

Recombination in Primeval Genomes: A Step Forward but Still a Long Leap from Maintaining a Sizable Genome

Mauro Santos,^{1,2} Elias Zintzaras,^{1,3} Eörs Szathmáry^{1,4}

¹ Collegium Budapest, Institute for Advanced Study, Szentháromság u. 2, H-1014 Budapest, Hungary

² Departament de Genètica i de Microbiologia, Grup de Biologia Evolutiva, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

³ Department of Biomathematics, School of Medicine, University of Thessalia, Papakyriazy 22, Larissa 41222, Greece

⁴ Department of Plant Taxonomy and Ecology, Eötvös University, 2 Ludovika tér, H-1083 Budapest, Hungary

Received: 10 February 2004 / Accepted: 3 May 2004 [Reviewing Editor: Dr. Niles Lehman]

Abstract. A stubborn question in early molecular evolution is how to sustain the “minimum” informational length required for the basic features of life with a putative low-copying fidelity RNA polymerase ribozyme. Proposals to circumvent the information crisis have primarily focused on networks of cooperative molecules or compartmentalization of non-cooperative unlinked templates, but success has been very limited so far. Lehman (2003) has recently suggested that recombination—a frequently ignored player in early evolution—could have been crucial to building up primeval genomes of sizable length. Here we investigate this claim by assuming (without loss of generality) that genes were already enclosed in a compartment (vesicle, protocell). The numerical results show a quite intricate interplay among mutation, recombination, and gene redundancy. Provided that the minimum number of gene copies per protocell was enough for recombination to recreate wild-type templates—but not too high to impose an unbearable burden of mutants—informational content could have increased by at least 25% by keeping the same mutational load as that for a population without recombination. However, the upper bound of informational length would still be far from the “minimal life” provisions.

Key words: Recombination — Origin of life — Error threshold — Compartmentalization — Stochastic corrector model — Theoretical biology

Introduction

The emergence of the first RNA-like replicators (i.e., the fundamental units of any process of natural selection [Dawkins 1976; but see below]) sets a boundary between prebiotic chemistry and the early dawn of self-sustained biological systems (albeit not-yet-living entities, since naked self-replicating molecules hardly satisfy the criteria for “minimal life” [cf. Gánti 1987, 2003; Joyce 1994; Luisi 1998; Szathmáry 2003a,b]). At that first level in the hierarchy of units of evolution there would have been primarily strong selection pressure to improve replication efficiency and accuracy (Scheuring 2000; Szabó et al. 2002), but without the aid of peptide enzymes the upper bound of copying fidelity per nucleotide per replication was likely around 96–99% (Friedberg et al. 1995; Johnston et al. 2001). With such a high mutation rate the number of mutated offspring molecules far exceeded the number of nonmutated ones. Consequently, a stable cloud of mutants (so-called “quasispecies” [Eigen and Schuster 1979, 1982; Szathmáry 1989; Eigen 1992; Nowak 1992; Maynard Smith and Szathmáry 1995]) formed around a “master se-

quence” as long as the maximum chain length was roughly equal to the inverse of mutation rate per site and replication (the “error threshold of replication”; i.e., a sharply defined threshold beyond which heredity breaks down and evolutionary adaptation becomes impossible as first emphasized by Eigen [1971] and Eigen and Schuster [1979]; for recent reviews see Joyce [2002b] and Stadler and Stadler [2003]). Experimental evidence in favor of such a persistent cloud of mutants is available from RNA viruses, whose populations seem to live near the critical value of replication accuracy (Steinhauer et al. 1989; Domingo 1996; Domingo and Holland 1997, 1998).

An era of molecular evolution on two-dimensional spatial structures (surfaces) likely preceded the compartmentalization of genes in reproductive vesicles (Ferris et al. 1996; Huber and Wachtershauser 1998; Czárán and Szathmáry 2000; Szabó et al. 2002; Franchi et al. 2003; Scheuring et al. 2003; Franchi and Gallori 2004). At that precellular phase the quasispecies, and not the individual replicating sequences, was almost certainly the unit of selection as originally proposed by Eigen and Schuster (1979; see also Schuster and Swetina 1988; Wilke 2001; Wilke et al. 2001). Once compartments (protocells) replaced surface dynamics they created a higher-level evolutionary unit and were, by clonal selection, not only the best countermeasures against parasites (a problem first recognized by Maynard Smith [1979], who defined a parasitic molecule as one whose replication gets catalytic support from other molecule but does not itself reciprocate to “the common good”), but the best vehicles for accumulating information and for forming a catalytic network that otherwise could be unstable in homogeneous solution (Bresch et al. 1980; Luisi 1998; Szostak et al. 2001; Matsuura et al. 2002; Szathmáry 2003a, b). Two types of compartmentalization scenarios are conceivable: (i) a package of a truly preexisting hypercyclic system (i.e., a cyclically coupled system of autocatalytic and cross-catalytic molecular mutualists, where each member helps the following member and receives help from the preceding one [see Eigen 1971; Eigen and Schuster 1979; Eigen et al. 1981; Zintzaras et al. 2002]) and (ii) a group of competing genes that happened to be together, namely, the “stochastic corrector model” (SCM) as first defined by Szathmáry and Demeter (1987) and further studied by Grey et al. (1995) and Zintzaras et al. (2002). The latter authors have shown that a population of SCM protocells can tolerate higher deleterious mutation rates and reaches an equilibrium mutational load lower than that in a population of protocells hosting hypercycles. However, compartmentalization *per se* was not sufficient to overcome the information bottleneck imposed by the error threshold, and hence, genome size was also limited by replication accuracy (Zintzaras et al. 2002;

Hogeweg and Takeuchi 2003). To summarize, a working model for the protocell scenario in the origin of life would enclose a persistent cloud of mutants around master sequences of genes competing for within-group common resources. Among-group (clonal) selection would favor those lineages with noncompetitive molecular assemblies, thus stabilizing the population against random loss of essential genes after compartment fission (see Szathmáry and Demeter 1987; Grey et al. 1995; Zintzaras et al. 2002; Santos et al. 2003).

Lehman (2003) has recently challenged the basic premise that asexual reproduction was a primitive condition (presumed also in the SCM; but see Santos et al. 2003). As an alternative, he presents four lines of evidence to support the hypothesis that recombination (i.e., the exchange of genetic information between two sources) was in fact the ancestral state: (i) the antiquity of enzymatic machinery involved in replication and recombination; (ii) the potential for nonrecombining asexual lineages to deteriorate due to the irreversible accumulation of deleterious mutations (Muller’s ratchet: see Muller [1964], Felsenstein [1974], and the cautionary remark at the end of the paragraph); (iii) the fact that in the RNA world scenario (Gilbert 1986; Poole et al. 1998; Bartel and Unrau 1999; Gesteland et al. 1999; Joyce 2002a; Doudna and Cech 2002; but see also Orgel and Crick 1993; Segré and Lancet 2000; Sowerby and Petersen 2002; Orgel 2003) a parent molecule would not direct the synthesis of a whole complementary daughter strand but would instead grow after ligation of short oligomers from a common pool; and (iv) the fact that assembling a complex genome when the copying fidelity per nucleotide was so low would be feasible only from a parallel process of polymerization of short polymers (see also Segré et al. 2001). A caveat: The occurrence of thresholds for error propagation in asexually replicating RNA-like molecules was originally derived as a deterministic kinetic theory that is valid only in the limit case of an infinite number of molecules. Nowak and Schuster (1989; see also Alves and Fontanari 1998) have extended this to finite populations, where replication has to be more accurate because of the additional problem of random loss of the master sequence (i.e., the operation of Muller’s ratchet [Haigh 1978; Bell 1988; Charlesworth and Charlesworth 1997]).

Points (i), (iii), and possibly (iv) are not fundamentally challenging. Actually, in the article that coined the phrase “the RNA world,” Gilbert (1986) wrote, “The RNA molecules evolve in self-replicating patterns, using recombination and mutation to explore new functions and to adapt to new niches.” More recently, Riley and Lehman (2003) have shown that *Tetrahymena* and *Azoarcus* ribozymes can promote RNA recombination. Anyway, the question

at the root of Lehman's (2003) scenario is whether or not recombination could significantly minimize the burden imposed by the error threshold. Bottom line, there are two possibilities for the putative advantage of genetic exchange: It was already favorable before primitive genomes were organized inside protocells, or compartmentalization was certainly a necessary condition. If sex is broadly defined as "the exchange of genetic material between genomes or between two sources" (Michod and Levin 1988; Lehman 2003), we must inevitably distinguish between two worlds: that of naked self-replicating molecules and the latter with compartmentalization of genes in reproductive vesicles. Thus, in a population of naked replicators (i.e., in the era of precompartmentalization of genes) recombination between molecules was obviously the only feasible mechanism for genetic exchange. However, once genes were enclosed within cells sex and recombination became concepts not necessarily interchangeable (aside from somatic recombination): Cavalier-Smith (2002) defines "true" sex as the combined presence of syngamy, nuclear fusion, and meiosis, a characteristic only found in eukaryotes (whose origin dates 1.5 Gyr ago at best [see Knoll 2003; Martin and Russell 2003]). We have, nevertheless, recently provided some reasons to think of sex as being very deeply rooted in the origin of life (Santos et al. 2003), but our basic aim there was to refute the (wrong) idea that recombinational repair was the major selective force for the emergence of sex (Bernstein et al. 1984, 1985). The rationale is simply that genetic systems that involve fusion between organisms (protocells) offer higher prospects for parasitic genes (see also Hoekstra 2003).

Here we analyze the effects of recombination by assuming that genes were already enclosed in compartments. Our approach does not escape the enquiry on the effectiveness of recombination in a spatially nonstructured population. Quite the opposite, it enables us to examine a wide range of potential interactions of selection levels. The remainder of the paper is organized as follows. First, we describe the Monte Carlo model at the core of the numerical results. Then, based on these numerical results, we discuss the potential benefits and limitations of recombination for protocells to cope with high mutation rates. Finally, we discuss the implications of our results in a wider scenario and provide some concluding remarks.

Deleterious Mutation and Recombination Within Compartments

We use the basis behind the SCM model prompted by Zintzaras et al. (2002) to introduce deleterious mutations that impair the fitness of the protocell as a whole. However, in that work we were chiefly inter-

ested in the comparison of mutational loads between "conceptually analogous" versions for the hypercycle and the SCM. Each template was assumed to have only three sites under selection, a clearly unworkable situation to numerically simulate the swapping of genetic information between molecules. Therefore, a number of adjustments in the more realistic and dynamically continuous case of our original Monte Carlo model were undertaken.

The behavior of the system depends on two types of stochasticity: (i) replication of templates within protocells and (ii) random assortment of templates into offspring protocells. Even though templates compete within compartments, selection on stochastically produced offspring variants ("between-protocell selection") can rescue the population from extinction, which reaches equilibrium with a constant frequency of the optimal protocell. As before, each protocell consists of two types of unlinked genes, M_1 and M_2 , with a metabolic function essential for survivorship that are replicated by genes (R) with a nonspecific replicase functionality. Gene redundancy is necessary so that the probability of transmission of at least some copies to each daughter protocell after replication and stochastic fission would be high enough for positive population growth (see also Niesert et al. 1981; Koch 1984; Reaney 1987; Santos et al. 2003).

Metabolic genes were organized as having a "target" region that defines an average affinity toward the replicase plus a sequence length of nt nucleotides involved in their functional activity. Similarly, the functionality of the gene acting as a common replicase also depends on nt nucleotides. Deleterious mutations were introduced at a rate equal to u per nucleotide per replication. Within a protocell there are two types of gene replication dynamics (i.e., "within-compartment selection"): one concerning the M_1 and M_2 genes and the other the replication of R. These dynamics and the probabilities of replication are given in the Appendix. Briefly, the growth rate of a given metabolic gene relies on its concentration, the concentration of the randomly selected replicase, and the replication rate, which is a decreasing sigmoid function depending on the number of mutant nucleotides in the replicase. The growth rate of a replicase from a certain mutant class also depends on its concentration, the concentration of the replicase that acts as catalysts, and the replication rate. It is important to remark that the catalysts replicase and the target gene are two physically independent molecules. Therefore, an R gene cannot replicate itself if there is only one copy within the protocell.

Protocell fitness ("between-compartment selection"; $0 \leq w \leq 1$) is an exponential function of the number of copies of the metabolic genes according to the various mutant classes present within the protocell (see Appendix). The simulations start with a

population of K protocells with n templates (genes) initially at equal concentrations (i.e., $M_1 = M_2 = R = n/3$ at t_0), and the flow protocol has four basic steps that together amount to one generation: (i) a protocell is chosen according to its relative fitness for template replication; (ii) a randomly chosen template is replicated according to its probability of replication and the protocell is then turned back to the population whenever its critical size is below a threshold (defined here as the doubling of the total number of genes); otherwise (iii) the protocell divides by randomly assorting its $2n$ templates into two offspring protocells; and (iv) the process continues until the population size is $2K$ and then half of the protocells are randomly discarded. Recombination was introduced in the flow protocol as follows: once a randomly chosen template was replicated according to its probability of replication as indicated above, with probability P_{rec} the daughter molecule was derived from a segment of length l_i ($1 \leq i \leq nt-1$) from that template and a segment of length l_{nt-i} from another randomly chosen copy of a functionally equivalent gene whenever the number of copies in the protocell was equal to or larger than two. This process somewhat mimics a “copy-choice” mechanism for recombination, but we did not allow for insertions or deletions and, hence, kept the total gene length constant (see Concluding Remarks).

In his sexual reproduction scenario Lehman (2003) correctly assumes that product longer and shorter than the parental molecules can result from an asymmetric exchange. Nevertheless, this sort of gene chimerization or, more generally, gain-or-loss-of-function process would introduce some hurdles in our protocell situation. For example, a gain of function may cause an increase in essential genes for survivorship and, hence, a higher assortment load; namely, the drop in average fitness due to the random loss of genes after stochastic assortment of templates in the two daughter protocells. (Conversely, a loss of function could be simply modeled as a decomposition route.)

For the time being, we stick to the aforementioned copy-choice mechanism to numerically explore the effect of recombination on protocell population fitness. Therefore, we do not deal with the possibility of gene chimerization resulting from illegitimate recombination, a putative source of major evolutionary innovations as discussed by Cavalier-Smith (2002). The simulation programs were implemented in MATLAB (V6; The MathWorks 2002), and in Compaq Visual Fortran90 (2000) using the IMS library.

Numerical Results and Discussion

Dynamical models using compartments of genes had long ago recognized that survival of protocells would

have dwelled somewhere between two Homer's *Odyssey* monsters: Scylla and Charybdis (Niesert et al. 1981; see also Niesert 1987). Decrease gene redundancy below a critical level or assume quite different growth rates within compartments and protocells become extinct (see also Santos et al. 2003). Increase the gene copy number and there is the risk that Darwinian selection would be stopped because of dilution of favorable mutations in an orgy of redundancy (Koch 1984). Perhaps not surprisingly, our numerical results also illustrate that there is a quite complex interplay among the number of gene copies (redundancy), mutation rate, and recombination.

Before the effect of recombination can be addressed, we first need a good balance between the sequence length (nt) of replicators and the per digit mutation rate (u). In our model an entire functional genome contains $3 \times nt$ sites under selection. Numerical results with no recombination indicate that (roughly speaking) the population will collapse when u is around $1/(3 \times nt)$ (assuming the most favorable case of a noncompetitive ensemble of molecules, i.e., equal target affinities for all templates), thus suggesting that compartmentalization *per se* does not substantially alleviate the burden imposed by the error threshold as previously argued. Therefore, Lehman's (2003, p. 773) proposal to explore mutation rates above 5% (which, in any case, seem to be too high even for a RNA polymerase ribozyme [see Johnston et al. 2001]) would only be feasible provided that recombination might render a fairly dramatic benefit. Unless otherwise stated, we discuss results obtained with $nt = 12$ but recall that qualitatively similar conclusions were reached with different combinations of parameter values. In addition, all simulations were based on an initial population size of $K = 750$ protocells and reasonably assumed that the population died out if its average fitness dropped below 0.06 (this criterion also avoids the problem of an intolerable waiting time before a chosen protocell has a relative fitness high enough for template replication and division). Considering a finite number of protocells will probably not modify our conclusions, but close to the error threshold there is the additional problem of random loss of the fittest.

Before describing the set of runs from the Monte Carlo model we look at the effect of recombination on the decline of protocell fitness as a result of the continuous input of deleterious mutations without introducing compartment selection. The starting population was composed of 750 protocells, hosting wild-type copies of all genes, and had previously reached its steady-state relationship between selection and random assortment of templates at an average number of copies (\pm SD) of 21 ± 6 for each gene. At each time step a template taken randomly from each cell was replicated according to its replication prob-

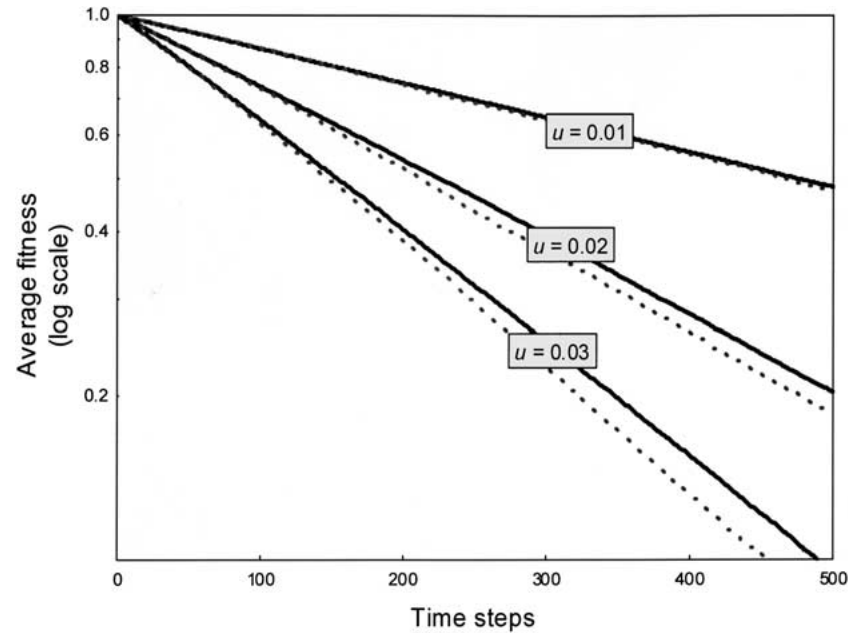


Fig. 1. Time course of decrease in the population mean fitness (averaged over 10 replicated runs) for the first 500 time steps with ($P_{\text{rec}} = 1$; gray dotted lines) or without (solid lines) recombination in an initial population of $K = 750$ protocells at steady state between selection and random assortment of wild-type templates. The average number of copies (\pm SD) for each gene at steady state was 21 ± 6 (target affinities were set to $\tau a_{M_1} = \tau a_{M_2} = \tau a_R = 0.5$, i.e., we assumed a non-competitive ensemble of replicators). One time step is defined as the replication of a randomly taken template (according to its replication probability) from each protocell with a mutation rate u per nucleotide per replication, which amounts to ~ 16 rounds of erroneous replication per template copy when adding both essential metabolic genes for protocell survival.

ability at mutation rate u , and the new daughter copy replaced the original one so that the protocell size was kept constant. (Notice that this template-number constraint is used here since we are interested in protocell fitness only as a function of the increasing mutation load in the initially wild-type templates.) A log-linear decrease in mean fitness was apparent when $P_{\text{rec}} = 0$ ($0.01 \leq u \leq 0.03$; 500 time steps). However, a slightly downward bending was observed at $P_{\text{rec}} = 1$, with a per generation decreasing ratio in average fitness compared to the fitness with no recombination of 0.003, 0.020, and 0.041% for $u = 0.01$, 0.02, and 0.03, respectively (Fig. 1). The reason is that a uniform distribution of deleterious mutants inside compartments along such a short gene sequence is swiftly approached at high mutation rates and relatively high redundancy levels, so recombination between a metabolically efficient and a random template can become a source more for functional deterioration than for mutational purging. Obviously recombination also increased (more than twice; results not shown) the variance in fitness on which between-compartment selection could have worked, but Fig. 1 suggests that recombination would hardly have any significant beneficial effect if there was a high level of gene redundancy within compartments.

Let us now focus on a mutation rate $u = 0.03$ (a reasonably high figure) by using protocells with different numbers of gene copies. All templates had the same replication rate, which assures that protocells will host throughout an even template composition as an average ($\eta_i \approx n/3$; $i = 1, \dots, 3$). The numerical results, depicted in Fig. 2 for some representative runs, are quite complex and somewhat amazing. Be-

fore going into the details it is worth mentioning that Fontanari et al. (2004), by considering the limit case of an infinite number of compartments, each one carrying a finite number of uncorrupted (wild-type) templates that can be mutated to the so-called error tail with probability u , have recently found that for low u increasing redundancy is favored as expected because random loss of essential templates is obviously attenuated. However, for large u the situation is reversed and low-capacity compartments can tolerate much higher mutation rates. Our numerical results somewhat validate those claims in the more complex but likely more realistic scenario of a finite number of continuously growing compartments that host different types of templates of length nt nucleotides. In addition, it should be mentioned here that the baseline survival probabilities—the average population fitness at the steady state between selection and random assortment of templates with $u = 0$; for a particular protocell, its fitness is 1 if at least one copy of each template's class is present, and otherwise the fitness is 0 and the cell is set as dead—are $w_0 \approx 0.61$, 0.84, 0.97, and 1 for initial redundancy values of $\eta_i = 3, 4, 6$, and 15, respectively. From these baseline fitness values we can estimate the mutational load (Haldane 1937; Crow 1970) as $L_{\eta_i} = (w_0 - w_1)/w_0$, where w_1 is the average fitness when $u > 0$.

With no recombination $\eta_i > 4$ is necessary to assure population survival (Fig. 2) but average fitness seems to be maximized at $\eta_i \approx 6$ (i.e., average fitness at $\eta_i = 8$ or 10 is ~ 0.15 ; results not shown). Increasing initial redundancy to $\eta_i = 15$ and $P_{\text{rec}} = 0$ lowers the fitness slightly below 10%, with the resulting outcome that populations faced a relatively high risk of extinction, probably because of

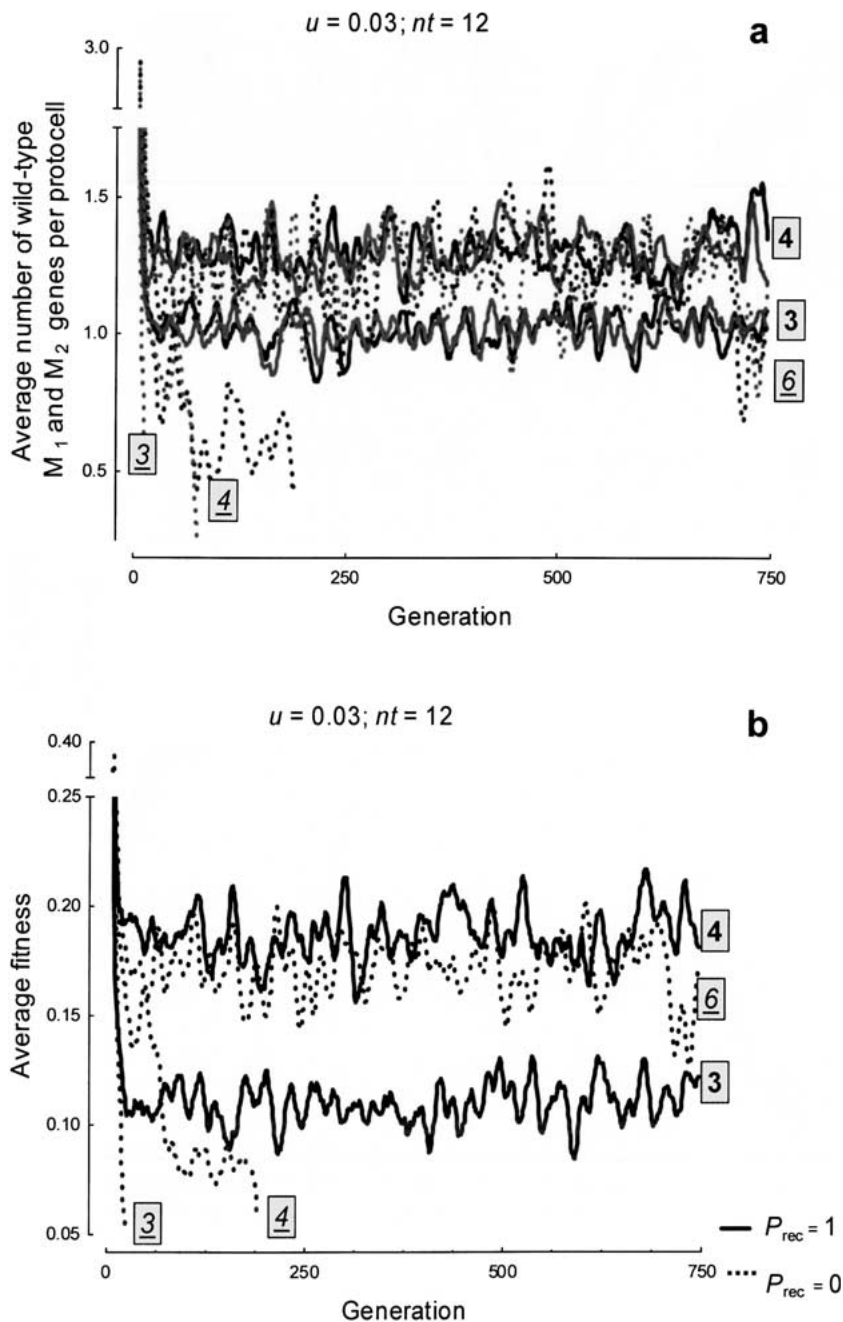


Fig. 2. Sample simulations showing (a) the average number of wild-type copies of essential metabolic genes (M_1 in black, M_2 in gray) per protocell as a function of recombination ($P_{rec} = 0$, dotted lines; $P_{rec} = 1$, solid lines) for different values of initial gene redundancy (an even composition of $\eta_i = 3, 4, 6$ wild-type copies for each gene at t_0 in all $K = 750$ protocells: $P_{rec} = 0$, underscored italic numbers; $P_{rec} = 1$, boldface numbers). Each template had $nt = 12$ sites under selection and target affinities were set to $ta_{M_1} = ta_{M_2} = ta_R = 0.5$. Deleterious mutation rate per nucleotide per replication was $u = 0.03$. The minimum amount of redundancy for population survival ($P_{rec} = 1$) was 3 copies. At $P_{rec} = 0$ and low redundancy (3, 4) the most likely result was, first, the loss of all wild-type copies of one of the two metabolic genes (either M_1 or M_2 with equal probability), followed by the extinction of the population. Increasing redundancy to 6 copies at $P_{rec} = 0$ allows the protocells to host wild-type metabolic genes at average numbers slightly lower than those for 4 copies at $P_{rec} = 1$. Recombination obviously helps the coexistence of wild-type templates inside protocells within certain redundancy levels. (b) At $P_{rec} = 1$ the population can survive whenever the initial redundancy $\eta_i \geq 3$. However, at a very high redundancy ($\eta_i = 15$) all populations died out quickly for $P_{rec} = 1$ but not for $P_{rec} = 0$ (results not shown). To enhance visibility all jagged line patterns were transformed into smoother curves by using a moving average of 10 generations.

random loss of the fittest protocells (approximately half of the independent runs went extinct; results not shown). On the other hand, recombination ($P_{rec} = 1$) can rescue the population from extinction at fairly low redundancy levels: $\eta_i = 3$ was already sufficient for the population to stay alive, i.e., at least two copies of a functionally equivalent gene, not counting the successful template to be replicated, were required. Figure 2a nicely illustrates the quasispecies situation within compartments (note that the original definition of quasispecies excludes recombination of any sort). Thus, when the coexistence of just about one wild-type copy (“master sequence”) of each metabolic gene essential for protocell survival is

possible, and the number of mutated templates is not too high (i.e., a small “cloud of mutants”), the continued existence of the population was guaranteed. Increasing initial redundancy to $\eta_i = 6, 8$ and $P_{rec} = 1$ did not raise average fitness above the level already attained at $\eta_i = 4$ (Fig. 2). However, with high redundancy ($\eta_i = 15$), recombination ($P_{rec} = 1$) doomed the population to extinction at a relatively fast rate (within the first 100 generations; results not shown). This last finding is in sharp contrast with the results from $P_{rec} = 0$ and agrees with the qualitative reasoning offered for Fig. 1 (see above).

Considering only those conditions in Fig. 2 for which the population can survive, the mutational

load was $L_{\eta_i=6} \approx 0.82$ for $P_{\text{rec}} = 0$ and $L_{\eta_i=3} \approx 0.83$ and $L_{\eta_i=4} \approx 0.78$ for $P_{\text{rec}} = 1$. Obviously the mutational loads are quite high and it appears that recombination in protocells would have had only a slight beneficial effect at best. However, since the consequence of changing protocell redundancy on the survival probability is quite complex (see also Fontanari et al. 2004), there seems to be no single answer to the question: What is the error threshold for a protocell population? Here we approached the problem from a different angle. Thus, drawing on $u = 0.03$ our question was: Are there any conditions under which a protocell population could increase its informational content by, say, 25%? The answer turns out to be yes, provided that we allow for recombination. Assuming $nt = 15$ and $P_{\text{rec}} = 0$, population extinction was swift at all initial redundancy levels ($3 \leq \eta_i \leq 15$). On the other hand, $P_{\text{rec}} = 1$ can rescue the population from extinction whenever the initial redundancy is low ($3 \leq \eta_i \leq 6$), and the maximum attained average fitness was $\sim 12\%$ ($L_{\eta_i=4} \approx 0.86$, $L_{\eta_i=6} \approx 0.88$; see Fig. 3). A (meaningful) further increase in informational content ($nt = 18$) imposed an intolerable burden to the population regardless of the initial redundancy (results not shown).

So far we have numerically explored the effects of recombination between finite populations of functionally equivalent genes enclosed within autonomous protocells. A feasible way of looking at the potential benefits of recombination in a nonstructured population of replicators is to simultaneously allow for protocell fusion, that is, by considering a series of ever-changing, swapping committees of proto-organisms that exchanged genetic information to a large extent (Santos et al. 2003). But at this point it should be clear that this sexual scenario introduces more harms than benefits because (i) fusion between protocells offers higher prospects for parasitic replicators; and (ii) a continuous reshuffling of templates from an effectively very large—or even infinite—pool of short molecules with an approximately uniform distribution of deleterious mutations along the sequence was likely more a source for functional deterioration than for mutational purging (Fig. 1; and numerical results with $u = 0.03$, $\eta_i = 15$, and $P_{\text{rec}} = 1$). Incidentally, point (ii) introduces an additional challenge to the once conjectured benefits for the origin of sex (see Santos et al. 2003).

Finally, a few words about the fitness function are in order. It is basically assumed that, other things being equal, a decreasing number of mutations in a gene increase protocell fitness exponentially, until it attains its maximum. The rationale for this choice is as follows. When all sites are mutant it is safe to assume that catalytic activity is reduced to zero. Since we are assuming an effect on kinetics, this approach

must be smooth in the end (as long as the discreteness of the underlying mutational steps allows for this). Conversely, we have maximum efficiency if all digits are the best for the case. The approach to this must be quasi-continuous also. Continuity in between has to hold as well. All this amounts to taking a sigmoid function of decreasing gene contamination for granted. Now, the most efficient end is irrelevant, since in the presence of recurrent, deleterious mutations the system cannot be maintained at the upper plateau of the sigmoid function anyway. So there will be a part of the sigmoid function, between its inflection point and the lower plateau, where it can, without generality, be approximated by an exponential function.

Concluding Remarks

It is now 50 years since the acclaimed paper by Miller (1953) showed the synthesis of important biological compounds using conditions thought (wrongly?) [see Kasting 1993; Kasting and Brown 1998]) to have existed on the primitive Earth and more than 30 years since the pioneering works of Eigen (1971) called attention to the fact that the length of selectively maintained genetic information is limited by the copying fidelity. However, the important gap between an obtainable mixture of some basic biological building blocks such as amino acids and the prebiotic synthesis of complex polymers still remains to be experimentally filled. Furthermore, even if problems related to prebiotic RNA synthesis and stability could eventually be solved, we still do not know how to get around the so-called Eigen's paradox (or the "Catch-22" of the origin of life [Maynard Smith 1983; Maynard Smith and Szathmáry 1995]) to even remotely approach the 10,000–15,000 bp mentioned by Jeffares et al. (1998) as the minimum genome size of the putative last ribo-organism. As stated by Orgel (1998, p. 495), "We are very far from knowing whodunit. The only certainty is that there will be a rational solution."

We have explored here Lehman's (2003) proposal that recombination might have been a prerequisite to building up primeval genomes of sizable length. We have assumed that recombination in protocells took place via copy-choice means, i.e., that the replicase switched between RNA-like templates. This assumption is reasonable since it is well known that template switching occurs frequently in RNA viruses and is crucial for retroviral replication during reverse transcription (e.g., Temin 1993; Negroni and Buc 2000; Hwang et al. 2001; Moumen et al. 2001; Cheng and Nagy 2003). It has also been suggested that the extensive genetic variation found in retroviruses is a direct consequence of that tem-

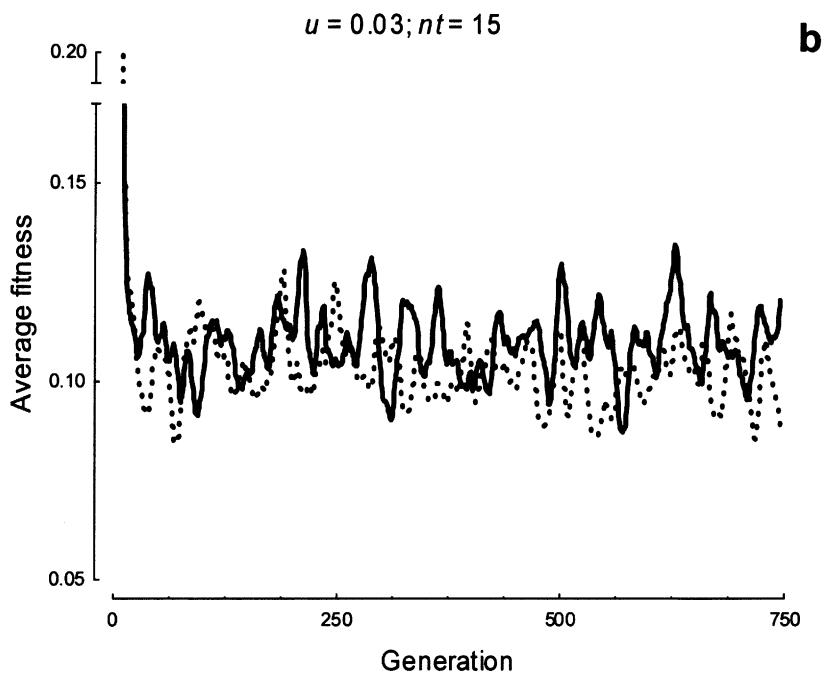
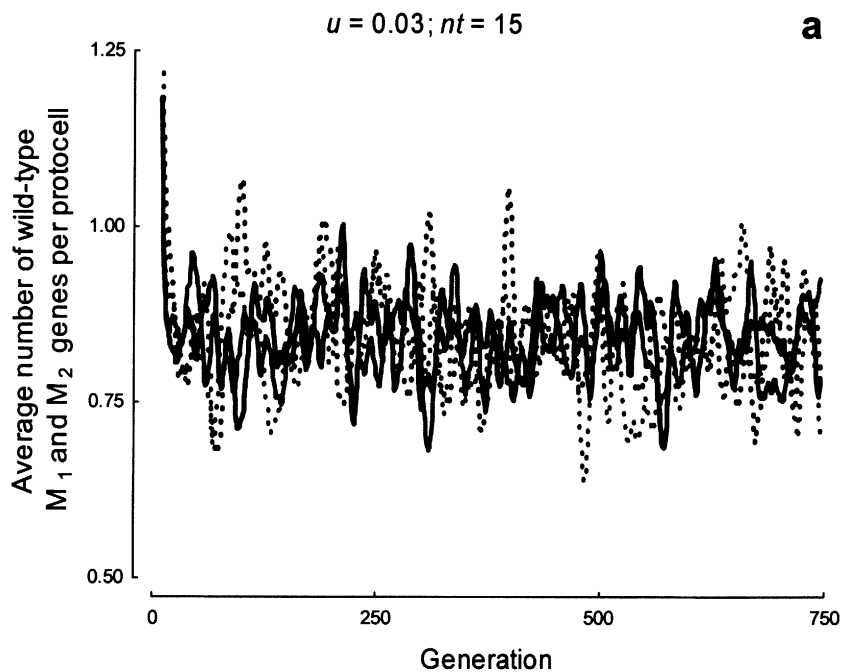


Fig. 3. Same as Fig. 2 for $P_{\text{rec}} = 1$ and $nt = 15$. Solid lines, $\eta_i = 4$ (M_1 in black, M_2 in gray); dotted lines, $\eta_i = 6$.

plate switching (Temin 1993). Frameshifts and indels commonly occur in retroviral replication; however, these “nuisances” would introduce many technical and theoretical problems in our protocell scenario and were neglected in the Monte Carlo model.

Target affinities were also assumed to be equal for all templates, i.e., protocells hosted a noncompetitive ensemble of molecules. Internal competition among unlinked templates has been regarded as the major difficulty for conserving a complete set of genes in compartments (Niesert et al. 1981; Niesert 1987; Su-

zuki and Ono 2003). We have, however, previously argued that the snag with this criticism is that it overlooks evolution—in a realistic package model scenario, only those lineages with a reduced variance in growth rates among unlinked genes would eventually survive (i.e., a sort of bet-hedging [see Santos et al. 2003; Fontanari et al. 2004]). Our assumption should then be taken as the limit when the variance in growth rates for templates is zero. Anyhow, increasing the variance in growth rates mainly affects the assortment load, and our conclusions would remain qualitatively the same.

The numerical results showed a somewhat intricate interplay among mutation, recombination, and number of gene copies. High levels of redundancy increased the mutational load—in agreement with the recent analytical results of Fontanari et al. (2004)—and eventually pushed the population to extinction no matter what the probability of recombination was. Provided the minimum number of gene copies per protocell was enough for recombination to recreate wild-type templates (i.e., $\eta_i \approx 3, 4$), an augmentation in informational content of 25% at least could have been achieved at about the same price in mutational load currency.

Compartmentalization of unlinked, competing templates could be thought of as a spatially structured population where members of each finite subpopulation (vesicle) can “reproduce” (replicate) by “apomixis” ($P_{\text{rec}} = 0$) or “amphimixis” ($P_{\text{rec}} = 1$). With $nt = 12$, $\eta_i \approx 3, 4$, and $u = 0.03$ the average number of new deleterious mutations after replication and just before vesicle fission is $\sim 3\text{--}4$ (assuming that any new mutant does not “hit” a formerly mutated position). Therefore, heuristic arguments rooted in formal population genetics suggest that the advantage of recombination in our protocell system lies in the nonindependent distribution of different deleterious mutants generated by drift and/or between-group (protocell) selection (see Kondrashov 1993). There are, however, some specific features in the system we are exploring. First, protocell selection is dynamically analogous to that in trait groups (vesicles) in completely isolated structured demes, i.e., there is no migration among lineages (see Santos et al. 2003). Second, the groups in question may have irregular ploidies and no rules analogous to those of meiosis, thus making a perfect fit with classical approaches in population genetics difficult. Although $\eta_i \geq 3$ and $P_{\text{rec}} = 1$ guarantee a quasispecies situation within compartments (Fig. 2), high levels of redundancy were costly because they greatly increased the mutational load. Hence, sampling within groups with relatively low ploidies invariably leads to genetic drift and within-group associations. Redundancy costs with a continuous input of deleterious mutations may not sound surprising at all from basic population genetics: We already know that, other things being equal, diploids end up being worse off than haploids (Crow and Kimura 1965). In the simulations we also assumed a constant K throughout generations and, therefore, did not allow the population size to decline as a consequence of the resulting reduction in average fitness (i.e., “the mutational meltdown” [see Lynch et al. 1993]).

Also interesting here is the contrast between the effect of recombination and the step-by-step increase in copying fidelity per digit (Scheuring 2000) to circumvent the informational limit. By using Eq. 23 [i.e.,

$q(N) = e^{\alpha/N}$] of Scheuring (2000) to estimate the per unit copying fidelity ($q = 1 - u$) increase for a sequence of length N bases ($3 \times nt$ in our case), the parameter α is ≈ -1.1 with $u = 0.03$ and $nt = 12$ (Fig. 2). The necessary increase in copying fidelity per digit to augment the informational content by 25% ($nt = 15$; Fig. 3); i.e., $q(3 \times nt + 9) - q(3 \times nt)$, is 5.9×10^{-3} , or about 0.6%. In other words, recombination in protocells could have been roughly equivalent to a 1% increase in copying fidelity per nucleotide. This may seem a very small increase, but *in silico* evolution in the RNA world points to a plateau for fidelity of replication (at $u \approx 0.02$ for the parameter setting) after a boost from very low values (Szabó et al. 2002). Even if we assume that a copying fidelity of $q = 0.98$ was already achieved in the era that preceded compartmentalization of unlinked templates—i.e., the period of molecular evolution on surfaces—the information content per gene (nt) could not have been longer than ~ 25 nucleotides at $P_{\text{rec}} = 1$ (results not shown). The RNA polymerase ribozyme obtained by Johnston et al. (2001) is ~ 10 times longer and it is not nearly capable of catalyzing the synthesis of additional copies of itself. Even if the estimate of 40–60 nucleotides as the minimum size for a replicase ribozyme (Joyce and Orgel 1999) is experimentally proved to be right, we are still a long leap from that scope.

In summary, recombination in the protocell scenario could have provided fairly significant benefits to somewhat alleviate the burden imposed by the error threshold. However, a “minimal life” protocell would entail a three-component system integrated by a lipid boundary, a metabolic subsystem for growing, and a genetic or informational subsystem (Gánti 1987, 2003; Luisi 1998; Szathmáry 2003a, b). The upper bound of ~ 75 nucleotides reached in the present work is still far from those minimal life provisions. It is frustrating to contrast the rich state of knowledge in molecular evolution with the stubborn problem of the informational crisis in prebiotic evolution. We are obviously missing some basic ingredients—“There will be a rational solution.” Attempts to find it are under way.

Appendix (Monte Carlo Model)

Between-Group Selection

The fitness function w at the protocell level is defined as

$$w = \left(\sum_{i=1}^d \frac{1}{g_i} \right) / \left(\sum_{i=1}^d \frac{1}{A_i} \right)$$

where d is the number of different metabolic gene types, g_i is the number of copies of the i th metabolic

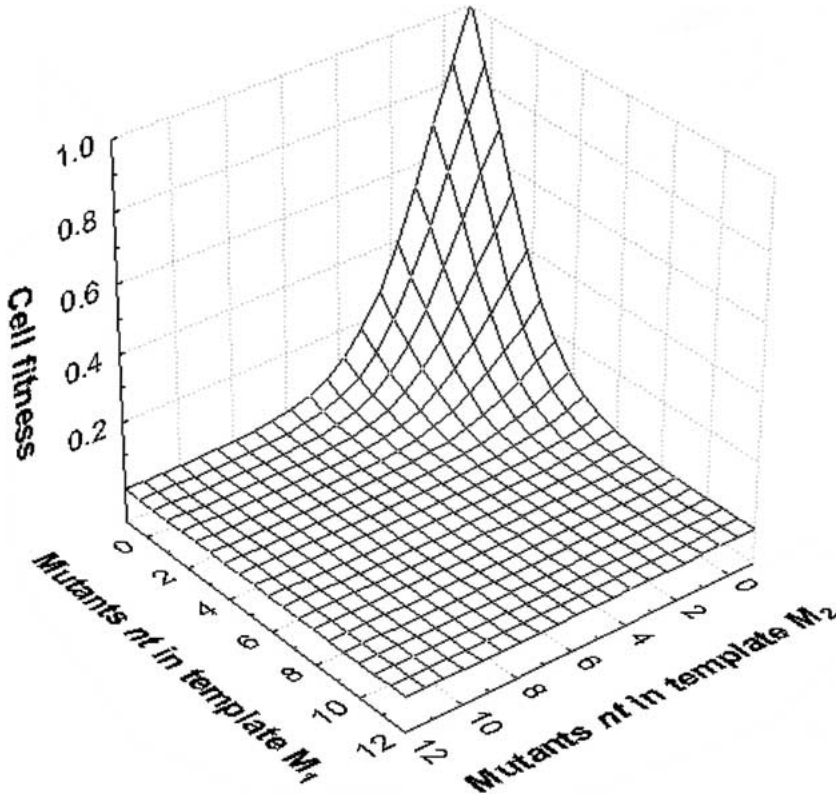


Fig. A1. Fitness function of a protocell according to the number of deleterious mutant nucleotides nt in metabolic genes essential for survivorship.

gene type, and $A_i = \sum_{j=1}^{g_i} A_{ij}$ contribution of all its j variants to protocell fitness.

$$A_{ij} = \left(\frac{1}{e^{(nt-c_{ij})^2}} - \frac{1}{e^{(nt^2)}} \right) / \left(1 - \frac{1}{e^{(nt^2)}} \right)$$

where nt is the number of nucleotides under selection for gene i , and c_{ij} is the number of nonmutated nucleotides. The fitness of the protocell exponentially decreases from $w_{\max} = 1$ to 0 depending on the number of deleterious mutations per gene (Fig. A1; see also Zintzaras et al. 2002).

Beyond the qualitative reasoning in the main text, there is a straightforward rationale behind the construction of the fitness function, which was used first in essentially this form by Szathmary (1992) for the fitness of a ribo-organism. First, we assume that fitness is—as usual for microbes—essentially determined by the flux F of a pathway unsaturated by enzymes. Metabolic control theory shows that for such a case

$$F = C / \sum_i \frac{1}{e_i}$$

where e_i is the enzymatic efficiency of enzyme i and C is a constant (Kacser and Burns 1973). It holds that

$$e_i = c_i E_i / K_i$$

where c_i and E_i are the catalytic efficiency and concentration, respectively, of enzyme i in the pathway and K_i is the equilibrium constant for the respective reaction. Catalytic efficiency is an exponential function of the binding strength G between the enzyme and the substrate of the catalyzed reaction (Kacser and Beeby 1984), and the latter scales linearly with the degree of match between residues (here nucleotides) in the active site and the substrate. We just assume that each deleterious mutation considered here decreases G by the same decrement. These considerations, together with the constraint that w is bounded between 0 and 1, imply the fitness function applied.

Within-Group Selection

The growth rates of metabolic and replicase templates are given by

$$\begin{aligned} \frac{d(M_{i(\eta M)} | R_{(\eta R)})}{dt} &= \mu_{ij(\eta M, \eta R)} M_{i(\eta M)} R_{(\eta R)}; \\ & \quad i = 1, 2; j = 1, 2; i \neq j; \\ & \quad \eta M = 0, \dots, nt; \eta R = 0, \dots, nt \\ \frac{d(R_{(\eta R)} | R_{(\eta Rc)})}{dt} &= \mu_{s(\eta R, \eta Rc)} R_{(\eta R)} R_{(\eta Rc)}; \\ & \quad s = 1, 2; \eta Rc = 0, \dots, nt \end{aligned}$$

where $M_{i(\eta M)}|R_{(\eta R)}$ denotes that M_i has ηM deleterious mutations and is replicated by a replicase gene R with ηR mutations. The replication rates (μ s) are defined by the product of the target affinity of the template (i.e., ta_{M1} , ta_{M2} , and ta_R for M_1 , M_2 , and R , respectively) times the traveling speed of the replicase along the template (i.e., the replicase activity). The target affinities for each kind of template were kept constant regarding the number of deleterious mutants, i.e., the target affinities are unrelated to the metabolic activity of the template. There are two basic reasons for this assumption. First, it obviously prevents selecting the least-loaded mutant copies twice: on the basis of both gene selection and protocell selection. Second, it is likely that the ribogenes were replicated in a manner similar to present-day Q β phage RNA with tRNA-like 3' genomic tag but using a ribozyme as the replicase (Weiner and Maizels 1987). Hence, a more realistic organization for a template would differentiate between two regions: the genomic tag, which must conserve its integrity so any nucleic acid belonging to the set must be replicated, and the purely enzymatic region.

Deleterious mutations were assumed to impair the replicase activity according to the decreasing sigmoid function:

$$f(\eta R) = 1 - \frac{\eta R^4}{100 + \eta R^4}$$

The probabilities of replicating a template depend on the previous growth rates and were calculated as follows:

$$P(M_{i(\eta M)}|R_{(\eta R)}) = \frac{d(M_{i(\eta M)}|R_{(\eta R)})}{dt} \bigg/ \sum_i \sum_{\eta M} \sum_{\eta R} d(M_{i(\eta M)}|R_{(\eta R)}) + \sum_{\eta M} \sum_{\eta R} d(R_{(\eta R)}|R_{(\eta R)})$$

$$P(R_{(\eta R)}|R_{(\eta R)}) = \frac{d(R_{(\eta R)}|R_{(\eta R)})}{dt} \bigg/ \sum_i \sum_{\eta M} \sum_{\eta R} d(M_{i(\eta M)}|R_{(\eta R)}) + \sum_{\eta M} \sum_{\eta R} d(R_{(\eta R)}|R_{(\eta R)})$$

Acknowledgments. Many thanks go to J. F. Fontanari for insightful discussions during the preparation of the manuscript. The paper greatly benefited from the critical comments and suggestions of the Associate Editor and two anonymous referees. M.S. is supported by a fellowship (PR2003-0431) from the Programa Nacional de Ayudas para la Movilidad de Profesores de Universidad, Ministerio de Educación, Cultura y Deporte (Spain), and by Fundación Ramón Areces (Spain). E.Sz. is supported by the Hungarian Scientific Research Found (OTKA). The sponsorship of COST-European Cooperation in the field of Scientific and Technical Research, Internal Project D27/0003/02, "Emergence and Selection of Networks of Catalytic Species" (<http://cost.cordis.lu/src/extranet/publish/D27WGP/d27-0003-02.cfm>), is very much appreciated.

References

- Alves D, Fontanari JF (1998) Error threshold in finite populations. *Phys Rev E* 57:7008–7013
- Bartel DP, Unrau PJ (1999) Constructing an RNA world. *Trends Biochem Sci* 24:M9–M12
- Bell G (1988) Recombination and the immortality of the germ line. *J Evol Biol* 1:67–82
- Bernstein H, Byerly HC, Hopf FA, Michod RE (1984) Origin of sex. *J Theor Biol* 110:323–351
- Bernstein H, Byerly HC, Hopf FA, Michod RE (1985) Genetic damage, mutation, and the evolution of sex. *Science* 229:1277–1281
- Bresch C, Niesert U, Harnasch D (1980) Hypercycles, parasites and packages. *J Theor Biol* 85:399–405
- Cavalier-Smith T (2002) Origins of the machinery of recombination and sex. *Heredity* 88:125–141
- Charlesworth B, Charlesworth D (1997) Rapid fixation of deleterious alleles can be caused by Muller's ratchet. *Genet Res* 70:63–73
- Cheng CP, Nagy PD (2003) Mechanism of RNA recombination in carmo- and tombusviruses: Evidence for template switching by the RNA-dependent RNA polymerase in vitro. *J Virol* 77:12033–12047
- Compaq Visual Fortran Professional Edition 6.6.0 (2000) Polyhedron Software Ltd., Witney, UK (<http://www.polyhedron.co.uk/>)
- Crow JF (1970) Genetic loads and the cost of natural selection. In: Kojima KI (ed) *Mathematical topics in population genetics*. Springer-Verlag, Berlin, pp 128–177
- Crow JF, Kimura M (1965) Evolution in sexual and asexual populations. *Am Nat* 99:439–450
- Czárán T, Szathmáry E (2000) Coexistence of replicators in prebiotic evolution. In: Dieckmann U, Law R, Metz JAJ (ed) *The geometry of ecological interactions: Simplifying spatial complexity*. IIASA and Cambridge University Press, Cambridge, pp 116–134
- Dawkins R (1976) *The selfish gene*. Oxford University Press, Oxford
- Domingo E (1996) Biological significance of viral quasispecies. *Viral Hepat Rev* 2:247–261
-
- Domingo E, Holland JJ (1997) RNA virus mutations and fitness for survival. *Annu Rev Microbiol* 51:151–178
- Domingo E, Holland JJ (1998) High error rates, population equilibrium and evolution of RNA replication systems. In: Domingo E, Holland JJ, Ahlquist P (eds) *RNA genetics*, Vol 3. CRC Press, Boca Raton, FL, pp 3–36
- Doudna JA, Cech TR (2002) The chemical repertoire of natural ribozymes. *Nature* 418:222–228
- Eigen M (1971) Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* 58:465–523
- Eigen M (1992) *Steps towards life: A prospective on evolution*. Oxford University Press, Oxford
- Eigen M, Schuster P (1979) *The hypercycle: A principle of natural self-organization*. Springer-Verlag, Berlin

- Eigen M, Schuster P (1982) Stages of emerging life-five principles of early organization. *J Mol Evol* 19:47–61
- Eigen M, Schuster P, Gardiner W, Winkler-Oswatitsch R (1981) The origin of genetic information. *Sci Am* 244:78–94
- Felsenstein J (1974) The evolutionary advantage of recombination. *Genetics* 78:737–756
- Ferris JP, Hill AR Jr, Liu R, Orgel LE (1996) Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* 381:59–61
- Fontanari JF, Santos M, Szathmáry E (2004) Coexistence and error propagation in package models: A group selection approach (submitted)
- Franchi M, Gallori E (2004) Origin, persistence and biological activity of genetic material in prebiotic habitats. *Orig Life Evol Biosph* 34:133–141
- Franchi M, Ferris JP, Gallori E (2003) Cations as mediators of the adsorption of nucleic acids on clay surfaces in prebiotic environments. *Orig Life Evol Biosph* 33:1–16
- Friedberg EC, Walker GC, Siede W (1995) DNA repair and mutagenesis. ASM Press, Washington
- Gánti T (1987) The principle of life. OMIKK, Budapest
- Gánti T (2003) The principles of life. Oxford University Press, Oxford
- Gesteland RF, Cech TR, Atkins JF (1999) The RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Gilbert W (1986) The RNA world. *Nature* 319:618
- Grey D, Hutson V, Szathmáry E (1995) A re-examination of the stochastic corrector model. *Proc R Soc Lond B* 262:29–35
- Haigh J (1978) The accumulation of deleterious genes in a population-Muller's ratchet. *Theor Popul Biol* 14:251–267
- Haldane JBS (1937) The effect of variation on fitness. *Am Nat* 71:337–349
- Hoekstra RF (2003) Power in the genome. Who suppresses the outlaw? In: Hammerstein P (ed) Genetic and social mechanisms of cooperation. MIT Press, Cambridge, MA, pp 257–270
- Hogeweg P, Takeuchi N (2003) Multilevel selection in models of prebiotic evolution: Compartments and spatial self-organization. *Orig Life Evol Biosph* 33:375–403
- Huber C, Wächtershäuser G (1998) Peptides by activation of amino acids with CO on (Ni,Fe)S surfaces: Implications for the origin of life. *Science* 281:670–672
- Hwang CK, Svarovskaia ES, Pathak VK (2001) Dynamic copy choice: steady state between murine leukemia virus polymerase and polymerase-dependent RNase H activity determines frequency of in vivo template switching. *Proc Natl Acad Sci USA* 98:12209–12214
- Jeffares DC, Poole AM, Penny D (1998) Relics from the RNA world. *J Mol Evol* 46:18–36
- Johnston WK, Unrau PJ, Lawrence MS, Glasner ME, Bartel DP (2001) RNA-catalyzed RNA polymerization: Accurate and general RNA-templated primer extension. *Science* 292:1319–1325
- Joyce GF (1994) Foreword. In: Deamer DW, Fleischaker GR (eds) Origins of life: The central concepts. Jones and Bartlett, Boston, MA, pp xi–xii
- Joyce GF (2002a) The antiquity of RNA-based evolution. *Nature* 418:214–221
- Joyce GF (2002b) Molecular evolution: Booting up life. *Nature* 420:278–279
- Joyce GF, Orgel LE (1999) Prospects for understanding the origin of the RNA world. In: Gesteland RF, Cech TR, Atkins JF (eds) The RNA world, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp 49–77
- Kacser H, Beeby R (1984) Evolution of catalytic proteins or on the origin of enzyme species by means of natural selection. *J Mol Evol* 20:38–51
- Kacser H, Burns JA (1973) The control of flux. *Symp Soc Exp Biol* 27:65–104
- Kasting JF (1993) Earth's early atmosphere. *Science* 259:920–926
- Kasting JF, Brown LL (1998) The early atmosphere as a source of biogenic compounds. In: Brack A (ed) The molecular origins of life. Cambridge University Press, Cambridge, pp 35–56
- Knoll AH (2003) Life on a young planet. The first three billion years of evolution on earth. Princeton University Press, Princeton, NJ
- Koch AL (1984) Evolution vs the number of gene copies per primitive cell. *J Mol Evol* 20:71–76
- Kondrashov AS (1993) Classification of hypothesis on the advantage of amphimixis. *J Hered* 84:372–387
- Lehman N (2003) A case for the extreme antiquity of recombination. *J Mol Evol* 56:770–777
- Luisi PL (1998) About various definitions of life. *Orig Life Evol Biosph* 28:613–622
- Lynch M, Bürger R, Butcher D, Gabriel W (1993) The mutational meltdown in asexual populations. *J Hered* 84:339–344
- Martin W, Russell MJ (2003) On the origins of cells: a hypothesis for the evolutionary transitions from abiotic geochemistry to chemoautotrophic prokaryotes, and from prokaryotes to nucleated cells. *Philos Trans R Soc Lond B Biol Sci* 358:59–85
- The MathWorks Inc. (2002) MATLAB V6. The language of technical computing (<http://www.mathworks.com>)
- Matsuura T, Yamaguchi M, Ko-Mitamura EP, Shima Y, Urabe I (2002) Importance of compartment formation for a self-encoding system. *Proc Natl Acad Sci USA* 99:7514–7517
- Maynard Smith J (1979) Hypercycles and the origin of life. *Nature* 280:445–446
- Maynard Smith J (1983) Models of evolution. *Proc R Soc Lond B* 219:315–325
- Maynard Smith J, Szathmáry E (1995) The major transitions in evolution. WH Freeman, Oxford
- Miller SL (1953) A production of amino acids under possible primitive earth conditions. *Science* 117:528–529
- Michod RE, Levin BR (1988) The evolution of sex: An examination of current ideas. Sinauer Associates, Sunderland, MA
- Moumen A, Polomack L, Roques B, Buc H, Negroni M (2001) The HIV-1 repeated sequence R as a robust hot-spot for copy-choice recombination. *Nucleic Acids Res* 29:3814–3821
- Muller HJ (1964) The relation of recombination to mutational advance. *Mutat Res* 1:2–9
- Negroni M, Buc H (2000) Copy-choice recombination by reverse transcriptases: reshuffling of genetic markers mediated by RNA chaperones. *Proc Natl Acad Sci USA* 97:6385–6390
- Niesert U (1987) How many genes to start with? A computer simulation about the origin of life. *Orig Life* 17:155–169
- Niesert U, Harnasch D, Bresch C (1981) Origin of life: Between Scylla and Charybdis. *J Mol Evol* 17:348–353
- Nowak MA (1992) What is a quasispecies? *Trends Ecol Evol* 7:118–121
- Nowak M, Schuster P (1989) Error thresholds of replication in finite populations mutation frequencies and the onset of Muller's ratchet. *J Theor Biol* 137:375–395
- Orgel LE (1998) The origin of life—A review of facts and speculations. *Trends Biochem Sci* 23:491–495
- Orgel LE (2003) Some consequences of the RNA world hypothesis. *Orig Life Evol Biosph* 33:211–218
- Orgel LE, Crick FHC (1993) Anticipating an RNA world. Some past speculations on the origin of life: Where are they today? *FASEB J* 7:238–239
- Poole AM, Jeffares DC, Penny D (1998) The path from the RNA world. *J Mol Evol* 46:1–17
- Reaney DC (1987) Genetic error and genome design. *Cold Spring Harb Symp Quant Biol* 52:751–757
- Riley CA, Lehman N (2003) Generalized RNA-directed recombination of rRNA. *Chem Biol* 10:1233–1243
- Santos M, Zintzaras E, Szathmáry E (2003) Origins of sex revisited. *Orig Life Evol Biosph* 33:405–432

- Scheuring I (2000) Avoiding catch-22 of early evolution by step-wise increase in copying fidelity. *Selection* 1:135–145
- Scheuring I, Czárán T, Szabó P, Károlyi G, Toroczka Z (2003) Spatial models of prebiotic evolution: Soup before pizza? *Orig Life Evol Biosph* 33:319–355
- Schuster P, Swetina J (1988) Stationary mutant distributions and evolutionary optimization. *Bull Math Biol* 50:635–660
- Segré D, Lancet D (2000) Composing life. *EMBO Rep* 1:217–222
- Segré D, Shenhav B, Kafri R, Lancet D (2001) The molecular roots of compositional inheritance. *J Theor Biol* 213:481–491
- Sowerby SJ, Petersen GB (2002) Life before RNA. *Astrobiology* 2:231–239
- Stadler BMR, Stadler PF (2003) Molecular replicator dynamics. *Adv Complex Syst* 6:47–77
- Steinhauer DA, de la Torre JC, Meier E, Holland JJ (1989) Extreme heterogeneity in populations of vesicular stomatitis virus. *J Virol* 63:2072–2080
- Suzuki H, Ono N (2003) Segregational instability under cell selection pressure—Many chromosome analyses. *Proc First Australian Conf Artif Life (ACAL-2003)*
- Szabó P, Scheuring I, Czárán T, Szathmáry E (2002) *In silico* simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. *Nature* 420:340–343
- Szathmáry E (1989) The integration of the earliest genetic information. *Trends Ecol Evol* 4:200–204
- Szathmáry E (1992) What determines the size of the genetic alphabet? *Proc Natl Acad Sci USA* 89:2614–2618
- Szathmáry E (2003a) Concepts and dynamics: A theoretical issue of OLEB. *Orig Life Evol Biosph* 33:313–317
- Szathmáry E (2003b) The biological significance of Gánti's work in 1971 and today. *The principles of life (by Gánti T)*. Oxford University Press, Oxford, pp 157–168
- Szathmáry E, Demeter L (1987) Group selection of early replicators and the origin of life. *J Theor Biol* 128:463–486
- Szostak JW, Bartel DP, Luisi PL (2001) Synthesizing life. *Nature* 409:387–390
- Temin HM (1993) Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proc Natl Acad Sci USA* 90:6900–6903
- Weiner AM, Maizels N (1987) tRNA-like structures tag the 3' ends of genomic RNA molecules for replication: implications for the origin of protein synthesis. *Proc Natl Acad Sci USA* 84:383–7387
- Wilke CO (2001) Selection for fitness vs. selection for robustness in RNA secondary structure folding. *Evolution* 55:2412–2420
- Wilke CO, Wang JL, Ofria C, Lenski RE, Adami C (2001) Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature* 412:331–333
- Zintzaras E, Santos M, Szathmáry E (2002) “Living” under the challenge of information decay: The stochastic corrector model vs. hypercycles. *J Theor Biol* 217:167–181