the kinetic behaviors of ensembles of molecules are translated into differential equations based on the laws of mass action and average statistical properties. However, in some cases, the fluctuations related to stochastic effects of single molecule transformations become significant, as in the realm of mesoscopic physics, defining a high level of detail. The phenomena studied here - similarly to present-day cellular life - lie at the border between these two levels (Bartholomay, 1962; Bolhuis and Frenkel, 1997; Shenhav et al., 2002).

• A viable way of coping with the ignorance about the detailed composition of a primordial mixture is to use the statistical knowledge available about large repertoires of molecules, and their mutual interactions. In this respect, the emergence of life could be thought of as a planet-scale random chemistry experiment. The Receptor Affinity Distribution (RAD) model provides a statistical basis for modeling the probability of graded catalytic efficiencies among random molecules (Lancet et al., 1993, Rosenwald et al. 2002).

• Continuity of functions, rather than continuity of molecular structures, has been used as a criterion for depicting a self-organization scenario (Morowitz, 1992; Morowitz et al., 1988). Some simple features of modern cell biochemistry, such as the widespread prevalence of hydrophobic forces (Tanford, 1978), or the ubiquity of networks of weak catalytic effects (Altreuter and Clark, 1999; Cousins et al., 2000) could be readily obtained in an unevolved prebiotic scenario. Therefore such features should be included in a model for early evolution.

• A model aimed at explaining the emergence of self-reproduction need not include molecular autocatalysis and self-reproduction in its foundations. Rather, self-reproduction should be pursued as an emergent phenomenon (Dyson, 1999; Fontana and Buss, 1994; Kauffman, 1993; Morowitz et al., 1988; New and Pohorille, 2000; Oparin, 1957; Segrè and Lancet, 2000). This behavior may be derived from a large number of mutually catalytic interactions among different molecules, none of which is necessarily self-replicating. This view of self-reproduction requires a thorough quantitative description, similar to classical molecular autocatalysis (Eigen, 1971; Eigen and Schuster, 1979).

• Good candidate molecules for a chemical embodiment of early self-reproduction should (i) be among the possible products of abiotic synthesis, (ii) entail self-aggregation properties and soft-matter attributes, and (iii) possess the capacity to form spatially confined domains potentially capable of growing and dividing. Lipid-like amphiphilic molecules seem the best organic candidates. However, these could be rather different from present-day lipids, and could have diverse chemical properties, particularly in their polar head group (Segrè et al., 2001).

4.2. THE GARD SCENARIO

The GARD model was introduced in four fundamental forms: a) the Monomer-GARD, in which N_G kinds of monomeric species A_j can inter-convert through a common precursor molecule A_0. Molecular species are enclosed in a compartment, or vesicle; b) the Dimer-GARD, in which $N_G=F^2$ kinds of dimeric species D_{ij} are formed from F different monomeric precursors M_i; c) the Amphiphile-GARD, in which N_G kinds of amphiphilic

monomers participate in reactions of spontaneous aggregation into noncovalent assemblies, such as micelles and vesicles. In this case the enclosure and the set of mutually catalytic molecules are the same entity; and d) the Chiral-GARD, for which the N_G Monomer-GARD molecular species have a chiral center, and can appear in two enantiomeric forms (A_{jL} and A_{jD}) (Kafri, 2002).

4.2.1. Monomer GARD

The Monomer-GARD model provides a framework for a thermodynamic and kinetic analysis of mutually catalytic assemblies combined with statistical tools. Similarly to other scenarios (Dyson, 1999; Eigen and Schuster, 1979; Kauffman, 1993) it assumes a finite enclosure (e.g. an amphiphilic vesicle), containing the catalytic set members, and absorbing energy-rich chemical precursors ("foodstuff") from the external environment, represented by a single species A_0. GARD's chemical and physical rules are computationally implemented by numerical solution of differential equations or by Monte Carlo simulation procedures. The flow of free energy that maintains the system far from thermodynamic equilibrium involves a volume expansion, combined with the supply of high free energy precursor molecules. The dynamics of the molecular species is dominated by the β matrix of kinetic parameters for mutual catalysis, rather than by the thermodynamic equilibrium constants. Despite the simplicity of its reaction topology, the Monomer-GARD model allows one to quantify the homeostatic capacity of molecular ensembles and relate it to the microscopic kinetic properties of intermolecular interactions. 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. Such preservation is a basic hallmark of life forms, and a prerequisite for more elaborate characteristics, such as reproduction and selection. The constant supply of free energy-rich precursors, together with the process of vesicle expansion, maintain the GARD system far from thermodynamic equilibrium. This is somewhat analogous to Eigen's assumption in the quasi-species scenario, involving a combination of a monomers inflow and a polymers outflow. This property 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. The λ_C measure introduces for the first time a quantitative, graded scale for the efficiency of a catalytic network. The strength of this criterion for measuring homeostatic efficiency is demonstrated by its capacity to lead to complex network self-organization, when used as a fitness parameter for stochastic simulations of GARD evolution. In such simulations, based on a Metropolis algorithm for mutation and selection, stable compositions are obtained, which are resistant throughout millions of mutation steps, and possess highly organized topologies of catalytic interconnections. The evolutionary time curve of the fitness parameter λ_C displays long plateaus of invariant states, alternated to burst of changes and avalanches of reorganization steps. The distribution of the lengths of the plateaus is shown to obey a power law with exponent –1, possibly revealing a self-organized critical behavior (Bak and Sneppen, 1993; Bak et al., 1987), or other kinds of known self-organization dynamics (Newman and Sneppen, 1996).

The GARD formalism does not describe a specific molecular scenario, and may therefore apply equally well to different kinds of chemistries (e.g. peptides, nucleotides or

amphiphiles) and reaction schemes (isomerizations, covalent dimerization or noncovalent associations, as formulated in the different versions of GARD). In the context of Kauffman's definition of catalytic closure (Kauffman, 1986), GARD is formally always catalytically closed, but it is possible to define within it realms, through the λ_C parameter, the *degree* of catalytic closure for every conceivable assembly. Scenarios for prebiotic mutually catalytic sets (Kauffman, 1993, Dyson, 1982) had only included a qualitative discussion of the link between the catalyzed replenishment of a molecular ensemble subject to dilution, and the self-reproduction potentially ensuing. Our work has provided for the first time a detailed mathematical and computational description of this principle.

4.2.2. Dimer-GARD

The generation of diversity through combinatorics is one of the basic features of the chemistry of life. In the Monomer-GARD model (as well as in the Amphiphile-GARD), the diversity of monomer kinds is an assumption rather than a result, and further diversity is obtained at the level of molecular ensembles, through the combinatorics of different compositions. The fact that these compositions can be inherited constitutes one of the main implications of this part of my work. The Dimer-GARD goes one step further, by incorporating the covalent combinatorial diversity of dimers within the setup of a composition-based assembly.

One of the exciting prospects of the GARD view is the possibility of exploring a stepwise transition from such primordial form of homeostasis and implicit inheritance, to more advanced systems, in which the initially available monomers can bind covalently to each other to form longer and longer polymers. The Dimer-GARD constitutes a first step in this direction. Dimer-GARD was shown to be formally equivalent to Monomer-GARD, and therefore to share with it the property of homeostatic preservation during growth. It is expected that collections of even longer oligomers would behave similarly, and that more powerful computer simulations will allow to track in a graded fashion the transition from assemblies of monomers to complex mixtures containing longer polymers with specialized catalytic roles. It should be stressed that in an alternative interpretation, the transition from Monomeric to Dimeric GARD may be viewed as a progression from an absolutely heterotrophic to a somehow autotrophic organism (Maden, 1995), namely from a system which only recruits available components from its surrounding to one that performs covalent reactions for generating some of its constituents (Ben-Eli, 2000; Shenhav et al., 2003).

4.2.3. Amphiphile-GARD

Through the Amphiphile-GARD (A-GARD, Segrè and Lancet, 1998; Segrè et al. 1998; Segrè et al. 2000) a self-consistent and comprehensive scenario for the emergence of primordial self-reproduction is proposed. This brings together for the first time the concept of mutually catalytic networks (Dyson, 1982; Kauffman, 1986; Morowitz et al., 1988), the fundamental role of self-assembling molecules (Bachmann et al., 1992; Deamer, 1986), a mathematical formulation of chemical evolution through chemical kinetics equations (Eigen, 1971), and the stochastic nature of mesoscopic processes (Bartholomay, 1962; Gillespie, 1977). This synthesis is mainly due to the following innova-

tive features with respect to the Monomer- and Dimer- GARD: (i) A natural process of assembly growth is introduced. In the first two GARD embodiments, as well as in other models, the coupling between the potential growth of a vesicle-like enclosure and the internal catalytic activity is not clearly defined. In Amphiphile-GARD it is importantly assumed that the same molecules that form the self-enclosed 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. (ii) Amphiphile assemblies have a facile capacity to undergo spontaneous, physically-dictated splitting, generating primitive progeny. The homeostasis ensuing from steady concentration of mutually catalytic molecular species within an expanding boundary, nicely devised by Dyson and Kauffman, is subject, in A-GARD, to the splitting events, which are chance-regulated sampling processes. This allows a natural definition of inheritance, which is only implicit in the previous GARD versions. One may ask whether a given assembly is "robust" enough to survive and keep propagating its composition despite the fluctuations due to such splitting events. The Morowitz boundary formula (Morowitz, 1967) (see 2.10 above) provides an approximate answer to this question. In addition, the newly defined Heritability measure η (Segrè et al. 2001a) represents a more advanced tool, which allows one to evaluate the probability of reaching different degrees of fidelity in compositional inheritance fidelity during the sampling process. This allows one also to express in a rigorous way the fact that an assembly, whose composition vector is highly biased, has higher probability of a faithful transmission of its molecular components. (iii) In A-GARD equations and simulations, an assembly is described through the counts of molecules of different types, rather than through their concentrations. In this setup, the kinetics of the catalyzed self-assembly process is governed by a stochastic single-molecule discrete approximation of the Monomer-GARD differential equations. Therefore the compositional mutations, which in other settings (including the evolutionary process for Monomer-GARD) are ad hoc externally imposed discontinuous events, become natural consequences of the dynamics of the system. This results in a situation where the dynamic model for primordial evolution may begin in a condition of near-randomness, with compositional mutations being the rule rather than the exception, and end up in a stationary state characterized by a relatively small mutation rate. (iv) A-GARD simulations provide a pragmatic way of addressing the propagation of information along generations of growing and splitting assemblies. A new definition for assembly similarity H (Segrè et al. 2000) is employed to follow both the changes in an assembly's composition with respect to its past configurations (cross correlation), and the resemblance to reference compositions, such as the eigenvectors of the linear operator used for the linearized GARD equations. In the observed quasi-stationary states (QSSs) assemblies behave homeostatically despite the on-going growth and splitting processes. In most cases these states have a finite lifetime, and decay spontaneously into other states, as a consequence of large compositional fluctuations. In other cases the stationary states seem extremely stable. When homeostasis prevails for numerous A-GARD cycles we say that a rudimentary form of inheritance is taking place. What is inherited is the specific composition of the assembly, which may be regarded as a *Compositional Genome*. The different QSSs observed during A-GARD simulations correspond to different assortments that have the capacity of being propagated faithfully for a detectable number of generations. Such different states are called Composomes.

These features of A-GARD allow one to tackle profound questions about the properties that characterize prebiotic self-reproducing assemblies. Populations of assemblies may be simulated, so that different composomes can compete and give rise to a compositional genome-based evolutionary process. We envisage that in simulations with larger N_G a process of Darwinian evolution could take place under these conditions.

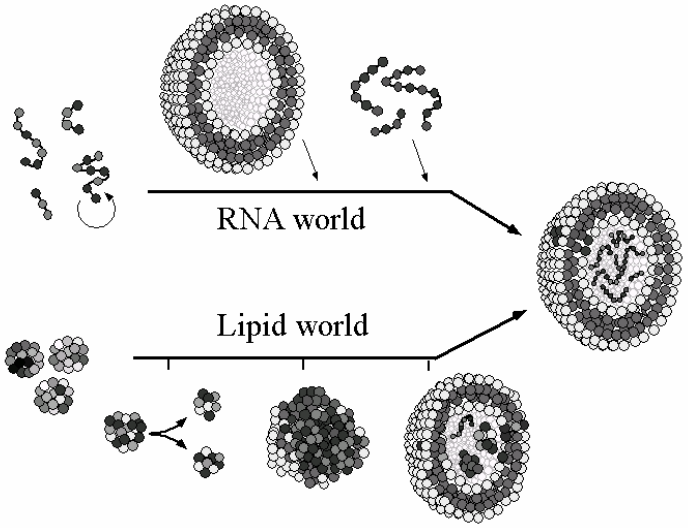

Figure 2 Comparison of two possible views for the emergence of a reproducing protocell: The RNA world (see chapter 2.6) and the Lipid World (chapter 4.3).

4.3. THE LIPID WORLD

The experimental evidence that lipid assemblies may entail many of the properties required by the A-GARD model led to the formulation of a novel scenario for the origin of life, "The Lipid World" (Segrè et al. 2001). As part of the Lipid World view, several novel concepts were proposed as fundamental unexplored aspects of early prebiotic evolution. One notion relates to the possible combinatorial nature of lipid-like chemicals. Thousands of different configurations can be achieved with minimal variations on phospholipid-like structures. The Lipid Word View is compatible with the finding that molecular species found in the Murchison meteorite display aggregation properties in water (Deamer, 1985), as well as with experimental proofs that amphiphiles could form under likely pre-biotic conditions (McCollom et al., 1999).

In analogy to the concept of ribozyme (Lohse and Szostak, 1996), emphasizing the catalytic role of RNA, we coined the term lipozyme (Segrè et al. 2001), by which we indicate any noncovalent assembly, whose components exert catalytic action on any chemical reaction. An autocatalytic lipozymes represents a specific case that can synthesize more of its constituents. It should be noted that while this concept is compatible with the idea of mutually catalytic network, it does not involve necessarily autocatalytic molecules. This is supported by statistical analysis of a large amount of data for micelle and

liposome catalysis, extracted from a list compiled by Janos Fendler's (Fendler and Fendler, 1975; Fendler, 1982; Segrè et al. 2001). It was found that the distribution of catalytic rate enhancement factors obeys the lognormal shape predicted by the RAD model (Lancet et al., 1993, Rosenwald et al. 2002) and used throughout GARD simulations.

The fluid nature of lipozymes may be responsible for the formation of transient efficient catalytic sites on a micelle or membrane, due to random encounter of molecules freely diffusing on the interface (Fendler, 1987; Schneider-Henriquez et al., 1992). This might be an intermediate step towards the formation of covalent bonds, which would entail more specific and efficient, but less flexible catalytic functions. Tests of some of these ideas could derive from extensions of Luisi's experiments with autocatalytic micelles and vesicles (autopoietic units (Walde et al., 1994)), whereby many different molecular kinds could be used. Technical problems, especially related to the detection of single assemblies compositions, need to be overcome in order for such an experiment to be useful.

4.4. FUTURE PROSPECTS

An understanding of the evolutionary significance of the idea of a primordial inheritance of a compositional genome requires a thorough description of large heterogeneous populations or assemblies of interacting molecules. GARD and other simulations mostly embodied up to N_G=100 different types of molecular species. Realistic prebiotic mixtures, in contrast, may contain a much larger number of different molecules, in the range of 10^6-10^9 (cf. Morowitz, 1992). Larger N_G values are expected to generate a remarkable assortment of different meta-stable compositions, many of which might arise only under certain sets of initial conditions, or as a consequence of very rare compositional mutations. Large N_G values, however, impose serious computational limitations. For example, the β matrix for catalytic enhancement factors grows quadratically with N_G.

Two main possible approaches can be envisaged for these simulations. The first, and so far most productive, has been the one based on a fixed number (N_G) of chemical kinetics differential equations. This includes GARD kinetics, stochastic Amphiphile-GARD simulations, and many other kinetic models of mutually catalytic networks. In a second possible approach the repertoire of molecular species available is potentially unlimited. This is because the composition of an assembly is represented by a complete list of its molecular components, and no constraints exist on the number of possible kinds. Such an implementation requires however a consistent description of catalytic rate acceleration factors for any possible molecular interaction. This could be obtained by defining a set of fundamental recognition rules for a given set of N_G monomers, and implementing a universal catalytic interaction function, based on string matching rules analogous to the RAD formalism (Bagley and Farmer, 1991; Shenhav et al., 2003), for computing the expected catalytic rate exerted between oligomers.

A new generation of computer simulations, based on stochastic dynamics rather than on rigidly defined differential equations, should be designed in order to cope with such an unlimited number of different possible molecular species. One problem with such an embodiment is that small molecular assemblies might wander indefinitely in the unlimited compositional space without ever falling into a stationary state, and without displaying

any visible sign of organization. Only relatively large assemblies may be expected to contain a sufficient number of molecular components, in accordance with the optimal N/N_G ratio (Segrè et al. 2001a).

Extended computations should also include more detailed physico-chemical features, and conform accurately to standard physical chemistry rules for self-assembly, as previously explored in different contexts (Goetz and Lipowsky, 1998; Nekovee et al., 2000). An ultimate GARD setting would be represented by a mesoscopic-scale molecular dynamics simulation (Shenhav et al., 2002), in which membrane and polymer formation would be the consequences of elementary interaction rules. A richer description of the dynamics of GARD populations, involving splitting and fusion of assemblies of different sizes, may pave the way to a novel interpretation of early prebiotic evolution phenomena, whereby single molecule compositional mutations could be extended to more elaborate methods for the generation of compositional diversity, in analogy to the role of genetic recombination.

Under these circumstances, it is conceivable that large-scale supercomputer simulations, analogous to the ones currently utilized for astrophysics and particle physics realistic simulations (Norman, 1998) may in the future open new horizons in our understanding of early molecular evolution. One should bear in mind that such large-scale simulations would produce an enormous amount of complex data, whose interpretation would require a major theoretical effort, and perhaps the invention of new computational techniques, a counterpart of the experimental ones used by biologists for deciphering the real biochemical world. It is becoming also increasingly apparent that interesting connections exist between the study of present-day biochemical networks (Segrè et al., 2002) and the mutually catalytic networks explored in origin of life research (Shenhav et al., 2002). The study of prebiotic evolution and the rising discipline of Systems Biology, both involved in an effort to uncover general principles of biological organization, may end up displaying deep and perhaps yet unknown links.

5. References

Allamandola, L.J., Hudgins, D.M., Bauschlicher Jr., C.W. and Langhoff, S.R. (1999) Carbon chain abundance in the diffuse interstellar medium. Astron. Astrophys., 352, 659-664.

Altreuter, D.H. and Clark, D.S. (1999) Combinatorial biocatalysis: taking the lead from nature. Curr. Opin. Biotechnol., 10, 130-136.

Alves, D. and Fontanari, J.F. (1998) Error threshold in finite populations. Phys. Rev. E, 57, 7008-7013.

Anders, E. (1989) Prebiotic organic matter from comets and asteroids. Nature, 342, 255-257.

Arnett, D. (1996) Supernovae and nucleosynthesis. Princeton Univ. Press, Princeton, NJ.

Arnett, D. and Bazan, G. (1997) Nucleosynthesis in stars: recent developments. Science, 276, 1359-1362.

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

Bagley, R.J. and Farmer, J.D. (1991) Spontaneous emergence of a metabolism. In: C.G. Langton, C. Taylor, J.D. Farmer and S. Rasmussen (eds.), Artificial Life II. Addison-Wesley, Vol. X, pp. 93-140.

Bagley, R.J., Farmer, J.D. and Fontana, W. (1991) Evolution of a metabolism. In: C.G. Langton, C. Taylor, J.D. Farmer and S. Rasmussen (eds.), Artificial Life II. Addison-Wesley, Vol. X, pp. 141-158.

Bak, P. and Sneppen, K. (1993) Punctuated equilibrium and criticality in a simple model of evolution, Phys. Rev. Lett., 71, 4083-4086.

Bak, P., Tang, C. and Wiesenfeld, K. (1987) Self-Organized Criticality: An explanation of 1/f Noise, Phys. Rev. Lett., 59, 381-384.

Bartel, D.P. and Unrau, P.J. (1999) Constructing an RNA world. TIBS (Millennium Issue), 24, M9-M13.

Bartholomay, A.F. (1962) A stochastic approach to statistical kinetics with application to enzyme kinetics. Biochemistry, 1, 223-230.
Ben-Eli, D. (2000) M.Sc. Thesis. The Weizmann Institute of Science, Rehovot, Israel.
Bernstein, M.P., Sandford, S.A., Allamandola, L.J., Gilette, J.S., Clemett, S.J. and Zare, R.N. (1999) UV irradiation of polycyclic aromatic hydrocarbons in ices: production of alcohols, quinones, and ethers. Science, 283, 1135-1138.
Bolhuis, H., van, V.H., Poolman, B., Driessen, A.J. and Konings, W.N. (1997) Mechanisms of multidrug transporters. Fems Microbiol. Rev., 21, 55-84.
Bolhuis, P.G. and Frenkel, D. (1997) Microscopic and mesoscopic simulation of entropic micelles. Physica A, 244, 45-58.
Bolli, M., Micura, R. and Eschenmoser, A. (1997) Pyranosyl-RNA: chiroselective self-assembly of base sequences by ligative oligomerization of tetranucleotide-2',3'-cyclophosphates (with a commentary concerning the origin of biomolecular homochirality). Chemistry & Biology, 4, 309-320.
Bonaccio, S., Walde, P. and Luisi, P.L. (1994) Liposomes containing purine and pyrimidine bases: stable unilamellar liposomes from phosphatidyl nucleosides. J. Phys. Chem., 98, 6661-6663.
Buck, L. and Axel, R. (1991) A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. Cell, 65, 175-187.
Buhse, T., Pimienta, V., Lavabre, D. and Micheau, J.-C. (1997) Experimental evidence of kinetic bistability in a biphasic surfactant system. J. Phys. Chem. A, 101, 5215-5217.
Burnet, F.M. (1963) Cold Spring Harbor Symp. Quant. Biol., 32, 1.
Cairns-Smith, G. (1982) Genetic takeover and the mineral origins of life. Cambridge University Press, Cambridge, UK.
Cech, T.R. (1986) A model for the RNA-catalyzed replication of RNA. Proc. Natl. Acad. Sci. U S A, 83, 4360-4363.
Chou, H.-H., Reggia, J.A., Navarro-Gonzalez, R. and Wu, J. (1994) An extended cellular space method for simulating autocatalytic oligonucleotides. Computers Chem., 18, 33-43.
Chyba, C.F. and Sagan, C. (1992) Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: An inventory for the origin of life. Nature, 355, 125-132.
Collins, J. (1997) Phage display. In: A.D. Ellington, W.H. Moos, M.R. Pavia and B.K. Kay (eds.), Annual Reports in Combinatorial Chemistry and Molecular Diversity. ESCOM, The Netherlands, Vol. 1, pp. 210-262.
Cousins, G.R., Poulsen, S.A. and Sanders, J.K. (2000) Molecular evolution: dynamic combinatorial libraries, autocatalytic networks and the quest for molecular function. Curr. Opin. Chem. Biol., 4, 270-279.
Crick, F. (1981) Life itself. Simon and Schuster, New York.
Cupp, M.J. and Tracy, T.S. (1998) Cytochrome P450: new nomenclature and clinical implications. Am. Fam. Physician, 57, 107-16.
Dawkins, R. (1996) The River out of Eden. Phoenix Press.
de Duve, C. (1995) Vital Dust. Harper Collins Publishers.
de Graaf, R.M., Visscher, J. and Schwartz, A.W. (1995) A plausibly prebiotic synthesis of phosphonic acids. Nature, 378, 474-477.
Deamer, D.W. (1985) Boundary structures are formed by organic components of the Murchison carbonaceous chondrite. Nature, 317, 792-794.
Deamer, D.W. (1986) Role of amphiphilic compounds in the evolution of membrane structure on the early earth. Origins of Life, 17, 3-25.
Doty, S.D. and Leung, C.M. (1998) Detailed chemical modeling of the circumstellar envelopes of carbon stars: application to IRC +10216. The Astrophys. J., 502, 898-908.
Dyson, F. (1999) Origins of Life. Cambridge University Press, Cambridge.
Dyson, F.J. (1982) A model for the origin of life. J. Mol. Evol., 18, 344-350.
Dyson, F.J. (1985) Origins of life. Cambridge University Press, Cambridge.
Ehrenfreund, P. and Charnley, B. (2000) Organic molecules in the interstellar medium, comets and meteorites: a voyage from dark clouds to the early Earth. Annu. Rev. Astron. Astrophys., 38, 427-483.
Eigen, M. (1971) Selforganization of matter and the evolution of biological macromolecules. Naturwissenschaften, 58, 465-523.
Eigen, M. (2000) Natural selection: a phase transition? Biophys. Chem., 25, 101-123.
Eigen, M. and Schuster, P. (1979) The Hypercycle. Springer Verlag, Berlin.
Ellington, A.D., Robertson, M.P. and Bull, J. (1997) Ribozymes in Wonderland. Science, 276, 546-547.

Farmer, J.D., Kauffman, S.A. and Packard, N.H. (1986) Autocatalytic replication of polymers. Physica, 22D, 50-67.
Fendler, H.J. and Fendler, E.J. (1975) Catalysis in Micellar and Macromolecular Systems. Academic Press, New York.
Fendler, J.H. (1982) Membrane Mimetic chemistry. Wiley, New York.
Fendler, J.H. (1987) Atomic and Molecular Clusters in Membrane Mimetic Chemistry. Chem. Rev., 87, 877-899.
Fontana, W. and Buss, L.W. (1994) What would be conserved if "the tape were played twice"? Proc. Natl. Acad. Sci. USA., 91, 757-761.
Fox, S.W. (1976) The evolutionary significance of phase-separated microsystems. Origin Life Evol. Biosphere, 7, 49-68.
Fox, S.W. (1991) Synthesis of life in the lab? Defining a protoliving system. Q. Rev. Biol., 66, 181-5.
Fraser, C.L. and Folsome, C.E. (1975) Exponential kinetics of formation or organic microstructures. Orig. Life, 6, 429-33.
Fry, I. (2000) The Emergence of Life on Earth : A Historical and Scientific Overview. Rutgers Univ. Press.
Gesteland, R.F., Cech, T.R. and Atkins, J.F. (1999) The RNA world, Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
Gilbert, W. (1986) The RNA world. Nature, 319, 618-618.
Gillespie, D. (1977) Exact stochastic simulation of coupled chemical reactions. J. Phys. Chem., 81, 2340-2360.
Glusman G., Yanai I., Rubin I., Lancet D. (2001) The complete human olfactory subgenome. Genome Res. 2001, 11, 685-702.
Goetz, R. and Lipowsky, R. (1998) Computer simulations of bilayer membranes: Self-assembly and interfacial tension. J. Chem. Phys., 108, 7397-7409.
Gold, L., Brown, D., He, Y.-Y., Shtatland, T., Singer, B.S. and Wu, Y. (1997) From oligonucleotide shapes to genomic SELEX: Novel biological regulatory loops. Proc. Natl. Acad. Sci. USA, 94, 59-64.
Hoogenboom, H.R. (1997) Designing and optimizing library selection strategies for generating high-affinity antibodies. Trends Biotechnol., 15, 62-70.
Horowitz, N.H. (1945) On the evolution of biochemical syntheses. Proc. Natl. Acad. Sci. USA, 31, 153-157.
Huber, C. and Wachtershauser, G. (1997) Activated acetic acid by carbon fixation on (Fe, Ni)S under primordial conditions. Science, 276, 245-247.
Inman, J.K. (1978) The antibody combining region: Speculations on the hypothesis of general multispecificity. In: G.I. Bell, A.S. Perelson and G.H.Jr. Pimbly (eds.), Theoretical Immunology. Marcell Dekker, New York, pp. 243-278.
Irvine, W.M. (1998) Extraterrestrial organic matter: a review. Origins Life Evol. Biosphere, 28, 365-383.
Jain, S. and Krishna, S. (1998) Autocatalytic sets and the growth of complexity in an evolutionary model. Phys. Rev. Lett., 81, 5684-5687.
Janda, K.D., Lo, L.C., Lo, C., Sim, M.M., Wang, R., Wong, C.H. and Lerner, R.A. (1997) Chemical selection for catalysis in combinatorial antibody libraries. Science, 275, 945-8.
Joyce, G.F. (1991) The rise and fall of the RNA world. The New Biologist, 3, 399-407.
Kafri, R. (2002) M.Sc. Thesis, Weizmann Institute of Science, Israel.
Kauffman, S. (1996) Even peptides do it. Nature, 382, 496-497.
Kauffman, S.A. (1986) Autocatalytic sets of proteins. J. Theor. Biol., 119, 1-24.
Kauffman, S.A. (1993) The origins of order, Oxford University Press, Oxford.
Kauvar, L.M., Higgins, D.L., Villar, H.O., Sportsman, J.R., Engqvist-Goldstein, A., Bukar, R., Bauer, K.E., Dilley, H. and Rocke, D.M. (1995) Predicting ligand binding to proteins by affinity fingerprinting. Chem. Biol., 2, 107-18.
Kerr, R.A. (1996) Ancient life on Mars? Science, 273, 864-866.
Koch, A.L. (1985) Primeval cells: possible energy-generating and cell-division mechanisms. J. Mol. Evol., 21, 270-277.
Kondepudi, D.K. and Nelson, G.W. (1985) Weak neutral currents and the origin of bio-molecular chirality. Nature, 314, 438-441.
Kuppers, B.-O. (1983) Molecular theory of evolution. Springer-Verlag, Berlin-Heidelberg.
Lahav, N. (1999) Biogenesis: Theories of Life's origin. Oxford University Press, NY.
Lancet, D. (1986) Vertebrate olfactory reception. Annu. Rev. Neurosci., 9, 329-55.
Lancet, D. and Ben-Arie, N. (1993) Olfactory receptors. Current Biology, 3, 668-674.

Lancet, D., Horovitz, A. and Katchalski-Katzir, E. (1994a) Molecular recognition in biology: Models for analysis of protein/ligand interactions. In: J.-P. Behr (ed.), Perspectives in supramolecular chemistry. J. Wiley New-York, pp. 25-71.

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

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

Lee, D.H., Granja, J.R., Martinez, J.A., Severin, K. and Ghadiri, M.R. (1996) A self-replicating peptide. Nature, 382, 525-528.

Lee, D.H., Severin, K., Yokobayashi, Y. and Ghadiri, M.R. (1997) Emergence of symbiosis in peptide self-replication through a hypercyclic network. Nature, 390, 591-594.

Levitan, B. (1997) Models and search strategies for applied molecular evolution. In: A.D. Ellington, W.H. Moos, M.R. Pavia and B.K. Kay (eds.), Annual Reports in Combinatorial Chemistry and Molecular Diversity. ESCOM, The Netherlands, Vol. 1, pp. 95-152.

Levitan, B. (1998) Stochastic Modeling and Optimization of Phage Display. J. Mol. Biol., 277, 893-916.

Li, T. and Nicolaou, K.C. (1994) Chemical self-replication of palindromic duplex DNA. Nature, 369, 218-221.

Li, W.-H. and Graur, D. (1991) Fundamentals of Molecular Evolution, Sinauer Associates, Sunderland, MA.

Lifson, S. (1996) On the crucial stages in the origin of animate matter. J. Mol. Evol., 44, 1-8.

Lissauer, J.J., Marcy, G.W. and Ida, S. (2000) Extrasolar planets. Proc. Natl. Acad. Sci. USA, 97, 12405-12406.

Lohse, P.A. and Szostak, J.W. (1996) Ribozyme-catalysed amino-acid transfer reactions. Nature, 381, 442-444.

Lorsch, J.R. and Szostak, J.W. (1994) In vitro selection of RNA aptamers specific for cyanocobalamin. Biochemistry, 33, 973-982.

Luisi, P.L. (1997) Self-reproduction of chemical structures and the question of the transition to life. In: C.B. Cosmovici, S. Bowyer and D. Werthimer (eds.), Astronomical and Biochemical Origins and the Search for Life in the Universe, Editrice Compositori, Bologna, pp. 461-468.

Luisi, P.L., Walde, P. and Oberholzer, T. (1999) Lipid vesicles as possible intermediates in the origin of life. Curr. Opinions in Colloid & Interface Sci., 4, 33-39.

Macià, E., Hernandez, M.V. and Oró, J. (1997) Primary sources of phosphorus and phosphates in chemical evolution. Origins Life Evol. Biosphere, 27, 459-480.

Macken, C.A. and Perelson, A.S. (1991) Affinity maturation on rugged landscapes. In: A.S. Perelson and S.A. Kauffman (eds.), Molecular Evolution on Rugged Landscapes: Santa Fe Institute Studies in the Sciences of Complexity. Addison-Wesley, Reading, MA, Vol. 9, pp. 93-118.

Maden, B.E. (1995) No soup for starters? Autotrophy and the origins of metabolism. TIBS, 20, 337-341.

Maher, K.A. and Stevenson, J.D. (1988) Impact frustration of the origin of life. Nature, 331, 612-614.

Mason, S.F. (1984) Origins of biomolecular handedness. Nature, 311, 19-23.

McCollom, T.W., Ritter, G. and Simoneit, B.R.T. (1999) Lipid synthesis under hydrothermal conditions by Fisher-Tropsch-type reactions. Origins Life Evol. Biosphere, 29, 153-166.

McDonald, G.D., Whited, L.J., DeRuiter, C., Khare, B.N., Patnaik, A. and Sagan, C. (1996) Production and chemical analysis of cometary ice tholins. Icarus, 122, 107-117.

McKay, C.P. (1997) Life in comets. In: P.J. Thomas, C.F. Chyba and C.P. McKay (eds.), Comets and the origin and evolution of life. Springer Verlag, New York, pp. 273-282.

McKay, D.S., Gibson Jr., E.K., Thomas-Keprta, K.L., Vali, H., Romanek, C.S., Clemett, S.J., Chillier, X.D.F., Maechling, C.R. and Zare, R.N. (1996) Search for past life on Mars: possible relic biogenic activity in Martian meteorite ALH84001. Science, 273, 924-930.

McMullin, B. (1995) Replicators Don't! In: F. Moran, A. Moreno, J.J. Merelo and P. Cachon (eds.), Advances in Artificial Life. Springer, Granada, Spain, pp. 158-169.

Miller, S.L. (1953) A production of amino acids under possible primitive earth conditions. Science, 117, 528-529.

Miller, S.L. and Urey, H.C. (1959) Organic compound synthesis on the primitive Earth. Science, 130, 245-251.

Miyakawa S., Yamanashi H., Kobayashi K., Cleaves H.J. and Miller SL. (2002) Prebiotic synthesis from CO atmospheres: implications for the origins of life. Proc. Natl. Acad. Sci. U S A, 99, 14628-14631.

Mojzsis, S.J., Arrhenius, G., McKeegan, K.D., Harrison, T.M., Nutman, A.P. and Friend, C.R. (1996) Evidence for life on Earth before 3,800 million years ago. Nature, 384, 55-59.

Mojzsis, S.J., Krishnamurthy, R. and Arrhenius, G. (1998) Before RNA and After: Geophysical and Geochemical Constraints on Molecular Evolution. In: R.F. Gesteland, T.R. Cech and J.F. Atkins (eds.), The RNA world, Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, pp. 1-47.

Moore, J.C., Jin, H.-M., Kuchner, O. and Arnold, F.H. (1997) Strategies for the in vitro evolution of protein function: enzyme evolution by random recombination of improved sequences. J. Mol. Biol., 272, 336-347.
Morowitz, H.J. (1967) Biological self-replicating systems. In: F.M. Snell (ed.), Progress in Theoretical Biology. Academic Press, Vol. 1, pp. 35-58.
Morowitz, H.J. (1979) Energy flow in biology. Academic Press, New York.
Morowitz, H.J. (1992) Beginnings of cellular life. Yale University Press, London.
Morowitz, H.J., Heinz, B. and Deamer, D.W. (1988) The chemical logic of a minimum protocell. Origins Life Evol. Biosphere, 18, 281-287.
Muller, H.J. (1966) The gene material as the initiator and the organizing basis of life. Am. Nat., 100, 493-517.
Nekovee, M., Coveney, P.V., Chen, H. and Boghosian, B.M. (2000) Lattice-Boltzmann model for interacting amphiphilic fluids. Phys. Rev. E, 62, 8282-8294.
New, M.H. and Pohorille, A. (2000) An inherited efficiencies model of non-genomic evolution. Simulation Practice and Theory, 8, 99-108.
Newman, M.E.J. and Sneppen, K. (1996) Avalanches, scaling and coherent noise. Phys. Rev. E, 54, 6226-6231.
Nicolis, G. and Prigogine, I. (1977) Self-organization in nonequilibrium systems. Wiley & Sons, Toronto.
Niesert, U., Harnasch, D. and Bresch, C. (1981) Origin of life between scylla and charybdis. J. Mol. Evol., 17, 348-353.
Norman, M. (1998) Cosmos in a Computer. Science, 280, 1522-1523.
Oberholzer, T., Wick, R., Luisi, P.L. and Biebricher, C.K. (1995) Enzymatic RNA replication in self-reproducing vesicles: an approach to a minimal cell. Biochem. Biophys. Res. Commun., 207, 250-257.
Oberhummer, H., Csoto, A. and Schlattl, H. (2000) Stellar production rates of carbon and its abundance in the universe. Science, 289, 88-90.
Oparin, A.I., and Gladilin, K., L.,. (1980) Evolution of self-assembly of probionts. BioSystems, 12, 133-145.
Oparin, A.I. (1957) The origin of life on the earth. Oliver and Boyd, London.
Orgel, L. (1998) The origin of life - a review of facts and speculations. TIBS, 23, 491-495.
Orgel, L.E. (1968) Evolution of the genetic apparatus. J. Mol. Biol., 38, 381-393.
Orgel, L.E. (1992) Molecular replication. Nature, 358, 203-209.
Orò, J. (2000) Organic matter and the origin of life in the solar system. Sheridan Books, Michigan.
Plunkett, M.J. and Ellman, J.A. (1997) Combinatorial chemistry and new drugs. Sci. Am., 276, 68-73.
Ponnamperuma, C., Mariner, R. and Sagan, C. (1963a) Formation of adenosine by ultraviolet irradiation of a solution of adenine and ribose. Nature, 198, 1199-1200.
Ponnamperuma, C., Sagan, C. and Mariner, R. (1963b) Synthesis of adenosine triphosphate under possible primitive earth conditions. Nature, 199, 222-226.
Rebek, J.J. (1994) Synthetic self-replicating molecules. Sci. Am., July, 34-40.
Reid, C. and Orgel, L.E. (1967) Synthesis of sugars in potentially prebiotic conditions. Nature, 216, 455.
Richards, F.F., Konigsberg, W. H., Rosenstein, R. W., and Varga, J. M. (1975) On the specificity of antibodies. Science, 187, 130-137.
Rizzotti M. (1996) Defining Life: the central problem in theoretical biology, University of Padova.
Robertson, M.P. and Miller, S.L. (1995) Prebiotic Synthesis of 5-Substituted Uracils: A Bridge Between the RNA World and the DNA-Protein World, Science, 268, 702-705.
Rosenwald, S. (1998) M.Sc. Thesis. The Weizmann Institute of Science, Rehovot, Israel.
Rosenwald, S., Kafri, R. and Lancet, D. (2002) Test of a statistical model for molecular recognition in biological repertoires, J. Theor. Biol., 216, 327-36.
Sagan, C. (1963) Primordial ultraviolet synthesis of nucleoside phosphates. In: S.W. Fox (ed.), The origin of prebiological systems, pp. 207-219.
Salam, A. (1991) The Role of Chirality in the Origin of Life. Journal of Molecular Evolution, 33, 105-113.
Schlesinger, G. and Miller, S.L. (1983) Prebiotic synthesis in atmospheres containing CH4, CO, and CO2. I. Amino acids. J. Mol. Evol., 19, 376-382.
Schneider-Henriquez, J.E., Denicourt, N. and Fendler, J.H. (1992) Surface Forces in Supported Spherical Lipid Membranes. Langmuir, 8, 3063-3072.
Schopf, J.W., Barghoorn, E.S., Maser, M.D. and Gordon, R.O. (1965) Electron microscopy of fossil bacteria two billion years old. Science, 149, 1365-1357.
Schultz, P.G. and Lerner, R.A. (1995) From molecular diversity to catalysis: lessons from the immune system. Science, 269, 1835-42.
Schwartz, A.W. (1996) Did minerals perform prebiotic combinatorial chemistry? Chemistry and Biology, 3, 515-518.

Schwartz, A.W. and Orgel, L.E. (1985) Template-directed synthesis of novel, nucleic acid-like structures. Science, 228, 585-587.
Sciama, D.W. (1971) Modern Cosmology. Cambridge University Press, Cambridge, UK.
Scott, J.K. and Smith, G.P. (1990) Searching for peptide ligands with an epitope library. Science, 249, 386-390.
Segrè, D., Pilpel, Y., Glusman G. and Lancet, D. (1997) Self-Replication and Evolution in Primordial Mutually Catalytic Sets. In: C.B. Cosmovici, S. Bowyer and D. Werthimer (eds.), Astronomical and Biochemical Origins, Editrice Compositori, Bologna, pp. 469-476.
Segrè, D. and Lancet, D. (1998) Mutually catalytic amphiphiles: simulated chemical evolution and implications to exobiology, In: J. Chela-Flores and F. Raulin (eds.), Exobiology: Matter, Energy and Information in the Origin and Evolution of Life in the Universe, Kluwer Ac. Pub., 123-131.
Segrè, D., Pilpel, Y. and Lancet, D. (1998) Mutual catalysis in sets of prebiotic organic molecules: evolution through computer simulated chemical kinetics, Physica A, 249, Nos. 1-4, 558-564.
Segrè, D., Kedem, O., Lancet, D. and Pilpel, Y. (1998a) Graded Autocatalysis Replication Domain (GARD): kinetic analysis of self-replication in mutually catalytic sets, Origins Life Evol. Biosphere, 28, 501-514.
Segrè, D. and Lancet, D. (1999) A statistical chemistry approach to the origin of life, Chemtracts - Biochemistry and Molecular Biology, 12 (6), 382-397.
Segrè, D. and Lancet, D. (2000) Composing Life, EMBO Reports, 1(3), 217-222.
Segrè, D., Ben-Eli, D. and Lancet, D. (2000) Compositional genomes: prebiotic information transfer in mutually catalytic non-covalent assemblies, Proc. Natl. Acad. Sci. USA, 97(8), 4112-4117.
Segrè, D., Ben-Eli , D. and Lancet, D. (2000a) Prebiotic evolution of amphiphilic assemblies far from equilibrium: from compositional information to sequence-based biopolymers. In: C.B. Cosmovici, S. Bowyer and D. Werthimer (eds.), Astronomical Society of the Pacific Conference Series Vol.213, Sheridan Books, Michigan, pp. 373-378.
Segrè, D., Ben-Eli, D., Deamer, D.W. and Lancet, D. (2001) The Lipid World, Origins Life Evol. Biosphere, 31, 119-145.
Segrè, D., Shenhav, B., Kafri, R. and Lancet, D. (2001a) The molecular roots of compositional inheritance, J. Theor. Biol., 213, 481-491.
Segrè, D., Vitkup, D. and Church, G.M. (2002) Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. U S A, 99, 15112-15117.
Shapiro, R. (1984) The improbability of prebiotic nucleic acid synthesis. Origins Life Evol. Biosphere, 14, 565-570.
Shapiro, R. (2000) A replicator was not involved in the origin of life. IUBMB Life, 49, 173-176.
Shenhav, B., Segrè, D., and Lancet, D. (2002) Mesobiotic emergence: molecular assemblies that self-replicate without biopolymers. Advances in Complex Systems, submitted.
Shenhav, B., Kafri, R. and Lancet, D. (2003) The transition from compositional to polymer based replication. Origins Life Evol. Biosphere, in press.
Sievers, D. and von-Kiedrowski, G. (1994) Self-replication of complementary nucleotide-based oligomers. Nature, 369, 221-224.
Smith, T. and Morowitz, H. (1982) Between history and physics. J. Mol. Evol., 18, 265-282.
Soai, K., Shibata, T., Morioka, H. and Choji, K. (1995) Asymmetric autocatalysis and amplification of enantiomeric excess of a chiral molecule. Nature, 378, 767-768.
Stadler, P.F., Fontana, W. and Miller, J.H. (1993) Random catalytic reaction networks. Physica D, 63, 378-392.
Stadler, P.F., Schnabl, W., Forst, C.V. and Schuster, P. (1995) Dynamics of Small Autocatalytic Reaction Networks - II. Replication, Mutation and Catalysis. Bulletin of Mathematical Biology, 57, 21-61.
Swetina, J. and Schuster, P. (1982) Self-replication with errors - A model for polynucleotide replication. Biophys. Chem., 16, 329-345.
Szathmary, E. (1999) Chemes, Genes, Memes: a revised classification of replicators. Lectures on mathematics in the life sciences, 26, 1-10.
Szathmary, E. and Maynard Smith, J. (1997) From Replicators to Reproducers: the First Major Transitions Leading to Life. J. Theor. Biol., 187, 555-571.
Szostak, J.W. (1992) In vitro genetics. TIBS, 17, 89-93.
Szostak J.W., Bartel D.P. and Luisi P.L. (2001) Synthesizing life. Nature, 409 (Suppl.), 387-90.
Tanford, C. (1978) The hydrophobic effect and the organization of living matter. Science, 200, 1012-1018.
Tjivikua, T., Ballester, P. and Rebek, J. (1990) A self-replicating system. J. Am. Chem. Soc., 112, 1249-1250.
Trimble, V. (1997) Origin of the biologically important elements. Origins Life Evol. Biosphere, 27, 3-21.
Unrau, P.J. and Bartel, D.P. (1998) RNA-catalysed nucleotide synthesis. Nature, 395, 260-263.

Vant-Hull, B., Payano-Baez, A., Davis, R.H. and Gold, L. (1998) The mathematics of SELEX against complex targets. J Mol Biol, 278, 579-97.
Wachtershauser, G. (1988) Before Enzymes and Templates: Theory of Surface Metabolism. Microbiol. Rev., 52, 452-484.
Walde, P., Wick, R., Fresta, M., Mangone, A. and Luisi, P.L. (1994) Autopoietic Self-Reproduction of Fatty Acid Vesicles. J. Am. Chem. Soc., 16, 11649-11654.
Wills, P.R., Kauffman, S.A., Stadler, B.M.R. and Stadler, P.F. (1998) Selection Dynamics in Autocatalytic Systems: Templates Replicating Through Binary Ligation. Bull. Math. Biol., 60, 1073-1098.
Wilson, E.O. and Bossert, W.H. (1971) A primer of population biology. Sinauer Assoc., Sunderland, Mass.
Wimberly, B.T., Brodersen, D.E., Clemons, W.M.J., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T. and Ramakrishnan, V. (2000) Structure of the 30S ribosomal subunit. Nature, 407, 327-339.
Woolfson, A. (2000) Life without genes. Harper Collins, London.
Wright, M.C. and Joyce, G.F. (1997) Continuous in vitro evolution of catalytic function. Science, 276, 614-617.
Yarus, M. (1999) Boundaries for an RNA world. Curr. Opin. Chem. Biol., 3, 260-267.
Yockey, H.P. (1992) Information theory and molecular biology. Cambridge University Press.
Yockey, H.P. (2000) Origin of Life on Earth and Shannon's Theory of Communication. Computers and Chemistry, 24, 105-123.