important in social life. Even a chimpanzee's deceptions depend on its ability to predict how best to deceive its intended dupe, which appears to presuppose an ability to imagine how it would respond to its acts were it in the dupe's place (Jolly 1991).

Concluding Remarks

Adaptive evolution presupposes modular organization. Indeed, a precise understanding of the nature and history of the modular organization of living things is needed to assess their potential for adaptive evolution and may reveal that living things are organized to facilitate their evolution by natural selection of "random" mutations (Leigh 1987).

The most objective mark of evolutionary progress is the series of evolutionary transitions where parts combined to form larger, more effective wholes (Maynard Smith 1988). Each such transition involved potential conflict between different levels of selection. These conflicts, and the ways they are resolved, comprise one of the grand unifying themes of biology.

Parts join to form larger wholes only if there is a genuine community of interest among the parts, and if circumstances allow the enforcement of this common interest. In the major transitions of evolution, community of interest plays the same crucial role as in Aristotle's *Politics*.

The traces of the means by which conflicts between levels of selection are resolved in favor of the higher level represent unmistakable footprints of the decisive role of natural selection in macroevolution. These traces are instances where evolutionary history testifies to evolutionary mechanism.

Finally, evolutionary studies of social animals suggests that truth, beauty, and goodness are not totally beyond the reach of evolutionary biology.

3 The First Replicators

Eörs Szathmáry

The replicator concept of Dawkins (1976) has turned out to be extremely useful in analyzing evolutionary questions. Here I follow the definition of Hull (1980), who emphasized that replicators must pass on their structure largely intact. Although selection acts on them directly, interactors (such as organisms) do not qualify as replicators because their structures are not copied. I shall come back to this important conceptual issue at the end of this chapter because organisms usually qualify as reproducers.

My primary interest here lies in the origin of the earliest units of evolution. Entities qualify as units of evolution if they meet the following criteria (Maynard Smith 1987):

1. *Multiplication*. Entities should give rise to more entities of the same kind.
2. *Heredity*. Like begets like; A-type entities produce A-type entities; B-type entities produce B-type entities, etc.
3. *Variability*. Heredity is not exact; occasionally A type objects give rise to A′ type objects (it may be that A′ = B).

If objects of different types have a hereditary difference in their fecundity and/or survival, the population undergoes evolution by natural selection.

To explain the origin of life, we need to explain the origin of heredity in terms of chemistry. Heredity merely means that like begets like. This, in turn, requires variation: multiplication of an entity that can exist in only one form does not constitute heredity and could not form the basis of evolution by natural selection. I argue that mere heredity is not enough. Ongoing evolution requires "unlimited heredity," that is, the existence of replicators that can exist in an indefinitely large number of forms. Although, as I outline below, heredity with a small number of possible types can exist without copying, it seems very probable that unlimited heredity requires template copying of replicators with a modular structure.

The first experiment relevant to the origins of replicators (and life in general) was carried out more than a hundred years ago by Butlerov, a Russian chemist. He found that, if formaldehyde is kept in a reaction vessel for a few hours under moderately alkaline conditions, sugars readily form. Nowadays the "formose reaction" appears to be a formidable network of interconversions of sugars, among them ribose, which is a building block of RNA (e.g., Cairns-Smith and Walker 1974). Even more interesting is the fact that, given a sufficient amount of formaldehyde, the accumulation of sugars follows exponential kinetics, indicating that something is replicating in the solution. It is the sugars that replicate: The "hard core" of the reaction is the cyclic,

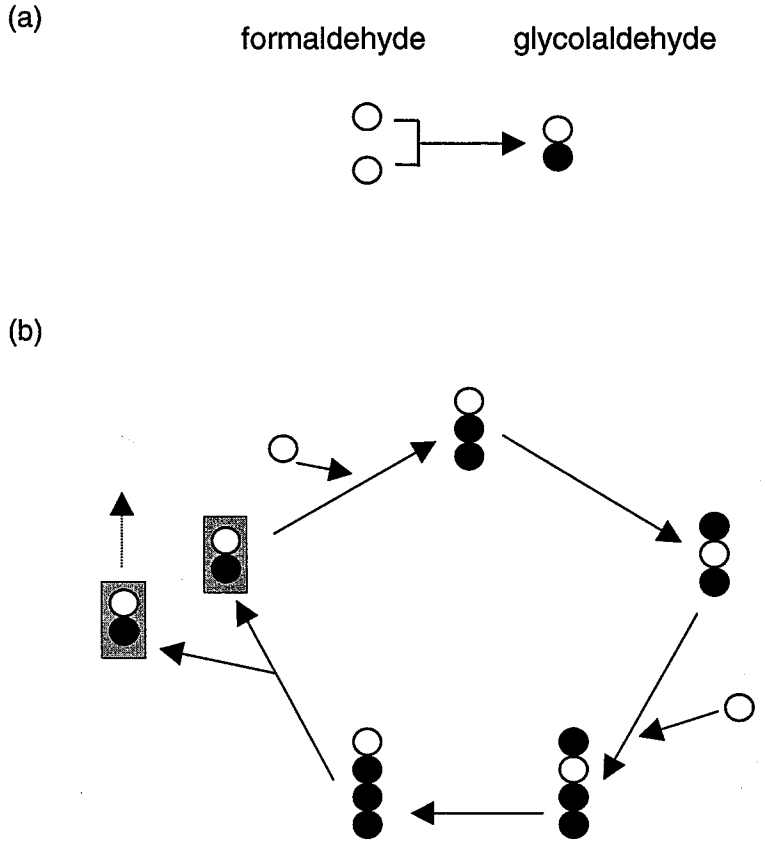

FIG. 3.1. The autocatalytic core, or seed, of the formose reaction. (a) The "spontaneous generation" of the core autocatalyst. (b) The central autocatalytic cycle. Each circle represents a chemical group including one carbon atom.

autocatalytic formation of glycolaldehyde at the expense of continuous formaldehyde consumption (fig. 3.1). (An autocatalyst catalyses its own formation.)

It takes some time until the first glycolaldehyde molecule is formed. The coupling of two formaldehyde molecules is the spontaneous generation, as it were, of the first replicator. This teaches us a general lesson: The first evolutionary units must have arisen, logically, by means other than evolution by natural selection.

There is practically no heredity in this system: We do not yet know of alternative, repetitively replicating variants of this network, but this may not be the final situation. It is noteworthy that the Calvin cycle, fixing carbon dioxide in plants, is also autocatalytic (Gánti 1979) at the level of the sugar phosphates: If one starts with three molecules of glyceraldehyde-3-phosphate, then after one turn of the cycle, one has four molecules. It is potentially confusing that this system is called autocatalytic, because every step in it is catalyzed by a specific enzyme, not produced by the cycle itself. The explanation is that there are two levels of catalysis in the system: one at the level of small molecules (constituting the cycle as we draw it), and another at the level of enzymes (also operating in cycles, of course).

For our topic we need examples of autocatalysts replicating without the aid of enzymes, because the latter are evolved latecomers. Günter Wächtershäuser, chemist and patent lawyer, proposed a hypothetical system of this sort, which may have been an archaic variant of the contemporary reductive citric acid cycle. This cycle fixes carbon in a variety of bacteria (organic acids are produced at the expense of the incorporation of carbon dioxide by the cyclic transformation of the same organic acids; the amount of organic acids thus increases). (The reductive citric acid cycle is almost the exact reverse of the well-known Krebs [citric acid] cycle, producing carbon dioxide by breaking down activated acetic acid.) Note that Wächtershäuser produced an elaborate scenario of prebiotic metabolism on the surface of pyrite (FeS_2, fool's gold; Wächtershäuser 1988, 1992). His primary motivation was that many students in the field became disenchanted by the classical "primordial soup" scenario for the origin of life (see Maynard Smith and Szathmáry 1995). He envisaged a protracted phase of evolution of what I call replicators with limited heredity. Indeed, he suggested a series of alternative chemical cycles, mutants of his archaic reductive cycle. At one time, evolution must have given rise to replicators that were copied digit by digit, in the fashion of contemporary nucleic acid molecules.

The crucial difference in the mode of replication of the aforementioned autocatalytic small molecules and polynucleotides can be visualized as follows. If one takes a snapshot of polynucleotide replication, one can easily decide about the degree of completion. It is perfectly sensible to say, for example, that replication is "half-way through"; this means that all nucleotides in the first half of the molecule pair with their complementary nucleotides. If one were to cut the polynucleotide in two halves, one could say that one of the halves is completely copied, whereas the other one awaits replication. Polynucleotides are copied digit by digit, or modularly; the nucleotides serve as chemical modules. This does not hold true for molecules like the intermediates of the Calvin cycle. If one chooses an intermediate, one can see all sorts of chemical transformations acting on it until after a certain number of steps, the molecule suddenly falls into two identical pieces. The molecule is continuously modified and processed until replication is complete. Wächtershäuser has called this mode of replication "processive." Replication in a sense is holistic: Half of the initial molecule cannot replicate at all.

A general problem arises here concerning heredity in autocatalytic networks of small molecules. One molecule of glycolaldehyde is indistinguish-

able from the other. If so, where is heredity? Maynard Smith and Szathmáry (1995) suggested a resolution. It makes sense to distinguish between replicators with limited heredity and those with potentially unlimited heredity (see also Szathmáry and Maynard Smith 1993b). In the former case, the number of types is smaller than, or roughly equal to, the number of objects (individuals) because objects exist only in a few alternative types. A didactic example is the set of possible hexanucleotides. In contrast, unlimited hereditary replicators have (many) more types than the number of objects (individuals) in any realistic system. What distinguishes even small oligonucleotides from other autocatalysts like glycolaldehyde is that, for the latter, the production of hereditary variation is much more difficult. One could argue that for replicators that are not modularly replicated, variants are allowed to arise only through macromutations (cf. Wächtershäuser 1988). Hence, as limited hereditary replicators, members of autocatalytic cycles lack the ability to undergo microevolution: Heredity is almost always exact. (Of course, there can be strong nonheritable fluctuations, but this is a different issue.)

It seems reasonable to assume that autocatalytic cycles only come in a relatively small number of types; their intermediates are limited hereditary replicators (Szathmáry and Maynard Smith 1993b; Maynard Smith and Szathmáry 1995). I must stress that, of course, these systems have replication at the molecular level. As Orgel (1992, p. 203) wrote, "All replicating systems are, by definition, autocatalytic and all autocatalytic systems result, in some sense, in replication." The reproduction of molecules leads to the growth of the population.

Note that this departure from the traditional "gene-chauvinistic" view of replicators (cf. Dawkins 1976) is absolutely essential if one wants to understand the origin of living systems, for the following reasons. First, genes are too complex to start with. Second, autocatalysis, and what has been called the Darwinian dynamic (Bernstein et al. 1983), transcend traditional genetics and even biology proper. This is why it is important to define units of evolution as generally as possible.

Recognizing the importance of replicators other than genes, Szathmáry (1995) offered the following classification of replicators:

1. Limited hereditary replicators
   a. Processive (holistic)
      Example: formose reaction, archaic reductive citric acid cycle
   b. Modular
      Example: oligonucleotide analogues
2. Unlimited hereditary replicators
   a. Processive
      Example: unknown, probably impossible
   b. Modular
      Example: genes of extant organisms

The information carried by processive and modular replicators can be termed *analogue* and *digital*, respectively (Wächtershäuser 1994), although the terms *holistic* and *digital* seem more appropriate (Maynard Smith and Szathmáry in prep.).

According to our current view, evolution proceeded from holistic, limited hereditary replicators through digital, limited hereditary replicators to digital replicators with unlimited heredity. In chemical terms, digital replicators are likely to have emerged as by-products of holistically replicating autocatalytic cycles, on which they initially act like parasites (Wächtershäuser 1992).

The first digital replicators must have been relatively small molecules because long templates do not easily dissociate from the copies made. Von Kiedrowski (1986) synthesized the first nonenzymatically replicating oligonucleotide analogue; there were more to follow. These results clearly show that digital replication is possible without enzymes, but do not yield, by themselves, a chemically realistic scenario. One of the main problems is that we still do not know what came before Gilbert's (1986) RNA world. RNA seems chemically too complex to originate by simple chemical evolution. People in the field seem convinced that some form of replicators, incorporating nucleic acid bases, must have preceded RNA, but they do not really know what the nature of the primordial backbone, instead of the contemporary ribose-phosphate-ribose-phosphate . . . , could have been. Schwartz (1997) gave a current account of the many relevant speculations and the few experimental approaches.

The dynamics of growth of such replicators turns out to be crucial. Indeed, from a chemical standpoint, natural selection is simply the dynamics of replicators. I review the reasons, empirical and theoretical, for thinking that the growth of the simplest digital replicators would have been parabolic, that is, slower than exponential. This is important because our traditional analyses of selection processes hinge on the assumption that growth would be exponential (Malthusian) without abiotic and biotic limitations. We shall see that parabolic growth does not, under selective conditions, lead to survival of the fittest.

At some time in early evolution, replicators with an exponential growth tendency must have appeared. Around that time, nucleic acids occupied their paramount status as carriers of genetic information. Again, the first nucleic acids could not have been very long. In the absence of specific replicases, copying would have been inaccurate, and large molecules would have accumulated errors. Thus, it is unlikely that the earliest nucleic acids had the length of even the smallest present-day chromosomes. Primordial genomes therefore must have consisted of several smaller pieces of nucleic acids. This raises the central problem of how cooperating groups of small replicators could have arisen, and how they could have protected themselves against invasion by molecular parasites. The answer lies in two phenomena that underlie all increases in complexity and cooperation in evolution, from the

first populations of cooperating polynucleotides to the emergence of animal and human societies. These are synergism and genetic compartmentalization. That is, complementation of functions (division of labor) can result in strong synergistic fitness effects, and limited dispersal can result in different individuals bound to "sit in the same boat." I discuss the role of these two phenomena in the early evolution of life. Surface metabolism turns out to be important as a precursor of cellular organization in reducing dispersal and, hence, in favoring cooperation.

People have for a long time been puzzled by the proportion of traits that follow from some physical, chemical, or engineering constraint to those that are a result of historical contingency in biology. This question soon may become practical rather than purely theoretical. First, we may redesign life to some extent for biotechnological or mere scientific reasons. Second, we may find that life has arisen independently elsewhere in the solar system (e.g., on Europa, one of the Galileo moons of Jupiter). Would Europan organisms be running their heredity on digital replicators? Presumably yes. Would their hereditary material contain nucleotide bases or their analogues? Because such molecules are formed relatively easily under prebiotic conditions, the answer again is, presumably, yes. Would their genetic alphabet consist of four letters (analogous to our A, G, C, and U/T)? I argue that this is likely to be the case, if their hereditary material also had once been used for enzymatic purposes as well.

The last part of the chapter is concerned with the transition from replicators to reproducers. Existing organisms are not replicators; they do not reproduce by copying. Instead, they contain DNA that is copied, and that acts as a set of instructions for the development of the organism. Hence, reproduction requires both copying and development. Following Griesemer (1996), I outline the concept of a reproducer and show a particularly elegant model of it, conceived by the chemist Tibor Gánti (1971, 1975) more than two decades ago in order to understand what a minimum living system should look like. We shall understand the origin of life only when we are able to outline a convincing scenario for the origin of such a system—we are not quite there yet.

For obvious reasons, I lean heavily on the recent reviews of Maynard Smith and Szathmáry (1995; Szathmáry and Maynard Smith 1995, 1997).

Artificial Replicators and "Survival of Everybody"

Artificial Replicators

Von Kiedrowski (1986) and Zielinski and Orgel (1987) synthesized the first nonenzymatically replicating chemical species (in both cases, close chemical relatives to oligonucleotides). They found that, surprisingly, the growth dynamics of these replicators followed a slower than exponential time course.

An intuitive understanding of this phenomenon is possible if we understand why conventional replication of molecules is exponential. When RNA molecules are replicated in a test tube, a replicase enzyme efficiently separates the template and the copy so that both are ready to enter a subsequent round of replication. The same holds, in the absence of enzymes, for glycolaldehyde molecules in the formose network. In contrast, in the case of the von Kiedrowski-type replicators, template and copy do not fall apart so easily because they are bound together by hydrogen bonds. To the contrary, they associate spontaneously in solution at a certain rate. Now, it is important to realize that only the free templates can initiate a round of replication. Two templates (in this case identical to the template: copy complex) sticking together are replicationally inert. This amounts to self-inhibition: The net growth rate depends on the apparent equilibrium between association and dissociation of templates. In fact, the chemists found the growth rate to be proportional to the square root of the total concentration (all forms counted). In Malthusian growth, the rate is simply proportional to the total concentration.

Parabolic Growth: Survival of Everybody

The aforementioned self-inhibition leads to parabolic growth. Total concentration, were resources provided ad libitum, would reach infinity in infinite time, similar to exponential growth, but the growth curve follows a parabola. Contrary to the exponential case, parabolic growth entails "survival of everybody" in a competitive situation (Szathmáry and Gladkih 1989; Szathmáry 1991a; von Kiedrowski 1993) because e^{kt} (exponential) is a much steeper function than $(kt)^2$ (parabolic), where k is a rate constant, and t is time. As explained above, growth is slower than exponential because the replicators limit their own growth, but density reaches infinity in the limit because there are more and more replicators in the growing population. These two opposing effects lead to the parabolic nature of the growth curve.

Because self-limitation is based on molecular complementarity, *AA* and *BB* complexes (where *A* and *B* are two different replicators) are stronger than *AB* complexes. Hence, each species limits its own growth more strongly (by associating with itself). This condition for joint survival is also found in traditional Lotka-Volterra competitive systems. This is the ultimate cause for survival of everybody in parabolic systems (Szathmáry 1991a).

Since the pioneering work of von Kiedrowski, several replicators obeying the same type of growth dynamics have been constructed by Rebek (1994) and Sievers and von Kiedrowski (1995) among others. Growth of a recently synthesized self-replicating peptide (!) obeys similar kinetics (Lee et al. 1996). Von Kiedrowski (1993) worked out a detailed kinetic theory for parabolic growth of minimal replicators. It seems that the survival of everybody is a rather robust phenomenon among these replicators.

One of the important steps of prebiotic evolution thus must have been the emergence of replicators with exponential growth. An attempt to mimic such a process has been made in von Kiedrowski's lab (pers. comm.). In light of the general importance of surfaces in chemical as well as early biological evolution (see later), it is noteworthy that a certain surface plays a crucial role in this nonenzymatic, but exponential replication process: Sticking to the surface helps separate the strands.

Eigen's Paradox and the Importance of Population Structure in the Prebiotic Context

Serious considerations suggest that primordial nucleic acids (or their analogues) must have been rather short molecules because of excessive noise in their copying. If different replicators are thus needed to establish a primordial genome having the size of a few genes, some mechanism to ensure their dynamic coexistence must have operated. Various models show that if selfish mutants are taken into account, some form of population structure is mandatory for indefinite survival. As we shall see, in some cases, survival depends on genuine group selection.

The Error Threshold of Replication

Eigen (1971) called attention to the fact that the length of molecules (number of nucleotides) maintained in mutation-selection balance is limited by the copying fidelity. An intuitive appreciation of this point is readily obtained if one makes a simple calculation of the overall copying fidelity (Q) as a function of the copying fidelity per digit (in the concrete case, nucleotide, q) and the length (ν) of the molecule to be copied. If we assume that mistakes made during replication are independent, then standard probability calculus gives $Q = q^{\nu}$. Let us take the numerical case of $q = 0.99$; that is, one mistake in a hundred is made per nucleotide per replication. Then for a molecule with $\nu = 100$ we obtain $Q = 0.37$; that is, only about one-third of the copies will bear no mutations. Whether the error-free copies can still be maintained depends on the strength of selection. Calculations and analysis of replication of viruses revealed that, realistically, for the value of q given above one cannot go beyond $\nu > 100$ (Eigen 1971). This phenomenon is Eigen's *error threshold of replication.* Whenever the mutation rate is lower than the critical one for the catastrophe to occur, a population of molecules will be maintained by mutation-selection balance. This is hardly surprising to a population geneticist. After all, the so-called quasi-species model of such a population (Eigen and Schuster 1979) is isomorphic to a population genetical model of a haploid, asexual population with many alleles coupled

through mutations, subject to selection. In contrast, the deduction of the error threshold has been a genuine discovery. Manfred Eigen, who earned a Nobel Prize in chemistry, took a naive and fresh look at the mutational load, which enabled him to see something that the researchers within the field have missed.

Exponential replication, implying survival of the fittest, comes at a high price in early evolution. Early genomes must have consisted of independently replicating entities, but they would have competed with each other, and the one with the highest fitness would have won (Eigen 1971). Hence the "Catch-22" of molecular evolution: no enzymes without a large genome, and no genome without enzymes (Maynard Smith 1983b). We know, however, that evolution did not stop at the level of a few naked genes. Something that prevented this from happening must have occurred.

Molecular Hypercycles

Eigen (1971) thought to resolve this problem by proposing the hypercycle (fig. 3.2) as a model for molecular mutualists, coupled directly with mutual aid in replication. It is important to see that the hypercycle is a doubly autocatalytic system. First, each member serves as template in its own replication and is, therefore, autocatalytic. Second, each member receives help, through a replicase activity, from the one preceding it. Replication of each member thus depends on the product of its own concentration and that of the preceding one. This has been called second-order autocatalysis at the level of the system as a whole. Even if the replication rate constants of the members are different, dynamic coexistence is guaranteed because of the cyclic closure of replicational help (Eigen 1971).

Hypercyclic Illusions

I must digress by discussing mistakes in the literature concerning which systems can be regarded hypercyclic. This is important because once a misidentification occurs, people think that hypercycle theory (Eigen and Schuster 1979) becomes readily applicable for the described systems. Conversely, such misinterpretations strengthen the perceived applicability of hypercycle theory to real cases. Considerable turmoil has already resulted from such mistakes.

Ricard and Noat (1986), for example, thought that any link between two-chemical cycles results in a hypercycle. If a simple chemical cycle A (similar to the citric acid cycle) produces substance Z, which is consumed by cycle B, the two cycles are coupled, but are definitely nonhypercyclic. First, there is no autocatalysis whatsoever in the system; second, the coupling between the two cycles is not catalytic (Szathmáry 1988).

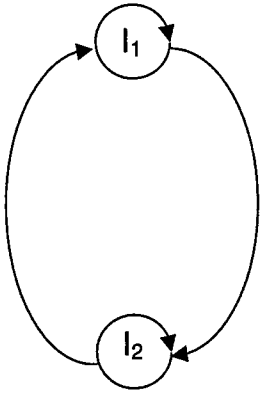

FIG. 3.2. The hypercycle, a model of molecular cooperation. Molecules I_i are auto- and heterocatalysts at the same time. Autocatalysis and heterocatalysis correspond to template and replicase activity, respectively. The number of members may be different from two.

The second case is much more confusing, because it is linked to an important experimental result, the publication of which bears the term "hypercycle" in its title, whereas the system made is not hypercyclic at all! Lee et al. (1997) presented fascinating experimental evidence of two self-replicating peptides that mutually aid the formation of each other. They made three claims.

1. Their system "constitutes a clear example of a minimal hypercyclic network."
2. "A large number of hypercycles are expected to be embedded within the complex networks of living systems."
3. Two previously presented experimental systems "may contain vestiges of a hypercyclic organization."

Unfortunately, none of these claims is correct.

Figure 3.3a shows a schematic diagram of the system synthesized by Lee et al. (raw materials are omitted throughout). Apparently, peptides I_1 and I_2 are both self-replicating, that is, autocatalytic. They are also heterocatalytic in that I_1 catalyzes the formation of I_2 and vice versa. This property suggested to Lee et al. that the peptides are molecular mutualists, and hence manifest two members of a minimum hypercycle. Like ecological mutualists, hypercycles are characterized by a heterocatalytic aid by one member of the system given to the *autocatalytic* replication, rather than mere formation, of the other, and vice versa. (Pollinators do not enhance the spontaneous generation of plants from inanimate matter, but they do help them in reproduction.) A truly hypercyclic variant of the system discussed is shown in figure 3.3b where first-order self-replication is combined with second-order hypercyclic coupling (the two processes run in parallel). When the first

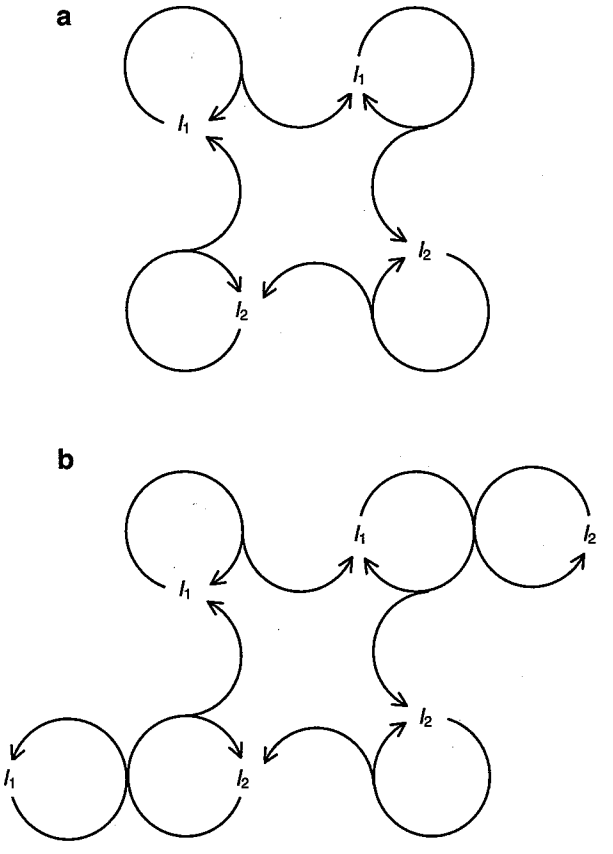

FIG. 3.3. A nonhypercyclic system (a) and a hypercyclic (b) system. (a) A schematic representation of the peptide network synthesized by Lee et al. (1997). *Open arrowheads:* stoichiometric chemical transformations.

type of process is omitted, one arrives at a truly minimum hypercycle, identical in schematic form to the one shown in figure 3.2. Note that filled arrowheads (⟶) in the symbolic version represent catalytic action, rather than stoichiometric transformation (open arrowheads, ⟶). Consequently, the original system (fig. 3.3a) follows growth dynamics different from those of hypercycles. Although we are dealing with coupled replicators, the overall growth is still parabolic: The cycles are limited by product-inhibition because of the association between enzyme and product.

Evolutionary Instability of Naked Hypercycles

It is true that the hypercyclic link ensures indefinite ecological survival of all member replicators. Problems arise, however, when mutations are taken into

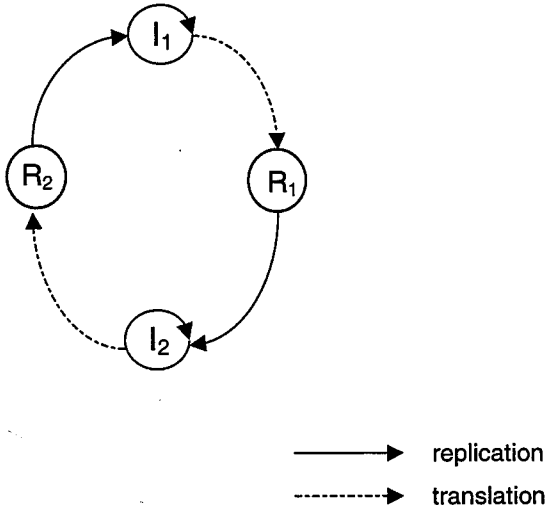

FIG. 3.4. The hypercycle with translation. R_1 is a replicase protein enzyme coded for by gene I_1.

account. In order to consider them, it is worthwhile to look at a diagram in which auto- and heterocatalytic aids are functionally clearly separate, such as in a hypercycle with protein replicases (fig. 3.4). Mutants providing stronger heterocatalytic aid to the next member are not selected for. In contrast, increased autocatalysis is always selected for, irrespective of its concomitant effect on heterocatalytic efficiency. This is the well-known problem of parasites in the hypercycle (Maynard Smith 1979). As Eigen et al. (1981) observed, putting hypercycles into reproducing compartments helps, because "good" hypercycles (with efficient heterocatalysis) can be favored over "bad" ones. Two questions arise out of this. (1) Are there other means whereby parasites can be selected against? (2) Are there nonhypercyclic systems that function well in a compartment context? The answers to both of these questions are "yes." I discuss them below.

MOLECULES AND THE STRUCTURED DEME

Michod (1983) was the first to argue that a looser form of population structure could have been important in the selection against selfish genes in a prebiotic context. Szathmáry (1992a) showed that the same mechanism could ensure coexistence of competitive, useful templates as well. Below I follow the short account of these models as given in Szathmáry (1994). Let us imagine the following situation: Templates replicate at a surface, maybe on pyrite (Wächtershäuser 1992), where they grow and interact in semi-isolated groups. They are regularly washed away, become perfectly mixed,

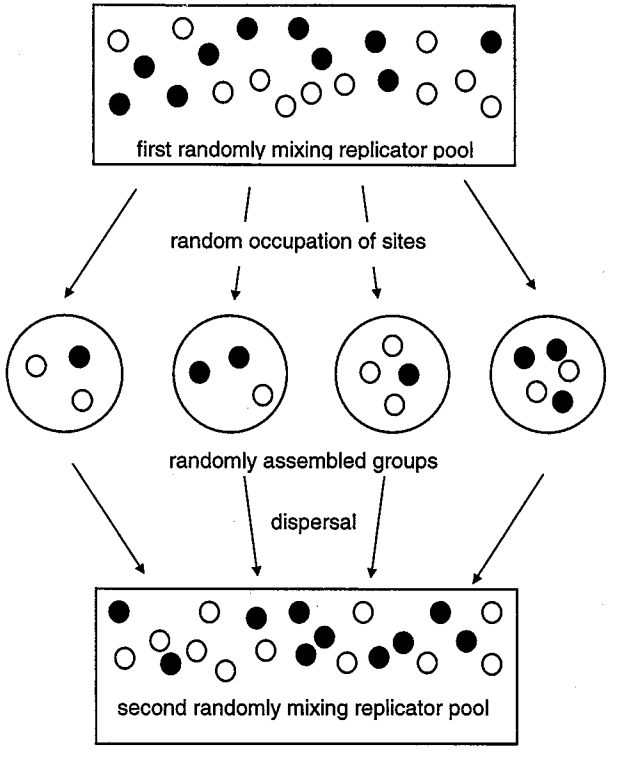

FIG. 3.5. The trait group model as applied to protobiology. *Open and closed circles:* two different replicators (an altruist and a parasite) or two complementing genes.

and are then readsorbed into the surface. Such a situation is similar to the one visualized in D. S. Wilson's (1980) trait-group (structured-deme) model (fig. 3.5).

The coupling among the templates is nonhypercyclic. Instead, we consider a so-called metabolic system (cf. fig. 45 in Eigen and Schuster 1978). The templates are assumed to contribute to metabolism via enzymatic aid, and metabolic products are in turn used up by the templates for replication at different rates. Although all templates contribute to metabolism ("the common good"), they are able to use it with different efficiency. Thus, in a spatially homogenous environment, competitive exclusion follows, despite the metabolic coupling (Eigen and Schuster 1978).

This is not so in the context of the structured-deme model. Local sites on which the relative proportions of templates are closer to optimal (hence metabolism is run more efficiently) yield more new templates, because more building blocks for reproduction are produced. Although differential replication rates of templates cause a shift in relative proportions, the shifted distribution will still not be too far from optimal. In contrast, local sites with a very unfavorable template composition will yield few new templates. In

sum, sites closer to the optimal template composition contribute more to the template pool of the new homogeneous phase. The outcome is a "protected polymorphism" of all different templates, deviating strongly from the fatal homogeneous case (Szathmáry 1992a).

Selection Dynamics in Two Dimensions: Hypercyclic and Metabolic Systems

Interesting selection dynamics also occurs when molecules are bound to the surface without being washed away regularly. This problem is modeled by the use of "cellular automata." Without becoming too technical, it suffices to say that each square of a grid is assumed to be occupied by a single molecule (template) or to be empty. Templates can do two things: replicate (put an offspring into a neighboring empty cell, if available) or hop away into empty sites nearby. Replication may depend on the composition of the few neighboring cells. In the case of a hypercycle, for example, the template and a specimen of the preceding cycle member must be present in the same small area if replication of the former is to occur. This, of course, makes perfect chemical sense.

Boerlijst and Hogeweg (1991) simulated hypercycles on a surface exactly in this way. They found that rotating spirals on the surface do appear, provided the hypercycle consists of more than four members. This is linked to the fact that such a hypercycle without population structure shows sustained oscillation in time. Each wing of a rotating spiral looks a bit like the arm of a galaxy and is dominated by templates of the same membership in the hypercycle. Parasites are unable to kill the hypercycle in that system. This finding was attributed to the dynamics of spirals. Two questions emerge: Are spirals necessary? What happens if one models other systems in the same way (i.e., by cellular automata)?

Czárán and Szathmáry (1999) managed to show that, given such a spatial setting, nonhypercyclic systems are once again viable alternatives. The fundamental difference between their model and that of Boerlijst and Hogeweg (1991) is that the dynamical link among the replicators is realized through a common metabolism instead of direct, intransitive hypercyclic coupling. Using the cellular automaton model of the metabolic system, the aim was to show that (1) metabolic coupling can lead to coexistence of replicators in spite of an inherent competitive tendency; (2) parasites cannot easily kill the whole system; (3) complexity can increase by natural selection. The result—there is coexistence without any conspicuous pattern (i.e., something like spirals)—is robust and counterintuitive. It results from the inherent discreteness (i.e., the corpuscular nature of the replicator molecule populations) and spatial explicitness of the model, which grasp essential features of the living

world in general, and macromolecular replicator systems in particular. An inferior (i.e., more slowly replicating) molecule type does not die out because there is an advantage of rarity in the system: a rare template is more likely than a common one to be complemented by a metabolically sufficient set of replicators in its neighborhood.

The general importance of surface dynamics seems more and more important for the origin of life in general. As Wächtershäuser (1992) pointed out, chemical evolution leading to more and more complicated networks, is likely to have taken place on the surface, especially on that of pyrite. Surface dynamics of replicators with indefinite heredity is a natural outgrowth of this "primordial pizza" dynamics (cf. Maynard Smith and Szathmáry 1995).

Protocells and Group Selection of Replicators

The phase of evolution just outlined refers to the precellular level. Later in evolution protocells must have appeared. Cellularization offers the most natural, and at the same time most efficient, resolution to Eigen's paradox. It also leads to the appearance of linkage, that is, the origin of chromosomes. The dynamics of genes encapsulated in a reproductive protocell are described by the *stochastic corrector* model (Szathmáry and Demeter 1987; Szathmáry 1989a,b), which rests on the following assumptions (fig. 3.6).

1. Templates contribute to the fitness of the protocell as a whole, and there is an optimal proportion of the genes. Concretely, we assume that genes encode enzymatic aid given to intracellular metabolism.
2. Templates compete with each other within the same protocell. As before, replication rates may differ from gene to gene.
3. Replication of templates is described by stochastic means. Since the number of genes in any compartment is small (up to a few hundred), their growth is affected by the plays of chance. Ecologists would express this as demographic stochasticity.
4. There is no individual regulation of template copy number per protocell.
5. Templates are assorted randomly into offspring cells upon protocell division.

I must emphasize that, in the stochastic corrector model, the templates are not coupled to one another through a reflexive (intransitive) cycle of replicational aid, because that would be a hypercycle. Instead, we assume that they contribute to the common good of the protocell by catalyzing steps of its metabolism. Within each compartment, the templates are free to compete, because they can reap the benefits of a common metabolism differently. (A

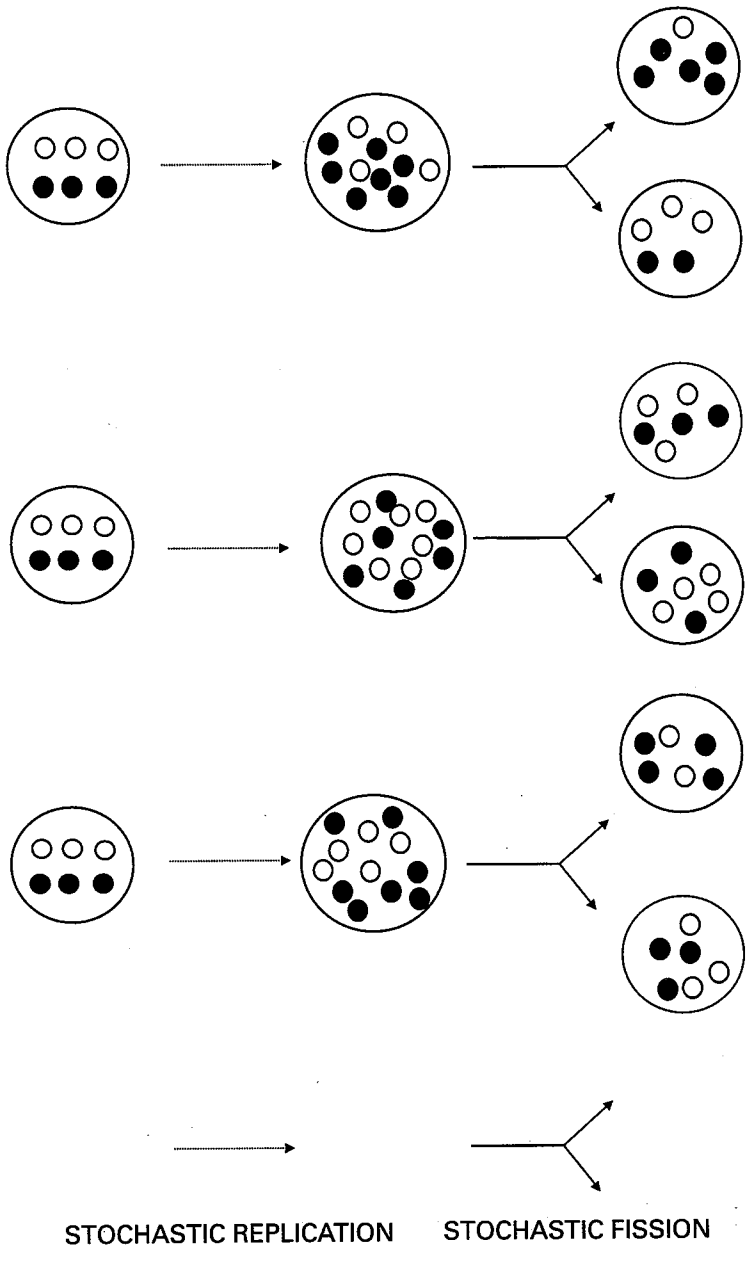

FIG. 3.6. The stochastic corrector model. *Open and filled circles:* two types of useful gene in the protocell, albeit with different within-cell replication rates. It is assumed that a 3:3 composition at the start of the protocell cycle is optimal for metabolism. Note that two such compartments (with bold cell envelopes) recur upon cell division, despite internal competition.

similar situation can arise among chromosomes and plasmids in contemporary bacteria.) Even though templates compete, the two sources of stochasticity generate between-cell variation in template copy number, on which natural selection (between protocells) can act. This is an efficient means of group selection of templates, because the protocells are the groups obeying the stringent criteria: (1) there are many more groups than templates; (2) each group has only one ancestor; (3) there is no migration between groups (cf. Leigh 1983). Grey et al. (1995) gave a fully rigorous reexamination of the stochastic corrector model. The two mentioned sources of stochasticity effectively lead to the correction of a malign within-protocell trend of harmful competition of the templates. It is interesting to see that genuine group selection is likely to have aided a major transition from naked genes to protocells. Group structure is provided by the physical boundaries of cells.

Within the same context, the origin and establishment of chromosomes (linked genes) in the population have also been analyzed (Maynard Smith and Szathmáry 1993). A chromosome consisting of two genes takes about twice as long to be replicated as do single genes. Chromosomes are strongly selected for at the cellular level even if they have this twofold within-cell disadvantage. Linkage reduces intracellular competition (genes are necessarily replicated simultaneously) as well as the risk of losing one gene by chance upon cell division (a gene is certain to find its complementing partner in the same offspring cell if it is linked to it). The molecular biology of the transition from genes to chromosomes has also been worked out (Szathmáry and Maynard Smith 1993a).

Simulations also show that sex is detrimental to protocells before the establishment of chromosomes because sex allows the horizontal spread of selfish genes (Szathmáry and Maynard Smith in prep.). Without sex, parasitic genes are passed on in a clonal manner and can be efficiently selected against. This is analogous to the problem of intracellular parasites and selfish organelles: There is selection for uniparental inheritance because it reduces intragenomic conflict (Eberhard 1980; Cosmides and Tooby 1981).

Hypercyclic versus Nonhypercyclic Compartments

Manfred Eigen aptly recognized the need for some kind of coupling among unlinked genes in early evolution. His concrete suggestion, the hypercycle, like other cooperative systems, is viable only with local interactions. One can justifiably ask the following questions. Do we need hypercycles or not? Did they play a role in prebiotic evolution?

The answer is that we do not know with certainty. For a long time, I have not favored hypercycles, partly because they have been oversold and sometimes misleadingly interpreted. More important, it seems that the stochastic corrector mechanism can solve the conundrum of dynamic coexistence of

unlinked genes, the very problem that initiated the development of hypercycle theory. One must bear in mind that hypercycles are ecologically stable, and thus they could present a within-protocell copy-number regulation mechanism. Such a mechanism would be favored by natural selection in a population of compartments described by the stochastic corrector model.

The snag is that *any* such mechanism would be selected for. There are literally an infinite number of ways whereby copy-number control could be acheived through density-dependent growth. As I pointed out before (Szathmáry 1989a), the hypercycle is a particularly wasteful means of information integration (or copy-number control), for the following reasons.

1. It can produce shortcut mutations that make the cycle arbitrarily shorter. This type of mutation simply does not exist for nonhypercyclic systems.
2. Even hypercycles would have had to escape into compartments sooner or later (Eigen et al. 1981). For the compartment to thrive, the members of the hypercycle would have had to encode a function additional to the replicase activity: This could have been enzymatic aid of metabolism (similar to the internal working of the stochastic corrector model). It is unlikely that in general the same gene could encode both functions. Thus, members of the hypercycle would have been twice as long as the genes of other systems. This would have entailed a much higher mutational load for hypercycles.
3. By the same token, the practical doubling in the number of genes would have led to an excessive metabolic burden also.

All in all, although hypercycles may have played an episodic, incipient role in early evolution, they are unlikely—in my opinion—to have had any decisive role, because alternative systems could have fulfilled the same role and at a reduced cost in terms of mutations and metabolites.

The Size of the Genetic Alphabet in a Metabolically Complex RNA World

An assumption of the stochastic corrector model is that the templates catalyze the reactions of intermediate metabolism, which in turn produces the monomers for template replication and the building blocks of the encapsulating membrane. It could be true that this catalytic help is indirect: Like contemporary nucleic acids, the templates could encode protein enzymes. This would, of course, require protein synthesis through translation. The problem is that translation is a highly evolved process: One does not easily see how a protocell could have had it to begin with. Happily, the idea of an "RNA world" (Gilbert 1986) brings us closer to bridging a gap between purely nonenzymatic systems and those based on protein enzymes. An extended version of this hypothesis—one that I favor—holds that the steps of intermediate metabolism were catalyzed by ribozymes (RNA molecules with enzymatic activity), an idea that goes back to suggestions by Woese (1967), Crick (1968), and Orgel (1968). If this hypothesis turns out to be true, it will bridge one of the most unpleasant gaps in evolutionary reasoning. Remarkably, it may also give an answer to the basic question, Why are there only four bases in the genetic alphabet?

The RNA World and Selection of Ribozymes

As I said, I favor the view that the RNA world was metabolically complex (Benner et al. 1987, 1989). Because there are almost no metabolic ribozymes known from contemporary organisms, what makes this suggestion nonetheless credible? Apart from arguments based on comparative biochemistry, it is the success of in vitro ribozyme genetics (for a review, see Szostak 1992) as was suggested for RNAs binding small ligands (Szathmáry 1984) and catalyzing reactions (Szathmáry 1989a, 1990a,b).

Very briefly, the experimental protocol yielding RNA molecules with specific functions consists of the following steps.

1. Choose a molecule (a so-called ligand) for which you want to generate RNAs that specifically bind to it. Simplifying somewhat, this molecule can be a substrate of a reaction that you want to be catalyzed by the RNAs generated.
2. Bind this molecule chemically to an indifferent material that will act as a vacuum cleaner collecting RNAs that bind the ligand.
3. Generate a pool of RNAs with different sequences.
4. Pour a solution of these RNAs over the material presenting the ligand.
5. Wash away RNAs that do not bind to the ligand.
6. Replicate those RNAs that bind the ligand.
7. If binding is strong enough, analyze the resulting RNA molecules; otherwise go to step 4.

By now, several different RNAs with enzymatic activity (ribozymes) have thus been selected (for a review, see Szathmáry and Maynard Smith 1997). It seems likely, therefore, that ribozymes were able to run a complex metabolism. Why then did protein enzymes replace most ribozymes? The consensus holds that the number of functional chemical groups provided by the 20 amino acids gives a definite advantage to proteins over ribozymes, the latter having only four building blocks (cf. Wong 1991). Obviously, one can make more versatile catalytic molecules with more building blocks.

There are two ways of increasing the catalytic potential of RNA-like molecules: (1) increasing the number of monomer types, and (2) post-synthetic modifications. I discuss these cases in turn.

Novel Nucleic Acid Building Blocks

The beautiful results of Piccirilli et al. (1990) show that one can have as many as 12 different bases, forming six Watson-Crick-type base pairs, if one varies the hydrogen acceptor and donor groups on the monomers (bases). One of the theoretically possible novel base pairs was synthesized by Piccirilli et al. (1990). Polymerase enzymes do accept it in template-dependent reactions! Why is it then that we find only two base pairs in contemporary RNAs? The answer, as pointed out by Orgel (1990), is that (1) either Nature has never experimented with more than two base pairs, or (2) she decided that two were enough. Although we may never know whether (1) holds, rather forceful arguments have been put forward in favor of (2).

According to the suggestions of Fontana et al. (1991) two base pairs in natural nucleic acids is seen as a compromise between stability against mutations and thermodynamic stability. In general, the secondary structures of GACU sequences are richer and more variable than the corresponding structures built exclusively of GC or AU sequences. Calculated replication rate constants are maximal for much shorter chains for GC than for GACU. Because GC sequences form base pairs more readily than GACU sequences, the phenotype (in the context of the model the two-dimensional structure) of the former is less stable against random mutations. In contrast, three base pairs make it difficult for random sequences to fold into stable structures (for a review, see Schuster 1993).

A snag with the above explanation is that evolution of functional RNAs probably did not always proceed from random RNA sequences. An alternative, complementary, approach considers the fitness of ribo-organisms as a function of the size of the genetic alphabet, noting that catalytic efficiency of ribozymes increases with the number of letters, whereas copying fidelity of such molecules must decrease with it. (This decrease is intuitively obvious considering that it is easier to err when more, closely related molecules from a set are present.) The increase is slower than, and the decrease is faster than, exponential. Hence, there is an evolutionary optimum at a certain number of base pairs (Szathmáry 1991b, 1992b); some considerations even suggest that this optimum may indeed lie at 2. Thus, this trait may be a footprint of the decisive role of natural selection (cf. Leigh 1995) in molecular evolution.

Replicators and Reproducers: From Simple Autocatalysts to Chemotons

I presented the problems associated with the origin of life in the broader context of the major evolutionary transitions (Maynard Smith and Szathmáry 1995; Szathmáry and Maynard Smith 1995). How do higher-level evolutionary units emerge from the ones at lower levels? How did the storage and

usage of hereditary information change? How did division of labor play an important role? As Griesemer (1996) aptly noticed, the frequent usage of the term "replicator" in the context of "comparative transitionology" (cf. Bonner 1995) is partly unwarranted. In many cases, the units that these authors were referring to were not replicators *sensu* Dawkins (1976) at all: Whole genomes, symbiotic organelles, cells within organisms, and sexual organisms within societies are certainly always vehicles, but rarely replicators. Their structure is usually not transmitted through copying. Reproduction of a whole mitochondrion is not replication.

In the present context, I prefer to use the term *reproducer* rather than *vehicle* (Dawkins 1976) or *interactor* (Hull 1980) because nothing in the latter concepts would suggest that organisms can be units of evolution. Concepts such as reproduction, heredity, and variability can be meaningfully applied to them. The term *reproducer* spells this out while emphasizing the difference to replication.

The origin of life itself (cf. Gánti 1997) is synonymous with the appearance of a certain type of chemical supersystem, the model of which is the chemoton (e.g., Gánti 1975, 1979, 1987). It consists of three autocatalytic subsystems: a metabolic network providing building blocks for the other two subsystems; a population of replicating templates; and an encapsulating membrane, essentially a protocell. The system as a whole is also autocatalytic and reproducing at the same time, but it is *not* a replicator. The only replicator playing a significant role from an orthodox Dawkinsian point of view is the template macromolecule. Although it is true that the intermediates of the metabolic cycle undergo processive replication at the molecular level, what is passed on to the offspring chemotons is a *population* of these replicators, that is by no means a replicator itself. By the same token, a similar reasoning applies to the genome of the chemoton: A bag of genes, undergoing random segregation into the offspring protocells does not undergo replication at the level of the bag. (Note, by the way, that when I conceived the stochastic corrector model, I had the chemoton in mind as a chemically detailed protocell model. Templates are imagined to catalyze the reaction steps of the metabolic cycle of A_i molecules.)

What would be the appropriate term to use in the context of transitionology? Griesemer suggests it is the reproducer. Reproducers often qualify as units of evolution, provided they have heredity and variation (for further discussion, see Szathmáry and Maynard Smith 1997). I think that the reproducer is a useful concept. It remains true that frequently a gene-centerd approach, like that of Williams (1966a) and Dawkins (1976), is extremely rewarding in the analysis of the spread of alleles in various contexts. It is also true, however, that (1) it is reproducers, rather than replicators, of a higher level that arose during the transitions; (2) when a higher-level reproducer appears, a novel type of development is worked out; and (3) rather old-fashioned replicators are packaged into novel reproducers.

One can summarize this chapter by considering the origins and functions of the subystems of the chemoton. The origin of the membrane has not been dealt with here in terms of chemistry (see Maynard Smith and Szathmáry 1995), but we have seen that it provides the most stringent version of group selection by acting as a physical barrier to gene flow and forcing local interactions on the constituent genes. Local interactions were important before cellularization as well. Molecular cooperation of naked, unlinked replicators is inconceivable without such an effect.

The first replicators are likely to have emerged on mineral (probably pyrite) surfaces. The templates in the chemoton are assumed to gain ribozymic function aiding metabolism of the compartment. Obviously, evolution must have started before enzymes and templates. Simple autocatalysts, replicating in a holistic manner and having limited heredity (and thus limited evolutionary potential) are likely to have been the first replicators. Replicators carrying information in digital form appeared when replication became modular, rather than holistic. We know from experiments that such replicators can form a growing population in the absence of enzymes, but we do not know the evolutionary pathway to the appearance of the first RNA molecules. Nucleic acids are important because they are replicators with unlimited heredity. There can be so many types (sequences) that evolution may go on indefinitely.

Metabolism of autotrophic protocells hinges on the presence of an autocatalytic network of small molecules, modeled by the central cycle of the chemoton. At some point in evolution, nucleic acid replication became grafted onto such a metabolic network. Systems in which nucleic acids acted as ribozymes had an advantage: Metabolism and replication proceeded faster. The enzymatic function of RNAs in the has-been RNA world probably led to the fixation of the size of the genetic alphabet; accuracy decreases too fast if the number of base pair types is increased. The advantage from an increase in catalytic potential cannot compensate for this adverse effect.

Many experiments along the lines suggested by the current theories have not been carried out yet. This is an important task for the future. One would like to see a sensible scenario for the chemical origin of nucleic acids. Then one would like to see the spontaneous formation of protocells (chemotons) in the lab. Once we have achieved this, we shall have understood how life originated.

Acknowledgments

This work was supported in part by the Hungarian Scientific Research Fund (OTKA) and by a grant from the Ministry of Culture for research in higher education.

4 Individuality, Immortality, and Sex

Richard E. Michod

The emergence of new and higher levels of organization during evolution provides a compelling context for understanding the relations among certain fundamental properties of life, such as individuality, immortality, and sex. By *immortality*, I mean following Weismann (1890) the never-ending cycle of life; by *individuality*, I refer to the familiar levels in the hierarchy of life and their capacity to function as units of selection during evolution (genes, cells, organisms, societies, species); and by *sex*, I mean breakage and reunion of DNA molecules from different individuals.

The evolution of multicellular organisms is the premier example of the integration of lower evolutionary levels into a new, higher-level individual. Explaining the transition from single cells to multicellular organisms is a challenge for evolutionary theory. Sex and individuality are in constant tension as new units emerge, because sex mixes elements from different individuals and naturally threatens the integrity of evolutionary units. Yet, sex is fundamental to the continued well-being of evolutionary units and the immortality of life (Michod 1995). Although sex by creating mixis would seem to undermine individuality, history shows that sex is reinvented as each new level of individuality emerges in the evolutionary process.

Cooperation and Conflict

The benefits of cooperation provide the imperative for forming new, more inclusive evolutionary units. Increments in fitness are traded among levels of selection through the evolution of behaviors that are costly to individuals yet beneficial to groups. Cooperation is necessary for the emergence of new units of selection, precisely because it trades fitness from the lower level (the costs of cooperation) for increased fitness at the group level (its benefits). In this way, cooperation can create new levels of fitness and individuality (see table 4.1). This trade, if sustained through group selection, kin selection, and conflict mediation, results in an increase in the heritability of fitness and individuality at the new higher level. In this way, new higher levels of selection may emerge in the evolutionary process.

Although fueling the passage to higher levels, cooperation provides the opportunity for its own undoing through the frequency-dependent advantage of defection. Selfish interactions (defection) reap the benefits of cooperation while avoiding the costs and, for this reason, can be expected to spread