

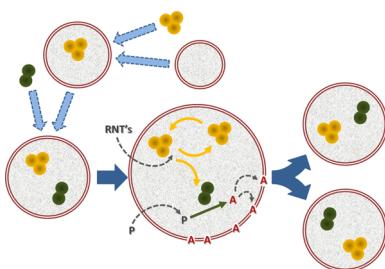
## Prebiotic Systems Chemistry: New Perspectives for the Origins of Life

Kepa Ruiz-Mirazo,<sup>†,||</sup> Carlos Briones,<sup>‡,||</sup> and Andrés de la Escosura<sup>\*,§</sup>

<sup>†</sup>Biophysics Unit (CSIC-UPV/EHU), Leioa, and Department of Logic and Philosophy of Science, University of the Basque Country, Avenida de Tolosa 70, 20080 Donostia—San Sebastián, Spain

<sup>‡</sup>Department of Molecular Evolution, Centro de Astrobiología (CSIC–INTA, associated to the NASA Astrobiology Institute), Carretera de Ajalvir, Km 4, 28850 Torrejón de Ardoz, Madrid, Spain

<sup>\*</sup>Organic Chemistry Department, Universidad Autónoma de Madrid, Cantoblanco, 28049 Madrid, Spain



### CONTENTS

1. Introduction	285	3.3.3. RNA Virus Quasispecies as a Model System for the RNA World	330
1.1. A Historically Controversial Field	286	4. Systemic Integration Approaches	331
1.2. Systems Chemistry as a New, More Encompassing Perspective	287	4.1. Template–Metabolism Integration	332
2. Chemical Pathways to Biomolecules	288	4.2. Boundary–Template Integration	333
2.1. Prebiotic Synthesis of Monomers: Lipids, Amino Acids, and Nucleotides	289	4.3. Metabolism–Boundary Integration	335
2.1.1. Synthesis of Lipids	289	4.4. Toward the Integration of Template, Metabolism, and Boundary	336
2.1.2. Synthesis of Amino Acids	290	5. Current Methodological Tools for the Challenge	338
2.1.3. Synthesis of Nucleotides	292	5.1. Dynamic Combinatorial Chemistry	338
2.1.4. Summary	297	5.2. High-Throughput Biochemical Techniques: Sequencing Technologies and Microarrays	342
2.2. Prebiotic Synthesis of Polymers: Peptides and Nucleic Acids	297	5.3. In Vitro Evolution of Nucleic Acids and Other Biomolecules	346
2.2.1. Synthesis of Peptides	297	5.4. Microfluidic and Nanofluidic Approaches	347
2.2.2. Synthesis of Nucleic Acids	300	6. Concluding Remarks and Open Questions	349
2.2.3. Coevolution in the Synthesis of Peptides and Nucleic Acids	301	Author Information	351
2.3. Origins of Homochirality	303	Corresponding Author	351
3. Complex Chemical Processes on the Way to Living Systems	308	Author Contributions	351
3.1. Emergence of Complex Chemical Behavior: Self-Organization and Self-Assembly	308	Notes	351
3.1.1. Oscillatory Reaction Dynamics	308	Biographies	351
3.1.2. Autocatalytic Networks and Protometabolic Cycles	310	Acknowledgments	352
3.1.3. Assembly of Amphiphilic Molecules into Protocellular Compartments	313	References	352
3.2. Reproduction and Replication Processes	315		
3.2.1. Reproduction of Vesicles	315		
3.2.2. Networks of Replicating Polymers	316		
3.3. Development of Biological Evolutionary Mechanisms: The RNA World	318		
3.3.1. Pre-RNA Worlds	322		
3.3.2. The RNA World Hypothesis: Combining Genotype and Phenotype at the Molecular Level	323		
	327		

### 1. INTRODUCTION

The origin of life is a fascinating, unresolved problem whose deciphering would have important scientific, epistemological and social implications. Over time, philosophers and scientists have proposed many different theories for the origin (or origins) of life, but we are still far from understanding which principles governed the transition from inanimate to animate matter. Among the scientific fields involved in this interdisciplinary endeavor (i.e., biology, chemistry, physics, and earth and planetary sciences), chemistry holds a most relevant position. After all, living beings are mainly composed of water and organic molecules, and any theory or model about the origins of life will ultimately have to be confronted with experimental data on the transition from complex enough inert chemical systems to sufficiently simple living ones.<sup>1</sup>

Numerous problems in the search for life's origins demand solutions that should come from chemistry.<sup>2</sup> First, plausible synthetic pathways toward the molecules of life (i.e., lipids, sugars, amino acids, nucleotides, etc.) or their precursors have

Received: December 30, 2011

Published: October 31, 2013

to be found and be compatible with the conditions that were possibly present on the early Earth, in the context of a young solar system, and within the general framework of astrophysical and astrochemical evolution. Second, prebiotic conditions for the efficient polycondensation of amino acids and nucleotides, in heterogeneous aqueous solutions or in interfaces with water-based media, are needed to explain the emergence, without the aid of biological catalysts, of the first functional biopolymers (e.g., polypeptides and polynucleotides). Third, the appearance and development of cellular membranes should be accounted for, including not only the physicochemical properties of supramolecular systems such as vesicles and micelles (i.e., noncovalent self-assembled aggregates of prebiotic amphiphilic molecules) but also more stable and elaborate cellular boundaries, made of internally synthesized lipids and other surfactant compounds (e.g., peptides or polycyclic aromatic hydrocarbons). These boundaries would allow the enclosed protocellular system to continuously grow and divide. Fourth, the ability of nucleic acid molecules or their precursor polymers to replicate themselves and be coupled to the reproduction of the compartment should also be subject to analysis by means of chemical and biochemical methods. Finally, mineral surface adsorption or catalysis may have been important and, thus, should also be considered.

In this review we summarize the main advances of almost a century of scientific research on this fundamental problem, highlighting at the same time the central questions that remain unanswered. In this way, the limitations of traditional synthetic chemistry approaches to the problem of the origins of life will become apparent, while the potential of new methodologies and perspectives, embraced in what is beginning to be called the systems chemistry approach,<sup>3,4</sup> will be explored and discussed with special interest.

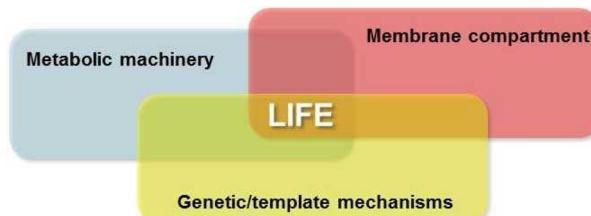
### 1.1. A Historically Controversial Field

The first major difficulty encountered when facing the problem of how life originated is that there is not a general consensus on what life is and on how it should be defined. It is clear, since the pioneering work by Woese and Fox,<sup>5</sup> that all the current biodiversity is the outcome of Darwinian evolution from a primitive cellular species, the so-called last universal common ancestor or LUCA. However, is it possible to define life scientifically, with strictly chemical and biological criteria?

The complete picture and implications of this issue come out only when we try to specify the requisites that, in principle, any type of system (i.e., not only an organic chemical one) should actually fulfill to be considered alive. Opening or generalizing the problem of the nature of life, and thus of its origins, makes it richer, wider, and more challenging, as can be reflected in the recent merging of the traditional field of origins of life and the younger ones of synthetic biology and astrobiology. Indeed, the main questions addressed by astrobiology are the origins, evolution, and distribution of life in the universe.<sup>6–8</sup>

However, we do not aim to discuss here in detail the issue of defining life: the reader is referred to a special issue of the journal *Origins of Life and Evolution of Biospheres*,<sup>9</sup> or to a comprehensive anthology of articles on this subject.<sup>10</sup> For the purposes of this review, we will assume that a general consensus is being reached, which allows us to consider that a living organism must possess, at least, the following three properties:<sup>11</sup> (i) the ability to process and transmit heritable information to progeny (i.e., a genetic mechanism); (ii) the ability to capture energy and material resources, staying away

from thermodynamic equilibrium (i.e., metabolic machinery); and (iii) the ability to keep its components together and distinguish itself from the environment (i.e., cell membrane). Joint consideration of these three properties (Figure 1) allows us to operationally define living systems as autonomous entities with the capacity for open-ended evolution.<sup>12</sup>



**Figure 1.** The three main interdependent components of life.

Nevertheless, even under this assumption, investigations on the origins of life are riddled with numerous controversial questions. Where on Earth did life emerge? Or did life arrive here from some extraterrestrial source instead? Did life originate only once, by accident, or is it the probable outcome of chemical evolution, which has frequently occurred elsewhere in the universe? Which property of living beings came first: their ability to reproduce and transmit information to progeny, their metabolic capacity, or their compartmentalization as individual entities? Were the first living organisms based on autotrophic or heterotrophic metabolic cycles (i.e., were they mainly feeding from inorganic or organic materials)? Resolving these questions and other related ones is extremely complicated because of our lack of knowledge about the conditions existing on Earth more than 3.5 billion years ago (the closest time estimated for the origins of life, from palaeontological evidence).<sup>13</sup> In fact, the range of possible values for many relevant physical and chemical parameters (temperature, radiation dose, pressure, amount of liquid water, pH, ionic strength, etc.) is remarkably wide in different proposed prebiotic scenarios, as we shall see. This, in practice, has forced researchers to turn the question around: instead of trying to determine the exact prebiotic conditions under which experiments should be conducted, they usually propose a set of working hypotheses and evaluate them according to the degree of success in their results (e.g., the yields of a compound from a certain synthesis). Indeed, when we review the different achievements in the field over the last few decades, the issue of compatibility becomes a major difficulty: diverse experimental settings, each relatively successful in dealing with a partial aspect of the problem, often turn out to be incongruent.

Some of the previous questions, in addition, constitute paradoxes of the “chicken-and-egg” type. The most paradigmatic of them is the controversy between nucleic acids-first and proteins-first scenarios, given that in current living systems proteins are needed for DNA replication and, in turn, nucleic acids are required for the biosynthesis of proteins. This was partially overcome with the advent of the “RNA world” hypothesis (see section 3.3), but most of the prebiotic pathways to the synthesis of RNA-like compounds have so far proved difficult. A related debate in the field has been whether replication<sup>14</sup> or metabolism<sup>15</sup> arose first. In order to solve such controversial questions, many different and sometimes antagonistic theories have been proposed during the second half of the 20th century. Most of them can be classified into two

main groups: gene-first (e.g., the aforementioned RNA world hypothesis)<sup>16</sup> and metabolism-first (e.g., Kauffman's autocatalytic networks)<sup>17</sup> theories. Much attention is also being paid, since the "lipid world" theory was put forward a decade ago,<sup>18</sup> to cellular boundaries, increasingly considered of crucial importance for the emergence of life, as Oparin originally claimed. Nevertheless, none of these views has yet gained a unanimous preferential position over the others, each having its own shortcomings.

When these various difficulties are considered, it is unlikely that scientists will ever know which exact synthetic itinerary led to the first forms of life. A nonhistorical point of view might be more fruitful, the target of research turning to be the general physicochemical processes that could trigger the transition from a nonliving chemical system into a protoliving one and, finally, into a living organism. Along those lines, only a fundamental understanding of the physics and chemistry that could make possible this transition will allow us to solve the major questions still lying at the heart of the origins-of-life mystery.

More precisely, several key problems faced by most theories on the origins of life are related to the different ways in which chemical and living systems deal with thermodynamic constraints. Even the simplest microorganisms known on Earth are breathtakingly complex. Indeed, the probability that a random sequence of physicochemical events would lead to a bacterium by spontaneous self-organization of biomolecules is negligibly low. This would also be the case if we tried to imagine the spontaneous synthesis of relatively simpler, nonautonomous replicative entities, such as viruses or viroids. The whole process of molecular self-organization toward protoliving or living systems seems to be against the second law of thermodynamics, although we know that an open system can generate ordered cycles at the expense of its increasingly disordered environment, that is, in continuous matter–energy flow conditions.<sup>19,20</sup> The organization of living beings is much more intricate than any other known spontaneously generated pattern of spatial/temporal order, as in the so-called "dissipative structures".<sup>21</sup> Life manages to stay far from equilibrium due to molecular mechanisms that are much more diverse, complex, and also based on quasi-equilibrium structures. Living beings are therefore able to combine self-organization and self-assembly processes,<sup>22</sup> keeping many of the resulting molecular ensembles (i.e., polymers, membranes, etc.) just at the edge of a transition phase, so that their behavior can be more easily adjusted to changes in the surrounding conditions.

Eschenmoser<sup>23</sup> and Stano and Luisi<sup>24</sup> have posed this question as how to operate a transition from thermodynamic to kinetic control. Similarly, Pross<sup>25,26</sup> states that the structural and reactivity features of inanimate matter are dominated by thermodynamic considerations, with kinetics remaining secondary attributes. In contrast, living systems are, essentially, a kinetic state of matter.<sup>27</sup> They do not tend toward equilibrium but rather maintain a far-from-equilibrium state by continuous exploitation of an external energy source. This property of living systems could be due to the combination of their self-replication capability with a metabolic apparatus.<sup>28</sup> The ability to self-replicate would confer dynamic kinetic stability to chemical systems, which become capable of exhibiting exponential growth rates even if they do not constitute the most favorable thermodynamic state.

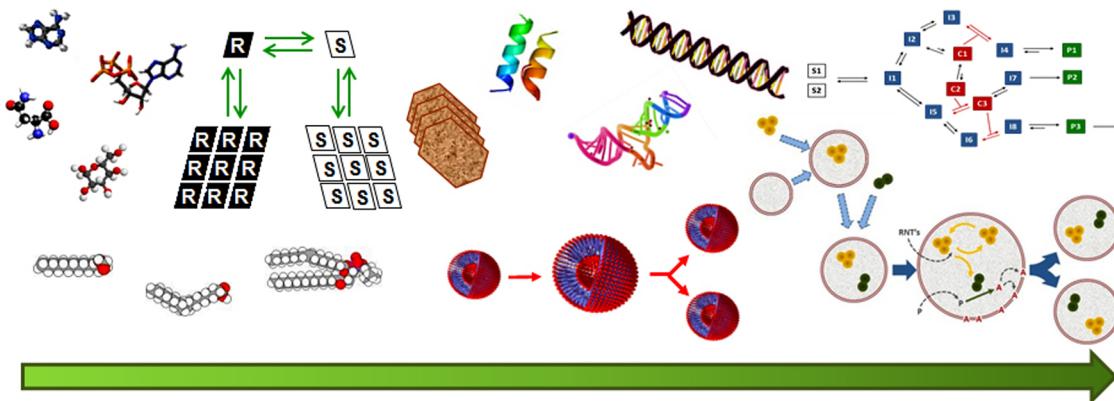
However, a simple chemical replicator is still a thermodynamically driven system, meaning that the replication reaction will continue only until the equilibrium concentrations of the

replicating molecule and its building blocks are reached. In contrast, living self-replicating systems, despite their high kinetic stability, keep far from equilibrium and are therefore thermodynamically unstable. It could well be, as Eschenmoser and Pross suggest, that the incorporation of metabolic capacity is what turns a replicator from being thermodynamically driven into kinetically driven.<sup>28</sup> This is exemplified by a metabolic molecular replicator or a bacterial cell that have consumed all nutrient resources: through an energy-gathering apparatus, both of them avoid the decay into a thermodynamic minimum and keep their replication capacity. Ganti<sup>11</sup> and Kauffman<sup>17</sup> would express this as the requirement for any living entity to perform work in order to maintain itself robustly in far-from-equilibrium conditions.

## 1.2. Systems Chemistry as a New, More Encompassing Perspective

In light of these ideas, presenting the problem of the origins of life as a dichotomy between a network of autocatalytic cycles (i.e., the basic idea of the metabolism-first theories) or a population of self-replicating molecules (i.e., the gene-first scenario) is perhaps misleading. The overall self-organization process of a complex mixture of molecules into a living chemical system must have been a stepwise dynamic process of kinetic aggregation, from which decay into thermodynamic minima was probably very easy at different stages. Theories proposing a purely metabolic or genomic starting point actually encounter numerous theoretical and experimental objections, mainly because the type of initial system they propose has a much higher probability to fall into equilibrium than to originate a sustainable, kinetically driven process leading to life. Similarly, theories that do not take into account cellular boundaries, or some sort of adsorption surfaces able to delimit individualized microenvironments, have to face important thermodynamic objections.<sup>29</sup> Dilution of a self-replicating molecule and its building blocks would reduce the rate of the replication reaction, ultimately returning the system to thermodynamic control. Furthermore, compartmentalization of replicating and metabolic systems within protocellular structures may have increased the selectivity of chemical reactions involved in these processes. Membrane-based compartments, in turn, could have taken advantage of ion gradients across the membranes, whose potential energy might have been further coupled to the production of chemical energy useful for an emerging metabolism.

In order to establish a thermodynamically plausible pathway toward the far-from-equilibrium complexity of the first cells, or protocells, a more logical approach would be to consider that a high diversity of precursor components was available on the prebiotic Earth and that these components could, by means of different physicochemical mechanisms and reaction pathways, progressively turn into, or be integrated into, primordial metabolic, self-replicating, and membrane-bounded subsystems that could combine into more complex systems. This view conceives the initial stages of the prebiotic Earth as a huge flow reactor containing an amazingly complex set of small molecules of different types, eventually establishing a wide variety of possible interactions among each other. Provided that there is an overall flow of matter and energy across such a supersystem, therefore sustaining far-from-equilibrium dynamics, this complex set could explore an immensely large number of possible reaction pathways. In practice, that exploration would be geochemically and geophysically constrained (by limitations in



**Figure 2.** General scheme of the review, from molecules to systems. Contents are structured following a tentative scheme according to which chemical complexity will unfold as biomolecules become involved in a higher number and wider diversity of transformation processes or network interactions.

resources and by the time required for different processes to actually take place), so kinetically driven pathways would have an advantage. One could even speak, as Pross<sup>26</sup> does, about a chemical selection process in which “fitness” means nothing but dynamic kinetic stability and “survival of the fittest” is simply the drive toward greater kinetic stability. In this context, reaction pathways in which constrained metabolic and self-replicating entities (e.g., individuals or networks) are formed would be kinetically selected over the others and the corresponding entities somehow amplified.

Such a scenario implies that finding thermodynamically plausible chemical pathways toward life necessarily requires looking into at least moderately complex mixtures of various potential molecular components of life, and studying the chemical and physical interactions among them (e.g., interconversion processes, condensation and polymerization reactions, supramolecular aggregation processes, propagating chemical oscillations, surface and colloidal effects, etc.).<sup>30–32</sup> This task has been no more than a dream for chemists for a long time, since complex chemical mixtures are very hard to work with experimentally. But what used to be absolutely intractable is becoming within reach: chemists now have a number of methodologies and conceptual frameworks, developed during the last 15 years, that allow us to think, for the first time, about manipulating and analyzing molecular mixtures of a fairly high degree of complexity. The emerging area of systems chemistry, which includes a number of subfields such as dynamic combinatorial chemistry, oscillating reactions, self-replicating networks, high-throughput biochemical analysis, microfluidics, *in vitro* evolution systems, and others, is opening new avenues of research that will shed light on the chemistry that occurs in multicomponent and heterogeneous mixtures.

The new conceptual framework of systems chemistry was raised by Ludlow and Otto in a review article in 2008,<sup>3,4</sup> although other articles that could be classified within this field had been disseminated in the literature over the previous decade. There are also a number of very well documented reviews, books, and various journals' special issues about the origin of life and the historical development of this field of research,<sup>33–44</sup> although many of them are biased by either the gene-first or the metabolism-first views. We believe that a comprehensive review, involving prebiotic chemistry directed to the synthesis, polymerization, and self-organization of all the possible components of a primordial cell, is required for a wider

and deeper understanding of the origins-of-life problem. A systematic evaluation of the thermodynamic and kinetic features, yields, and conditions necessary for the synthesis of amphiphilic molecules, sugars, amino acids and peptides, and nucleotides and nucleic acids may help to determine how these processes could have occurred and been integrated within a protocell under prebiotic conditions. Thus, a systems view will be adopted in this review to discuss the most recent advances in the field, with particular attention paid to the new set of methodologies that are enabling researchers to deal with complex mixtures of molecules. In any case, in order to set the stage for this approach, we will first summarize the main achievements in traditional prebiotic chemistry and the basic theories that have addressed the problem of the origins of life. Then, we shall move on toward experimental systems with progressively higher levels of chemical complexity, understood in terms of diversity of compounds and types of interaction processes involved (Figure 2). In that journey, we will also pay attention to prebiotic surface chemistry, since many of the aforementioned processes may have been assisted by mineral surfaces,<sup>45</sup> as we said earlier, and/or soft organic interfaces (e.g., colloids, water–lipid emulsions, etc.).<sup>46</sup>

## 2. CHEMICAL PATHWAYS TO BIOMOLECULES

All modern organisms contain three principal types of functional molecules: (i) nucleic acids, responsible for the storage and transfer of genetic information; (ii) proteins, which sustain metabolic networks by catalyzing their elementary reaction processes and regulating/coordinating global system behavior; and (iii) lipid molecules that form cell membranes, establishing boundaries with the environment as well as the necessary coupling mechanisms with it. In order to explain the origin of life, taking as an assumption that the first organisms were composed of the same types of molecules, one of the major goals of the field should be finding plausible pathways toward the constituents of nucleic acids, proteins, and membrane bilayers. Since the pioneering work by Miller and Urey in 1953,<sup>47</sup> many experiments have been devoted to the synthesis of biological monomers under prebiotic constraints. Although these efforts have led to impressive progress in the field, prebiotic chemistry still faces problems, especially in the case of nucleotides, for which efficient synthetic routes have not been found. In this section, we will focus on laboratory approaches to the synthesis of lipids, amino acids, and

nucleotides in prebiotically plausible conditions. Other biomolecules such as saccharides, cofactors, and pigments (e.g., flavins, quinones, porphyrins, etc.) will not be considered in so much detail here, though they may have been important for primitive metabolism. For the interested reader, several reviews covering these aspects of prebiotic chemistry are available.<sup>48–50</sup>

## 2.1. Prebiotic Synthesis of Monomers: Lipids, Amino Acids, and Nucleotides

The set of simple organic molecules that could have been involved in the abiotic synthesis of biomonomers is relatively small.<sup>51</sup> The main starting materials in prebiotic chemistry are one-, two-, and three-carbon atom molecules, such as hydrogen cyanide, cyanate, cyanogen, formaldehyde, formamide, formic acid, ammonium formate, ammonium cyanide, urea, acetaldehyde, cyanoacetylene, and cyanoacetaldehyde. These compounds can indeed be synthesized by spark discharge, UV irradiation, or shock waves in various gaseous mixtures of methane, carbon oxides (CO and CO<sub>2</sub>), ammonia, nitrogen, and water, and they can be interconverted in different number of steps through hydrolytic, redox, or photochemical processes.<sup>51</sup> But from this pool of simple molecules, a myriad of more complex compounds could have formed, some of which may have benefitted, via organocatalysis, from mutual enhancement of the efficiency of their respective synthetic processes.

Organocatalysis<sup>52</sup> is a fast-growing field that will probably yield important results and insights for prebiotic chemistry. The rate constants of chemical reactions can be considerably increased by organic molecules (i.e., the catalysts) that bind substrates and decrease their activation energy, modifying the transition state of the reaction in which they are involved. This resembles the way in which more elaborate enzymatic catalysts operate, with binding energy compensating for the entropy cost occurring when small molecules are linked to produce more complex compounds. Such similarity suggests that the functions of enzymes currently catalyzing the formation of biomonomers might have been carried out, on the prebiotic Earth, by much simpler organic molecules.

A major concern regarding the synthesis of biological monomers is the availability of precursor materials under primitive Earth conditions. In this respect, two possible sources are usually considered: endogenous (i.e., terrestrial) synthesis and exogenous formation, either at different places of the solar system or in the interstellar medium. The total amount of extraterrestrial organic matter delivered to Earth, during the 100 million year period between the late Hadean and early Archean eras (i.e., the 3.9–3.8 Ga period), has been estimated to be on the order of 10<sup>16</sup>–10<sup>18</sup> kg.<sup>53</sup> Given that there is about 6 × 10<sup>14</sup> kg of organic matter in the present biosphere, it is likely that the role of extraterrestrial organics was significant, although we cannot be certain about the relative amounts and differential composition of endogenously synthesized and externally delivered organics. Extraterrestrial infall occurred in the form of interplanetary dust particles, comets, and meteorites, all accreted from the same type of interstellar molecular clouds.<sup>54</sup> Presently retrieved meteorites can therefore be useful as guides to estimate the composition of extraterrestrial organic infall before life appeared on Earth. In the Murchison meteorite, a thoroughly analyzed example of carbonaceous chondrite (i.e., a class of chondritic stony meteorite), the most abundant organic material was found to

be a complex, highly aromatic hydrocarbon polymer (90% of the total organic content), accompanied by a variety of soluble organic acids, aliphatic and aromatic hydrocarbons, amino acids, urea, ketones, alcohols, aldehydes, and purines (Table 1).<sup>55–58</sup> Nevertheless, our current perspective on the

**Table 1. Soluble Organic Compounds in the Murchison Meteorite<sup>a</sup>**

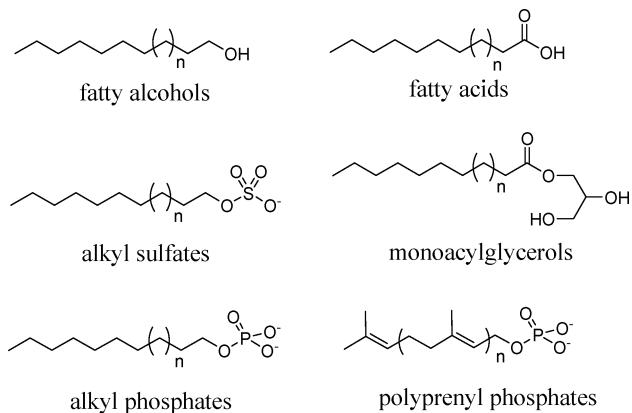
class of compounds	parts per million	<i>n</i> <sup>b</sup>
aliphatic hydrocarbons	>35	140
aromatic hydrocarbons	15–28	87
polar hydrocarbons	<120	10 <sup>d</sup>
carboxylic acids	>300	48 <sup>d</sup>
amino acids	60	75 <sup>d</sup>
imino acids	nd <sup>c</sup>	10
hydroxy acids	15	7
dicarboxylic acids	>30	17 <sup>d</sup>
dicarboximides	>50	2
pyridinecarboxylic acids	>7	7
sulfonic acids	67	4
phosphonic acids	2	4
N-heterocycles	7	31
amines	13	20 <sup>d</sup>
amides	nd <sup>c</sup>	27
polyols	30	19

<sup>a</sup>Data from ref 56. <sup>b</sup>Number of compounds identified with reference standards or with unequivocal matching of compound and library mass spectra. <sup>c</sup>Not determined. <sup>d</sup>Groups where a larger number of compounds are recognized by their mass spectra in the absence of standards or library references.

abundance and diversity of organic matter in the Murchison case could be biased by the way specific compounds have been searched for. A recent reevaluation of the overall composition of organics in the Murchison meteorite,<sup>59</sup> as well as the analysis of other meteorites (e.g., the Renazzo-like chondrites), shows that the abiotic organic chemical diversity in meteorites is much larger than previously assumed.<sup>57</sup> One important question to address, then, is what fraction of these substances survived the delivery process, becoming available for the construction of complex biochemical systems, and what fraction degraded into simple carbon compounds such as CO and CO<sub>2</sub>.

**2.1.1. Synthesis of Lipids.** Modern cell membranes consist of closed spherical bilayers made of different families of lipids, in which transmembrane proteins with diverse functions are embedded. The structure of most of those lipids and surfactant compounds (e.g., phospholipids, glycolipids, cholesterol, etc.) is in general quite complex, and the probability that they were formed prebiotically seems rather low. Despite the pioneering work of Hargreaves et al. in 1977,<sup>60</sup> who demonstrated that the synthesis of phosphatidic acid and other lipids could be achieved abiotically, it is considered very improbable that fatty acids, glycerol, and phosphate (i.e., the standard molecular components of a phospholipid) could have been present together in high enough concentrations on the primordial Earth.<sup>61,62</sup> Other amphiphiles, like terpenoid derivatives, do not seem easy to produce abiotically either, and as an indirect sign, they also have complex metabolic pathways for their synthesis.<sup>63,64</sup> For these reasons, it is generally assumed that the composition of primordial membranes was much simpler than that observed in modern cell membranes. Alkyl phosphates, alkyl sulfates, fatty acids, and polyprenyl chains have been

proposed as possible constituents of early membranes (Figure 3).<sup>65–67</sup>



**Figure 3.** General structure of different types of amphiphilic molecules that have been proposed as possible prebiotic lipids.

Several possible origins for lipidlike molecules, including terrestrial and extraterrestrial sources, have been explored.<sup>68–70</sup> Both long-chain monocarboxylic acids and polycyclic aromatic hydrocarbons (PAHs) with amphiphilic properties were extracted from the Murchison meteorite. Deamer et al.<sup>71,72</sup> have shown that these materials are able to form vesiclelike structures under specific conditions. Their formation could have occurred by irradiation of interstellar matter with UV light, as revealed by experimental irradiation of simulated cometary and interstellar ice.<sup>73</sup> Not all examined meteorites, however, contain long-chain amphiphiles.<sup>74</sup>

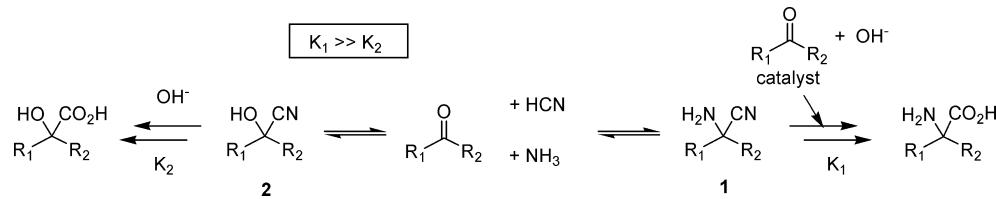
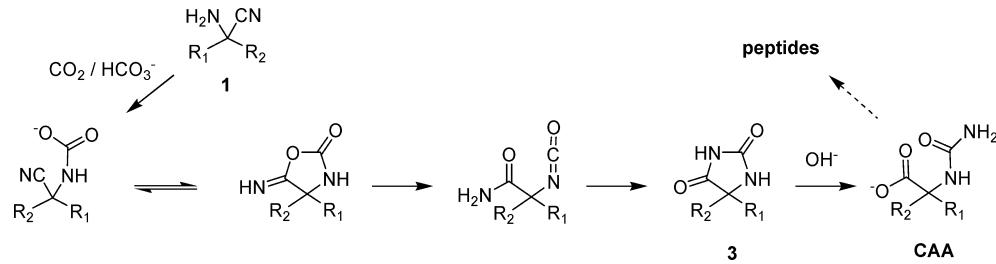
The Fischer–Tropsch synthesis, which is known to produce long hydrocarbon chains from carbon monoxide and hydrogen gases in the presence of a metal catalyst at high temperatures, is considered another possible source of fatty acids and fatty alcohols. In addition to the classical results on those lines by Oró and co-workers,<sup>75</sup> more recently Simoneit and co-workers<sup>76–78</sup> have conducted this reaction by heating oxalic acid solutions at temperatures that simulate the conditions at deep-sea hydrothermal vents. At the optimal temperature (150–250 °C), the lipid components ranged from C<sub>12</sub> to more than C<sub>33</sub> and included *n*-alcohols, *n*-alkanoic acids, *n*-alkyl formates, *n*-alkanals, *n*-alkanones, *n*-alkanes, and *n*-alkenes, all with essentially no carbon number preference. Remarkably, *n*-alkanoic acids increased in concentration while *n*-alcohols decreased when the temperature was raised above 200 °C. Sunlight, as a primary energy source in the present biosphere and most probably in the primordial Earth, has also been shown to drive the synthesis of long-chain hydrocarbon amphiphiles through the photochemical oxidation of alkanes in the presence of PAHs.<sup>79</sup>

A different approach, developed by Ourisson and Nakatani,<sup>66</sup> suggests that primitive membranes were composed of densely branched polyprenyl chains instead of long-alkyl amphiphilic hydrocarbons. Polyprenyl lipids could presumably have been formed by the acid-catalyzed polycondensation of isopentanol, followed by phosphorylation; while isopentanol, in turn, could be formed by the acid-catalyzed Prins reaction between isobutene and formaldehyde.<sup>80</sup> In any case, the prebiotic availability of isobutene and phosphorylating reagents at sufficient concentration is an open issue.<sup>63</sup> The absence of this type of lipids in meteorites, and the general finding that the

abundance of organic molecules decreases with increasing carbon number,<sup>81</sup> also questions the efficiency of their formation in comparison to fatty acids.

The combination of fatty acids with other prebiotic molecular building blocks is proving to lead to amphiphile assemblies in bilayers with improved stability and permeability properties. It has been demonstrated that alcohols and monoacylglycerols, for instance, broaden the pH range under which fatty acid vesicles are stable,<sup>82,83</sup> and the latter can be prepared by the reaction of fatty acids with glycerol under model hydrothermal conditions.<sup>84</sup> There are, in fact, several examples of condensation reactions that link fatty acids with other prebiotic molecules to form phospholipids.<sup>60,85–87</sup> In mixtures of lipids, amino acids, nucleotides, nucleosides, or nucleobases, this type of condensation chemistry may lead to peptidolipids and nucleolipids,<sup>88</sup> where a fatty acid would be connected to an amino acid, nucleotide, nucleoside, or nucleobase or even to an oligomer of these biomonomers. Peptidolipids could have played a role in the transport of substrates through primitive membranes. In turn, nucleolipids are interesting because they possess a nucleobase in their polar head, with H-bonding and π-stacking capabilities that could specifically base-pair with other nucleobases. The interplay between these specific, nucleobase-driven base-pairing interactions and the characteristic aggregation behavior of lipids leads to a combination of properties that might have been very relevant for the constitution of protocells. More than 20 years ago, Yanagawa et al.<sup>89</sup> put forward the possibility that phospholipid-nucleosides could have served as self-assembled nucleic acid template structures without the need to form covalent bonds, as in biopolymer replication processes. This suggestive idea was further followed by Bonaccio et al.,<sup>90</sup> who studied experimentally the assembly properties of this type of compounds. More recently, Rosemeyer<sup>91</sup> has proposed a model for the origin of life in which low molecular weight nucleolipids could have promoted early molecular evolution and could have acted as templates for the replication of simple RNA oligomers at the inner surface of a vesicle or liposome. According to this model, hybrid molecules of this kind would have been discarded later by nature, once more complex and efficient methods to replicate nucleic acids were developed inside the protocell, while nucleolipids found today in biology (e.g., nucleolipid antibiotics produced by certain bacteria) would be molecular relics from those ancient times.

**2.1.2. Synthesis of Amino Acids.** Amino acids were the first prebiotic molecules to be identified in the Miller–Urey experiment<sup>47,92</sup> and are probably the easiest biomonomers to synthesize, as their high abundance in some chondritic meteorites suggests.<sup>56–58</sup> Nevertheless, despite their relative ease of formation and the importance of proteins in biology, amino acids somehow lost part of their historical role in the origins of life research since the advent of the RNA world hypothesis (see section 3.3).<sup>42,93</sup> Yet this apparent loss of importance is surely misleading. In a prebiotic world where an ever-increasing diversity of organic species was being originated, it is more than reasonable to think that amino acids interacted with all the surrounding repertoire of molecules, including nucleotides and their precursors. Amino acids, and by extension their polymeric forms (i.e., peptides), have shown their ability to catalyze multiple reactions,<sup>94–96</sup> some of which could have led to intermediates with the right stereochemistry in the synthesis of nucleotides.<sup>97–99</sup> Common intermediates for the synthesis of peptides and nucleosides have

Scheme 1. Synthesis of  $\alpha$ -Amino Acids through the Strecker ReactionScheme 2. Bücherer–Bergs Hydrolysis of  $\alpha$ -Aminonitriles

been produced experimentally.<sup>100</sup> The possibility that life arose from an early interplay between amino acid/peptide and ribonucleotide/RNA chemistries has also been proposed as one possible reason for the universality of the genetic code.<sup>101,102</sup> Indeed, the hypothesis of a primordial ribonucleopeptide world, in which oligopeptides and oligonucleotides could cooperate in supramolecular assemblies or covalently bound conjugates, seems perfectly coherent with the plethora of ribonucleoprotein aggregates being currently discovered. These aggregates are involved in catalysis<sup>103</sup> and in the flow of genetic information within cells, the ribosome being the paramount example.<sup>104–106</sup>

As with other biomonomers, there are two prebiotically relevant sources of amino acids: endogenous and exogenous syntheses. From both pathways a wide variety of amino acids can be obtained,<sup>107</sup> but here we will focus on  $\alpha$ -amino acids, given their major relevance in biochemistry. The exogenous formation and delivery of amino acids have been evaluated by analyzing the composition of different carbonaceous chondrites. The amino acid set in this carbon-rich class of meteorites comprises more than 70 species [measured at the 10–100 ppm (w/w) level], with most of them being  $\alpha$ -amino acids and including at least eight proteogenic ones.<sup>56,57</sup> The decreasing amino acid abundance with the number of carbon atoms in homologous series, and the predominance of branched isomers, suggests that the chemistry involved in their extraterrestrial synthesis is at least partly based on nonselective photochemical and radical processes.

The endogenous production of amino acids on the primitive Earth has been investigated for the last six decades. Even if the particular conditions (e.g., the recreated reductive atmosphere) in which Miller's original experiments were carried out are eventually discarded as unrealistic, further studies have demonstrated that amino acids can still be synthesized under other various conditions (e.g., in weakly reducing or neutral atmospheres).<sup>108,109</sup> Most geochemists at present consider the primeval Earth atmosphere to have been, overall, non-reducing,<sup>109</sup> but reducing conditions could have been locally or transiently prevalent, for instance, near volcanic plumes.<sup>110,111</sup> In this respect, a group of scientists have recently analyzed samples from an experiment of Miller that recreated such environments and have shown the formation of 22 amino acids, most of which were not identified in Miller's original

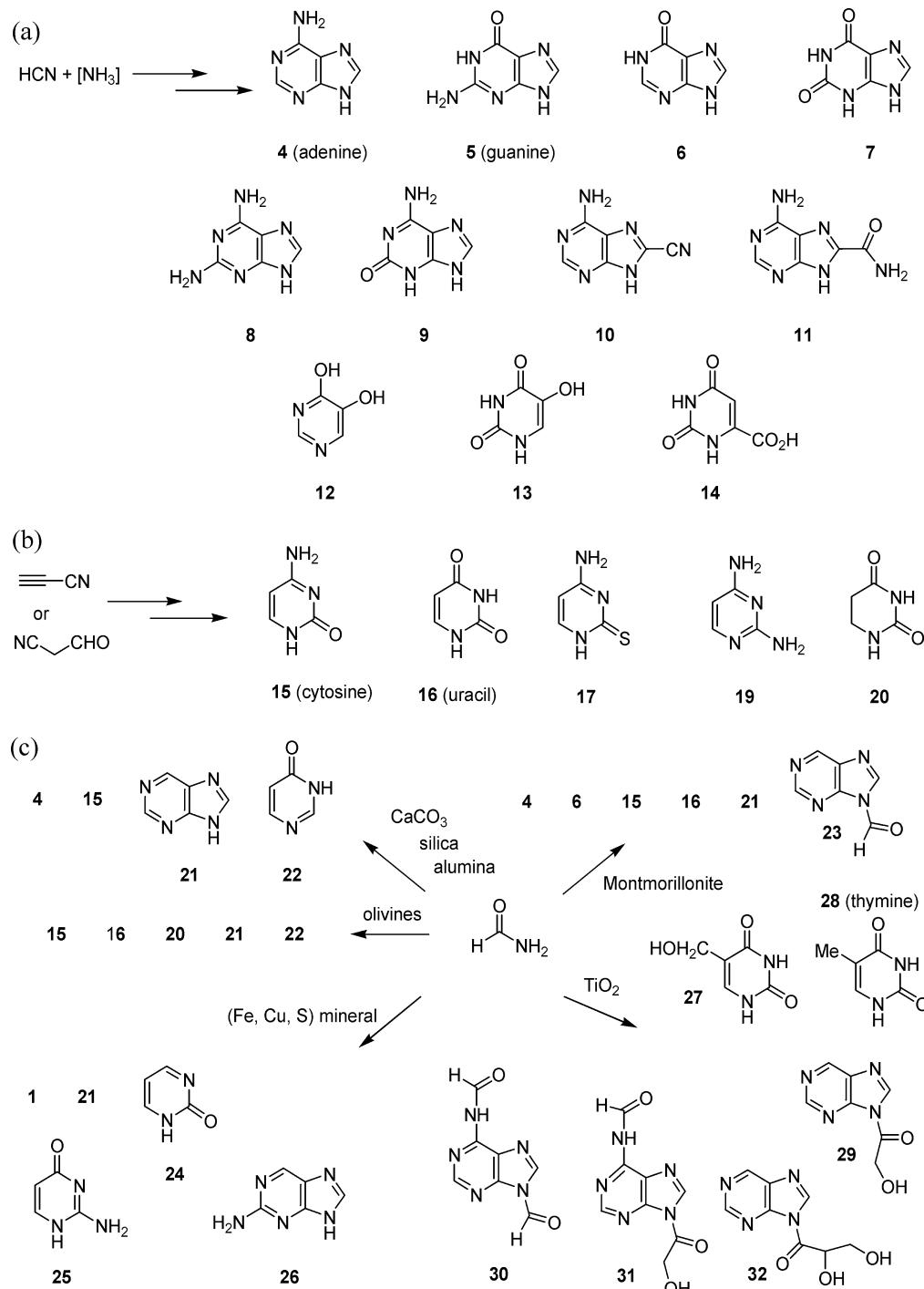
experiments.<sup>92,112,113</sup> The synthesis of  $\alpha$ -amino acids and  $\alpha$ -hydroxy acids under possible volcanic conditions has also proven to be feasible in solution by CO-dependent carbon fixation at temperatures between 80 and 120 °C, with nickel or nickel/iron precipitates as catalysts; carbonyl, cyano and methylthio derivatives as carbon sources; and calcium or magnesium hydroxide as buffering salts.<sup>114</sup>

Several other plausible prebiotic amino acid syntheses have been reported.<sup>115</sup> In endogenous syntheses, the Strecker reaction appears as the most straightforward route to obtain amino acids from carbonyl compounds, ammonia, and HCN.<sup>116</sup> The prebiotic relevance of this classical reaction<sup>117</sup> was pointed out by the production of aldehydes and HCN in Miller-type experiments. The reaction consists in the formation of  $\alpha$ -aminonitriles (1, Scheme 1) in moderately alkaline aqueous conditions, followed by hydrolysis of the nitrile group. Remarkably, some precursors of this reaction are also important substrates in the synthesis of nucleotides: HCN is the main starting material in the generally accepted prebiotic synthesis of purines, while aldehydes are components of the formose reaction, which is considered a prebiotic source of sugars (see section 2.1.3). Furthermore, the analysis of meteorite mineral matrices suggests that water has been present at some stage of their existence in the outer space, so Strecker-type chemistry might also explain to some extent meteoritic amino acid formation.

The first step in this reaction is an equilibrium by which, in addition to  $\alpha$ -aminonitriles (1), cyanohydrins (2) can be formed. Slow hydrolysis of the nitrile group leads to mixtures of  $\alpha$ -amino acids and  $\alpha$ -hydroxy acids, which are also abundant in meteorites and in Miller-type experiments. The irreversible hydrolysis of  $\alpha$ -aminonitriles can be accelerated several orders of magnitude by organocatalysis with aldehydes, being able to yield amino acids at low concentration of ammonia.<sup>118</sup> In the special case of a low, continuous flow of ammonia, this catalytic process displaces the cyanohydrin/aminonitrile equilibrium toward aminonitriles by Le Chatelier's principle. Cyanohydrins could thus be considered as a source of reactive compounds for the synthesis of amino acids under a much more stable form.

Another faster version of the hydrolysis of  $\alpha$ -aminonitriles is the Bücherer–Bergs reaction, in which bicarbonate/CO<sub>2</sub> promotes the transformation of unreactive nitrile groups into

Scheme 3. Nucleic Acid Bases That Have Been Synthesized under Various Conditions



hydantoins (3) at moderate pH (Scheme 2).<sup>118</sup> This reaction cannot be considered a catalytic one, however, since CO<sub>2</sub> is not released as a product. Hydantoins are precursors of amino acids through a two-step hydrolysis, while *N*-carbamoylamino acids (CAAs), the products of the first hydrolysis from hydantoins, may have prebiotic importance, playing the role of activated monomers toward the synthesis of peptides (see section 2.2.1).

**2.1.3. Synthesis of Nucleotides.** Ribonucleotides and deoxyribonucleotides, the building blocks of RNA and DNA, respectively, are composed of three different moieties: a nucleobase (adenine, guanine, cytosine, or uracil in the case of RNA; thymine instead of uracil in DNA), a sugar (ribose in

RNA, deoxyribose in DNA), and a phosphate group. For decades, the only approach followed for the synthesis of ribonucleotides was based on the assumption that their three components were synthesized separately and subsequently assembled.<sup>39</sup> Successful chemistries for synthesizing nucleobases, sugars, and condensed phosphates, under primordial conditions, have been developed over the last 40 years, though most of the individual synthetic steps face difficulties. For example, formaldehyde oligomerizes to form sugars, but complex mixtures are obtained in which ribose is a relatively minor product.<sup>119</sup> Hydrogen cyanide leads, in combination with water or ammonia, to purines,<sup>120</sup> while urea and

cyanoacetylene can react to produce pyrimidines.<sup>121–123</sup> The synthesis of purines is quite robust, but pyrimidines seem to require much higher concentrations of reactants. Another problem in nucleobase syntheses is the lack of selectivity toward any hydroxyl group (i.e., 2', 3', or 5') in the phosphorylation of sugars. The most frustrating problem in the road toward nucleotides concerns, however, the linkage of nucleobases to ribose.<sup>124</sup> This reaction gives very low yields in the case of purines and no nucleoside formation at all in the case of pyrimidines.

Given these considerations, the chemistry of nucleotides seems to be, by far, the most complicated one among the different biomonomers, involving many independent reactions that had to be optimized and coupled in order to give an overall efficient process. This is the reason why, up to now, modular approaches have failed to solve the question of whether it was possible that RNA monomers could be synthesized on the early Earth.<sup>39</sup> In contrast, recent experimental evidence shows that moving beyond individual reactions, and considering the collective behavior of reaction sets in mixtures of nucleotide precursors, provides solutions to the aforementioned limitations.<sup>125–127</sup> These systemic approaches are more realistic and prebiotically plausible, since they allow the system as a whole to contribute solutions for the component parts, leading to higher yields of all the individual reactions.

Because of their historical and methodological relevance, we will first review the classical approaches of prebiotic chemistry to the synthesis of nucleotides. In that tradition, many experiments have been devoted to the synthesis of nucleobases, or to detect them in meteorites,<sup>58</sup> as one of the main constituents of nucleic acids. The first synthesis of a purine was described in 1960 by Oró,<sup>120</sup> who heated ammonium cyanide for several days at 70 °C to give adenine in 0.5% yield. Since then, abiotic synthesis of adenine and guanine from the polymerization of HCN under various conditions has been achieved many times (Scheme 3a), the highest yields obtained being 12% for adenine and 3.3% for guanine.<sup>128,129</sup> In contrast, only a few synthetic routes to cytosine and uracil have been reported,<sup>130</sup> even though a large number of pyrimidine derivatives were found in the Murchison meteorite.<sup>55</sup> Most of them use cyanoacetylene (or its hydrolytic product, cyanoacetaldehyde) and cyanate ions, cyanogens or urea, as the starting materials (Scheme 3b). Cytosine is generally obtained in yields from low to moderate, while uracil can be generated by hydrolysis of cytosine.

The formation of nucleobases can occur under different conditions, such as in drying/wetting cycles,<sup>131</sup> eutectic phases,<sup>131,132</sup> ice matrices,<sup>133–136</sup> and aerosols.<sup>137,138</sup> As discussed in section 1, minerals and metal oxides could have also enhanced synthetic processes by providing local micro-environments able to concentrate reagents and protecting newly formed compounds from degradation.<sup>139</sup> To date, the role of mineral catalysis in the prebiotic synthesis of nucleobases has only been analyzed in detail for the case of formamide as precursor.<sup>140</sup> Formamide has a high boiling point (210 °C), with limited azeotropic effects. In contrast to HCN, it could have been easily concentrated by heating in lagoons and on drying beaches eventually present on the early Earth. Due to its prebiotic potential,<sup>141</sup> formamide has been extensively studied as a plausible source of nucleobases in different experimental settings.<sup>142–144</sup> The most remarkable results were obtained when formamide was heated at 160 °C, or in some cases irradiated with UV light,<sup>145</sup> in the presence of

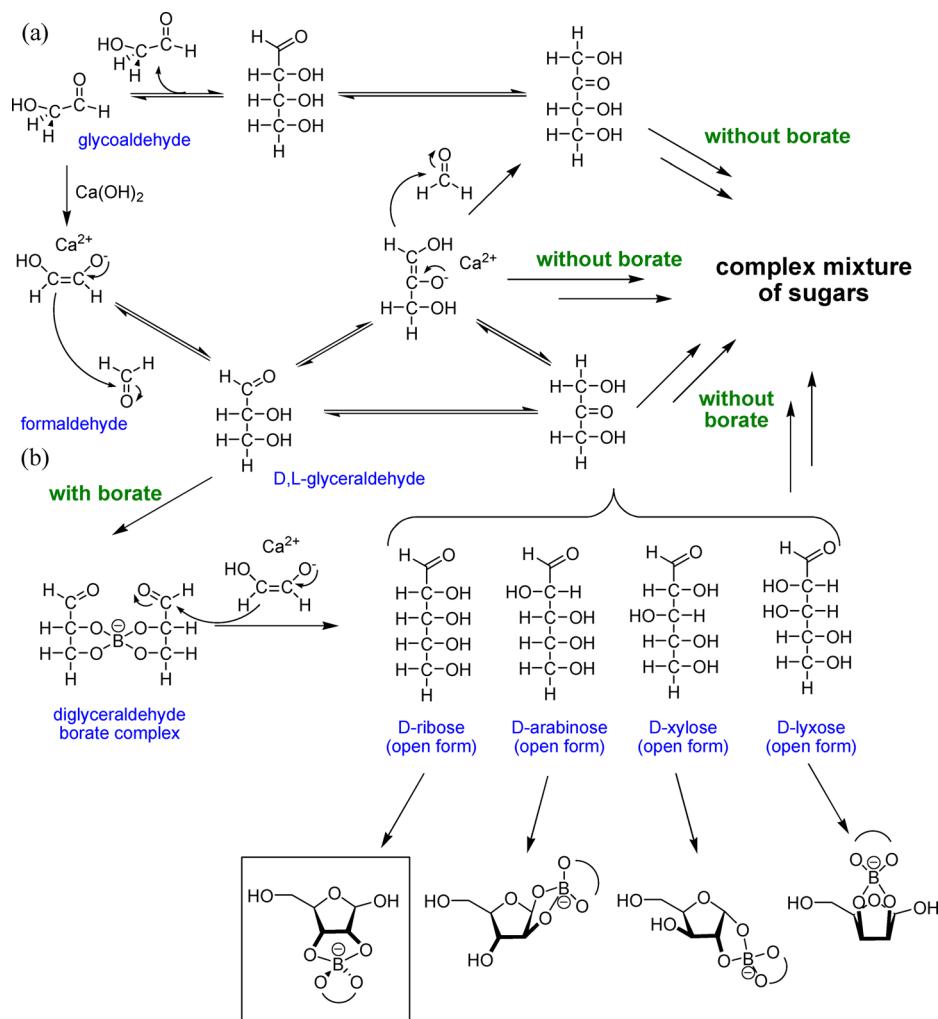
calcium carbonate, silica, alumina,<sup>146</sup> TiO<sub>2</sub>,<sup>147</sup> montmorillonite,<sup>148</sup> olivines,<sup>149</sup> iron sulfides,<sup>150</sup> zirconia,<sup>151</sup> or iron oxides<sup>152</sup> (Scheme 3c). Among the latter, montmorillonite, a phyllosilicate that likely formed large deposits on the early Earth, is particularly interesting as a prebiotic catalyst because this clay mineral is also able to catalyze the oligomerization of activated nucleotides (see section 2.2.2) as well as the spontaneous conversion of fatty acid micelles into vesicles (see section 3.1.3).<sup>153</sup>

The variety of possible starting materials, conditions, and catalysts that have been tested in prebiotic syntheses of nucleobases renders a large panel of purine and pyrimidine derivatives that could have been produced on the prebiotic Earth (see Scheme 3). The potential existence of such high diversity raises the question of whether other nucleobases different from the five canonical ones could have been initially involved as constituents of information-carrying polymers. Hud and co-workers<sup>154,155</sup> have demonstrated that noncanonical nucleobases, such as pyrimidinone, are able to form nucleosides more easily than the current bases in glycosidic bond formation reactions with D-ribose. It is intriguing, then, whether other informative elements, different from the biological nucleobases, could have been possible constituents of pre-RNA genetic polymers.<sup>40,156,157</sup> In order to study this possibility, Eschenmoser and co-workers<sup>158,159</sup> prepared, by conventional organic synthesis, oligo(dipeptide)s and oligo(dipeptoid)s as polymeric backbones bearing 1,3,5-triazine and 5-amino-1,3-pyrimidine nucleobases as recognition elements. The study of their base-pairing behavior revealed that these oligomers hybridized to themselves and also with RNA and DNA oligonucleotides. Nevertheless, this hybridization was not strong and selective enough to support a template-dependent replication process.

In other words, these results suggested the higher affinity of canonical nucleobases compared to other recognition elements of similar chemical complexity. Krishnamurthy<sup>160</sup> has actually related this behavior to the relationship between the pK<sub>a</sub> of nucleobases and the pH of the aqueous medium. More specifically, a difference of pK<sub>a</sub> – pH higher than 2 seems to favor, with few exceptions, nucleobase hydrogen bonding. At pH 7, the current types of nucleobase fulfill this condition and, consequently, polymers containing them could have been selected due to base-pairing efficacy. It is important to notice that the relative abundance of natural nucleobases in prebiotic scenarios would be uneven, since they follow specific synthetic pathways and are thus differentially affected by environmental factors. As an example, eutectic freezing of a dilute ammonium cyanide solution produced roughly 10 times more adenine than guanine, cytosine, and uracil.<sup>129</sup> Moreover, degradation affects nucleobases in different ways, and cytosine is particularly susceptible to spontaneous deamination.<sup>161</sup> The higher performance of genetic polymers based on the base-pairing capacity of nucleobases has also triggered, during the last four decades, the synthesis of various families of nucleobase-containing nucleic acid analogues. For some of them, leading roles in putative pre-RNA worlds have been proposed (see section 3.3.1).

Regarding the synthesis of ribose, the second component of ribonucleotides, a classical transformation such as the formose reaction has been proposed.<sup>119</sup> This reaction was originally discovered by Butlerow in 1861 and consists of the polymerization of formaldehyde in the presence of mineral catalysts (e.g., calcium hydroxide, Pb<sup>2+</sup> and Tl<sup>+</sup> ions) and, generally, alkaline medium. The Butlerow synthesis of sugars is a unique

**Scheme 4.** Simplified Representation of the Formose Reaction to Produce Sugars in the (a) Absence and (b) Presence of Borates<sup>a</sup>



<sup>a</sup>For simplicity, branched isomers are not drawn; yet, for a better understanding of the chemical complexity generated in the formose reaction, see ref 168.

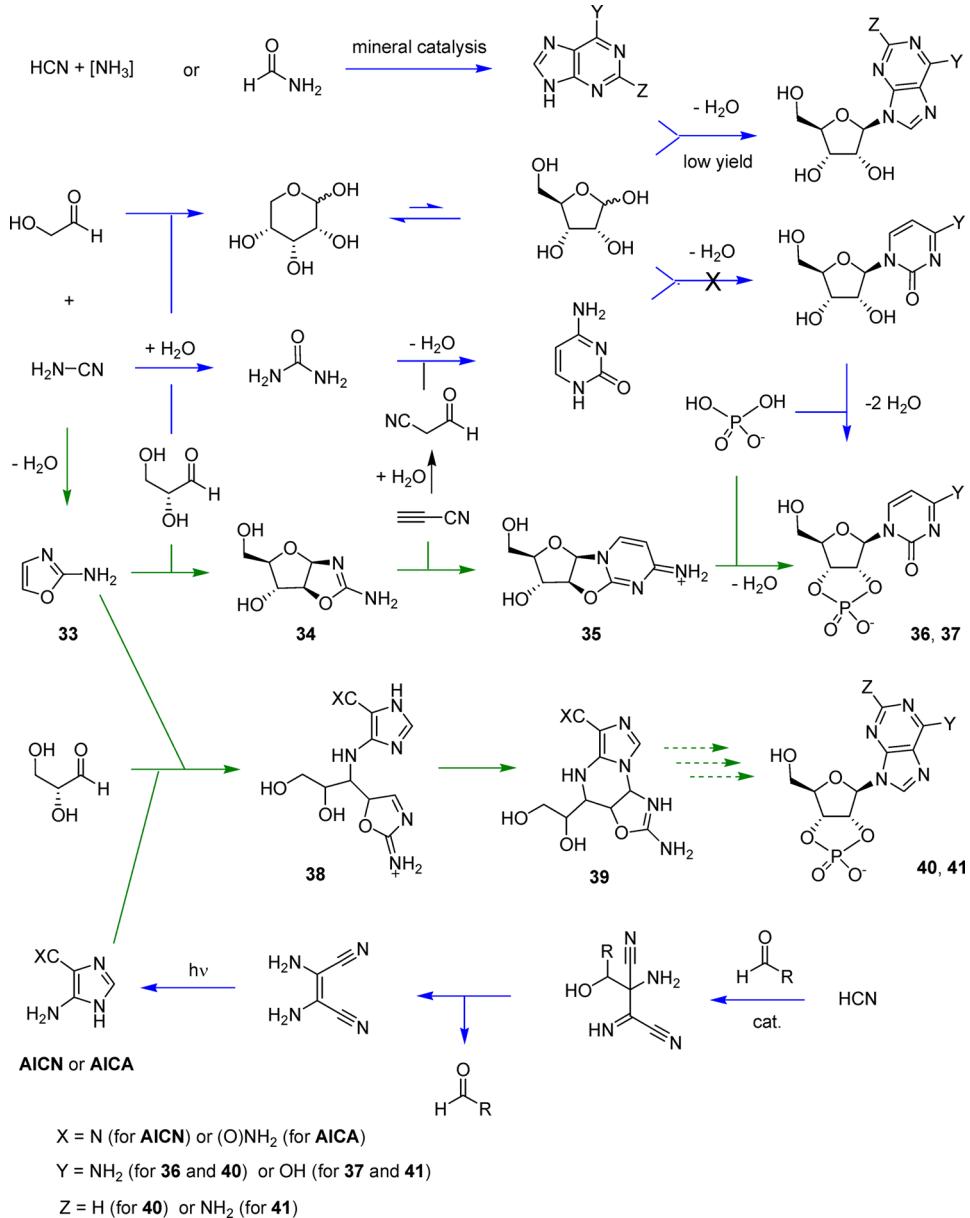
autocatalytic process, formed by many cycles (like the one depicted in Scheme 4a), in which glycoaldehyde acts as an efficient catalyst. As a consequence of all the possible cycles, a mixture of tetroses, pentoses, and hexoses is obtained. Some of these aldoses are building blocks of putative pre-RNA nucleic acid analogues (see section 3.3.1). It is also noteworthy that glyceraldehyde, one of the main products of this reaction, drives, in the presence of alanine amide (or ammonia), the synthesis of 3,5(6)-dimethylpyrazin-2-one, that is considered a potential pre-RNA nucleobase.<sup>162</sup> This is, again, an example of how rich the chemistry involved could be if prebiotic precursors from different synthetic routes were present in the same mixture.

Although the formose reaction leads to ribose and related aldopentoses, it is debatable whether it could have led to the accumulation of these particular sugars on the early Earth, since ribose is a notoriously minor and rather unstable product of the reaction. Its higher permeability through lipid bilayers could have been an advantage in that sense,<sup>163</sup> eventually promoting the concentration of ribose inside protocellular compartments. Besides, the assistance of reactive inorganic species may have enhanced the proportion of ribose in the reaction medium. The

pattern of products in the formose reaction can be simplified, for instance, if glycoaldehyde phosphate and glyceraldehyde 2-phosphate react in the presence of magnesium aluminum hydroxide as catalyst, leading to ribose 2,4-diphosphate.<sup>119</sup> Thanks to the catalyst, the reaction proceeds in dilute solutions at neutral pH. The phosphate groups from the starting materials prevent the rearrangement of trioses, tetroses, and pentoses, thus decreasing the characteristic complexity of sugars coming out of the formose reaction. Although a reaction where ribose 1,5-diphosphate or ribose 5-monophosphate were formed would be of much more prebiotic interest, Eschenmoser's work first illustrated how inorganic catalysts could appropriately direct chemistry toward the synthesis of nucleotides.

Later, Ricardo et al.<sup>164</sup> found that association with borate minerals stabilizes pentoses (i.e., ribose, xylose, lyxose, and arabinose) in the mixture of sugars resulting from a crossed aldol reaction between glycoaldehyde and glyceraldehyde, the four pentoses being obtained in similar amounts (Scheme 4b). This implies that, in the presence of borate minerals, ribose could have naturally accumulated, thus removing one of the major objections to the RNA world theory. Some articles

Scheme 5. Modular Synthesis (blue) versus Systems Synthesis (green) of Pyrimidine and Purine Ribonucleotides



followed in the years after this publication, claiming that other minerals (e.g., aqueous sodium silicates, which were undoubtedly abundant on the prebiotic Earth) could have also mediated aldol reactions from one- to three-carbon sugars (i.e., formaldehyde, glycoaldehyde, and glyceraldehyde), guiding the formose reaction toward the stabilization of tetroses and hexoses.<sup>165</sup> Such a statement provoked a debate between Kim and Benner<sup>166</sup> and Lambert et al.,<sup>167</sup> mainly focused on the relative effects, with regard to both stabilization strength and selectivity, that borates and silicates could have exerted on the sugar composition of primordial scenarios. More recently, Kim et al.<sup>168</sup> have conducted a detailed chemical analysis of the complexity of the formose reaction, with the aim of understanding more deeply the role of different minerals in guiding the final outcome of the process. It seems from these studies that both aldopentoses and aldötetroses form stable complexes with borates, yet tetroses were not accessible by pathways plausible under prebiotic conditions. An important question in this respect is the optimum reactivity of those

borate complexes, since too much stability would prevent carbohydrates from further reacting. A clue on this aspect is given by their finding that molybdate species could have catalyzed the interconversion of branched and unbranched carbohydrates without the need of intermediate aldol fragmentation, therefore preventing the formation of dead-end products. Unfortunately, although Kim and Benner<sup>166</sup> state that the presence of hydrated rocks undergoing subduction to form felsic magmas in the early Hadean era could have been a source of mineral borates and molybdates, this assumption remains far from being corroborated.

Nevertheless, being able to synthesize nucleobases and ribose does not imply that the route toward nucleotides is completely paved. Bringing all components together involves two difficult reactions: formation of a glycosidic bond, with the right stereochemistry linking the nucleobase and ribose, and phosphorylation of the resulting nucleoside. When ribose and a purine are heated in the presence of Mg<sup>2+</sup> and inorganic phosphate, only modest yields (i.e., 4–9%) of purine

nucleosides result.<sup>124</sup> The same reaction does not proceed at all if pyrimidines are used, due to their lack of the nucleophilic imidazoyl nitrogen characteristic of purines. Formamide, a very promising prebiotic precursor of nucleobases, has shown to yield acyclonucleosides,<sup>147</sup> possibly solving the problem of poor reactivity of purines and pyrimidines with sugars. In turn, the problem of nucleoside phosphorylation can also be solved by use of formamide as catalyst, which accelerates this reaction for adenosine and cytosine, using soluble phosphates or phosphate minerals as the phosphorus source.<sup>144,169</sup>

Interestingly, Sutherland and co-workers<sup>125,126,170–176</sup> have recently explored a radically different strategy that, as mentioned above, overcomes some of the problems of modular approaches in the synthesis of pyrimidine and purine ribonucleotides. In their systemic approach, nucleotides do not emerge from free sugar and nucleobase molecules as intermediates but from a common precursor that contains atoms that will later become part of the sugar and the nucleobase moieties. In this way, the classical separation between carbon–oxygen chemistry, responsible for the synthesis of sugars, and carbon–nitrogen chemistry, which leads to nucleobase formation, is eliminated. Although combining these two different chemistries would in principle produce a chemical combinatorial explosion,<sup>127</sup> with thousands of different organic compounds possibly being formed, these authors showed that the presence of phosphate tames this combinatorial explosion, allowing intermediates from different stages to interact fruitfully.

Sutherland's synthesis of  $\beta$ -ribocytidine 2',3'-cyclic phosphate starts from very simple building blocks: glycoaldehyde, glyceraldehyde, cyanamide, cyanoacetylene, and phosphate (Scheme 5).<sup>126</sup> The first key steps in the synthetic process are two successive condensation reactions: cyanamide condenses with glycoaldehyde, giving aminooxazole (33), which subsequently condenses with glyceraldehyde to yield cyanamidoarabinose (34), an intermediate that bears portions of both sugar and nucleobase. In these condensation reactions, the presence of phosphate buffer at pH 7 was crucial to control the reaction pathway, leading to high yields of products with the right stereochemistry. The use of phosphate with a dual function of pH and chemical buffer also allowed the high-yield production of arabinose anhydronucleoside (35) from the condensation reaction of 34 with cyanoacetylene. Another important finding was selective phosphorylation at the 3'-hydroxyl group of anhydroarabinocytidine (35), due to the unique steric environment of other hydroxyl functionalities. The phosphorylation is facilitated by the presence of urea, which comes from the phosphate-catalyzed hydrolysis of cyanamide. As a consequence of the phosphorylation reaction, cytidine 2',3'-cyclic phosphate (36) could be intercepted and subsequently converted photochemically into uridine 2',3'-cyclic phosphate (37). The photoconversion was shown to be clean because of the inert character of nucleoside 2',3'-cyclic phosphates against photodegradation, compared to other nucleosides and nucleotides, resulting in very little nucleobase loss. In this process, UV radiation is destroying multiple side products while directing the reaction to produce the second pyrimidine component of current ribonucleotides.

In order to overcome the unpaired development of prebiotic synthetic routes toward pyrimidine and purine nucleotides, Powner et al.<sup>175</sup> kept seeking for alternative manners to combine nitrogen chemistry with aldehydes in aqueous solution (Scheme 5). In particular, they found a high-yielding, one-pot,

multicomponent reaction that leads to plausible purine precursors (38) and, by a further 6-*exo-trig* intramolecular imidazole-iminium trapping reaction, to 39. These intermediates contain key nucleotide synthons such as 2-aminooxazole, whose synthesis has been described in the previous paragraph, and 5-aminoimidazole, which in turn can be obtained from HCN oligomers such as 5-aminoimidazole-4-carbonitrile (AICN) and its hydrolytic product (5-aminoimidazole-4-carboxamide, AICA). Moreover, some of the building blocks that constitute these masked sugar and nucleobase moieties are shared by both synthetic pathways to purine and pyrimidine nucleotides, revealing the close relationship between them. The proposal of these authors for the potential next steps from 39 toward adenosine 2',3'-cyclic phosphate (40) and guanosine 2',3'-cyclic phosphate (41) is also noteworthy, specially regarding the fact that activated nucleotide monomers would be obtained.

Interestingly, Powner et al.<sup>177</sup> have also described a related synthetic route in which the substitution of glycoaldehyde by  $\beta$ -mercaptoproacetalddehyde as one of the building blocks in the same type of multicomponent reaction leads to 2-aminothiazole (i.e., instead of 2-aminooxazole), an important intermediate toward masked 2'-thiosugars that could subsequently yield 2'-deoxynucleotides upon desulfurization. Another important finding, in relation to a plausible prebiotic scenario in which sugar and nitrogen chemistries are combined, is the recent demonstration of a Kiliani–Fischer-type synthesis of simple sugars (glycoaldehyde and glyceraldehyde, which are building blocks in the syntheses mentioned above) from hydrogen cyanide through photoredox cycling of cyanocuprate.<sup>178</sup> If the same kind of reaction is performed with  $H_2S$  as the ultimate reductant in the system, acetaldehyde is also obtained, with higher efficiency if thiocyanate is included in the reaction mixture.<sup>179</sup> Given that formaldehyde (i.e., obtained in its hydrated form as one of the reaction intermediates), glycoaldehyde, and acetaldehyde are the Strecker precursors of glycine, serine, and alanine, these results point to a possible linked origin of ribonucleotides and amino acids. The addition of HCN and excess ammonia to the system after 4 h of irradiation actually resulted in conversion of the three aldehydes into their corresponding  $\alpha$ -aminonitriles. As a side note, the use of  $H_2S$  as the reducing power to generate those important protobiomolecules is suggestive to establish a connection with the well-known scenario put forward by Wächtershäuser and co-workers<sup>29,180,181</sup> (see section 3.1.2).

In Sutherland's syntheses, phosphate compounds have several essential roles, ranging from buffering various steps to depleting undesired byproducts or saving crucial intermediates from degradation. Moreover, Powner and Sutherland<sup>125</sup> have found common conditions under which their synthetic modus operandi for the generation of nucleotides is compatible with the synthesis of lipids by phosphorylation of long-chain alcohols. Those conditions are plausibly prebiotic, for instance, in a postmeteoritic impact scenario. The presence of inorganic phosphate determines in these examples the right synthetic pathway, and this may be the case in many other cornerstones of prebiotic chemistry. Unfortunately, with the exception of volcanic vents,<sup>182</sup> plausible sources of condensed phosphates on the early Earth have remained elusive for a long time. The mineral apatite, which is the only significant terrestrial source of phosphate, in the form of orthophosphate, is problematic because of its low solubility and reactivity toward organic molecules. In recent years, some pathways toward useful forms

of phosphate have been demonstrated, and a possible extraterrestrial income of reactive phosphorus is also being examined.<sup>183–185</sup> The phosphorus problem seems no longer to be irresolvable, but it exceeds the scope of this review.

**2.1.4. Summary.** In 2004, Leslie Orgel wrote “There is at present no convincing, prebiotic total synthesis of any nucleotide. Many individual steps that might have contributed to the formation of nucleotides on the primitive Earth have been demonstrated, but few of the reactions give high yield of products, and those that do tend to produce complex mixtures of products.”<sup>39</sup> The reason for such a realization was that several steps needed for the synthesis of nucleotides are thermodynamically uphill. The problems highlighted by Orgel are still obstacles nowadays, but some radically new experimental strategies are shading optimism on the possibility of circumventing part of these thermodynamic constraints. Almost every step in the synthesis of nucleotides, including nucleoside formation and nucleoside phosphorylation, and of other prebiotically relevant monomers can proceed now with reasonable efficiency by coupling with thermodynamically downhill chemical processes and the action of various possible catalysts.

In fact, new catalytic effects induced by small organic molecules are being discovered every year in the field of organocatalysis. We know now that amino acids can assist the synthesis of sugars, aldehydes catalyze the formation of amino acids and drive the synthesis of some nucleobases, formamide is crucial in the linkage of nucleobases with sugars to produce nucleosides and in nucleoside phosphorylation, etc. Inorganic species such as phosphates and borates, as well as metal or mineral surfaces, also play a key role by stabilizing important intermediates, preventing their degradation, and even catalyzing other reactions. By any means, catalysis must have been pivotal in the route toward the molecules of life because it makes a wide variety of thermodynamically plausible chemical processes faster, increasing their probability to occur on much shorter time scales. Moreover, catalysis allows modification of the balance between thermodynamic and kinetic control over various possible, competing reaction pathways. In a mixture of chemical reagents under equilibrium, for instance, the distribution of products is established according to their relative thermodynamic stabilities. Hypothetically, there could be other thermodynamically stable products resulting from reactions between different components of the mixture, but they do not effectively form if the kinetic barriers of those reactions are too high at the existing conditions. If a catalyst were at some point introduced in the system (generated from the component mixture or recruited from the external environment), the energy landscape of the system would be altered and some of the kinetic barriers might be significantly lowered. This would then lead to new products whose formation was (thermodynamically possible but) kinetically impossible before the catalyst entered the scene.

Under this assumption, reaction rates were a determinant feature in prebiotic chemistry, due to the variability of physicochemical conditions and geological constraints on the early Earth. In the complex prebiotic scenario we are envisaging, at the same time as more and more new molecules were formed by means of catalysis and the coupling of thermodynamically unfavorable reactions with favorable ones, the existing chemical space would have increased, and so would the number of possible catalysts and reactions that could interact with each other. This would certainly lead to a

combinatorial explosion of molecular diversity. It has been shown, however, that such an explosion could be tamed by the presence of feedback loops and regulation mechanisms involving different intermediate species, redirecting synthetic processes in unexpected ways.<sup>127</sup> Chemical autocatalysis, in all its forms, may have also funneled the prebiotic combinatorial explosion toward molecules currently recognized as relevant for life. As a result, protometabolic networks of different kinds could have emerged out of these hypothetical, complex molecular mixtures (see section 3.1.2). This will be of crucial importance for our general conception of chemical evolution.

## 2.2. Prebiotic Synthesis of Polymers: Peptides and Nucleic Acids

Condensation reactions are among the most important processes in the pathway to life, since they contribute to the formation of biopolymers, including proteins and nucleic acids. The general problem regarding the condensation of small organic molecules to form macromolecules in an aqueous environment is the thermodynamically unfavorable process of water removal. In the current biosphere, these types of reactions are catalyzed by enzymes and energetically driven by pyrophosphate hydrolysis. Obviously, one cannot assume that biocatalysts and energy-rich inorganic phosphorus species were present on the Earth before life had actually originated. A number of proposals to overcome these difficulties have been made, taking into account possible scenarios where the abiotic polymerization of different biomonomers could have occurred. These proposed polymerization reactions can be divided into melting processes, hydration–dehydration cycles, heterogeneous systems involving other phases like lipid domains or mineral surfaces, and polymerizations induced by various coupling reagents in homogeneous solutions. In all cases, the starting materials would consist of an enormous amount of disparate molecules produced during the combinatorial explosion that supposedly characterized the synthesis of biomonomers (see section 2.1). In such complex polymerization systems, mutual catalytic and inhibitory effects, prevention of polymer degradation by minerals and preformed macromolecular species, deactivation of side reactions, etc., may have enhanced the oligomerization of amino acids and nucleotides to produce primitive peptides and nucleic acids, respectively. It should be kept in mind, though, that early biochemistry, being based on the continuous generation and breakdown of biomolecules, probably had more precarious mechanisms than today to avoid the spontaneous molecular decay of its main components. In the particular case of polymers, their endurance must be considered from a dynamic perspective, implying a steady state favorable to the emergence and evolution of new properties. Once biopolymers were long enough to fold into active structures endowed with catalytic properties, or even self-replication capabilities (despite the apparent problems associated with copying folded macromolecules; see sections 3.2 and 3.3),<sup>186</sup> the amplification of their relative concentration would have led to their later predominance in the precellular world.

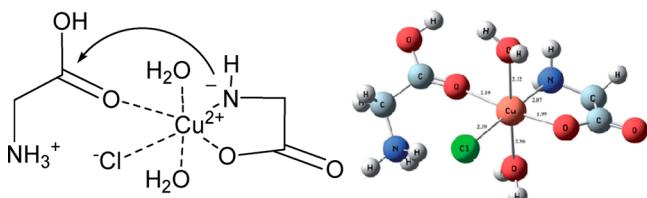
**2.2.1. Synthesis of Peptides.** As shown in a previous section, amino acids are probably the easiest biomonomers to synthesize. Therefore, it is likely that they were relatively abundant substrates on the prebiotic Earth. Nevertheless, the formation of peptide bonds in the presence of water is thermodynamically disfavored, and the free energy required to link two amino acids together in aqueous solution lies within

the range 2.5–3.6 kcal/mol.<sup>187</sup> Making peptide bond formation a thermodynamically downhill process requires a nonaqueous medium. Thus, removing products from solution by precipitation or performing melt polymerization reactions shifts the equilibrium toward the products, even in the absence of catalysis. This assertion is, however, only true within a limited pH range, where peptide bond hydrolysis is not chemically accessible.<sup>188,189</sup> Thus, the key is to provide conditions that make amide bond formation significantly faster than amide bond hydrolysis, which would in principle allow the synthesis of peptides.<sup>190</sup> For a detailed review on possible pathways toward peptide synthesis and evolution in a prebiotic environment, see ref 191.

Amino acids can be oligomerized by heating in the dry state<sup>192,193</sup> or in aqueous solutions at 150–250 °C.<sup>194,195</sup> Provided that polymer formation in the presence of water is a more plausible terrestrial prebiotic scenario than dry heating, most authors have chosen to work with high initial concentrations of amino acids to kinetically favor the process, as well as short heating times to avoid rapid amino acid degradation at high temperatures.<sup>196–198</sup> These conditions simulate submarine hydrothermal vents (SHVs), though given the expected low concentrations of amino acids and the long residence times and range of temperatures in SHVs, hydro-thermal peptide syntheses seem highly implausible, geochemically speaking.<sup>199</sup>

One possible way to favor the removal of water in peptide bond formation and, therefore, shift the equilibrium toward polymeric products, is through subtle physicochemical changes in the reaction medium. Fluctuating heating cycles of amino acid solutions in the presence of mineral catalysts have been shown to induce peptide synthesis.<sup>200</sup> The formation of peptides catalyzed by different types of clays (e.g., montmorillonite)<sup>201</sup> or silica or alumina,<sup>202</sup> as well as by recurrent wetting/drying cycles, works only for the simplest amino acids (i.e., glycine, alanine, etc.) and in very small yields. Nevertheless, these processes could have importance for chain elongation and stabilization against hydrolysis.<sup>203</sup> Interestingly, montmorillonite clay also catalyzes the synthesis of nucleobases (see section 2.1.3) and the oligomerization of activated ribonucleotides (see section 2.2.2). Thus, a prebiotic scenario where oligopeptides and oligonucleotides were synthesized by the same catalytic system (i.e., phyllosilicate surfaces or interlayers in contact with a bulk aqueous medium) could be envisioned.

Wetting and drying cycles have also been used in the so-called "salt-induced peptide formation" (SIPF), thanks to the dehydrating ability of concentrated NaCl solutions.<sup>204</sup> A decisive role in this reaction is played by Cu<sup>2+</sup>, which coordinates two amino acid ligands to form a copper complex intermediate (Figure 4). In this way, the reaction proceeds



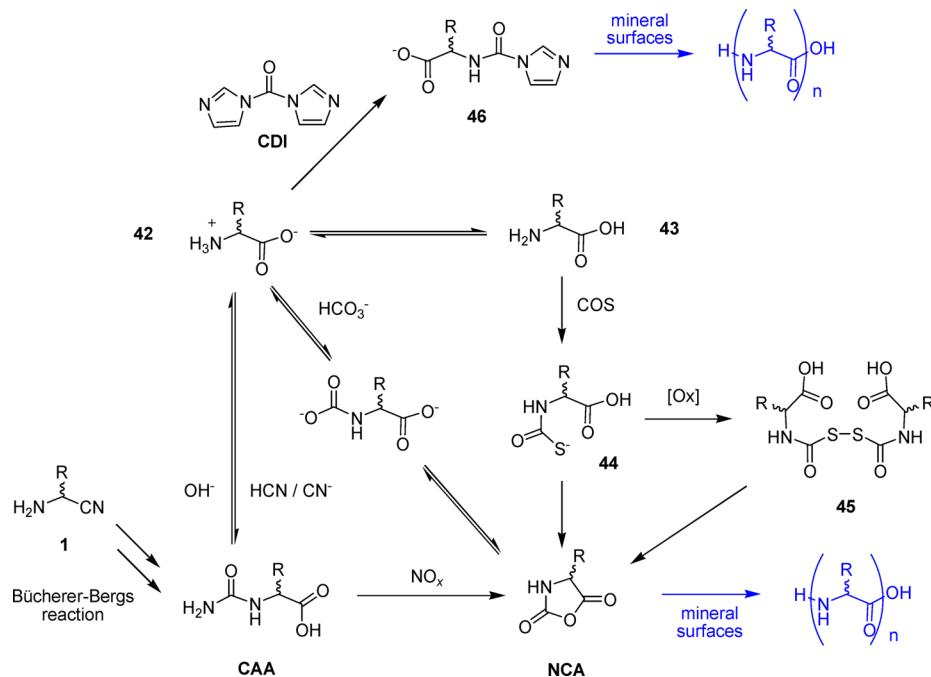
**Figure 4.** SIPF copper complex geometry with two glycine ligands, optimized by ab initio Hartree–Fock calculations. Reprinted with permission from ref 210. Copyright 2007 Wiley–VCH.

intramolecularly and with much more efficiency. If the reaction takes place in combination with a clay catalyst, longer peptides are obtained.<sup>205</sup> Furthermore, the presence of glycine<sup>206</sup> or histidine<sup>207</sup> accelerates the reaction rates with less reactive amino acids.<sup>208,209</sup> The Cu<sup>2+</sup>-catalyzed SIPF reaction represents the simplest way to obtain peptides in aqueous solutions and is compatible with conditions presumably present in the primordial environment.<sup>210,211</sup> Other important features of this reaction are its preference for natural  $\alpha$ - over  $\beta$ - or  $\gamma$ -amino acids, due to the better complexation of the former with Cu<sup>2+</sup>; its general applicability to most amino acids; and a slow racemization rate, especially at low concentrations of reagents (see section 2.3). As a drawback, only short oligomers [i.e., up to 6 amino acids (aa) long]<sup>205</sup> are obtained by this procedure.

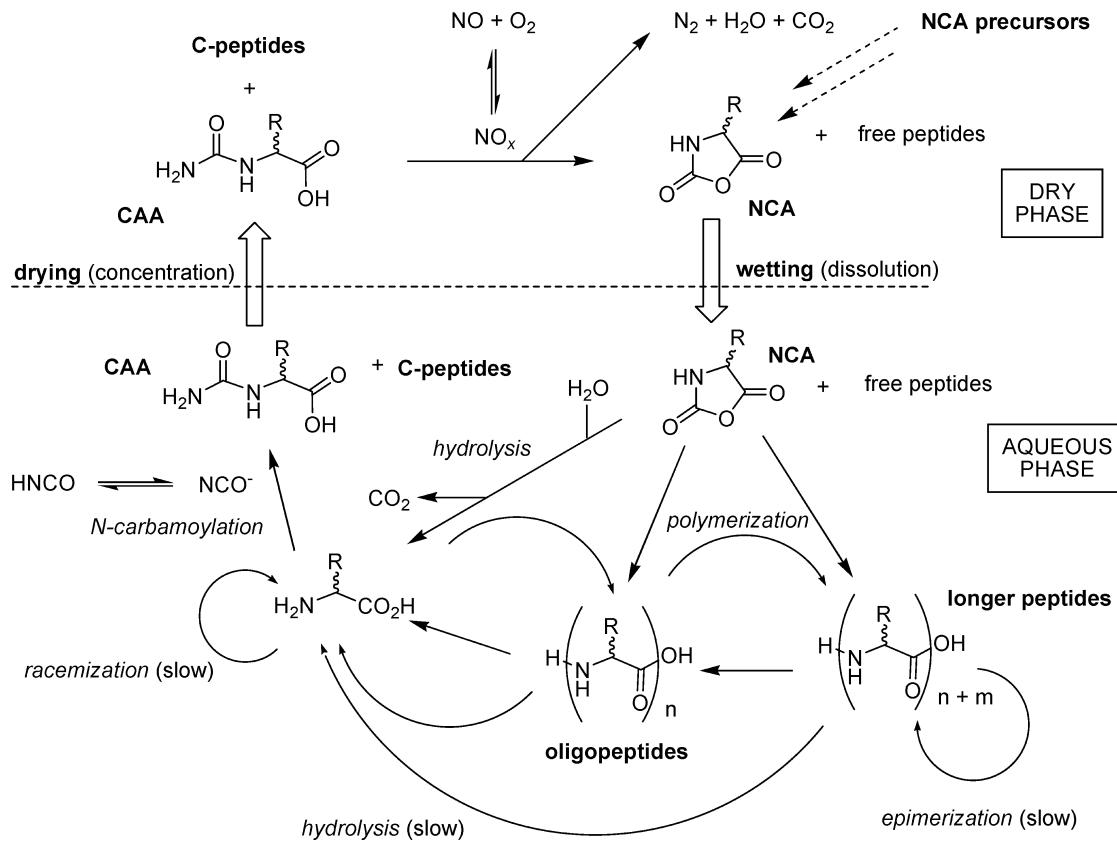
The above polymerization systems involve heating, which is problematic when a cold scenario for the origin of life is considered.<sup>42,212</sup> Yet peptide bonds can be formed from cooler, more dilute solutions, as has been demonstrated, for instance, using Langmuir-trough methods at the air–water interface with Cu<sup>2+</sup> molecules as catalysts.<sup>213</sup> Small organic activating molecules can also be used, in combination with mineral surfaces. The process resembles ribosome-catalyzed amino acid polymerization because the thermodynamically unfavorable peptide bond formation is coupled with an energetically favorable reaction. Typical examples of coupling reagents are imidazole and carbodiimides.<sup>214</sup> Orgel and co-workers<sup>215–217</sup> studied the oligomerization of several α- and β-amino acids on the surfaces of hydroxylapatite and illite, using 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDAC) and N,N'-carbonimidoyldiimidazole (CDI), respectively, as condensing agents. FeS<sub>2</sub> also catalyzes the CDI-induced oligomerization of arginine, although it does not adsorb oligoarginines. More recently, phenylalanine was polymerized with EDAC at 0 °C in the presence of Fe(OH)<sub>3</sub>.<sup>218</sup> The reaction is pH-dependent and seems to be based on an increase of substrate concentration in the adsorbed phase. All these studies are useful because they address different aspects regarding polymerization kinetics on a mineral surface, but their prebiotic significance is doubtful. The presence of carbodiimides in sufficient amounts on the prebiotic Earth has not been proved. Cyanamide tautomerizes in the presence of water to produce unsubstituted carbodiimide, but just at temperatures far below 0 °C ( $T < 100$  K).<sup>219</sup> Nevertheless, this process could be relevant as a possible source of carbodiimide derivatives in cometary and interstellar chemistry. In addition, Danger et al.<sup>220</sup> have recently claimed that cyanamide could promote peptide activation in aqueous media through formation of the cyclic intermediates 5(4H)-oxazolones.

One generally used procedure to accelerate the production of peptides over their hydrolysis is to convert the rather inert amino acids into an activated form, which is much more reactive than the amino acid. N-Carboxyamino acid anhydrides (NCAs), also named Leuchs anhydrides after their discoverer,<sup>221</sup> have led to high polymerization degrees (ca. 55-mers) upon adsorption on the surfaces of hydroxylapatite and illite minerals).<sup>222</sup> Excellent reviews on the chemistry of this universal amino acid activated form have been recently reported.<sup>93,223</sup> Chemically, NCAs offer the advantage of possessing an activated carbonyl group, while the amino group is protected. Moreover, they are neutral, in contrast to amino acids, which at neutral pH tend to adopt their charged zwitterionic form. NCAs are easily accessible through different transformations (Scheme 6), for example, by reaction of

Scheme 6. Different Synthetic Routes to N-Carboxyamino Acid Anhydrides (NCAs) and Other Amino Acid Activated Forms



Scheme 7. Primary Pump Model



zwitterionic amino acids (**42**) with CO<sub>2</sub>/HCO<sub>3</sub><sup>-</sup>, and subsequent dehydration.<sup>224</sup> NCAs can also be obtained by nitrosation of *N*-carbamoyl amino acids (CAAs),<sup>225,226</sup> a product of the Bücherer–Bergs hydrolysis of  $\alpha$ -aminonitriles (**1**) (see section 2.1.2). Activation of amino acids (**43**) by carbonyl sulfide (COS),<sup>227</sup> which is present in volcanic gases

although in a rather limited proportion ( $\leq 0.09$  mol %), leads to thiocarbamates (**44**), which can subsequently be converted into NCAs in good yields. These yields are particularly high when the reaction proceeds in the presence of an oxidizing agent, due to the formation of disulfide derivatives (**45**) as intermediates. The reaction can also take place with CO gas in the presence of

colloidal nickel or ferrous sulfide, producing carbonyl sulfide and, subsequently, NCAs.<sup>194</sup> Another volcanic gas that may activate amino acids for peptide formation is SO<sub>2</sub>.<sup>228</sup> Activation as imidoesters (**46**) with CDI results in peptide bond formation in aqueous solutions<sup>224</sup> and on mineral surfaces.<sup>216,222</sup> Finally, peptide synthesis from amino acid thioesters and thioacids has been shown to occur spontaneously.<sup>229,230</sup>

The main problem with all these scenarios is the obvious lack of a continuous source of activated monomers. Theoretical models involving protometabolic networks, by which a continuous feeding of activated reactants could be achieved, have been proposed.<sup>195,231</sup> One of these models is known as the primary pump (Scheme 7), developed by Commeyras et al.,<sup>232</sup> which could have operated in tidal shores or other fluctuating dehydrating/hydrating environments.

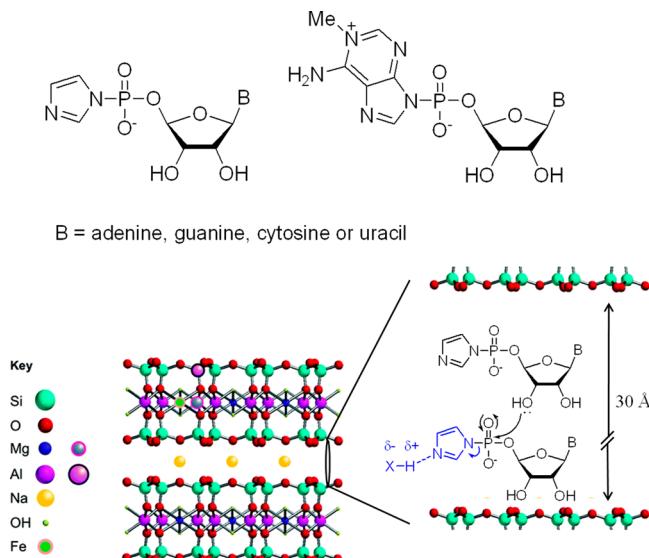
As stated above, these authors described *N*-carbamoyl amino acids (CAAs) that can be converted into NCAs, which readily polymerize under certain conditions. Since the aminolysis of NCAs is much faster than their hydrolysis, the formation of peptides is possible if the concentration of amine groups is enough and the pH is above 4. In the model, for the whole process of NCA synthesis and polymerization, a series of alternate chemical and physical processes need to be assumed, in order to ensure the constant income of CAAs by nitrosation of amino acids and to provide adequate polycondensation conditions. This could be achieved through dehydration/hydration cycles, slowing down competing reactions such as hydrolysis and epimerization. With these premises, if the time elapsed between each drying/wetting cycle was short enough so that peptide hydrolysis is not complete, peptides could in principle be accumulated in the medium. This model of protometabolism would also explain the emergence of homochirality without the need of autocatalysis (see section 2.3).<sup>233,234</sup> Unfortunately, the model does not have enough experimental support yet.

**2.2.2. Synthesis of Nucleic Acids.** The polymerization of current ribo- and deoxyribonucleotides in water is, as previously discussed for amino acids, a thermodynamically uphill process with free energy of ca. 5.3 kcal/mol<sup>235</sup> and does not happen spontaneously. Although heat drives the formation of phosphodiester bonds, mixtures of very short oligomers are obtained when ribonucleotides<sup>236</sup> [e.g., uridine 2'(3')-phosphate] are heated in acidic solutions until evaporation.<sup>237</sup> It has been shown that, in environments simulating hydrothermal vents, the rate of bond formation is faster than the rate of decomposition at 100 °C, but at higher temperatures (i.e., 200–300 °C) hydrolysis rates far exceed synthetic rates.<sup>238</sup> In this respect, the interface between hot and cold regions deserves special attention. Matsuno and co-workers<sup>239</sup> reported the first example of oligoribonucleotide synthesis at high temperatures from nonactivated nucleotides, but the oligomers obtained were short. Kawamura<sup>240</sup> has suggested that phosphodiester bond synthesis is possible in hot aqueous solutions, but coupling reagents, chemically activated monomers, or catalysts are required.

Coupling reagents such as cyanamide, water-soluble carbodiimides, etc., have failed to give oligomers longer than dimers and trimers.<sup>39</sup> The use of activated phosphates (e.g., polymetaphosphate, polyphosphoric acid or its salts) has resulted in the formation of short polyriboadenylates [poly-(A)], poly(C), and heterogeneous oligoribonucleotides.<sup>241</sup> The nonbiological 2'(3')-phosphodiester bond was predominantly present in the oligomeric chains obtained.<sup>242–244</sup> The most

utilized activation procedure for ribonucleotides has been their transformation into phosphoimidazolide derivatives (e.g., 5'-phosphoimidazolide of adenosine, ImpA).<sup>215,245</sup> With this type of activated monomer, both template-directed and non-templated polymerizations have been achieved. In this section, we will examine the case of nontemplated polymerizations, which give ca. 50-mers at maximum, while template-directed reactions will be revised in oncoming sections (sections 3.2 and 3.3, in particular).

The main leaving groups that have been used to activate ribonucleotides are nitrogen-containing heterocycles, such as imidazole and pyridine derivatives.<sup>246</sup> The reactivity of activated monomers is correlated with the pK<sub>a</sub> of the activating group and its ability to stabilize the positive charge formed during its displacement by the hydroxyl group of another monomer molecule.<sup>247</sup> The optimal pK<sub>a</sub> range for these leaving groups was found to be between 6 and 9, with aminopyridine derivatives being more effective than imidazoles. It is not clear, however, that imidazoles and methylpyridines existed in high amounts on the early Earth. According to this constraint, the most common heterocyclic compounds presumably present in the prebiotic world have been tested as activating groups, with special attention paid to adenine derivatives, which indeed contain an imidazole ring (Figure 5, top). Adenine derivatives



**Figure 5.** (Top) Structure of ribonucleotide 5'-phosphoimidazolides (left) and ribonucleotide 5'-phosphoro-1-methyladeninium (right). (Bottom) Unit cell of montmorillonite and phosphodiester bond formation within the clay interlayers, as proposed by Ferris and co-workers (right). XH, depicted in blue in the cartoon, is any undifferentiated protic species inside the clay galleries. Adapted with permission from ref 258. Copyright 2009 American Chemical Society.

such as 1-methyladenine showed similar or even better properties than imidazole for the activation of ribonucleotides (e.g., adenosine 5'-phosphoro-1-methyladenine, MeadPA).<sup>248</sup> This suggests how connecting building blocks from a heterogeneous mixture of prebiological molecules could have led to the right monomers for nucleotide polymerization. Yet the energy source that could have driven the linkage of adenine derivatives to the phosphate group of nucleotides is not obvious. The same driving forces as for condensation reactions, that is, hydrothermal heating and wetting/drying and freezing/thawing cycles, might have contributed to overcome the

energetic barrier of the reaction. Recently, eutectic phases in ice/water have also been utilized to promote the template-directed polymerization of imidazole-activated ribonucleotides in the presence of  $Mg^{2+}$  and  $Pb^{2+}$  metal ions.<sup>249</sup>

It has been proposed that inorganic surfaces could have acted as catalytic sites for the oligomerization of activated ribonucleotides.<sup>139,215</sup> Relevant experimental work performed by Ferris and co-workers<sup>250–253</sup> has shown that montmorillonite can concentrate and correctly orient ribonucleotide phosphoimidazolides, so that RNA-like oligomer chains of 6–14 monomeric units are synthesized in the absence of a template. Up to 50-mers are obtained if a 10-mer primer is added in the so-called “feeding reaction” or if 1-methyladenine instead of imidazole is used to activate the phosphate group. Montmorillonite is a phyllosilicate structured in parallel sheets, which contain two layers of tetrahedral silicon oxide sandwiching a layer of octahedral aluminum oxide (Figure 5, bottom).<sup>254</sup> There are, in the interlayers of montmorillonite, galleries with variable widths depending on the hydration state of the clay. The galleries contain cations that counteract the negative charges present, due to lattice defects, in the silicon oxide and aluminum oxide layers. The expanded interlayer in wet montmorillonites is wide enough to accommodate (DNA) polynucleotide chains,<sup>255</sup> and interlayer monovalent or divalent cations are known to mediate the reversible binding of RNA oligonucleotides.<sup>256</sup> In fact, oligomerization yields are enhanced by the presence of  $MgCl_2$  in the reaction mixture because of its effect on the binding of activated ribonucleotides and their oligomers to the clay surface.<sup>257</sup>

The mechanism responsible for the catalytic effect is unfortunately not well understood. However, polymerization reactions performed on montmorillonite are regioselective, preferring elongation through the 5'-phosphate of the ribonucleotide with the biologically relevant 3'-hydroxyl group of the elongating chain's ribose. This is related to the strength and orientation of ribonucleotide binding to the catalytic sites. Experimental<sup>258</sup> and computational<sup>259</sup> studies indicate that phosphodiester bond formation occurs at the interlayers between the clay's platelets. Such interlayers provide an adequate environment for the monomers to get close to each other, with the right orientation to yield preferentially 3',5'-phosphodiester bonds. The binding affinity of the 5'-phosphorimidazolide monomers to the montmorillonite determines the different reactivity between purine and pyrimidine derivatives, the relative polymerization rate of the activated ribonucleotides (as measured in binary reactions) being ImpA > ImpG > ImpC > ImpU.<sup>260</sup> This differential behavior introduces a bias in the diversity of the sequences generated,<sup>251</sup> which are not random but the result of a process endowed with a certain selectivity.<sup>252</sup> Molecular dynamics simulations have also been used to compare the probability of polymerization for L- and D-activated ribonucleotides, revealing that homochiral products are enhanced over the heterochiral ones.<sup>259</sup> Indeed, homochiral selectivity has been observed, to some extent, in experiments by Ferris and co-workers.<sup>261–263</sup>

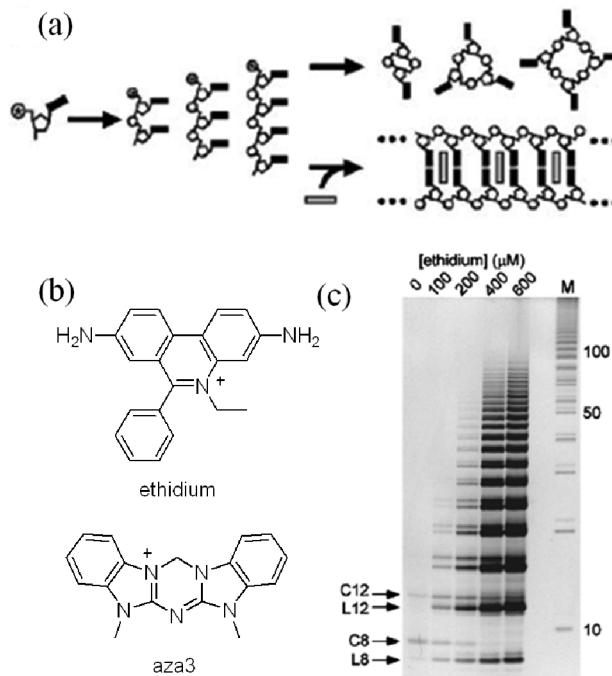
More recent approaches toward the polymerization of ribonucleotides in water have pointed their attention to some notions scarcely discussed in the literature until these new views arose. Costanzo et al.<sup>264</sup> have proved the synthesis of up to >120 ribonucleotide-long polymers from 3',5'-cAMP and 3',5'-cUMP, which can be obtained by phosphorylation of the corresponding nucleosides with a source of organic or inorganic phosphate in formamide.<sup>169</sup> These polymerizations occur

through two different types of reactions: an initial oligomerization of such activated ribonucleotides to yield up to 25- and 8-long oligomers for 3',5'-cAMP and 3',5'-cUMP, respectively, and a further chain extension by terminal ligation of oligomers. Moreover, the chemical ligation of RNA oligomers could be used to copy RNA molecules driven by sequence complementarity, and to increase the chemical complexity and information content of the pool of RNA molecules present in a certain prebiotic scenario, as theoretically illustrated by Briones et al.<sup>265</sup> and experimentally by Pino et al.<sup>266</sup> Szostak and co-workers<sup>268</sup> are also starting to explore the prebiotic synthesis of RNA by oligomer chemical ligation, which in the case of deoxynucleotides has been shown to proceed with high fidelity,<sup>267</sup> while good 3'-5' regiospecificities were obtained in the first template-directed ligations of oligoribonucleotides. Another advantage of this strategy is that the polymerization reaction would speed up due to the lower number of chemical steps necessary for polymer elongation. Anyhow, there are still many questions to be addressed about a scenario in which RNA polymers would grow and be copied in heterogeneous pools of nucleotide monomers and oligomers of differing lengths and sequences, as very well argued in a recent review by Szostak.<sup>31</sup>

A final comment about RNA polymerizations in water relates to the problem of strand cyclization. Activated short oligonucleotides have a strong tendency to cyclize efficiently, therefore reducing the efficacy of the polymerization process. Hud and co-workers<sup>269</sup> have developed a strategy to avoid oligomer cyclization by means of using nucleic acid intercalator agents such as ethidium salts or aza3 (Figure 6). These intercalators could be considered as cofactors assisting the chemical ligation of deoxyribonucleotide oligomers in the presence of the coupling reagent N-cyanoimidazole, a process generally favored by the capacity of the intercalator to bind to Watson–Crick base pairs. These results support the idea of Hud and co-workers<sup>269</sup> that some aromatic cofactors might have acted as “molecular midwives”, promoting the synthesis of duplex nucleic acids, up to 100 bp in length. However, this hypothesis needs to be confirmed for the case of ribonucleotide polymerizations, and the prebiotic plausibility of those aromatic cofactor molecules (prior to the existence of polymeric double-stranded nucleic acids) also needs to be addressed.

### 2.2.3. Coevolution in the Synthesis of Peptides and Nucleic Acids.

If the perspective that biomonomers existed in complex chemical mixtures is taken into account, various molecular species and aggregates could have driven the polymerization of amino acids and ribonucleotides. For instance, Luisi and co-workers<sup>270</sup> experimentally demonstrated that the polycondensation of NCAs could be assisted by the presence of liposomes, giving as a result up to 29-mers of Trp. Furthermore, they found that this could have interesting effects regarding the stereochemistry of the sequences obtained.<sup>271</sup> Lipid catalysis has also been proposed by Deamer and co-workers<sup>272</sup> for the synthesis of oligopeptides in the interior of vesicles. These authors proved that thioglutamic acid permeates into the vesicles at a rate sufficient to support its spontaneous, nonenzymatic polymerization, yielding peptides of up to 11 amino acid units. The same research group has used fluid lipid matrices, composed of present-day amphiphiles such as phosphatidylcholine, phosphatidic acid, and lysophosphatidylcholine, to polymerize nonactivated ribonucleotide 5'-monophosphate derivatives at high temperatures.<sup>273</sup> When a solution of monomers and lipid vesicles is heated and concentrated until total dehydration, vesicles fuse to form multilamellar films, with



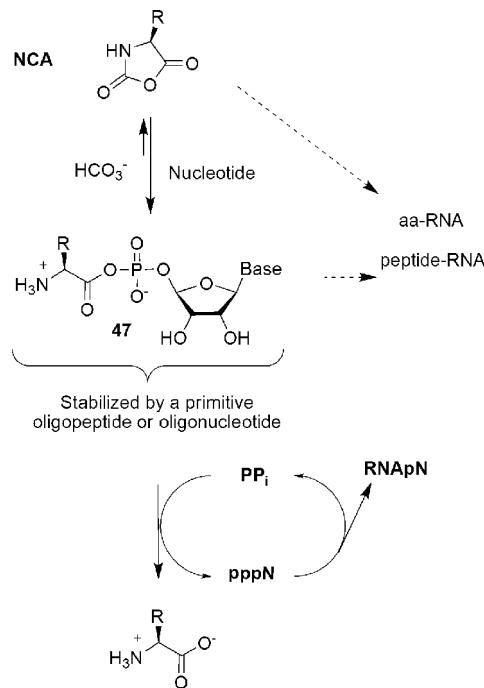
**Figure 6.** (a) Schematic representation of the problem of cyclization in the oligomerization of ribonucleotides and the chemical ligation of oligomers: in the absence of intercalators, oligomers cyclize (top route), while if intercalators are present, oligomers polymerize (bottom route). (b) Structures of two different intercalator agents used by Hud and co-workers.<sup>269</sup> (c) Polyacrylamide gel electrophoretic analysis showing the effect of ethidium in the polymerization of base-pairing oligodeoxyribonucleotides catalyzed by *N*-cyanoimidazole. In the absence of intercalator, the major products are cyclic octamer (C8) and dodecamer (C12). With increasing concentrations of ethidium present during the polymerization reaction, linear polymerization products up to 100 nucleotides long are promoted. Adapted with permission from ref 269. Copyright 2010 National Academy of Sciences.

monomer molecules encapsulated between the lamellae. This behavior led, after several dehydration/rehydration cycles, to RNA-like molecules containing up to 25–100 repeating units. Moreover, the same kind of alternating wet/dry prebiotic environment has enabled the nonenzymatic transfer of sequence information from a single-stranded (ss) DNA template into a complementary polymer of 2'-deoxyribonucleotides, with 5% yield with respect to the template and 9.9% fidelity of nucleotide incorporation during the replication process.<sup>274</sup> At the end of the reaction, the oligonucleotides are encapsulated in vesicles, which assemble after rehydration. It is noteworthy that the formation of vesicles can be catalyzed by montmorillonite and that RNA chains bound to the clay remain encapsulated within the vesicles.<sup>153</sup> Given the ability of montmorillonite to catalyze many prebiotic reactions, including the oligomerization of ribonucleotide phosphoimidazolides, a scenario where montmorillonite could have cooperated with lipid microenvironments in the synthesis of biopolymers gains significant appeal.

Regarding the activation of nucleotides, it is still uncertain whether amino acid derivatives or precursors could have acted as activating groups, leading to peptide–oligonucleotide interdependence.<sup>275</sup> As a relevant example of the systems chemistry approach, Mullen and Sutherland<sup>276</sup> have reported simultaneous ribonucleotide activation and synthesis of amino

acid amides in a potentially prebiotic multicomponent reaction. Amino acids in modern cells are usually activated with ATP, forming adenylate anhydrides, which are highly unstable and require stabilization. In current biochemistry, stabilization can be achieved by interaction with enzymes known as aminoacyl-tRNA synthetases, which covalently link the amino acid to its specific tRNA molecule, the specificity being defined by the genetic code. These aminoacyl-tRNA ester derivatives constitute the core of the translation apparatus in actual living organisms and may be a relic of an ancient combination of amino acid and nucleotide chemistries in the preribosome era. For the prebiotic existence of highly unstable aminoacyl adenylate anhydrides, which could result in the polymerization and coevolution of nucleic acids and peptides, Pascal et al.<sup>93</sup> have proposed a theoretical alternative mechanism in the absence of enzymes (Scheme 8), based on the chemistry of NCAs.

**Scheme 8. A Plausible Process of Ribonucleotide Activation Based on NCA Chemistry**



There is a difference in free energy of only 1.2 kcal/mol between NCAs and mixed anhydrides, so it is plausible that in the presence of a primitive oligopeptide or oligonucleotide that interacted favorably with the product (i.e., the ancestor of aminoacyl-tRNA synthetases),<sup>277</sup> such conversion could have taken place. Then the mixed anhydrides (47 in Scheme 8) may have converted ribonucleotide monophosphate into nucleoside 5'-triphosphates (pppN), for example, by reacting with pyrophosphate (PP<sub>i</sub>) or another source of activated phosphate. This reaction is just the reverse path of modern amino acid activation with ATP and is therefore thermodynamically downhill. The polymerization of ribonucleoside triphosphates to produce RNAPn would have regenerated pyrophosphate, giving rise to a metabolic pyrophosphate cycle, as shown in Scheme 8. Although this hypothesis still needs experimental confirmation, it makes conceivable an RNA world in which RNA chemistry would rely on the energy supplied by amino acid chemistry. Recent experiments suggesting that extremely

short peptides (e.g., dipeptides) could play a role in the formation of RNA phosphodiester bonds also point in this direction.<sup>278</sup> And the idea is compatible with the essential role that aminoacyl-tRNA molecules play in the interpretation of the genetic code.<sup>279</sup> Furthermore, peptidyl-tRNA conjugates may have served as anchors for RNA to be concentrated in aggregates of amphiphilic peptide molecules.<sup>280</sup>

An alternative approach to the coevolution of peptides and RNA has been followed by Yarus and co-workers<sup>281–284</sup> over the last three decades, through the investigation of interactions between RNA oligomers and individual amino acids or short peptides. The use of *in vitro* selection methods (see sections 3.3.2 and 5.3) has allowed characterizing specific RNA binding sites for a variety of amino acids with neutral, polar, charged, aliphatic, or aromatic side chains. In eight cases, a robust parallelism has been found between the nucleotide sequence of those RNA binding sites and the anticodon triplets of the corresponding tRNAs. This fact suggests that a “stereochemical era” might have existed during the evolution of the genetic code, relying on noncovalent interactions between amino acids and their binding RNA oligomers.<sup>281,282</sup> Furthermore, *in vitro* selection experiments made possible the isolation of RNA aptamers that bind to His-Phe dipeptides with much higher affinity than to each individual amino acid, and whose binding sites contain His anticodons.<sup>283</sup> This line of research has accumulated experimental support showing that most of the biochemical functions required for coded polypeptide biosynthesis, most likely, had precedents within the chemical repertoire of small RNA molecules present in the RNA world<sup>284</sup> and suggesting a prebiotically plausible, systems chemistry-based origin of coded peptide synthesis before the advent of ribosomes.

Despite the advances made in this avenue, from our point of view several challenges remain to be solved with regard to nonenzymatic, template-free polymerizations of amino acids and nucleotides. Although we have highlighted some experimental setups that lead to RNA and peptide oligomers, the lengths of the resulting molecules are not sufficient to ensure that at least some polymeric chains are replicated efficiently in a subsequent step. In addition, the different reactivity of the various activated amino acids or nucleotides determines the enrichment of preferential sequences in the corresponding oligomers. And regioselectivity (with regard to the formation of only 3'-5' phosphodiester bonds) has not been fully achieved in potentially prebiotic polymerizations of RNA. Moreover, the origin of stereoselectivity is also an open issue, although it is recognized that the generation of homochiral RNA and peptide molecules was a key step toward their templated replication. In fact, in the next section some of the main issues related to the origins of homochirality will be covered from a systems perspective. But in any case, the experimental efforts to tackle all these questions surely require prebiotically realistic scenarios containing higher levels of chemical diversity.

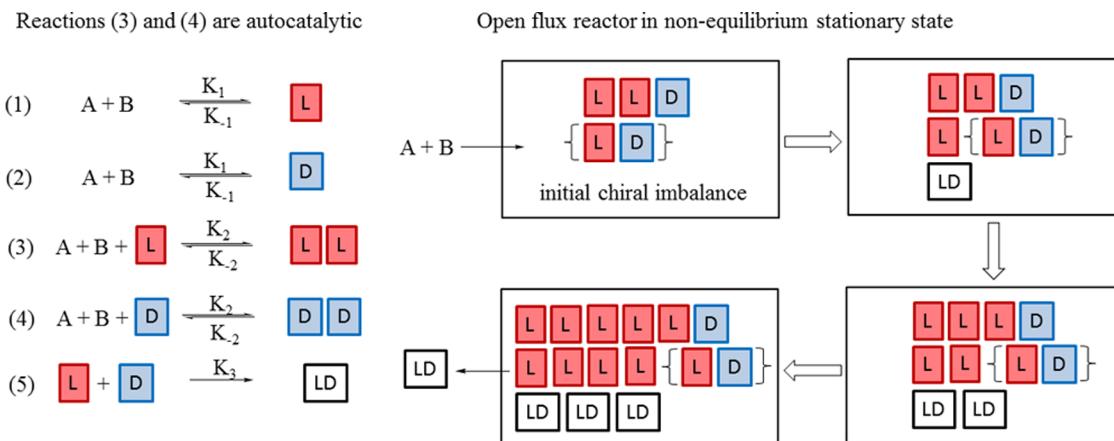
### 2.3. Origins of Homochirality

There is one general feature of the molecules constituting all known living systems on Earth, and in particular of biopolymers, which needs to be addressed and explained within the problem of origins: their homochirality.<sup>285–287</sup> Most molecules of life are homochiral, that is, they possess the same handedness or chirality. Biologically synthesized amino acids, for instance, occur exclusively in their levorotatory (*L*) form,

while the sugar constituents of nucleic acids are all dextro-rotatory (*D*). The question of whether homochirality was or was not an inevitable requisite for life is clearly open. The chemistry explaining how primitive homochiral peptides and RNA molecules could have been formed is not obvious, considering that most prebiotic routes toward nucleotides and amino acids start from achiral precursors such as formaldehyde, formamide, cyanoacetylene, etc. Moreover, there is also the possibility of finding elsewhere life forms based on biopolymers with opposite chirality to the ones present in our biological world. We do not know if nature's choice of *L*-amino acids and *D*-sugars was deterministic or accidental. We do not even know whether homochirality arose first for amino acids or sugars. The topic has been subject to several excellent reviews,<sup>288–294</sup> and this is not the first time that it has been approached from a systems chemistry perspective.<sup>295</sup> Due to its inherent complexity and the high level of technicalities present in the specialized literature, herein we will give only a brief overview of the main models and experiments around this intriguing issue.

Theories on the origin of homochirality in the living world can be classified into two major types: biotic and abiotic.<sup>296</sup> The first ones suggest that selection and amplification of one of the enantiomers of chiral biomolecules took place at an early stage of biological evolution. This view is, however, not consistent with the notion that biopolymers need to be composed of chiral monomers in order to perform their functions.<sup>297</sup> Proteins constituted by mixtures of *L*- and *D*-amino acids cannot form well-defined tertiary and quaternary structures. Similarly, ribose must have been in its *D* form for the first RNA molecules to adopt functional structures, which cannot occur with random mixtures of *D*- and *L*-nucleotides. An abiotic source of homochirality then seems more compatible with the principles of biology, but it implies the presumption of some kind of symmetry-breaking process leading to enantioenriched biomonomers.<sup>298</sup> It is also plausible that enantioenrichment could have happened along the synthesis of biopolymers rather than at the monomeric molecular level.<sup>272,299</sup> A compromise solution between both extremes would be that chiral monomers were only partially enantioenriched before they polymerized. The competition to build biopolymers would then be gradually won by the majoritarian enantiomer, leading to more efficient chiral selection as the complexity of biopolymers increased.

For symmetry reasons, chemical transformations starting from racemic mixtures of chiral compounds or from achiral substrates should statistically give nonchiral systems. The abiotic appearance of chiral materials hence entails deeper questions. How can asymmetry be originated in a universe governed by symmetrical laws? The unique possibility envisioned involves an initial symmetry-breaking event and the subsequent amplification of such asymmetry.<sup>290</sup> Experimental and theoretical studies have shown that initial nonzero enantiomeric excesses [ $ee = (D - L)/(D + L)$ , where *D* and *L* are the concentrations of both enantiomers] can happen by several means. A potential source of asymmetry relies on electroweak interactions and the parity violation principle, for which *L*-amino acids and *D*-sugars are slightly more stable than their corresponding enantiomers.<sup>300,301</sup> Yet the minute energy differences between them result in very low *ee* values, and experimental attempts to explain how chemical selection could have amplified such low excess have failed.<sup>302,303</sup> A second deterministic effect that could have induced mirror symmetry breaking is the presence of an asymmetric environment. Small *ee* values can be obtained by enantioselective degradation of



**Figure 7.** Schematic representation of Frank's model for chiral amplification. An open stationary system, maintained by the continuous input of achiral substrates A and B and the continuous pumping out (through precipitation, sublimation, reaction with another species, etc.) of the inactive LD product results in the amplification of one of the enantiomers if an initial imbalance is created from the racemic mixture, by either deterministic or stochastic mechanisms.<sup>292,324</sup> The L and D molecules that would react to form the inactive LD product are depicted in brackets.

amino acids with circularly polarized light (CPL).<sup>304,305</sup> Interestingly, organic materials extracted from the Murchison meteorite contain some nonnatural amino acids enriched with one of the enantiomers,<sup>306–308</sup> an effect that was proposed to be produced by CPL.<sup>309</sup> The Earth's rotation<sup>310</sup> and the asymmetry of quartz crystals<sup>311</sup> or meteoritic minerals<sup>312</sup> have been invoked as other possible sources of chirality from an asymmetric environment.

Stochastic processes can also explain the occurrence of "absolute asymmetric synthesis" (i.e., mirror symmetry breaking in the absence of an asymmetric physical or chemical driving force).<sup>313</sup> Since experimental chemical systems are of finite size, the probability that exactly equal numbers of molecules show opposite chirality is very low. For a system of  $N$  molecules, the standard deviation for the ee is equal to  $1/\sqrt{N}$ .<sup>314</sup> This nonzero ee is inherent to the discrete nature of matter, and in systems far from equilibrium, where phase transitions take place, the local microscopic bifurcation will always destabilize the racemic macroscopic state toward one of the two configurations. Amplification of the initial bifurcation is possible if it is coupled with some physical or chemical self-organization process by which enantiomers are separately organized into chiral structures.<sup>315</sup> There are several well-understood theoretical models based on this kind of stochastic induction of asymmetry, in either open or closed systems.<sup>316–318</sup> The initial and most elementary one was described by Frank,<sup>319</sup> and its confirmation happened to be experimentally feasible.<sup>320–322</sup> As stated in the literature, the asymmetry induced by electroweak interactions has shown a much lower intensity than statistical fluctuations, even for macroscopic systems.<sup>303,323</sup> The latter, therefore, might be more realistic under prebiotic conditions.

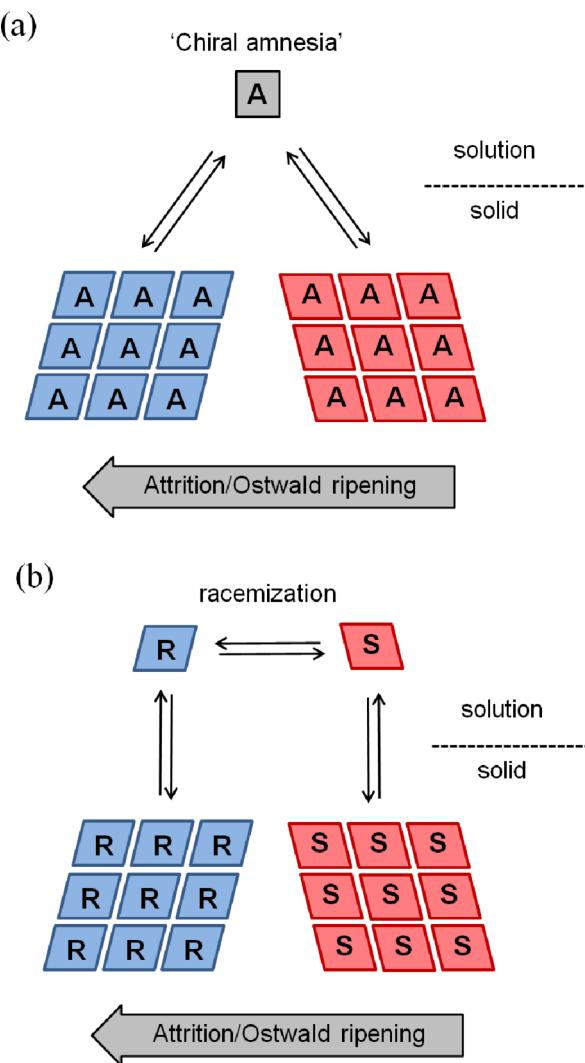
Most theories that aim at explaining how chirality can emerge from a nonchiral environment imply far-from-equilibrium conditions, with flows of energy continuously present. Energy flows are required as the driving force for self-organizing processes (i.e., crystallizations, polymerizations, supramolecular organizations, as we will analyze in more detail in coming sections) where chiral dissipative structures are formed.<sup>289</sup> If one of these structures catalyzes its own formation and inhibits the formation of the structure with opposite handedness, any stochastic mirror symmetry breaking fluctuation occurring in the system would be amplified. This behavior

is the basis of Frank-type models and is especially efficient in systems where autocatalysis is embodied in a complex network of reactions (Figure 7).<sup>292,324</sup> The potential appearance of different configurations at different locations of the system could be counteracted by the diffusion of molecules and spatial interactions between them, allowing a unique configuration to finally dominate the system.<sup>325,326</sup> This final state would then remain stable as long as the far-from-equilibrium regime is maintained.

As an alternative, some thermodynamic models based on the equilibrium phase behavior of crystal–solution eutectic systems<sup>327</sup> have been proposed for the enantioenrichment of amino acids<sup>328,329</sup> and nucleosides<sup>330</sup> in solution. The idea was first suggested by Morowitz<sup>331</sup> and has been implemented in recent years by crystal engineering of eutectics<sup>332,333</sup> or by exploiting the same type of phase behavior during the sublimation of amino acids.<sup>334–337</sup> Unfortunately, a solution enriched with one of the enantiomers of a chiral molecule by this type of mechanism would never be able to transfer and amplify the chirality acquired to other biomolecules or biopolymers, unless it is coupled with some chemical reaction network, preferably autocatalytic. As discussed in section 1, the route toward life is a kinetic self-organization process, for which the generation and maintenance of homochirality under far-from-equilibrium conditions seems in principle more coherent with the mechanisms by which the first living entities could have arisen.

Autocatalysis in Frank-type models is a very broad concept, which implies not only chemical transformations from one molecule to another molecule but also crystallizations, self-assembly processes, polymerization reactions, etc.<sup>290</sup> In any of these processes, the appearance of a certain chiral assembly of molecules and/or atoms may result in the enhancement of its own formation compared to the assembly of molecules of opposite handedness. As described in the model of frontal crystallization,<sup>338</sup> for instance, each growing crystal can be considered an open flow system, with a continuous transfer of matter and energy at the liquid/solid interface. Symmetry breaking can be observed in these systems due to small chiral imbalances and nonlinear autocatalytic effects during the incorporation of new molecules from solution into the growing crystal.<sup>339,340</sup> Such effects are particularly relevant for stirred

crystallizations, where the equilibrium between crushing of crystals and crystal growth is fundamental for the generation of single enantiomeric crystalline assemblies.<sup>341</sup> This phenomenon, first observed by Viedma<sup>342</sup> and later called “chiral amnesia”, has also been employed for deracemizing non-natural<sup>343–346</sup> and proteinogenic<sup>347,348</sup> amino acids that crystallize as conglomerates and can easily be interconverted in solution from one enantiomer to the other (Figure 8). The



**Figure 8.** (a) Complete deracemization of chiral crystals formed by an achiral molecule (e.g., NaClO<sub>3</sub>) through attrition-enhanced Ostwald ripening. (b) Complete deracemization of amino acids through attrition-enhanced Ostwald ripening of their conglomerate crystals and racemization in solution.

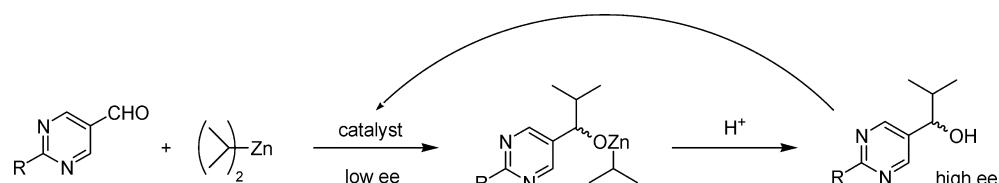
introduction of a small chiral bias (e.g., seeding with a small percentage of an enantiopure additive) allows direction of the

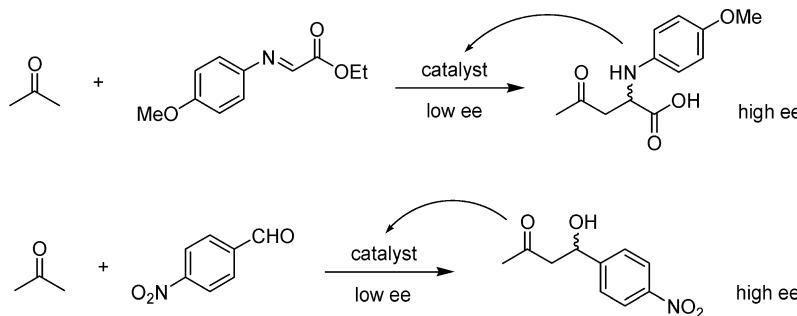
process to the formation of either pure L or D crystals.<sup>349,350</sup> Interestingly, mechanisms similar to these crystal resolutions operate in the generation of chiral supramolecular architectures from nonchiral molecules in solution,<sup>351</sup> which have been proved for different systems with potential prebiotic relevance, such as amino acid and porphyrin self-assemblies.<sup>352–355</sup> Particularly relevant among amino acid assemblies are the “magic number” clusters formed by eight units of serine (Ser<sub>8</sub>),<sup>356</sup> which have a remarkable preference for homochirality.<sup>357–359</sup>

A small chiral bias provoked by either deterministic or stochastic effects could also be amplified through coupling of the resulting enantioenriched system with efficient chemical, nonlinear autocatalytic processes. A paradigmatic example is the Soai reaction, in which a dialkylzinc operates as an alkylating reagent of prochiral aldehydes (Scheme 9).<sup>360,361</sup> The key aspect in this transformation is that the chiral alcohol obtained can associate with the organozinc compound and speed up the reaction. The alcohol product therefore acts as an autocatalyst, and its ee can be increased from an initial imbalance of, for example, 0.1% to values as high as 85%.<sup>362</sup> Moreover, chiral alcohols enriched in one of their enantiomers by photo-decomposition with CPL have been successfully used as catalysts.<sup>362,363</sup> Similar amplifications could be achieved by performing the reaction from racemic mixtures of the reactants in the presence of small quantities of chiral crystals of quartz,<sup>307</sup> DL-serine,<sup>364</sup> or adeninium dinitrate.<sup>365</sup> A striking example of the extraordinary power that the Soai reaction has for chiral amplification was the finding that it can even be triggered by the <sup>12</sup>C/<sup>13</sup>C carbon isotope chirality of the catalyst.<sup>366</sup> The mechanism of the reaction has been rationalized by Blackmond and co-workers,<sup>367–369</sup> who fit the experimental data to an extension of the model of Girard and Kagan.<sup>370</sup> Such kinetic model involved the coupling of autocatalysis with a form of “mutual antagonism”, therefore showing some resemblance with the model of Frank.

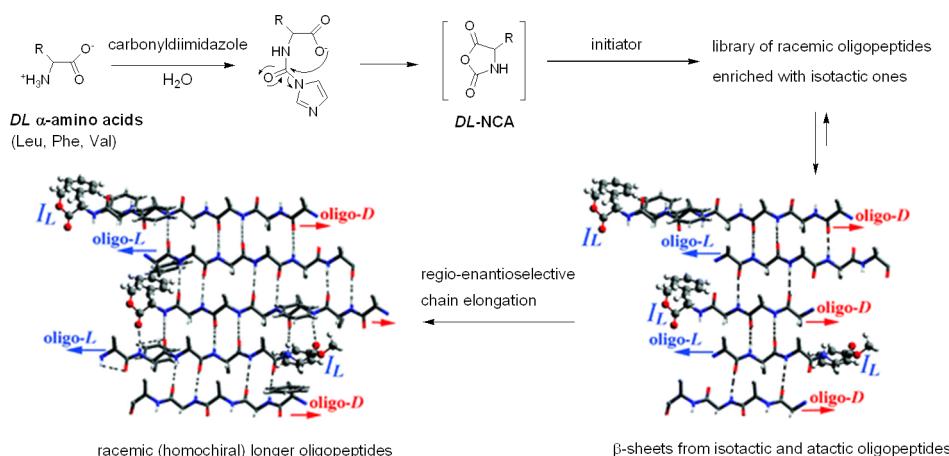
Although the Soai reaction is carried out in nonaqueous environments, and it is therefore not realistic from a prebiotic point of view, this process illustrates the feasibility of spontaneously driving nonchiral systems toward a single handedness by chemical autocatalysis. More recently, Tsogoeva and co-workers<sup>371–373</sup> have described purely organic reactions with similar properties, that is, autocatalysis and amplification of their ee. In particular, they described Mannich and aldol transformations (Scheme 10), which could have been likely involved in the prebiotic synthesis of amino acids and sugars, respectively. Mirror symmetry breaking in such organocatalytic systems has been explained by nonlinear reaction network kinetics and classical transition-state theory.<sup>374</sup> There have been, however, some controversies regarding the ee amplification mechanism,<sup>375</sup> and deeper mechanistic studies are underway in several laboratories. Deracemization by attrition-

#### Scheme 9. Soai Autocatalytic Reaction



Scheme 10. Mannich and Aldol Autocatalytic Reactions<sup>a</sup>

<sup>a</sup>Reported by Tsogoeva and co-workers.<sup>371–373</sup>



**Figure 9.** (Top) Schematic representation of regio- and enantioselective oligomerization of activated amino acids, like NCAs, and (bottom) proposed route for chain elongation via formation of racemic antiparallel  $\beta$ -sheets consisting of oligo-D- and oligo-L-amino acid chains, exemplified for the case of Phe-NCA. Adapted from ref 394. Copyright 2009 American Chemical Society.

enhanced Ostwald ripening of conglomerates is also possible with the products of these reactions.<sup>376</sup>

The above examples show that the absolute asymmetric synthesis of chiral materials, including amino acids, is possible. For the emergence of life, however, the synthesis of biopolymers is especially relevant. A crucial aspect of asymmetric crystallizations is that a large number of chiral interactions among molecules can be established within the forming crystal structure. Interestingly, similar cooperative interactions exist in polymer matrices.<sup>377</sup> Theoretical models state that, as in crystallizations, forward and backward reaction steps (i.e., polymerization/depolymerization) are required in order to establish real far-from-equilibrium conditions.<sup>233,378</sup> Again, the occurrence of autocatalysis or enantiomeric cross-inhibition under such circumstances can bias the chirality of polymerizing systems.<sup>379–382</sup> Although simple polymerization processes only lead to limited chiral amplifications, some experimental systems that amplify small chiral excesses in this way are starting to be developed.

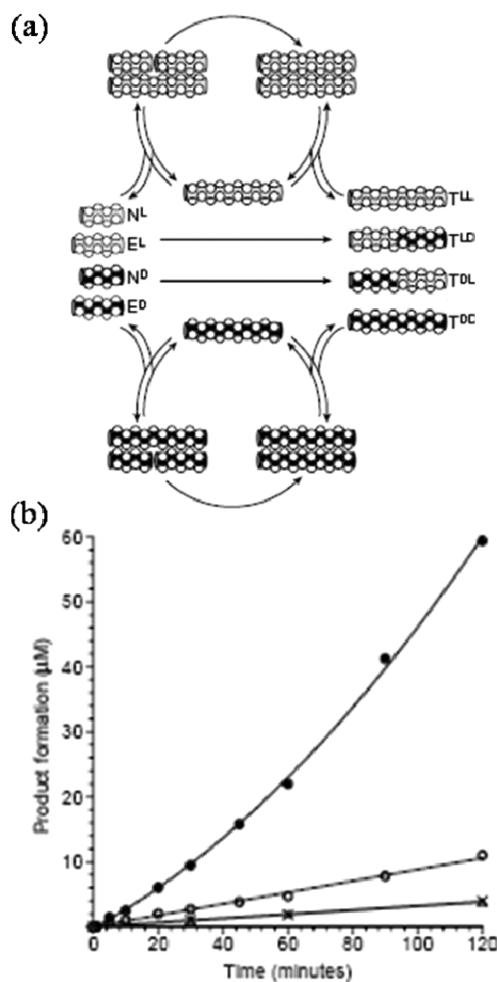
For example, a way to produce peptides of one handedness is by taking advantage of the supramolecular architectures (e.g.,  $\alpha$ -helices and  $\beta$ -sheets)<sup>383–385</sup> formed by growing peptides during their polymerization, which can serve as templates. One major limitation in this respect is the difficulty of obtaining isotactic chains of at least eight residues of the same chirality, the probability being as low as 1/2<sup>8</sup>. It is inferred from this that building up the right templates in a random polymerization of racemic amino acids is highly improbable. There is some

experimental support, however, for two approaches that could allow formation of homochiral short oligopeptides (Figure 9). First, during the polymerization of activated  $\alpha$ -amino acids, such as amino acid thioesters and NCAs, the N-terminal residues of growing oligopeptide chains have been shown to exert asymmetric induction, yielding libraries of short oligopeptides where the isotactic ones are expressed slightly further than those predicted from a random library.<sup>386,387</sup> Second, activated  $\alpha$ -amino acids can be organized in two- and three-dimensional (2D and 3D) crystalline surfaces and subsequently be subject to lattice-controlled polymerization at the crystal/aqueous interface.<sup>388–391</sup> After this initial induction of chirality,  $\beta$ -sheet colloidlike particles emerge and operate as stereoselective templates, or nuclei, in the formation of longer homochiral peptides and copeptides.<sup>392–395</sup> Such elongation of templates might have enjoyed a considerable enantioselective advantage in a prebiotic environment. The homochiral oligopeptides obtained by this means are unfortunately produced as racemic mixtures of oligo-L- and oligo-D-amino acids, and would therefore require further desymmetrization in subsequent aggregation or encapsulation steps.

Preferential formation of oligopeptides with homochiral sequences has also been observed on quartz<sup>396</sup> and in membranelike systems where phase separation of nonracemic activated monomers into racemic and enantiomeric 2D crystallites occurs.<sup>397,398</sup> Liposomes themselves have shown their ability to assist the stereoselective polymerization of amino acids and their NCAs.<sup>271,399</sup> Moreover, Luisi and co-

workers<sup>400</sup> described vesicles formed from homochiral methyl-fatty acids that catalyzed their own formation by hydrolysis of the precursor fatty acid anhydride, while the racemic vesicles were destabilized during hydrolysis and caused phase segregation. These results represent a model for the autopoietic self-reproduction of chiral prebiotic liposomes, which may have linked the emergence of homochiral compartments with the enantioselective synthesis of encapsulated biopolymers (see also section 3.2.1).

Once homochiral oligopeptides were obtained, Ghadiri and co-workers<sup>401–403</sup> have shown experimentally that they could have amplified their chirality by templated autocatalysis (Figure 10). Two short peptide fragments of 16 repeating units each,



**Figure 10.** (a) Schematic representation of the enantioselective replication cycles of homochiral peptides  $T^{LL}$  and  $T^{DD}$ , which are produced autocatalytically by template-directed ligation of a nucleophilic ( $N^L$  or  $N^D$ ) and an electrophilic ( $E^L$  or  $E^D$ ) 16-mer, while heterochiral peptides  $T^{LD}$  and  $T^{DL}$  result from uncatalyzed background reactions (light and dark backgrounds in the cartoon represent different regions of the sequence composed of L- and D-amino acids, respectively). (b) Rates of production of homochiral peptides ( $T^{LL}$  and  $T^{DD}$ , ●) and heterochiral ones ( $T^{LD}$  and  $T^{DL}$ , ○) as a function of time for a reaction mixture containing 100  $\mu\text{M}$  of each of the fragments  $E^L$ ,  $E^D$ ,  $N^L$ , and  $N^D$ . (X) Data from a similar reaction but carried out in a buffered solution with guanidine hydrochloride, which ensures the denaturation of catalytically competent intermolecular complexes. Reproduced from ref 403. Copyright 2001 Nature Publishing Group.

one nucleophilic and the other one electrophilic, were connected after their recognition on a template peptide of 32 complementary residues. The product of the coupling reaction was shown to have the same handedness as the template and thus could further catalyze its replication. The chiroselective replication cycles described by Ghadiri and co-workers possess a high fidelity, since autocatalysis is significantly diminished when only one out of the 16 amino acid repeating units in the peptide has opposite handedness. On the basis of this finding, self-replicating peptides (see section 3.2.2) could perpetuate homochirality.

The state of the art concerning research on the origin of nucleic acid homochirality is unfortunately less developed than for the case of amino acids and peptides, probably because a plausible prebiotic route for the synthesis of ribonucleotides is far from being resolved. In the prebiotic route toward nucleotides, Breslow<sup>404,405</sup> has suggested that the homochirality of (pre)biological sugars might not be independent of that of amino acids, based on the fact that a representative group of L-amino acids catalyze the synthesis of glyceraldehyde, from formaldehyde and glycoaldehyde, with D/L ratios of up to 60/40.<sup>406</sup> These ratios can later be amplified in water solutions to 92/8, by simply playing with the different solubilities of the D and DL forms. Similar conclusions can be extracted from the work of Weber and Pizzarello<sup>97–99</sup> on the enantioselective synthesis of tetroses and pentoses from glycoaldehyde and glyceraldehyde, catalyzed by chiral dipeptides. It is noteworthy, in the more interesting case of pentose production, that ribose was the only one to have a D ee, while lyxose and arabinose displayed L ee and xylose was roughly racemic.<sup>98</sup> Regarding further steps of nucleotide formation, the recent knowledge acquired by Powner et al.<sup>126,175</sup> on the synthesis of activated pyrimidine and purine ribonucleotides through aminooxazole chemistry (see section 2.1.3) has been utilized by Blackmond and co-workers<sup>407</sup> to develop a multicomponent reaction in which enantioenriched RNA precursors (with ee up to 80%) are produced from racemic starting materials. In particular, they make use of a small chiral imbalance of the amino acid proline (1% ee L-proline), which is amplified in a previous step by precipitation from organic solvent, to bias the reaction between racemic glyceraldehyde and 2-aminooxazole (compound 33 in Scheme 5) toward the corresponding enantioenriched aminooxazoline derivatives. Cooling the obtained mixture to 4 °C results in the subsequent formation of enantiopure riboaminooxazoline crystals, as previously described by Sutherland and co-workers.<sup>171</sup> Beyond all these important results, the homochirality of nucleic acid monomers is not strictly a requisite to explain certain enantiomeric excesses during the polymerization of nucleotides. During their investigations on montmorillonite-catalyzed polymerizations of activated ribonucleotides, Ferris and co-workers<sup>261–263</sup> reported small excesses of homochiral linear dimers, trimers, and tetramers, as well as of cyclic dimers, starting from racemic mixtures of ribonucleoside 5'-phosphoimidazolides.

In summary, given the large number of possible phenomena that have been described to generate chiral biomolecules from nonchiral matter, the assertion by Ball<sup>315</sup> that we are “spoilt for choice” among possible explanations for the origin of homochirality makes more sense every day. Mirror symmetry breaking events seem very likely in self-organizing chemical systems, from which it could be rationalized that homochirality was probably a general feature of the first biopolymers and

evolved together with their self-replication capabilities and the lipid compartments in which they were encapsulated.

### 3. COMPLEX CHEMICAL PROCESSES ON THE WAY TO LIVING SYSTEMS

In the previous sections we reviewed the main chemical pathways that have been discovered, so far, for the synthesis of monomers and polymers like those that living systems, as they are known on Earth, are made of. We also reviewed the literature about processes that could have led to one of the most striking features of those biomolecules, their homochiral nature. In the latter section it became apparent that the interaction processes and collective dynamics in which biologically relevant molecules are involved may have a crucial influence on their own properties as molecules, not only functional but also structural (e.g., their chirality). That brings us to consider now the more general question: how do biomolecules come together and modify or affect each other in the context of larger assemblies? How do prebiotic compounds become part of the complex network of interactions and transformation processes that constitute any living organism? The following two sections of this review will essentially deal with this question, approached from different angles. First, in this section, our focus will be on the interactions and self-organization/self-assembly/self-replication dynamics of biomolecules of one particular kind. Later, mostly in section 4, more heterogeneous mixtures will be analyzed, in which compounds or aggregates of various kinds are combined.

#### 3.1. Emergence of Complex Chemical Behavior: Self-Organization and Self-Assembly

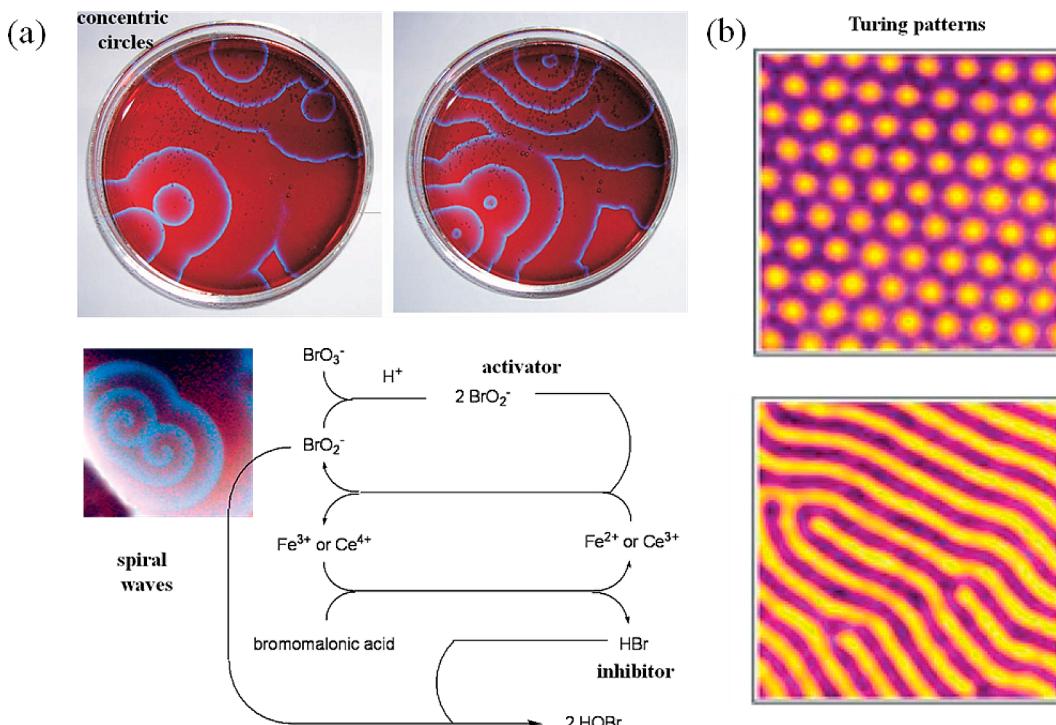
Prebiotic chemistry, whether it occurred in a “prebiotic soup”, in volcanic hydrothermal vents, in the early Earth’s atmosphere, or in outer space, is supposed to have involved many molecules interacting with each other in different ways. When a system is composed of a great diversity of components with multiple interactions between them, its global properties and dynamic behavior become much harder to explore and understand. Heterogeneous mixtures do not always produce interesting results, though; they may just lead to a collection of molecular structures clumped together in a more or less complicated arrangement. However, under certain conditions, they can trigger complex behaviors and emergent properties at a collective level, as we will see below. Even when chemical heterogeneity is not present in the system but it consists of a large ensemble of molecules under adequate physical conditions, complex patterns of dynamic behavior can originate, typically in a far-from-equilibrium thermodynamic regime (e.g., Bénard convection cells, hurricanes, or other types of dissipative structures).<sup>21</sup> This kind of phenomenon has been interpreted as self-organizing because the collective pattern spontaneously arises and maintains itself, provided that the boundary conditions that keep the system away from equilibrium are also maintained within a specific range of environmental parameters (e.g., in convection cells, a certain thermal gradient and a particular distance between the surfaces within which the liquid is contained). The temporal and spatial scales of these collective patterns are typically orders of magnitude larger than those of the constitutive parts. Therefore, nonlinear mechanisms must be present, so that microscopic fluctuations are amplified and stabilized into long-range correlations among many of the elementary constituents of the system. However, it is not easy to give a more detailed or

predictive explanation of the transition from the microscopic dynamics of those individual constituents to the global patterns generated, given the difficulties of applying methods from statistical physics in such far-from-equilibrium conditions. What has been successfully determined, for many cases, are the critical macroscopic parameters and range of values that rule bifurcations or changes in the various patterns of dynamic behavior observed.

With chemical systems the situation becomes even more intricate. As will be described in more detail next, the patterns of collective behavior in this context usually imply spatial and/or temporal synchronization of reactive processes, making it possible that under specific physicochemical conditions the concentrations of the various chemical species involved describe periodic or oscillatory trends. In fact, in chemistry nonlinearities abound (e.g., whenever autocatalytic reaction mechanisms are present, as we already saw in section 2.3), which may amplify microscopic fluctuations into long-range stable correlations, just as in physical systems. But the possibilities in terms of diversity of compounds and specific interactions among them (heterocatalysis, feedback loops, more indirect regulatory mechanisms, couplings between various autocatalytic cycles, etc.) are much wider and richer than in purely physical systems. In addition, other processes that involve many component parts but lead to quasi-equilibrium structures could be occurring locally at the same time. This second type of processes, when they happen spontaneously and are not dependent on the formation of covalent bonds but on weaker types of interaction (e.g., van der Waals, hydrophobic, medium-range ionic forces), are generally conceived as self-assembly and included within the area of supramolecular chemistry.<sup>408</sup> Although the latter bring about closer-to-equilibrium structures, with shorter-scale regularities or characteristic periodicities and a micro/macro relationship that can be easier to deal with theoretically, in practice they are very complex systems too.<sup>409</sup>

Interestingly, current living beings make extensive use of both self-organization and self-assembly processes at different levels, to support intra-, inter-, and multicellular behavior, but under the exquisite control of highly sophisticated genetic and enzymatic machinery.<sup>410</sup> In principle, these types of complex collective behavior do not require such a sophisticated molecular apparatus, as we will see in more detail below. However, it is not easy to find out the conditions under which self-organizing processes could provide a way toward more robust self-productive (metabolic) and reproductive/evolutionary (genetic) systems or be somehow recruited by the latter.

**3.1.1. Oscillatory Reaction Dynamics.** Oscillatory behaviors are so widespread in biological systems that they appear intrinsic to them.<sup>411–413</sup> Fundamental cellular functions where oscillatory processes play a role include metabolic cycles, such as respiration and carbohydrate synthesis in cyanobacteria,<sup>414</sup> immune responses,<sup>415,416</sup> apoptosis phenomena,<sup>417</sup> development<sup>418</sup> and growth hormone secretion.<sup>419,420</sup> There are two main mechanisms by which biochemical systems oscillate, involving either endogenous or exogenous processes. Endogenous processes arise from internal loops and feedbacks between the different components of metabolic networks (i.e., metabolites and enzymes).<sup>421</sup> A good example for this case relates to the periodic fluctuations exhibited by ATP and ADP in the glycolytic cycle.<sup>422</sup> Furthermore, studies on circadian oscillations of gene expression in various organisms<sup>423–425</sup> and



**Figure 11.** (a) Concentric circular and spiral wave fronts for the BZ reaction, whose core scheme is represented for the case of malonic acid as the organic substrate. The color waves are due to oscillatory changes in concentration of the metal complex redox states. (b) Turing patterns obtained for the CIMA oscillating reaction: (top) honeycomb and (bottom) labyrinthine structures. Adapted from ref 437. Copyright 2006 Wiley–VCH.

metabolic rhythms in yeast<sup>426</sup> suggest that cyclic changes in metabolic networks might be the underlying basis of many biological oscillating events, including circadian rhythms, hibernation, and the sleep–wake cycle.<sup>427</sup> If this is true, the roots of periodic biological phenomena could be deeply connected with prebiotic oscillating reaction cycles.

The second general type of oscillation mechanism in biology arises from external fluctuations in the environment. There are many sources of cyclicity (temperature, pH, humidity, illumination, UV irradiation, astronomic cycles, etc.) that might have contributed to the establishment of oscillatory behaviors in the first living entities.<sup>428–430</sup> In fact, oscillations often have more effective or significant outcomes than constant stationary states. The mathematical concept known as Parrondo's paradox, named after the Spanish mathematician Juan Parrondo, which postulates that two losing games can produce a winning result if appropriately alternated,<sup>431,432</sup> has actually been applied to model computationally chemical and biochemical systems.<sup>433,434</sup> The conclusion from these studies is that some kinetic systems counterintuitively yield higher concentrations of product(s) if some external parameters (e.g., temperature) are cycled than they do under steady-state conditions.

With the recent boom of studies on biological networks, some chemical systems that mimic biological oscillations have gained popularity as models to understand their biological counterparts.<sup>435–438</sup> The first oscillating reaction was discovered in the 1950s by Belousov<sup>439</sup> but encountered many objections from the scientific community to be published, until Zhabotinsky<sup>440,441</sup> became involved years later. The mechanism of this reaction was not fully understood until the 1970s,<sup>442,443</sup> and a decade later, similar oscillating reactions could already be designed.<sup>444,445</sup> Chemical oscillators are characterized by being far-from-equilibrium systems. In addition, two key features of

these complex networks of reactions are the autocatalytic production of at least one intermediate molecule and the presence of at least a negative feedback that consumes the autocatalyst(s). On these bases, we suggest the hypothesis that protometabolic networks could have presented similar basic properties, particularly when the ubiquity of autocatalytic and feedback loops in any current metabolic pathway is considered, together with the fact that their oscillatory character was probably ancient. As long as they keep far from equilibrium, for instance by the constant supply of a chemical fuel, oscillating reactions show characteristic temporal rhythms, often coupled to spatial wave patterns. They can be viewed as a chemical mimic of highly evolved biological behaviors including growth of bacterial colonies, morphogenesis in organisms, neuronal transmission signals through the nerves, and circadian rhythms.

Since the Belousov–Zhabotinsky (BZ) reaction is one of the most thoroughly studied cases, we will briefly focus on it, as an illustrative example of oscillatory reactions. In its classical form, this system comprises a one-electron metal redox catalyst (usually  $\text{Ce}^{3+}/\text{Ce}^{4+}$  or  $\text{Fe}^{2+}/\text{Fe}^{3+}$ ), an organic substrate (usually malonic acid or citric acid) and  $\text{BrO}_3^-$  ions, to give  $\text{Br}^-$ ,  $\text{CO}_2$ , and water. When all the components are dissolved in nitric or sulfuric acid and poured into a Petri dish, blue concentric and spiral circles form and propagate as waves (Figure 11a).<sup>446,447</sup> Under different experimental conditions, for instance, with water-in-oil microemulsions, the reaction can also bring about other interesting stationary patterns, such as Turing structures that are either stationary in time and periodic in space or periodic in both time and space.<sup>448–451</sup> Other common oscillating chemical systems are the chlorite/iodide/malonic acid (CIMA)<sup>452</sup> and ferrocyanide/iodate/sulfite<sup>453</sup> reactions. The CIMA reaction, and numerous modified versions of it, give Turing patterns in both solution and gel (Figure 11b).<sup>454</sup> For this type of patterns to be formed, it is necessary that the active

species in the system (i.e., the autocatalyst or activator) and its inhibitor diffuse at markedly different speeds. This can be adjusted in various ways, depending on the experimental setting and conditions of the system (e.g., through the percentage of water in oil, in microemulsions, or the strength of species interaction with the gel matrix, in gels).

As Alan Turing suggested in his revolutionary article to explain the phenomenon of morphogenesis,<sup>455</sup> chemical oscillations result from coupling the diffusion of reaction intermediates (i.e., what he called the morphogens) and their chemical transformation. The same mechanism was suggested for the formation of stripes in certain mammals, and the shape and arrangement of leaves in plants.<sup>456</sup> Similar analogies are commonly drawn in the field. For instance, a modification of the BZ reaction, consisting of loading a cation-exchange resin with the metal-redox catalyst ferroin  $[\text{Fe}(\text{phen})^{2+}]$ , allowed the synchronization of a large population of chemical oscillators.<sup>457</sup> The authors claim that there is a resemblance of this behavior with the quorum sensing transition observed in some bacteria.<sup>458</sup> The autocatalytic formation of  $\text{HBrO}_2^-$  is mediated by ferroin, hence this transformation is confined into the resin. In solution,  $\text{Br}^-$  ions inhibit the production of  $\text{HBrO}_2^-$  and therefore lower its concentration. The overall result is that the oscillations become synchronized over a certain concentration of  $\text{HBrO}_2^-$ , which is intimately related to the resin particle density. The system was implemented later by using  $[\text{Ru}(\text{bpy})_3]^{2+}$  instead of ferroin as catalyst, for which a smooth Kuramoto synchronization is observed (i.e., synchronization occurs by reaching a coupling strength threshold instead of a population density threshold).<sup>459</sup> This form of organization, together with communication ability between oscillator clusters, may play a role in the functioning of natural systems such as cellular differentiation. As shown by these examples, emergent behaviors could have originated in the first protocells not only from oscillations but from synchronization of various spatially confined, chemically oscillating systems.<sup>460</sup>

This active area of research will surely lead to further insights with biological relevance. In particular, the relationship between self-replicating chemical systems (see upcoming section 3.2) and the emergence of chemical oscillations and Turing patterns should be explored, following some initial theoretical results.<sup>461</sup> On parallel lines, it would be very interesting to establish possible chemical mechanisms to link primitive, oscillating reaction networks and protocellular compartments. Although no oscillating reaction that involves organic molecules of prebiotic significance has been reported up to now, this should not be an obstacle to the hypothesis that oscillatory systems were probably involved in the origin of the first protocells. Similar approaches to those described in the previous paragraphs could be used to discover prebiotically relevant reactions presenting spatial and/or temporal oscillations. In a 2007 article, Strier and Dawson<sup>462</sup> studied the oscillations observed in glycolysis. Turing patterns with a length scale smaller than the size of a cell could be induced in the system and might be responsible for cell division. Moreover, following a top-down approach, synthetic oscillators have been constructed through the integration of transcriptional and metabolic pathways observed in *Escherichia coli*.<sup>463,464</sup> In one of them it is the glycolytic flux, when its rate exceeds a critical value, that pushes the oscillation of acetyl phosphate.<sup>464</sup> Although the components used in these systems come from the current biological world, the search for prebiotic reactions

with analogous behavior may give us some surprises in the near future.

**3.1.2. Autocatalytic Networks and Protometabolic Cycles.** Autocatalysis is a very common mechanism in chemistry and is considered crucial in all scenarios of the origins of life, whether these are gene-first or metabolism-first views. We showed in the previous section that autocatalysis is at the basis of oscillatory behavior in many different reactions, not necessarily prebiotic but not necessarily involving complex molecules either. In addition, it appears pivotal to explaining the origin of homochirality, as discussed in section 2.3. But, beyond its relevance in those issues, autocatalysis is a fundamental concept from the point of view of any system that ought to maintain, grow, and produce more copies of itself, generating a collection of systems of similar—and potentially higher—complexity. Therefore, the transition from chemical systems toward biological ones seems simply unfeasible without molecular autocatalysis.

Nevertheless, the set of autocatalytic mechanisms that could be relevant for the origins of life is quite diverse. For instance, one should distinguish between network or cyclic autocatalysis, in which there is net production of (at least) one intermediate reactant within a closed cycle of chemical transformations (e.g., in the formose reaction network or in metabolic pathways), and molecular or replicative autocatalysis, where one of the products of a reaction serves as a catalyst for that very reaction (e.g., in template-directed RNA-dependent RNA replication). Biologically relevant autocatalysis often involves the net creation of bonds, but it may also be the result of cleavage reactions (e.g., hydrolysis of a precursor molecule). In addition, other forms of global autocatalytic behavior should be distinguished, like those deriving from cross-catalytic or cascade catalytic mechanisms in complex nonenzymatic molecular networks.<sup>465</sup> And there could even be more types of collective autocatalysis still not amenable to *in vitro* experimental exploration, like the one originally proposed by Kauffman in his reflexive autocatalytic sets<sup>466</sup> and more recent versions of similar ideas (e.g., the mutually catalytic noncovalent assemblies of Lancet and co-workers),<sup>18</sup> which are based on theoretical modeling and computer simulations.

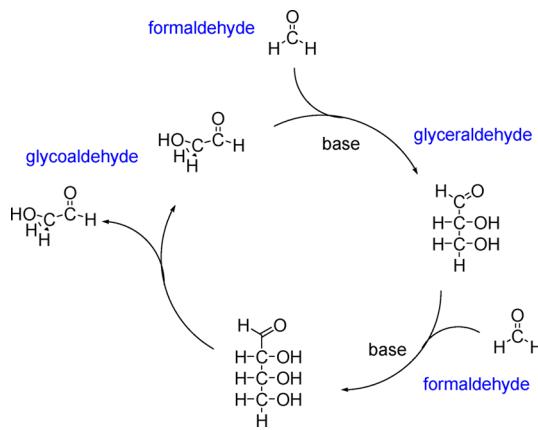
In the latter models, the global autocatalytic behavior of a network of molecules, be they peptides (in the case of Kauffman)<sup>466</sup> or lipids (in the case of Lancet and co-workers),<sup>18</sup> stems from assumptions made about the catalytic properties of those types of molecules when gathered in large ensembles of their respective kinds. The main claims of those theoretical proposals still require empirical confirmation, but the development of fields like organocatalysis<sup>52</sup> and the new avenues of research that are being opened to deal experimentally with increasingly complex nonenzymatic molecular networks<sup>465,467</sup> could soon provide additional support for them. Although the hypotheses made in that sense have been demonstrated in rather small networks so far (see section 3.2.2), progressive extensions to larger collections of molecules, particularly if these are spontaneously generated, would be real breakthroughs in the field of origins of life (see, for example, very interesting recent results on cooperative RNA networks, discussed in section 3.2.2.2).<sup>468</sup>

In this section we will focus our attention on network autocatalysis and review different proposals to establish a possible link between chemistry and primitive metabolic pathways, which is not an easy task to address experimentally.<sup>14</sup> For a recent specialized review on the topic, see ref 469. The

extensive work carried out on replicative autocatalytic networks will be covered in section 3.2.2. An important aspect of metabolic networks is their autocatalytic nature, although catalysis in a prebiotic context needs to be understood from a wide perspective.<sup>40,470–472</sup> One of the few currently uncontroversial assumptions in the field of origins of life is the fact that protein enzymes, the key catalysts in all metabolic pathways as we know them today, could not be initially present. However, alternative mechanisms to increase reaction rates have been suggested, without the aid of enzymes (e.g., organocatalysis, mineral catalysis, etc.).<sup>52,473</sup>

Furthermore, it is important to highlight that the concept of autocatalytic cycle does not necessarily involve some form of catalysis for each reaction step, as in present enzymatically controlled metabolisms, or a set of molecules where the formation reaction of any member of the set is catalyzed by another member of the same set, as in Kauffman's idea of reflexive autocatalysis. Strictly speaking, a protometabolic cycle could achieve self-sustenance simply by stoichiometric means.<sup>474</sup> In cycles whose elementary reactions are stoichiometric, rather than catalytic, the cycle itself can act as a catalyst if some byproduct of the cycle is produced in excess. An example corresponding to the formose reaction, in which two molecules of glycoaldehyde are formed for each molecule entering the cycle, is depicted in Scheme 11, but the same kind

**Scheme 11. One of the Possible Autocatalytic Cycles in the Formose Reaction**



of behavior is observed in Calvin or citric acid cycles. The emergence of catalytic pathways—for instance, by adjoining from the surrounding chemical milieu mineral catalysts, organocatalysts, or eventually ribozymes that catalyze one or various steps of the cycle—obviously would offer a kinetic and, thereby, evolutionary advantage to the cycle. In fact, this represents the most logical and probable evolution, from elementary autocatalytic cycles toward more intricate network topologies.<sup>475</sup>

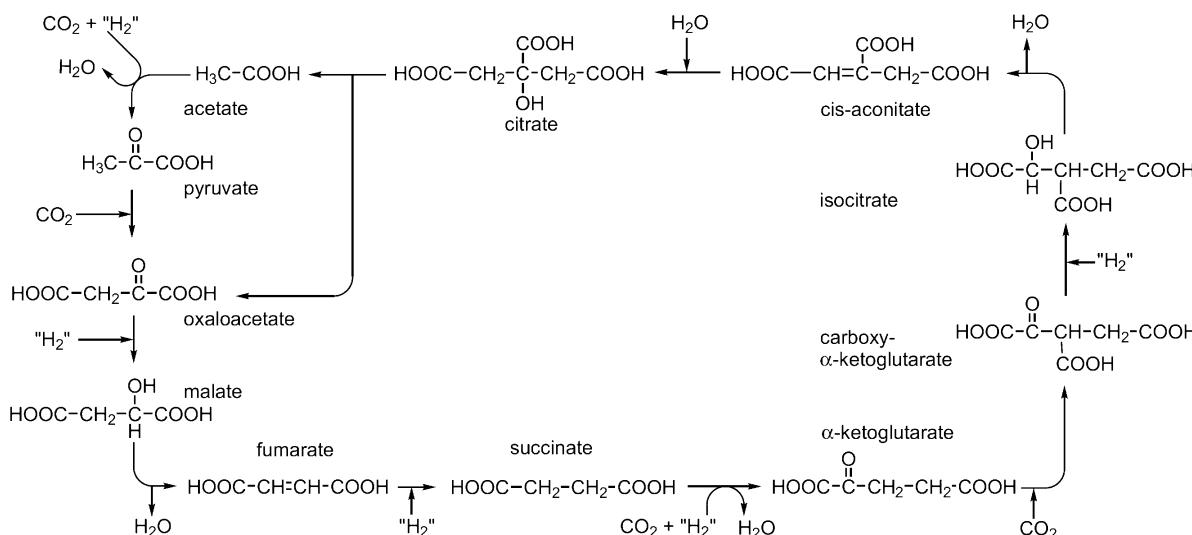
One of the main problems of this type of approach searching for protometabolic cycles is that, except for the already mentioned formose reaction, there are no other known examples of prebiotically plausible autocatalytic reaction cycles that could help to build a natural bridge between organic chemistry and biochemistry. From a perspective that cares about the overall energetic viability of prebiological systems,<sup>476–478</sup> and also reasoning from general thermodynamic grounds,<sup>20,479</sup> it is logically compelling to support the view that simple autocatalytic cycles had to come, in one way or another,

before the appearance of complex molecules like biopolymers (RNA, DNA, or proteins). On those lines, energetically plausible theoretical proposals for an “oligomer world”<sup>478</sup> or a “multimer–thioester world”<sup>476</sup> have been suggested. Another hypothesis, perhaps with stronger impact in the field, came from the study of intermediary metabolism and the universality and robustness of the reductive citric acid cycle. The main claim made by Morowitz and co-workers<sup>480,481</sup> is that this autocatalytic cycle (Scheme 12) could run without the assistance of enzymes and, given its central position in metabolism, could have been the starting point from which other metabolic pathways would unfold. Sharing similar concerns but resorting to the possible contribution of mineral surface catalysis in the process, Wächtershäuser and co-workers<sup>29,180,181</sup> proposed an alternative theoretical scheme for the development of autocatalytic cycles fixing carbon, from CO<sub>2</sub> or CO, through the reductive power of pyrite formation from Fe<sup>2+</sup> and H<sub>2</sub>S. More recently, Eschenmoser<sup>23,40</sup> has put forward another scheme that, starting from the chemistry of HCN and its oligomers, may build autocatalytic cycles and, hypothetically, lead to the prebiotic production of at least some of the constituents of the reductive citric acid cycle.

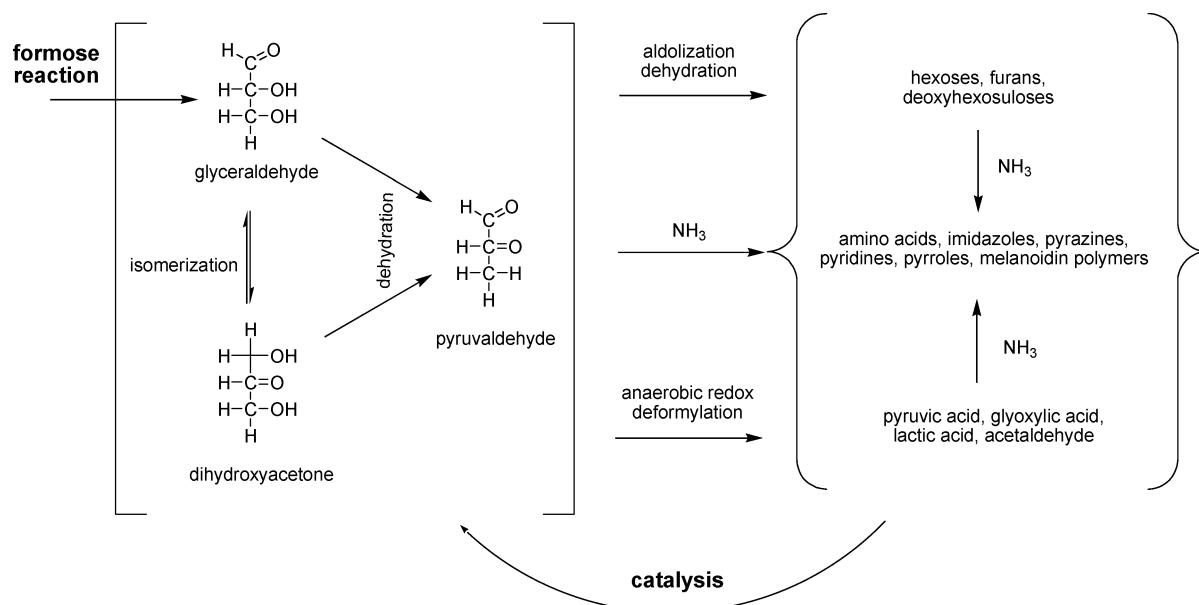
The hypotheses presented above, although very carefully conceived and with some empirical support for part of their claims, lack a solid body of experimental results behind them. Therefore, important objections have been raised against the metabolism-first scenario.<sup>14,482,483</sup> Any organic chemist would argue that the chemistry involved in these protometabolic processes would not be robust enough in the absence of enzymes. Furthermore, side reactions in the network would normally act as drains of intermediate molecules until the cycle turnover stops.<sup>484–486</sup> Nevertheless, experimental work to gain further insights into the chemical constraints that could have governed the emergence of metabolic networks has not stopped.<sup>194,487</sup> In this respect, Weber<sup>98,488</sup> has conducted a research program focused on understanding the thermodynamics and kinetics of prebiotically relevant organic transformations under mild aqueous conditions. Analysis of free energies for different transformations of prebiotic organic substrates, containing C, H, and O atoms, suggests that primitive chemometabolic pathways would progressively become more restricted, with very little ability to redistribute energy and promote kinetically unfeasible processes. In other words, irreversible reactions would provoke a continuous loss of energy, and stable functional groups that cannot undergo further transformations would be inevitably created and accumulated.

In this regard, Weber<sup>488</sup> has proposed that the most plausible pathway to build larger molecules from smaller substrates in aqueous medium, without the aid of either a high-energy supply or a catalyst, seems to be the addition of a carbonyl group to a nucleophile, such as in the aldol reaction. This suggestion strengthens the importance of the formose reaction, mainly based on aldol transformations from formaldehyde, as a primordial source of protometabolic chemistry. Sugars are interesting substrates from a thermodynamic point of view,<sup>489</sup> and their rich chemistry with amines or ammonia could have represented a path of energetic activation.<sup>490</sup> For instance, if they react with ammonia and a thiol under plausible prebiotic conditions,  $\alpha$ -hydroxy and  $\alpha$ -amino thioesters are produced,<sup>491</sup> showing an interesting link with the sulfur-based chemistry proposed by de Duve.<sup>477,492</sup> In addition, sugars spontaneously yield, in the presence of amines, numerous aliphatic and

Scheme 12. Reductive Citric Acid Cycle



Scheme 13. Chemistry of the Sugar Model: The Triose–Ammonia Reaction

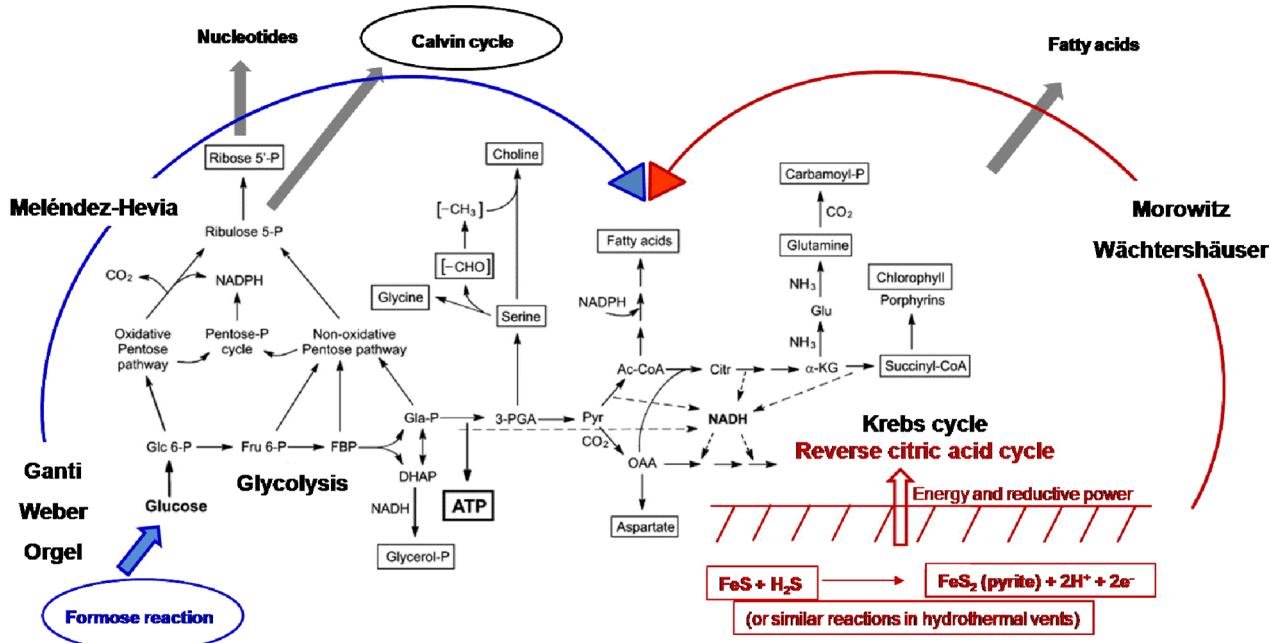


aromatic compounds (Scheme 13)<sup>162</sup> as well as melanoidin polymers, which have the ability to form cell-like aggregates.<sup>493</sup> Moreover, some of the compounds produced in transformations from sugar–ammonia reactive systems are catalysts of the same reaction, making it an autocatalytic one.<sup>494</sup> Such a rich chemistry favors the hypothesis that glycolysis was an ancient metabolic pathway, thus indirectly supporting a heterotrophic origin of life. In this respect, it is noteworthy that in the current biosphere some anaerobic microorganisms ferment sugars as the carbon and energy source to produce most of the essential biochemicals for their sustenance.<sup>495</sup>

There are also retrospective analyses of the current metabolic map that come in line with the previous proposal.<sup>496</sup> Following an alternative theoretical approach to the one taken by Morowitz, Meléndez-Hevia et al.<sup>497</sup> have arrived at the conclusion that the central, and eventually primordial, metabolic pathways are glycolysis and the Krebs or citric acid cycle (Figure 12). Their source of substrate would be ensured, since glucose is the most abundant product of the formose

reaction. From glycolysis and the Krebs cycle, the whole metabolic map and its possible chronology can be deduced. Taken together, these pathways are redox-independent and provide energy and reductive power, so they could have been chemically selected over multiple prebiotic processes. Again, this is just a tentative model that would need further experimental exploration, but it certainly seems worth trying.

One question that is relevant to ask, given the difficulties of this type of metabolism-first theories, is whether the initial chemistry involved not only should require mineral surfaces to proceed but also could rely on the presence of organic interfaces, like lipidic phases. Colloidal systems, in principle, make the scenario more messy and difficult to analyze from the point of view of a traditional synthetic chemist. Yet many oligomerization processes that are endergonic in free aqueous solution could be thermodynamically favored in the presence of these hydrophobic phases, as we mentioned earlier (see section 2.2). The influence of lipidic surfaces and compartments on prebiotically relevant processes could go beyond their role as



**Figure 12.** (Black) Minimal metabolic map, constructed by simplifying present-day cellular metabolisms. (Blue) The clockwise sense of metabolic evolution in the scheme of Meléndez-Hevia et al.<sup>497</sup> gives the formose reaction a prominent role as the first metabolic cycle, as Weber, Meléndez-Hevia, or Ganti proposed. (Red) The counterclockwise sense of metabolic evolution, according to the same scheme, would come from considering the reverse citric acid cycle as the first metabolic cycle, as Morowitz or Wächtershäuser have defended. In that case, energy and reductive power could be provided by redox reactions occurring on mineral surfaces (e.g., FeS, NiS) in hydrothermal vents, for instance. Adapted from ref 497. Copyright 2008 Elsevier.

concentrating agents or adjuvants of condensation reactions, as is typically assumed and has already been proved experimentally.<sup>273</sup> Some lipid bilayers could also have directed catalytic effects on protometabolic reactions taking place at the boundary with the aqueous solution. Despite the fact that relevant experimental evidence on those lines has been reported,<sup>271,272</sup> this avenue of research remains largely unexplored.

In the next section, we will review what kind of processes could underlie the formation of supramolecular lipidic structures in prebiotic conditions and whether it is reasonable to include that type of processes and structures in a hypothetical scenario for the origins of either biologically relevant protometabolic reaction pathways or genetic molecules.

**3.1.3. Assembly of Amphiphilic Molecules into Protocellular Compartments.** As was discussed in section 2.1.1, different surfactant molecules made of a polar head and a nonpolar, hydrophobic tail could be produced in certain prebiotic scenarios. They include single-chain fatty acids, isoprenoids, and, with increasing difficulty, their glycerol derivatives or more complex lipids (e.g., double-chain amphiphilic molecules with an additional phosphate group in their polar head).<sup>61,78</sup> Nevertheless, it seems reasonable to assume that the latter appeared biosynthetically at a later stage.<sup>64,498,499</sup> All these amphiphilic compounds tend to assemble into supramolecular aggregates of various types (e.g., micelles, rods, vesicles, etc.) when they come in contact with water. The aggregation process is mainly ruled by the so-called “hydrophobic effect” but also by other various weak interactions, like van der Waals and electrostatic forces.<sup>62</sup> Thus, in contrast with polymerization reactions, which require the formation of covalent bonds, the self-assembly of these

supramolecular structures typically relies on multiple weak interactions acting in parallel across the numerous components of the system. It is, therefore, a complicated process that depends not only on the properties of the amphiphiles involved (molecular geometry, packing parameter, size and possible charge of the polar head, and length and number of unsaturations of the hydrophobic chain) but also on their respective concentrations, on how they are combined (relative ratios and spatial distribution), on their interaction with solvent molecules (monomer solubility), and on the physicochemical parameters of the system (temperature, pH, ionic strength, etc.).<sup>61,500,501</sup>

Of all the possible supramolecular structures into which amphiphilic molecules can organize, here we will concentrate on vesicles (closed bilayers that contain and are surrounded by an aqueous environment), although micelles and other less finely ordered structures (e.g., simple lipid droplets) could also be important agents for prebiotic chemistry to develop.<sup>502</sup> Furthermore, the interactions between micelles and vesicles may prove crucial to understand the dynamics of the latter. For instance, micelles have actually been shown to play a role in the process of vesicle growth and division,<sup>503</sup> so the mechanisms involved could potentially allow the delivery of micellar contents into the lumen of vesicles. This type of process would represent a link between the chemistry operating in prebiotic micelles with the chemistry that might take place within vesicles. However, due to space constraints, this section will just cover protocellular model systems based on vesicles because they show a more obvious continuity with biological cellularity. For a wider historical perspective, see a recent review by Hanczyc.<sup>504</sup> In addition, we will leave aside investigations on standard liposomes (closed membrane structures whose main components are complex phospho- and glycolipids and whose

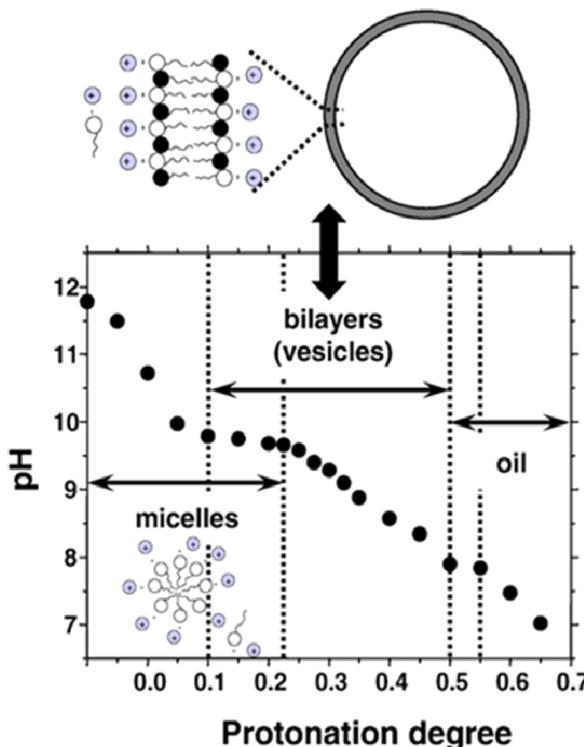
main explanatory targets are proper biological membranes),<sup>505</sup> in order to focus on approaches that address more specifically the question of origins.

As we mentioned in section 2.1, two alternatives have been proposed as the first biologically relevant amphiphilic molecules that could form stable bilayer compartments: simple isoprenoid lipids, made of single polyprenyl chains,<sup>506</sup> versus fatty acids, also single-chain but based on alkyl hydrocarbons.<sup>507</sup> The experimental confirmation of one model or the other, or a combination of both, would have important implications for our view on the origin of biomembranes and, more precisely, on how to solve the intriguing evolutionary question of the “lipid divide” (i.e., the fact that archaea and bacteria developed membranes made of radically different types of lipids).<sup>508,509</sup>

It has been experimentally demonstrated that both types of single-chain amphiphiles can self-assemble into vesicles.<sup>510–512</sup> Nevertheless, the lower polar head complexity required for self-assembly in the case of fatty acids, together with their apparently easier synthesis (see section 2.1.1), has led the community of origins-of-life researchers working on compartments to favor the latter, providing relatively less attention to isoprenoids. The issue is still under debate,<sup>513</sup> although the presence in Martian meteorites of fatty acids that do form vesicles<sup>71,72</sup> is a significant piece of evidence in support of the fatty acids-first scenario over the terpenoid-first model. Yet the hypothesis favoring primordial vesicle formation by fatty acids also faces important difficulties,<sup>67</sup> like their apparent lack of stability to changes in pH and osmotic pressure, as well as their relatively high critical aggregation concentration (cac) values.<sup>514</sup>

Indeed, the pH sensitivity of vesicles made of fatty acids reflects the fact that the aggregation state of this kind of molecules strongly depends on the degree of protonation/deprotonation of their terminal carboxylic groups (Figure 13).<sup>515,516</sup> Fatty acid molecules have an inherent conical geometry, which favors their self-assembly into micelles, and only if there is a fine balance between protonated and deprotonated molecules (i.e., if the pH of the solution is around the  $pK_a$  of the fatty acid in the aggregated state) do vesicles spontaneously form. This is thanks to hydrogen bonding between pairs of fatty acid and fatty acid salt.<sup>517,518</sup> These pairs become then more cylindrical, somehow mimicking a double-chain lipid that naturally tends to assemble into flat bilayers. Furthermore, Monnard et al.<sup>519</sup> experimentally demonstrated that these vesicles are also strongly sensitive to increases in the ionic content of the aqueous solution where they form, precipitating at salt concentration values well below those estimated for primitive ocean conditions, which are roughly similar to the present ones (i.e., ~35 g/L, or 0.46 M chloride ions, although this value could be lower for the first stages after the formation of oceans).

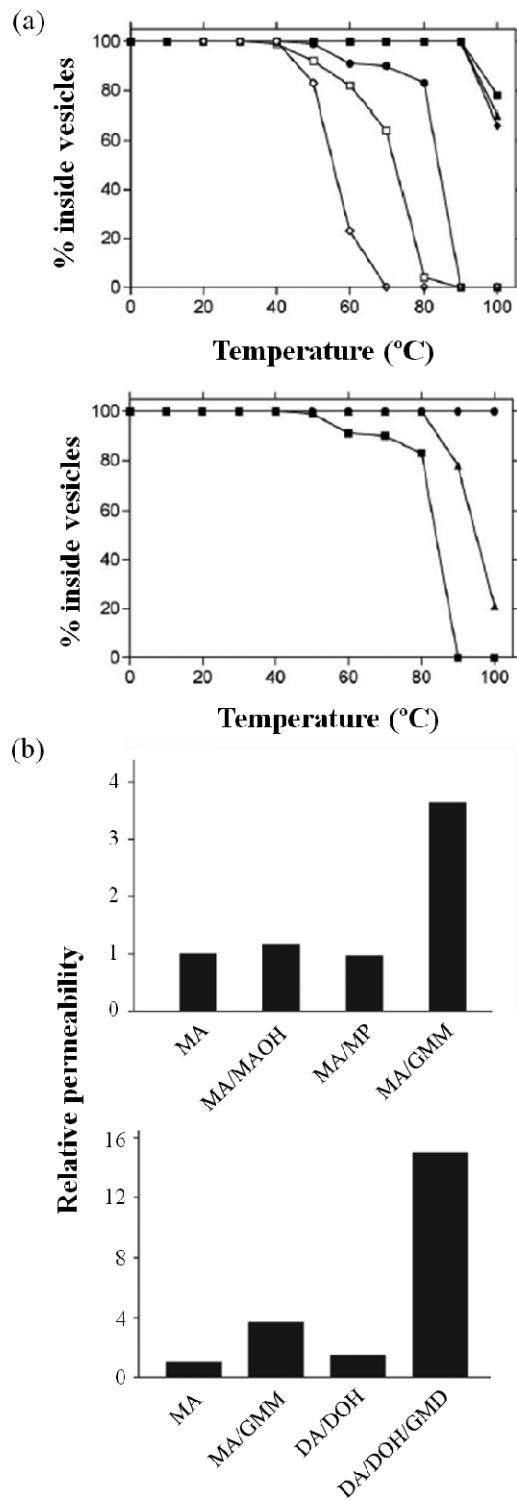
However, the situation changes quite remarkably when, instead of pure fatty acid vesicles, mixtures of different fatty acids or mixtures of fatty acids with other surfactant compounds are used. This, apart from being a more plausible premise of work in any realistic prebiotic setting,<sup>31</sup> has been intensively explored by different groups in recent years, providing a clear way out of the aforementioned problems. For instance, Apel et al.<sup>82</sup> showed that short-chain fatty acids mixed with alcohols not only widen the pH range under which vesicles are stable but also induce a significant decrease in the typical critical vesicle concentration (cvc) values of the system (see also ref 520 for an extension of this work to longer-chain fatty acid vesicles). Namani and Deamer<sup>521</sup> actually confirmed



**Figure 13.** Example of phase behavior of fatty acids in water, depending on pH: titration curve for 80 mM oleic acid/sodium oleate. The pH ranges where micelles, vesicles, or oil droplets form are indicated. Reproduced with permission from ref 501. Copyright 2007 Elsevier.

that fatty acid vesicles, if combined with an adequate surfactant, withstand even extreme conditions of pH and much higher ionic strength than the pure ones. Maurer et al.,<sup>83</sup> in turn, have recently explored mixed systems of fatty acids with their corresponding glycerol derivatives, showing that cvc values consistently decrease, whereas thermal stability is enhanced. In this context, PAHs, which have been proposed to be delivered in large amounts to the early Earth during the heavy bombardment phase of the young solar system (and which could have played a prebiotic role, for example, as energy transduction elements),<sup>522</sup> have also been incorporated into decanoic acid vesicle bilayers in molar ratios up to 1:10 (PAH:decanoic acid).<sup>523</sup> In particular, the insertion of oxidized PAH derivatives such as 1-hydroxypyrene and 9-anthracenecarboxylic acid produced an up to 4-fold decrease in permeability of the vesicles to small solutes, suggesting a potential cholesterol-like behavior of PAHs in primordial protocells.

Nevertheless, the work with fatty acid vesicles that is probably having a stronger impact in the field of origins of life in recent years is that carried out in Szostak's lab. The main insights gained from that line of research include the competitive dynamics that could be implemented in populations of this type of vesicles<sup>524</sup> and their reproductive capacities,<sup>153,503</sup> to be reviewed more specifically in section 3.2.1. Mansy and Szostak<sup>525</sup> have also shown that vesicles composed of single-chain fatty acids, fatty alcohols, and fatty acid glycerol esters are stable at temperatures up to 100 °C (Figure 14a). These compartments are able to retain nucleic acid oligomers in their lumen at those elevated temperatures, which has been claimed to make feasible the process of strand



**Figure 14.** Thermostability and permeability of model protocell membranes. (a) Plot of leakage of fluorescein-labeled dA<sub>10</sub> oligonucleotide from vesicles of different composition. Top, comparison of influence of acyl chain length: (◊) 2:1 decanoic acid/decanol; (●) myristoleic acid; (▲) palmitoleic acid; (◆) oleic acid; (■) linoleic acid; (□) 2:1 myristoleic acid/farnesol (farnesol is an isoprenoid derivative). Bottom, comparison of influence of headgroup: (■) myristoleic acid; (▲) 2:1 myristoleic acid/myristoleyl alcohol; (●) 2:1 myristoleic acid/glycerol monoester of myristoleic acid. Adapted with permission from ref 525. Copyright 2008 National Academy of Sciences. (b) Permeability values to ribose for different model protocell membranes, comparing the influence of (top) polar

Figure 14. continued

headgroup and (bottom) length of acyl chain. DA, decanoic acid; DOH, decanol; GMD, glycerol monoester of decanoic acid; GMM, glycerol monoester of myristoleate; MAOH, myristoleyl alcohol; MP, myristoleoyl phosphate. Adapted with permission from ref 526. Copyright 2008 Nature Publishing Group.

separation in double helices of nucleic acids, potentially encapsulated within a primitive protocell (see section 3.2.2).

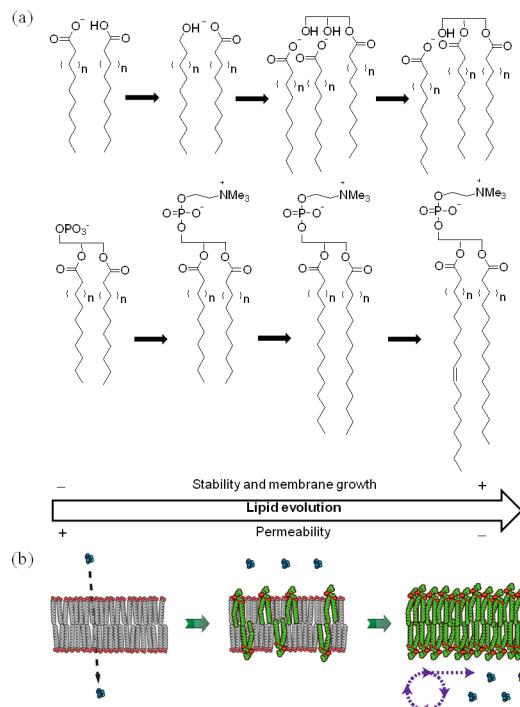
To be particularly highlighted here are their results regarding the permeability of mixed fatty acid vesicles,<sup>526,527</sup> because this marks a very important difference with current phospholipid vesicles. The emerging idea is that fatty acid vesicles, even if they share some of the fundamental properties of liposomes, constitute compartments that are much more permeable than liposomes, especially for fatty acid membranes of mixed compositions (Figure 14b). This, apart from other possible implications (e.g., for the debate on the autotrophic versus heterotrophic origin), entails an important general message for researchers in the field of origins of life: the first prebiotic membranes did not necessarily make so complicated the accessibility of molecules to their lumen and, in general, the exchange of material resources in and out of the system.

The main question arising out of this whole new perspective on the origins of the first self-assembled closed bilayer compartments is how the spontaneously formed fatty acid vesicles were progressively taken over by systems that produced their own, more complex lipids (i.e., proper protocells), which surely conferred on the compartments more stability but, at the same time, lower permeability (Figure 15). Apel and Deamer<sup>528</sup> suggested that glycerol derivatives of single-chain fatty acids might have represented a first step in this direction. A second step, more complicated from the metabolic point of view, should have been the internal production of double-chain (phospho)lipids, most likely made out of precursor fatty acids and glycerol available in the environment. The possible physicochemical driving forces and evolutionary advantages of this lipid takeover process in protocell membranes are currently being investigated.<sup>529</sup>

In any case, since completing the transition from single-chain lipid vesicles to actual biomembranes apparently implies a gradual decrease of permeability, such a transition would have surely involved the emergence of a relatively complex protometabolism inside vesicles, among other possible requirements (e.g., parallel development of cell walls).<sup>513</sup> That protometabolism should provide other types of functional molecules to be included in the membrane. In this context, the coevolution of increasingly complex lipid membranes with oligo- and polypeptides needs to be further explored,<sup>530–532</sup> given that the latter might have functioned as pore inducers or precursor-mediated transport mechanisms anchored on them. The incorporation of primitive nucleic acids or nucleic acid analogues into the membrane<sup>533–535</sup> and lumen<sup>526,536</sup> of vesicles, and their capacity to perform their role as information carriers (see section 3.3.2), should also be considered. All this brings us to a scenario of higher molecular heterogeneity and more complex integration mechanisms that will be discussed in later sections (in particular, sections 4.2 and 4.3).

### 3.2. Reproduction and Replication Processes

After reviewing how relatively simple organic—and even inorganic—compounds can become engaged in collective



**Figure 15.** Schematic representation of membrane evolution. (a) Possible sequence of chemical transitions from simpler to more complex lipid molecules (from left to right in the diagram). (b) The increase in complexity of lipids in model membranes is shown to produce slower amphiphile desorption rates, which would lead to faster growth of the protocell, as well as a decrease in permeability, which acts as a selective pressure for the emergence of internalized metabolic and transport machinery in the system. Adapted with permission from ref 529. Copyright 2011 National Academy of Sciences.

chemical dynamics and interactions that capture some features of biological phenomenology (compartmentation and metabolic aspects), we will focus now on processes of chemical interaction underlying other fundamental properties of living systems (reproductive aspects). Again, autocatalytic mechanisms will be very often involved in these processes. However, the result of autocatalysis in the cases we will explore now is not so strictly linked with the way in which complex, far-from-equilibrium reaction networks or (supra)molecular chemical patterns could spontaneously emerge and be held together (the target of the previous section) but rather with the capacity of chemical systems to grow, multiply in number, and change over time. This capacity is of paramount importance for life, in particular for the evolvability of living systems. Depending on how chemical systems grow and multiply and—no less important—on how similar “mother and daughter” systems are after reproduction, the characteristics of the resulting population and their evolutionary dynamics in a given environment will be determined.

It is patent that biological systems do grow and multiply, having managed to keep evolving over geological time on the Earth. Nevertheless, they have done so by means of rather sophisticated molecular mechanisms, way beyond what one would expect to find in a chemical system. Those biological, genetically controlled mechanisms ensure highly reliable multiplication and transmission of traits across generations, yet they leave room for some variation. Furthermore, they make it possible that variation, or at least part of it, can act as the

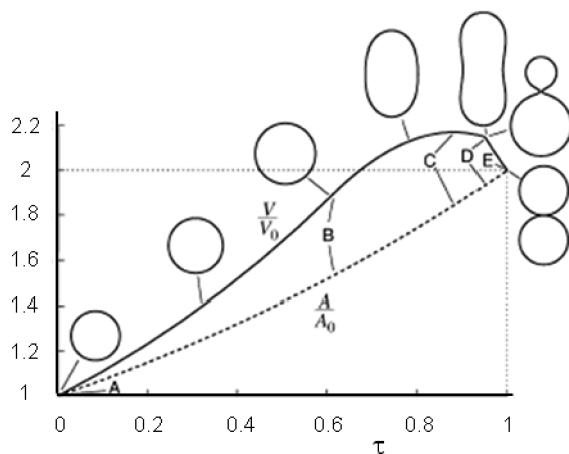
substrate for evolution by means of natural selection. Nonetheless, the nature of those intricate mechanisms is inherently chemical: they operate at the molecular level. From this point of view, how was their appearance possible?

Since the advent of molecular biology in the second part of the last century, various lines of research have tried to address this issue, analyzing the chemical roots and origins of biological evolution. In this and the following section (section 3.3),<sup>537</sup> we will try to cover those approaches. As in the rest of the review, we will not organize this section chronologically but thematically, and we will pay special attention to empirical results, even if we acknowledge the importance of theoretical models for both design and interpretation of experiments. Reproduction and replication dynamics in minimal (i.e., single-type component) chemical systems will be first discussed. Then some brief remarks on more complex (auto- and cross-catalytic) networks will follow, although the integration of diverse-component systems is left for a later part of the review (section 4).

The distinction between reproduction and replication is a relevant one to make here in order to clarify the different types of processes to which we will be referring, as well as their final outcomes. We follow Dyson’s insight<sup>538</sup> on this point: the term “replication” will convey any reliable copying process, taking place at the molecular level, that gives as a result new molecules that conserve the specific sequence of a pre-existing one (commonly called the template); in turn, the term “reproduction” will be used as a more general concept that involves the spatial multiplication or division of a whole system and is not necessarily reliable, in a statistically meaningful sense, as far as the production of identical copies is concerned.

Many researchers in the field of origins of life, particularly those who are skeptical about the possibility of protometabolic cycles spontaneously organizing into minimally robust systems,<sup>14</sup> have explored the potential of molecular replication mechanisms to generate growing populations of chemical species that, faced with limitations, could start competing for resources and evolving in a Darwinian sense.<sup>539</sup> Natural selection, in that scenario of molecular evolution, is conceived as a necessary driving force, from the very beginning, to overcome bottlenecks and reach increasing levels of diversity and complexity, much in the same way as it does later, once full-fledged living organisms take the stage. Accordingly, as we will show in sections 3.2.2 and 3.3, some kind of template replication mechanism is generally assumed as a basic requirement to drive chemical evolution in a productive, biologically relevant direction.<sup>540,541</sup> However, as we will explain next, that prebiotic evolutionary scenario is being enlarged in recent years through the implementation of competitive-dynamics experiments for other types of growing and potentially reproducing systems: lipid vesicles.

**3.2.1. Reproduction of Vesicles.** Theoretical models demonstrate that the reproduction of simple vesicles is, in principle, a feasible event (see Figure 16).<sup>542,543</sup> In practice, however, the spontaneous division of a lipid vesicle is not a trivial issue.<sup>70</sup> These supramolecular structures always have a critical minimal size, which depends on the geometry and molecular properties of the amphiphilic/surfactant molecules forming the bilayer (see section 3.1.3), but it is not so straightforward to set an upper limit of stability. Vesicles in aqueous solution, under conditions of pH, temperature, and ionic strength in which the bilayer lamellar phase is thermodynamically favored, are relatively stable at a wide



**Figure 16.** Dependence of (—) relative vesicle volume ( $V/V_0$ , where  $V_0 = 4\pi R_0^3/3$ ) and (---) relative membrane area ( $A/A_0$ , where  $A_0 = 4\pi R_0^2$ ) on the reduced time ( $\tau = t/T_d$ ), together with a schematic representation of the corresponding shape transitions in the pathway to division.  $R_0$  represents the initial radius of the vesicle and  $T_d$  is the doubling time. Reproduced with permission from ref 542. Copyright 2004 Springer.

range of sizes; their final size depends mainly on the preparation method.<sup>544</sup> Therefore, vesicles could grow without any apparent need to divide, remaining roughly spherical, unless some perturbation/constraining factor appears in the process, changing the volume-to-surface relationship. In living organisms, cellular division occurs very regularly, after a growth phase, but this is a genetically controlled process, which relies on a complex membrane of diverse composition and, once more, on a suite of concerted macromolecular mechanisms in action. These mechanisms not only take care that two dividing cells develop almost identically but also make sure through a septal ring that the fission of the membranes actually happens exactly when and where it must happen. So our question can be posed in the following terms: under what experimental conditions would much simpler cellular compartments (vesicles, in particular) show growth and reproductive behavior, even if it is not so regular and accurate?

Researchers working with giant vesicles, whose size is typically larger than 1  $\mu\text{m}$  in diameter ( $\Phi$ ) and are termed GUVs (if unilamellar) or GMVs (if multilamellar), have managed to exert very precise control over them, inducing their reproduction in various ways: fission, budding, birthing, etc.<sup>545</sup> However, from an origins-of-life perspective, such a strong, externally directed control of the vesicle division process is not really an advantage. A more interesting approach, like the one followed by Sugawara and co-workers,<sup>546</sup> is based on giant vesicles that can grow autonomously and “self-divide”. The advantages of using giant vesicles can then be exploited: optical microscopy allows observation of different stages and shapes during division in real time and with an individual specimen, whereas smaller conventional vesicles are more difficult to monitor and require electron microscopy or indirect methods, like dynamic light scattering (DLS). Unfortunately, the work done to date with giant vesicles is not so relevant for our purposes here: the preparation method for their initial formation remains rather artificial and their components have seldom been of the prebiotically plausible type, a few exceptions aside.<sup>547</sup> Therefore, we will focus on vesicles of more conventional sizes (i.e., from 100 to 500 nm  $\Phi$ ), and especially

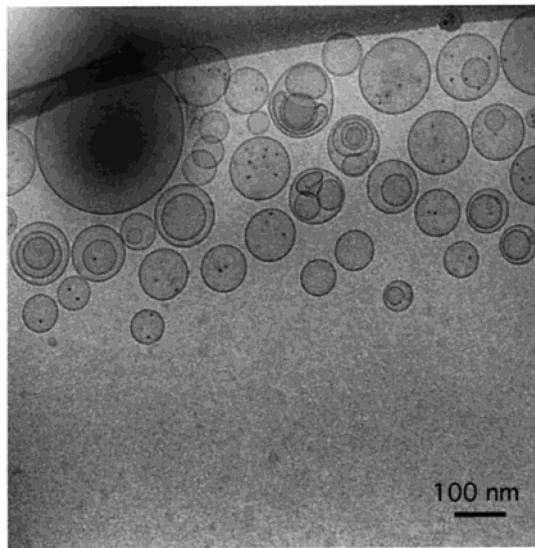
on those made of fatty acids (see sections 2.1.1 and 3.1.3), despite the methodological challenges to follow and capture their division processes, which are still apparent.

Most of the experimental work on fatty acid vesicle reproduction so far has been carried out by Luisi and co-workers.<sup>548</sup> Back in the 1990s, they devised two general methods to trigger off the growth and potential division of this type of self-assembled structures. First, they used chemical means: the experiment starts from a precursor of the amphiphile, normally the fatty acid anhydride, which is present in a separate organic phase. The anhydride is allowed to hydrolyze in contact with an aqueous solution until enough amphiphilic monomers are formed to saturate the solution and begin to self-assemble, or assemble into pre-existing hydrophobic structures that would therefore grow. This type of approach was initially carried out with fatty acids assembling into micelles<sup>549,550</sup> but was soon extended to vesicles.<sup>551</sup> Interestingly, the process was found to be autocatalytic, in the sense that formation of the supramolecular structures (i.e., micelles or vesicles) made it possible for more precursor molecules to come in contact with water and, in that way, speeded up their hydrolysis, generating more structures.

The second method used by Luisi and co-workers<sup>552</sup> involved direct physical addition of the fatty acid, typically in the form of micelles, to an aqueous solution that either contained preformed vesicles or was prepared in conditions that favored the transition from micelles to vesicles. Again, it was found that the micelle-to-vesicle transformation was accelerated by the presence of preformed vesicles, and the sigmoidal kinetic curves characteristic of autocatalytic behavior were reported. More detailed kinetic analyses of this process, carried out by Chen and Szostak,<sup>553</sup> showed that if the amount of micelles added was low in comparison to the number of pre-existing vesicles, growth was simply exponential. On the contrary, if the ratio of micelles-to-vesicles was bigger than ca. 0.4, a two-phase kinetic profile appeared: first a rapid process, interpreted as the coating by micelles of pre-existing vesicles, followed by a slower one in which the micelle-to-vesicle transformation apparently takes place in an independent way, creating vesicles *de novo*.

Indeed, it is not simple to determine, under these conditions in which the dynamics of fatty acid monomers, micelles, and vesicles are intermingled, whether vesicle growth leads effectively to division, whether the fresh amphiphiles added to the solution create new vesicles, or whether the two processes happen, to some extent, simultaneously. Furthermore, when the initial aqueous solution contains a population of pre-existing vesicles with a homogeneous size (i.e., extruded vesicles), their presence seems to have an effect (so-called “matrix effect”)<sup>552</sup> on the size of the newly formed vesicles which, if the stoichiometry is correctly kept, turn out to be roughly of the same size. In order to explore these issues, Luisi and co-workers<sup>554</sup> developed the strategy of internally labeling the pre-existing population of vesicles with a water-soluble probe (i.e.,  $\text{Fe}^{3+}$ -containing ferritin protein) and analyzing, through cryo-transmission electron microscopy (cryo-TEM), the resultant distribution of this probe in the population of vesicles after addition of either a fatty acid or its precursor anhydride. After a series of experiments, it was concluded that the final distribution depended on the method employed. If the new oleic acid molecules came from the hydrolysis of oleic anhydride added to the solution and uptaken by the pre-existing vesicles, then vesicle growth, without division, and de

novo formation of vesicles with no ferritin inside were the most prominent pathways (as reflected in Figure 17). If, instead, the new oleic acid molecules came from the direct addition of micelles, results suggested that the division pathway was favored.



**Figure 17.** Cryo-TEM micrograph of a population of oleate vesicles after complete hydrolysis of oleic anhydride in the presence of preformed oleic acid vesicles containing ferritin (black dots). The total concentration of oleic acid + oleate before oleic anhydride hydrolysis was 5.5 mM, whereas after hydrolysis it was 28 mM. Reproduced with permission from ref 554. Copyright 2001 American Chemical Society.

In any case, the previous type of results can provide only indirect evidence about vesicle reproduction. One never knows exactly what happened to get that new distribution of ferritin within the vesicle population. In a more recent article by Luisi and co-workers,<sup>555</sup> freeze-fracture electron micrographs of the vesicle population immediately after addition of fresh oleate micelles were reported, in which transition vesicular structures called “twin vesicles” are shown (Figure 18). These structures



**Figure 18.** Freeze-fracture electron micrographs of intermediate structures formed during vesiculation of oleate micelles in the presence of preformed extruded oleate vesicles. Samples were taken 40 s after injection of the oleate micelles. The bar represents 200 nm. Reproduced with permission from ref 555. Copyright 2006 IOP Science.

are clearly not the result of aggregation processes and are never found in the final vesicle populations. Therefore they are the most direct evidence we have, to date, of fatty acid vesicle “self-reproduction”.

The coupling of growth and reproduction of fatty acid vesicles has also been tackled in recent years by Szostak and co-workers but from a somewhat different approach. Less priority has been given to search for possible self-production and self-

reproduction processes, so relevant in Luisi’s “autopoietic” way of looking into the problem.<sup>556</sup> In their initial chief work on those lines, Szostak and co-workers<sup>153</sup> put the focus on proving that encapsulation, growth, and division could occur in repetitive cycles through purely physicochemical forces, even if some of them were somewhat artificial (implying using vesicle extrusion as a reproduction method) and disruptive (e.g., loss of encapsulated content). More recent progress on similar lines has involved changing the system from small unilamellar vesicles to large (~4 μm Φ) multilamellar ones, which contain a reservoir of lipids and have the additional technical advantage that their dynamics can be followed by epifluorescence microscopy.<sup>503</sup> Under those conditions, when vesicles are fed with micelles, an unexpected concentration-driven growth mechanism was discovered:<sup>557</sup> elongated structures (i.e., treadlike vesicles) come out of the multilamellar system and then divide into multiple daughter vesicles, just as a result of gentle agitation and without much loss of their internal contents. In this alternative scenario, multiple cycles of growth and reproduction were therefore achieved. Such filamentous vesicles have also been shown to pearl and divide, induced by photochemically driven redox chemistry in the presence of thiols.<sup>558</sup> Thus, the possible prebiotic relevance of both large multilamellar vesicles and this type of filamentous growth and division should be further explored.

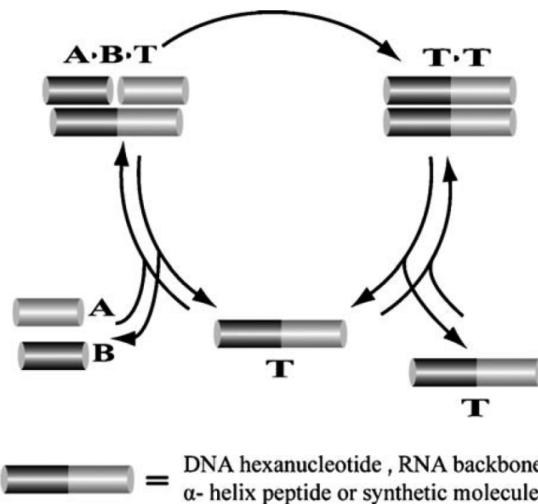
In any case, the longer-term and further-reaching contribution of these studies, beyond the specificities that will certainly teach us more about the reproductive processes in one or another type of vesicle, is the wider perspective they open for the field of origins of life. Evolution through growth and multiplication is not just relevant in the domain of biopolymer chemistry. In that sense, experiments of competition among populations of vesicles of different types and with different physicochemical properties should be more vigorously encouraged, following previous (still rather marginal) examples.<sup>524,539</sup> In particular, the finding that fatty acid vesicles could endure moderate levels of osmotic tension without breaking apart, and that this could be an important kinetic factor for their growth, was quite a remarkable advance.<sup>524</sup> Regarding the latter, Szostak and co-workers highlighted that vesicles containing osmotically active compounds, including oligonucleotides, would take up the available fatty acid faster, and so would grow and potentially divide more quickly, thus outperforming others in the population. Therefore, osmotic tension, initial size, composition, lamellarity, overall charge, permeability, and other properties (e.g., ability to generate pH gradients<sup>560</sup> or photochemically induced redox processes)<sup>558</sup> are bound to play key roles in the growth and reproductive success of prebiotic vesicles. This avenue of research, indeed, should not be overlooked.

**3.2.2. Networks of Replicating Polymers.** Even in the most optimistic and promising scenarios for the development of vesicle populations that could somehow self-produce, grow, and reproduce, transferring various properties to the offspring (e.g., their own composition or “composome”),<sup>561</sup> these systems would face a tremendous bottleneck in terms of preserving the complexity that, occasionally, might come out of their underlying chemistry. Let us imagine that one of these vesicular systems generates, by either pure chance or a rare combination of factors, an oligopeptide with a new structural function or with catalytic properties that somehow allow an unprecedented enhancement in the rate of one or various encapsulated reactions. This vesicle may perform very well for a

certain time, but the evolutionarily relevant question is, what prevents the degradation of that new component within the enclosed system and its almost irreversible loss? Although similar oligopeptide components could of course arise again, it is not guaranteed when/whether this would actually happen. Moreover, nothing ensures that, if similar molecules do arise, they will have the same effects on the dynamics of the system, because those effects could easily depend on the specific sequence of amino acids forming the original oligopeptide.

In order to overcome such a bottleneck, mechanisms for preserving the complexity and functionality of molecular species generated in a reproducing system must be implemented. Interestingly, if those mechanisms are based on the autocatalytic/template-replicative properties of the actual biopolymers, involving the conservation of their specific sequence, these could more easily spread, amplify, and be transmitted to new generations of similar systems. Thus, a scenario based on the evolutionary dynamics of populations of replicator molecules, like the one envisioned by Eigen, Orgel, and Schuster in the 1970s<sup>539,562,563</sup> (although not devoid of its own set of difficulties, which have been explored both theoretically and experimentally),<sup>541,564–568</sup> has the advantage of providing a potential solution to two fundamental problems faced by prebiotic systems at the same time: (i) the physiological problem of overcoming the degradation tendency of any complex molecule, like an oligomer or a polymer; and (ii) the evolutionary problem of transmitting the selective advantages of that complex molecule to the offspring. Template, sequence-conserving mechanisms (together with the corresponding dynamic properties of the molecules involved) could be implemented in vesicles but could also develop independently (e.g., in solution or associated with microenvironments present on mineral surfaces) and be later recruited by protocellular systems. In fact, most of the research carried out on the possible origins of molecular mechanisms of sequence preservation (i.e., templates), has considered only systems in bulk solution, with the aim to avoid possible additional hurdles (e.g., accessibility of monomers and other necessary materials to the local reaction domain).

**3.2.2.1. Minimal Replicating Systems.** A common strategy to tackle the problem of molecular sequence preservation, applied in particular to biopolymers, has been to simplify it as much as possible and to specialize on minimal self-replicating systems,<sup>569,570</sup> using a single type of molecule at a time: oligonucleotides (i.e., DNA or RNA strands), oligopeptides, or alternative abiotic organic polymers. These minimal systems typically consist of two oligomer molecules, A and B, and a longer template molecule T, which is somehow complementary to A and B and facilitates the irreversible ligation reaction between them. If the product of the reaction is T itself (i.e., the sequence of monomers in the ligated polymer, A–B or B–A, is equal or fully complementary to T), then the process is autocatalytic and the system is usually called “self-replicating”, to distinguish it from other chemical autocatalytic systems that do not necessarily involve sequence conservation. In each cycle of replication, T binds A and B to form the ternary complex A·B·T, in a way that the covalent bonding between A and B is subsequently promoted. The template–product complex T·T must then dissociate to provide two free copies of T, which can independently enter the next cycle of ligation-based replication (Figure 19). As we will see below, the base-pairing mechanism present in oligonucleotides also makes possible finer-grained experimental designs, like the sequential incorporation of



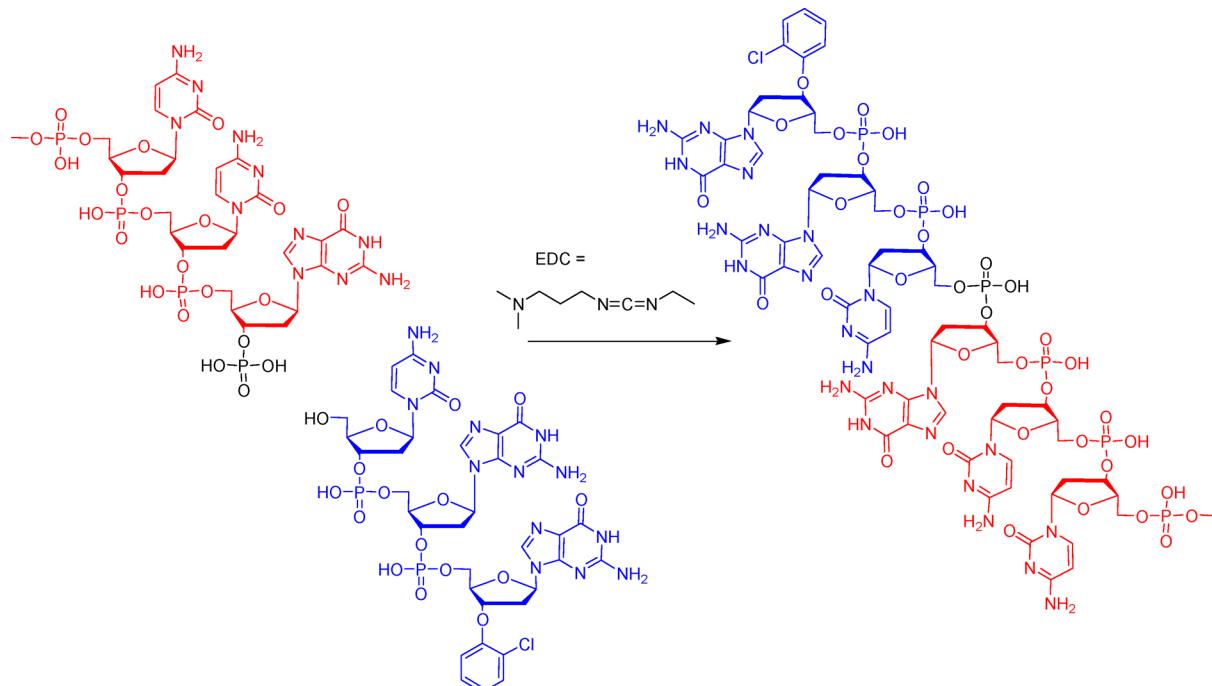
**Figure 19.** Schematic representation of a minimal ligation-based self-replicating system. Further details are described in the text. Reproduced with permission from ref 465. Copyright 2008 Wiley–VCH.

monomer units (i.e., template-directed polycondensation reactions), but in this section we present the general scheme applicable to most minimal replicating systems.

The efficacy of the autocatalytic loop in these minimal self-replicating systems can be measured through two main parameters:<sup>570,571</sup> (i) autocatalytic efficiency,  $\varepsilon$ , which is the ratio between template-directed and template-independent rates of the reaction  $A + B \rightarrow T$  (i.e., a measure of the speeding up of the reaction attributable to T); and (ii) order of the reaction,  $p$ , which reflects how easily  $T\cdot T$  dissociates to  $2T$ . If  $p = 0.5$  it means that the dissociation is rate-limiting, whereas  $p = 1$  corresponds to maximum dissociation, allowing for complete autocatalysis (i.e., exponential amplification and sigmoidal kinetic profile). Depending fundamentally on this  $p$  value, a wide range of kinetic behaviors, from parabolic to exponential growth dynamics, has been obtained by different authors over the years, using two biologically relevant systems (oligonucleotides and oligopeptides) under numerous experimental conditions.<sup>465,569,570,572,573</sup>

As we mentioned above, a general premise of this approach in the field of prebiotic chemistry is the absence of enzymes. Orgel and co-workers<sup>245,574–576</sup> were the first to tackle the nonenzymatic template-directed synthesis of DNA oligonucleotides, but their system was not really minimalistic. Instead, what they were implementing was a polycondensation reaction from activated mononucleotide precursors rather than a ligation. In this regard, although they succeeded in running the reaction with the assistance of templates under various conditions, they found other difficulties and did not manage to complete an autocatalytic replication cycle.<sup>39,577</sup> It was von Kiedrowski and co-workers<sup>578–580</sup> who reported the first nonenzymatic self-replicating system, based on a palindromic DNA hexanucleotide which assisted, in the presence of the water-soluble carbodiimide EDC, the ligation of two trideoxynucleotides, each of them complementary to half of the hexadeoxynucleotide sequence, and adequately protected at 5'- or 3'-positions, respectively, to prevent further polymerization (Figure 20).

Soon afterward, Zielinski and Orgel<sup>581</sup> implemented a different minimal replicating system in which two diribonucleotide analogues were linked, again in the presence of EDC,



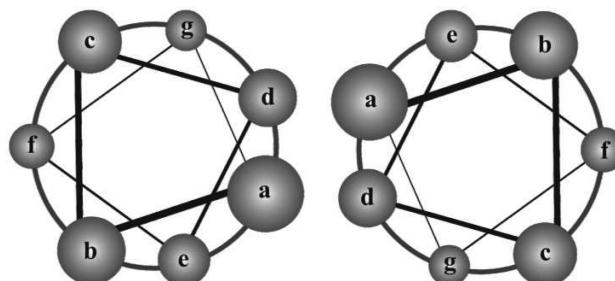
**Figure 20.** Self-replicating construct of von Kiedrowski and co-workers,<sup>578–580</sup> whose modus operandi is based on DNA base pairing for formation of the ternary complex in the presence of EDC. The template molecule is not depicted, but the product of the reaction is inverted in the figure in order to show its nucleobase complementarity with the two fragments from which it is synthesized. Adapted with permission from ref 580. Copyright 2007 ARKIVOC.

under the cyclic template action of a self-complementary tetraribonucleotide. This, in turn, was the first demonstration of self-replication of nucleic acid-like oligomers with an artificial backbone structure. The systems of both von Kiedrowski and Orgel, in those initial realizations, displayed square root law kinetics (i.e., parabolic growth), reflecting the effects of product inhibition on the autocatalytic process. However, in subsequent years, various optimization strategies were applied to obtain highly improved similar oligonucleotide constructs, whose autocatalytic behavior was truly exponential.<sup>579,582</sup> Some of these works will be reviewed in section 3.3.2, in the context of the research carried out around the RNA world hypothesis, because they usually involve ribozymes.<sup>583</sup>

In the meantime, Rebek and co-workers<sup>584–586</sup> proved that effective autocatalytic template replication could be achieved, as well, with even simpler organic compounds as an alternative to known biopolymers. They devised a system in which the template–substrate molecular recognition mechanism, to form the ternary complex, implied the formation of hydrogen bonds but not through standard nucleobase pairing. The ligation reaction, in turn, involved nucleophilic attack of an amine group of one of the reactants on an activated carboxyl ester group of the other, forming an amide bond. The subsequent dissociation step of the self-complementary template duplex was thermodynamically downhill and provided high educt concentrations of the final product. Despite the controversy raised about the actual pathway or molecular mechanism operating in the system,<sup>587,588</sup> this case illustrates the potential of bioorganic chemistry, not necessarily constrained by current biochemistry, for the development of alternative pathways toward synthetic systems with lifelike properties, as later demonstrated by other groups as well.<sup>589,590</sup>

A further-reaching landmark in this area of research was achieved when Ghadiri and co-workers<sup>401</sup> managed to

implement the same kind of minimal self-replicating system employing oligopeptides. The peptides employed in this system belong to a rather special though relatively abundant motif in nature: they contain heptad repeats in which the first and fourth positions are occupied by hydrophobic amino acids (a and d in Figure 21). If they are long enough (e.g., a 32-aa



**Figure 21.** Schematic (top-down) view of a coiled-coil structure, with the hydrophobic residues, a and d, of the corresponding  $\alpha$ -helices facing each other. Electrostatic interactions between amino acids in the e and g positions also contribute to the assembly. Reproduced with permission from ref 580. Copyright 2007 ARKIVOC.

oligopeptide was originally used as template), such sequences form  $\alpha$ -helices that may assemble into coiled-coil structures. Therefore, this molecular recognition mechanism was based on a combination of both hydrophobic and electrostatic interactions between amino acids in the a and d positions (hydrophobic) and in the e and g positions (electrostatic) (see Figure 21). The reacting fragments (a 17-mer and a 15-mer in the original setting) formed an amide bond through a Kent type of ligation (a thioester-promoted native ligation) favored by the presence of the 32-mer template.

The first reports of this system revealed parabolic growth, as a result of the slow release of the newly formed product. However, by reducing the length of the template to the minimum (i.e., 26 amino acids), Issac and Chmielewski<sup>591</sup> managed to destabilize the template–product complex, significantly raising the reaction order ( $p = 0.91$ ) and the autocatalytic efficiency for the replication cycle ( $\varepsilon = 1.0 \times 10^5$ ). In the same group, an alternative strategy was tried through the substitution of a Glu by a Pro at the fifth position (e in Figure 21) of the heptad repeat, to distort the coiled-coil structure of the template–product complex and facilitate the release of the latter, achieving similar outstanding results.<sup>592</sup>

An interesting question to be addressed within nonenzymatic peptide replication systems is how to overcome the minimal length limits imposed by the use of  $\alpha$ -helices and switch to  $\beta$ -sheet conformations that also show template capacity. Ashkenasy and co-workers<sup>593</sup> reported some advance on those lines, employing Glu-(Phe-Glu)<sub>n</sub> analogues, which readily self-assemble into soluble one-dimensional  $\beta$ -sheet aggregates of different sizes. Although the replicative cycle mechanism involved is not fully ascertained yet, it was demonstrated that the presence of these aggregates, which are formed by subsequent antiparallel additions of the peptide analogue, significantly increased the rate of a covalent condensation reaction leading to the product. Actually, the possibility that amyloid fibril supramolecular structures could have played an important role in the origins of life is gathering increasing interest, also due to the realization that their noncovalent assembly processes tend to be autocatalytic.<sup>594–596</sup>

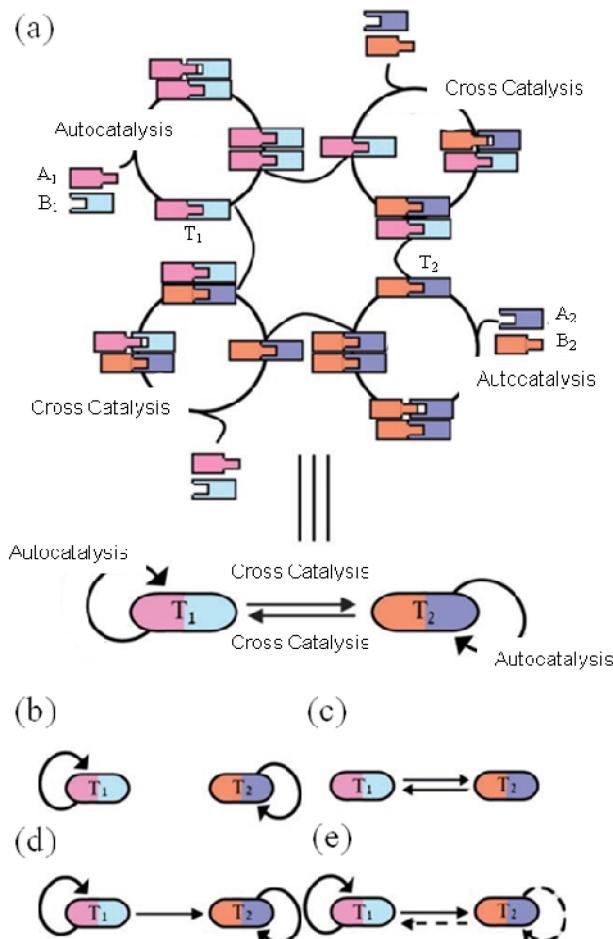
**3.2.2.2. Molecular Replication Networks.** The last example discussed also opens the door to replicating systems in which templates are not just sequences of monomeric units but have an oligomeric structure and in which higher-order catalytic relationships established among the diverse components of the set need to be determined.<sup>465,597</sup> As the number of components of the set increases, so do the possibilities of cross-catalytic and inhibitory or competitive pathways within the system, and “network behavior” is more likely to appear; that is, situations in which the characterization of catalytic pathways in isolation and the expected relationships among them differ from the actual complete system dynamics. The main reason behind that unforeseeable collective behavior lies in the fact that those pathways take place in parallel and their relative importance, which determines the final topology of the network, is difficult to tell in advance. This is because they are typically operated by molecules that participate in more than one equilibrium process. Since these cases are of particular relevance for the present review, we will briefly take them into consideration next.

In each of the three major lines of research within this area (oligonucleotides, oligopeptides, and alternative organic polymers), various extensions of the minimal replicating systems described above have been implemented. The group of von Kiedrowski, for instance, studied cross-catalysis in a quaternary system based on their hexadeoxynucleotide analogues involving, as templates, the four possible combinations of two different trideoxynucleotides (CCG and CGG). In this case the goal was to find a suitable set of molecules and conditions so that the template-directed ligation occurs between complementary DNA-like strains (and not self-complementary, as in direct autocatalysis). In their system, two of the hexadeoxynucleotide analogues behaved as autocatalytic replicators and two formed a cooperative cross-catalytic team. Sievers and von

Kiedrowski<sup>598,599</sup> demonstrated that the system as a whole could grow and undergo replication cycles if the reciprocal, cross-catalytic template effects are similar in efficiency to the autocatalytic syntheses of self-complementary products and if the reaction system is not dominated by the latter, which occur in parallel. The same group also devised a self-replicating system starting from three building blocks (A, B, and C) in which the copy of the template was formed in two reaction steps ( $A + B \rightarrow AB$ ;  $AB + C \rightarrow T$ ). Four additional smaller pieces made of only two fragments were found to be involved in autocatalytic and cross-catalytic activity, which amounted to six different pathways.<sup>600</sup> With this experimental setting, von Kiedrowski and co-workers could also explore the minimal length required for DNA molecules to start showing template effects.

More recently, Lehman and co-workers<sup>468,601</sup> have reported very interesting and promising results in terms of the spontaneous formation of collective replicator networks using the *Azoarcus* ribozyme system. They performed experiments with different mixtures of fragments of these RNA molecules, which not only manage to (covalently) self-assemble into self-replicating ribozymes but also establish additional cooperative catalytic cycles among them. From an initial system specifically designed with three-membered cooperative ribozymes (each one fragmented into two pieces), they extend their results to a wider, randomized network of fragments leading to 48 possible RNA outcomes. The authors demonstrate in both cases the selective advantage (i.e., faster rates achieved) when these cooperative catalytic loops are included in the system, in addition to the basic (selfish) dynamics of each replicator. These results have been theoretically interpreted and further elaborated in the context of a formal model of Kauffman-type autocatalytic sets.<sup>602</sup>

Reciprocal template catalysis, beyond minimal replicative autocatalysis, was also investigated by Rebek and co-workers,<sup>603,604</sup> using synthetic organic molecules such as triacid-based replicators. More recently, a system of two mutually complementary templates, consisting of different organic abiotic molecules that replicate through Diels–Alder reactions, was also reported by Kassianidis and Philp.<sup>590</sup> Ghadiri and co-workers,<sup>402</sup> in turn, developed a similar line of research toward progressively more complex networks, making use of specifically designed oligopeptides with coiled-coil aggregating properties. A relatively simple binary network obtained by mixing two of their originally self-replicating peptides, but with slightly modified hydrophobic interfaces for molecular recognition, allowed the construction of a system that included a cross-catalytic loop as well as the two autocatalytic ones (Figure 22). This was due to the not-so-high specificity of the hydrophobic areas that help to form the coiled-coils, which can accommodate different residues, both the one belonging to the self-complementary structure and to a complementary one. They also worked with a ternary system, in which an original replicator was mixed with two different mutants, each containing a single Leu → Ala mutation in the hydrophobic core. The coiled-coil assemblies containing the Ala mutants are less stable and thus lead to less efficient template-assisted ligations. It was shown that, in a system seeded with the three molecules, the original replicator (the one formed exclusively by Leu residues) could self-replicate without catalyzing the formation of the others. The mutants, on the contrary, do not replicate, yet each of them can enhance the synthesis of the original sequence. This was claimed to be an error-correction



**Figure 22.** Graph of possible catalytic pathways for the simplest network case: a binary system. Panel a represents all options, whereas in panels b–e, different specific examples are depicted (b, extreme autocatalysis; c, cross-catalysis; d and e, asymmetric cases in which the cross-catalytic pathway is just available, or stronger, in one direction). Reproduced with permission from ref 465. Copyright 2008 Wiley–VCH.

mechanism dynamically embedded in the system, since the two mutants are not selfish autocatalysts but contribute to the maintenance of the initial molecular replicator.<sup>605</sup> In a later implemented ternary network with similar topology, the autocatalytic behavior of one of the templates was switched off in the presence of the other two.<sup>606</sup> That silencing was proved to be a clear network effect, since it did not manifest itself when the different pathways were analyzed separately. Accordingly, many of the pathways that have been found in isolated studies of these replicating systems could actually remain latent or dominated by others in larger networks.<sup>465</sup>

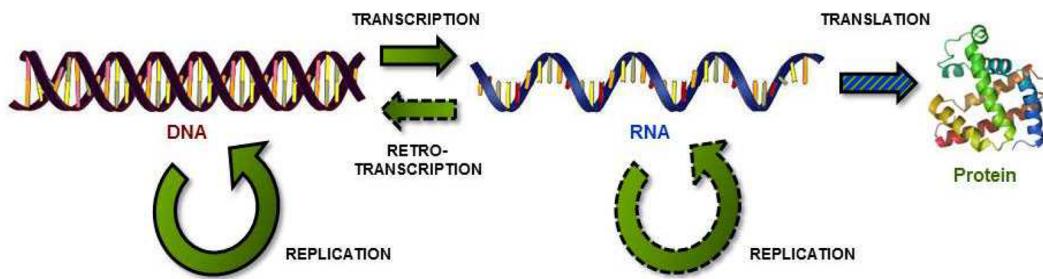
It is also important to mention that this kind of network behavior can be highly sensitive to experimental conditions. For instance, Chmielewski and co-workers<sup>607</sup> implemented a quaternary network of coiled-coil peptides in which both autocatalytic and cross-catalytic loops were present under physiological conditions. One of the two templates originally involved in the cross-catalytic pathway switched to autocatalytic behavior when the ionic strength of the solution was notably increased, whereas the complementary member of the couple did the same when the pH decreased. As could be expected, things become much worse, in terms of analyzing the system

susceptibility to experimental conditions, when the complexity of the network goes up. Ghadiri and co-workers<sup>467</sup> designed a network of 81 peptide components, some of which were self-replicating and some cross-replicating. It had the general form:  $\{A_1, A_2, \dots, A_9\} + B \rightarrow \{T_1, T_2, \dots, T_9\}$ ; that is, they employed nine different reactants plus a common substrate to give nine different template products. In principle, three of the templates behaved autocatalytically and 22 cross-catalytic interactions were identified (including two-, three- and four-member groups or cycles), but the most interesting feature was the fact that the topology of the network could substantially change depending on (i) which templates were initially present in the mixture, (ii) whether the pH of the solution was lower than 5, and (iii) whether metal salts were introduced or not. All this variability and potential mechanisms of regulation within the system, showing strong or latent catalytic pathways as a function of its initial and boundary conditions, open very interesting research possibilities (e.g., implementing Boolean logic operations with chemically interacting networks *in vitro*) and, indeed, get much closer to *in vivo* biochemical networks within current cells.

An obvious expansion of these approaches would be to work with template-replicating autocatalytic systems in which different types of biopolymers or biologically relevant molecules are simultaneously present. Several examples of this kind of exploration have already been tried, for instance, mixing DNA with RNA<sup>608,609</sup> or peptides with RNA.<sup>610</sup> In addition, the possible amplifying effects of encapsulation and release have been studied.<sup>611,612</sup> However, those cases will be discussed later, in the section on systemic integration approaches (section 4). Before doing so, we will pay particular attention to a scenario that has had an enormous impact in the field of origins of life and is also essentially based on the properties of a single type of molecule: RNA.

### 3.3. Development of Biological Evolutionary Mechanisms: The RNA World

So far in this review we have focused on bottom-up approaches, from chemistry toward biology. But the field of origins and early evolution of life has also benefited from the so-called top-down approach, in which present-day organisms are compared in order to extract their common features that, following simple parsimonious rules, can be considered the most ancient ones. In our current biosphere, nucleic acids store genetic information in all organisms and provide the chemical basis for inheritable traits to be passed on to the progeny. Cellular life uses DNA as genetic material, and the information content of DNA sequences is decoded by the system in two main steps: transcription of DNA into RNA, followed by translation of RNA into proteins. This basic two-step strategy is assumed to have already been present in the last universal common ancestor (LUCA).<sup>5,613,614</sup> In the case of viruses, some current viral families are DNA-based while others bear RNA genomes. Both cells and DNA viruses replicate their genomes by means of DNA-dependent DNA polymerases, protein enzymes that catalyze the copying of DNA → DNA. In turn, the replication cycle of RNA viruses involves either a viral-encoded RNA-dependent DNA polymerase (also called reverse transcriptase, RT, which catalyzes the retrotranscription RNA → DNA in retroviruses), or a viral RNA-dependent RNA polymerase (also known as RNA replicase, able to copy RNA → RNA in all remaining RNA viruses, called riboviruses).<sup>615</sup> Therefore, the updated version of the gene expression pathway, previously known as the “central dogma of molecular biology”, is



**Figure 23.** Schematic representation of flow of genetic information in the current biosphere. Processes catalyzed by either cellular enzymes (green) or ribonucleoprotein aggregates (green/blue) are depicted by arrows with solid outlines, while those performed only by viral enzymes are shown by arrows with dashed outlines.

schematized as  $\text{DNA} \leftrightarrow \text{RNA} \rightarrow \text{protein}$ : the two types of nucleic acids can be converted into each other, while the translation of genetic information from RNA to proteins is an irreversible step in terms of sequence coding capacity. Somewhat paradoxically, the whole decoding process would not run without proteins in present biochemistry, since they are the main effector molecules for all those steps that lead to their own production. The translation process is, somehow, an exception, since RNA has been shown to play the main catalytic role in the ribosome (see below). A comprehensive picture of the flow of genetic information, as currently understood, is represented in Figure 23.

RNA occupies a central position in the overall process, since this macromolecule is the genetic material in RNA viruses and viroids (see section 3.3.3), and messenger RNA (mRNA) is the intermediate molecule in the expression of DNA-encoded genes. Additionally, an increasing diversity of roles played by RNA has been revealed in recent decades, including its ability to catalyze certain biochemical reactions within cells. Thus, while in the current biosphere DNA acts as the major archive of coded information (the genotype), with proteins being the main functional or structural molecules (the phenotype), RNA can play both biological roles. Due to the versatility of RNA, it was proposed in the 1960s that, prior to the appearance of cellular life (i.e., before LUCA), a period might have existed when both the storage/transmission of genetic information and the catalysis of biochemical reactions were performed solely by RNA.<sup>575,616,617</sup> That putative period in the early history of Earth, dominated by hypothetical RNA molecules acting simultaneously as genotype and phenotype, was termed “the RNA world”,<sup>618</sup> whose main features will be discussed in section 3.3.2. Nevertheless, several pitfalls, including the difficulty of synthesizing ribose in prebiotic conditions (see section 2.1.3) and the lability of RNA in solution, might have hindered the *de novo* establishment of an RNA world from the prebiotic mixture of monomers, mineral surfaces, and amphiphile-based compartments. Therefore, it has been claimed that different polymeric molecules analogous to natural nucleic acids could have preceded RNA at the first stages of evolution of genetic molecules, constituting putative “pre-RNA worlds”.

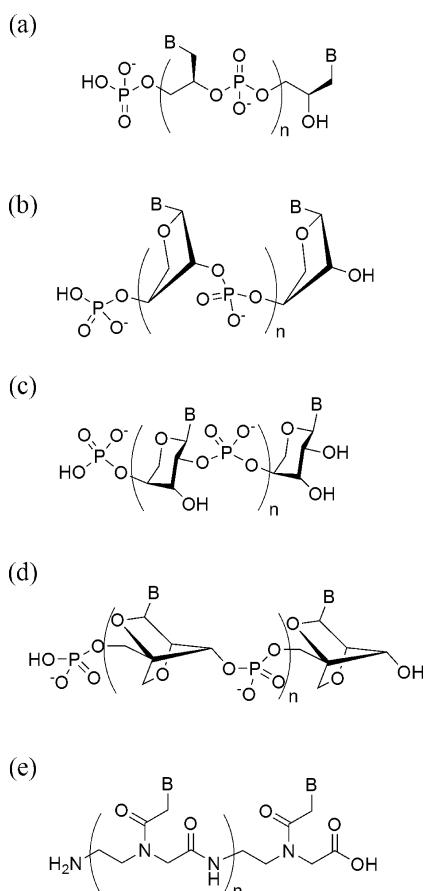
**3.3.1. Pre-RNA Worlds.** The information-bearing molecules that might have bridged the gap between prebiotic chemistry and the RNA world should have exhibited the physicochemical features of current nucleic acids, regarding information storage and potential for template-based replication. Moreover, such nucleic acid analogues should have been more stable and chemically simpler than RNA. Finally, their monomer units should be the outcome of plausible prebiotic

pathways, with their polymerization being a thermodynamically downhill process in certain conditions. The systematic investigation on nucleic acid analogues has proceeded in two ways. First, RNA and DNA mimics have been developed by incorporating artificial nucleobase derivatives into their natural backbones. A second strategy has relied on the synthesis of polymeric structures in which the ribose–phosphate backbone of natural nucleic acids is substituted with either combinations of other sugars and linkage isomers or short linear motifs of glycerol/glycine derivatives.

Certain RNA molecules, in particular transfer RNA (tRNA), contain nucleobase analogues derived from posttranscriptional ribonucleotide modification.<sup>619</sup> Structural investigation of such naturally modified nucleic acids, together with advances in synthetic organic chemistry, inspired the development of nucleic acid mimics containing different analogues of purines and pyrimidines.<sup>620</sup> Nucleobase pairing follows two rules of complementarity: hydrogen bonding (H-bond donors pair with H-bond acceptors) and size (larger purines pair with smaller pyrimidines).<sup>621</sup> While the natural nucleobases exploit mainly two pairing possibilities, six canonical pairing schemes have been investigated by use of carbon and nitrogen cyclic molecules isosteric to A, C, G, and T/U.<sup>622</sup> The use of a broad range of alternative nucleobases has allowed an artificial expansion of the genetic alphabet of nucleic acids.<sup>623,624</sup> This fact, together with the discovery of new nucleobase analogues in carbonaceous meteorites,<sup>625</sup> unveils the intriguing possibility that nucleic acids with an extended nucleobase repertoire might have simultaneously polymerized in isolated or interconnected prebiotic scenarios, prior to the establishment of the A/C/G/U-bearing RNA. Also, as previously discussed (see section 2.2.2), the different nucleobase-containing nucleotides could have polymerized as a combinatorial mixture of 2'- $\rightarrow$ 5', 3'- $\rightarrow$ 5', and 5'- $\rightarrow$ 5' phosphodiester linkages, D- and L-stereoisomers of the ribose, and a variable number of phosphate groups between consecutive sugar motifs.

In turn, the investigation of nucleic acid analogues with alternative molecular backbones was initiated in the 1980s, by combining some of the successful synthetic pathways previously developed for amino acid, sugar, and nucleotide chemistries (see section 2.1). It was soon found that certain modifications of the nucleic acid backbone that were designed to be subtle proved to dramatically change nucleobase pairing, thus preventing the formation of heteroduplexes with RNA or DNA. Those nonhybridizing analogues were discarded as potential prebiotic candidates to have transmitted genetic information to RNA.<sup>622,626</sup> One of the first analogues successfully developed was a glycerol-derived nucleic acid (GNA), with a backbone composed of phosphodiester-linked

acyclic glycerol units (Figure 24a).<sup>627</sup> The potential for nucleic acid replication on templates containing acyclic nucleotides has



**Figure 24.** Molecular backbones of five nucleic acid analogues relevant as possible genetic molecules in a pre-RNA world: (a) glycerol-derived nucleic acid (GNA); (b) threose nucleic acid (TNA); (c) pyranosyl-RNA (p-RNA); (d) locked nucleic acid (LNA); (e) peptide nucleic acid (PNA). Letter B denotes the position of the nucleobase.

been assessed by the nonenzymatic oligomerization of guanosine 5'-phosphoro-2-methylimidazolide on DNA hairpin molecules containing glycerylcytosine (glyC) residues.<sup>628</sup> Also, the ability of a DNA polymerase to use glycerol-nucleoside triphosphates as monomers for GNA synthesis on DNA templates has been reported.<sup>629</sup> Additionally, GNA analogues with N2'→P3' phosphoramidate linkages (npGNA) have been synthesized.<sup>630</sup> This polymer can form stable duplexes with itself and with GNA, and it can be assembled by nonenzymatic template-directed ligation of 3'-imidazole-activated 2'-amino GNA dimers (while monomers undergo rapid intramolecular cyclization). In turn, template-free polymerization of either GNA or npGNA has not been reported so far.

The search for maximum chemical simplicity in the sugar motif of the nucleic acid backbones led to the synthesis of threose nucleic acid (TNA), an analogue based on  $\alpha$ -L-threofuranosyl units (Figure 24b).<sup>631</sup> Since threose is one of the two four-carbon monosaccharides or aldotrehoses and can polymerize only by means of 3'→2' phosphodiester bonds, TNA is the simplest of all potential sugar-containing nucleic acids. TNA hybridizes with RNA and DNA in a sequence-specific manner, and it can be enzymatically polymerized by a DNA polymerase with DNA as the template molecule.<sup>632</sup>

TNA-based in vitro evolution experiments, using DNA and proteins for the amplification process, have recently demonstrated the ability of TNA to fold into three-dimensional structures that can elicit complex biochemical functions such as specific binding to ligands.<sup>409</sup> These features suggested that a template-free, nonenzymatic TNA polymerization might have eventually operated in prebiotic conditions<sup>633,634</sup> and that TNA could have served as a primordial molecule endowed with genotype and phenotype.<sup>409</sup>

Pyranosyl-RNA (p-RNA) contains six-membered, 4'-2'-linked,  $\beta$ -D-ribopyranosyl units instead of ribofuranosyl ones.<sup>635</sup> p-RNA forms duplexes with itself in antiparallel orientation,<sup>636</sup> and p-RNA oligomers are capable of secondary structure formation, including hairpin loops analogous to those ubiquitously found in RNA.<sup>637</sup> Interestingly, 8-mer oligomers of p-RNA can be formed in aqueous solutions by nonenzymatic ligation of two 2',3'-cyclophosphates of 4-mer p-RNAs, with a complementary p-RNA molecule in antiparallel orientation as template.<sup>638,639</sup> This nucleic acid analogue exemplifies how the ring size of the sugar-phosphate backbone can expand the structural and functional diversity of nucleic acids and suggests possible alternative pathways in the pre-RNA world.<sup>640</sup>

In turn, nucleic acid mimics formed by polymerization of 2'-amino-2',3'-dideoxynucleotide-5'-phosphorimidazolides have been developed. The activated monomers (G and C being more efficient than A and T) polymerize by forming N2'→P5' phosphoramidate linkages, which are favored in nonenzymatic polymerization reactions on short homopolymer templates of DNA, RNA, and LNA (see below).<sup>526,641</sup> Szostak and co-workers<sup>642</sup> have recently described the synthesis of N3'→P5'-linked phosphoramidate DNA (3'-NP-DNA) by template-directed polymerization of activated 3'-amino-2',3'-dideoxyribonucleotides. These monomers are more reactive than the corresponding ribonucleotides and, in contrast to the activated 2'-aminonucleotides, the four activated 3'-amino-2',3'-dideoxyribonucleotides polymerize efficiently on different RNA templates (which are superior to DNA templates). The activation of 3'-aminonucleotides with 2-methylimidazole results in a polymerization rate 1 order of magnitude greater than that obtained with imidazole-activated monomers.<sup>642</sup> Although prebiotic pathways for the synthesis of these monomers have not yet been reported, phosphoramidate nucleic acids are currently used, together with the above-described npGNA and other analogues that contain aminated backbones, as model polymers for studying nonenzymatic nucleic acid replication in solution and inside vesicles (see section 4).

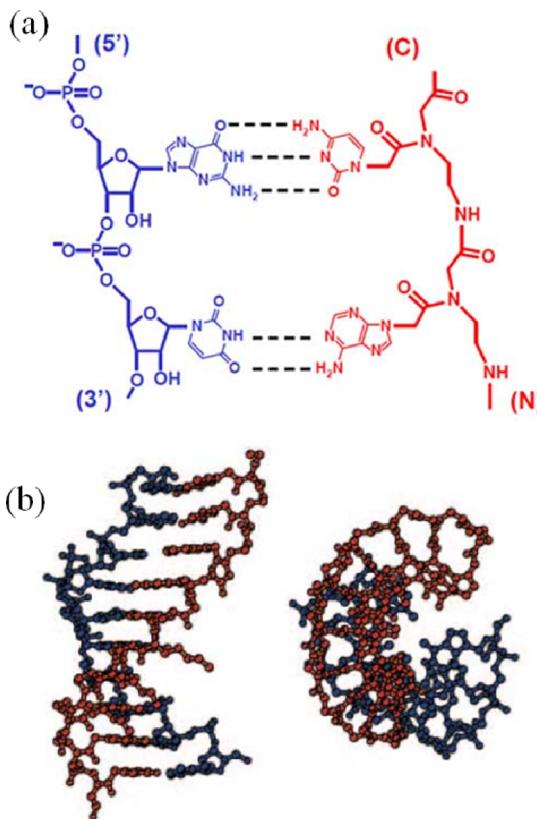
A different, successful approach to design nucleic acid mimics was based on the use of conformationally restricted sugar motifs, including the bicyclo-DNA and tricyclo-DNA families.<sup>643,644</sup> The most relevant analogue of this kind is the so-called locked nucleic acid (LNA), a polymer of 2'-O-4'-C-methylene-linked  $\beta$ -D-ribonucleotide monomers (Figure 24d).<sup>645–647</sup> The linkage of 2'-O and 4'-C atoms via a methylene unit restricts the ribofuranose in the 3'-endo conformation, responsible for the A conformation of LNA/RNA and LNA/DNA duplexes.<sup>648</sup>  $\beta$ -D-Ribo-LNA forms the strongest duplexes with complementary RNA so far described.<sup>649,650</sup> As mentioned above, nonenzymatic polymerization of phosphoramidate DNA from activated monomers (e.g., 2'-amino-2',3'-dideoxyribonucleotide-5'-phosphorimidazolides) has been achieved with LNA homopolymers as

templates.<sup>641</sup> Nevertheless, the synthesis of LNA monomers under prebiotically plausible conditions has not been reported.

The latest achievement in the field of sugar-modified nucleic acid analogues has been reported by Holliger and co-workers,<sup>651</sup> who showed that up to six artificial polymers can be used to store and propagate genetic information. The backbone of these xenonucleic acids (XNAs) contains either different sugar motifs (i.e., threose in TNA, the locked ribose analogue in LNA, arabinose in ANA, and 2'-fluoroarabinose in FANA) or six-membered cyclic structures (i.e., 1,5-anhydrohexitol in HNA and cyclohexene in CeNA). The amplification system developed for XNAs relies on their “reverse transcription” to DNA catalyzed by engineered variants of a natural DNA polymerase enzyme, followed by DNA amplification via conventional polymerase chain reaction (PCR), and “transcription” from DNA to XNA by different engineered DNA polymerases. Therefore, the process is strictly dependent on proteins and DNA, and none of the XNAs can still be considered as a primordial genetic system alternative to RNA. Nevertheless, different XNA molecules able to bind their targets with high affinity and specificity (analogous to RNA or DNA aptamers; see section 3.3.2) have been obtained by such an amplification protocol.<sup>409,651</sup> This demonstrates that certain nucleic acid analogues are capable of information storage and (DNA/protein-mediated) Darwinian evolution. Thus, they could have operated in a putative “XNA world” provided that still-unknown “XNAzymes” were available.<sup>652</sup>

The repertoire of alternative backbones for nucleic acid analogues is increasingly diverse. Besides the aforementioned analogues of ribose, the phosphate group has also been replaced by pyrophosphate, polyphosphate, and alkylphosphate motifs, as well as by sulfones and other sulfur-containing linkers.<sup>156,653</sup> A radically different approach was followed for synthesizing peptide nucleic acid (PNA), a nucleic acid mimic that combines RNA and DNA features with others typical of proteins. PNA was obtained by polymerization of *N*-(2-aminoethyl)glycine, a monomer that replaces the sugar–phosphate units constituting the backbones of natural nucleic acids and most of their artificial analogues (Figure 24e).<sup>654</sup> In PNA, the nucleobases are connected to the polyamide structure by methylenecarbonyl linkages.<sup>655</sup> PNA exhibits unique physicochemical properties, being an achiral and uncharged polymer of high chemical and biological stability,<sup>656</sup> capable of specific binding to complementary targets (RNA, DNA, or PNA) in both parallel and antiparallel orientations (Figure 25).<sup>657</sup> The PNA/RNA and PNA/DNA heteroduplexes are right-handed double helices with structural features intermediate between those of A- and B-form double-stranded (ds) DNA.<sup>658,659</sup> The high affinity of PNA for complementary RNA or DNA molecules is mainly due to the lack of electrostatic repulsion between the uncharged PNA backbone and that of the natural nucleic acid.<sup>654,656,657</sup> Furthermore, the interaction of PNA with RNA or DNA is highly specific because for virtually all base-pair mismatches, the decrease in thermal stability is greater for the PNA/RNA or PNA/DNA heteroduplexes than for the corresponding homoduplexes.<sup>655</sup>

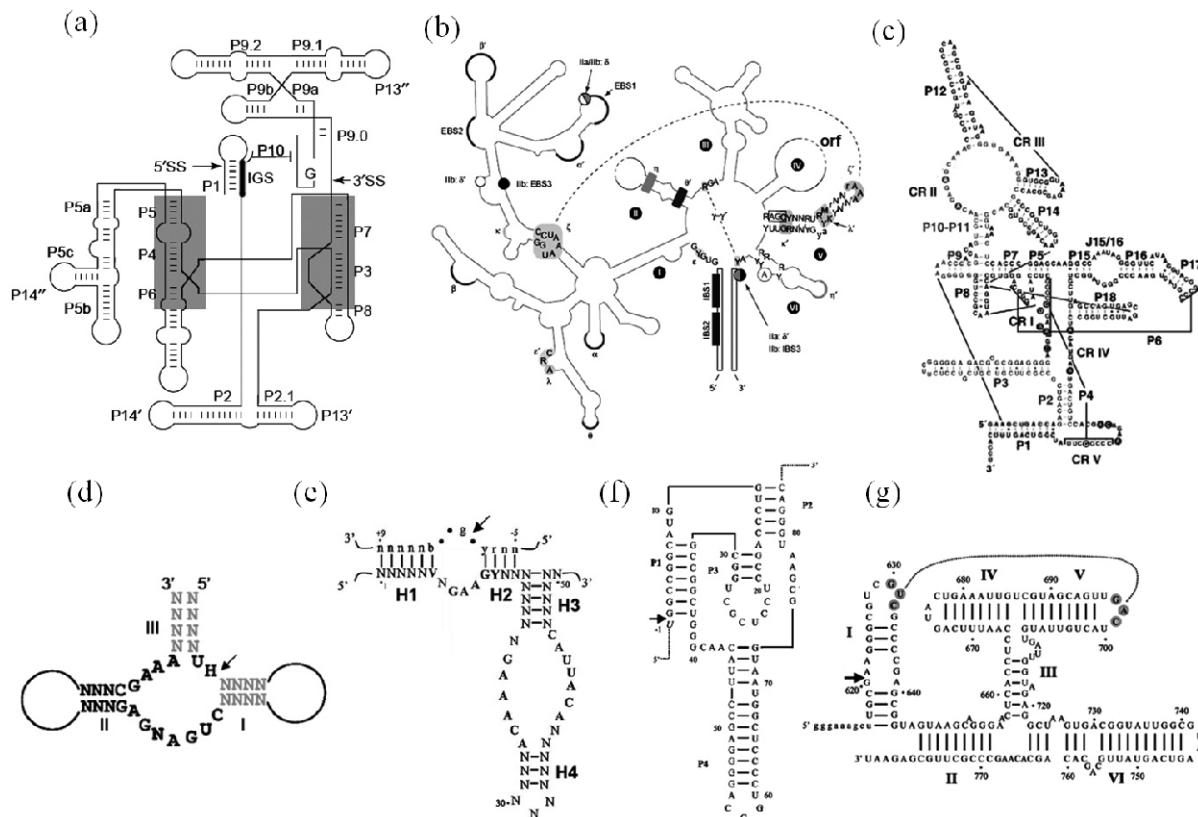
PNA is considered as a prebiotically relevant RNA analogue because aminoethylglycine has been synthesized in spark discharge reactions from a mixture of CH<sub>4</sub>, N<sub>2</sub>, NH<sub>3</sub>, and H<sub>2</sub>O.<sup>660</sup> Additionally, diamino acids and other PNA-related molecules, though not including the aminoethylglycine monomer, have been identified among the soluble organic material present in the Murchison meteorite.<sup>661</sup> The polymer-



**Figure 25.** (a) Schematic model of a dimer of RNA (blue, sequence 5'-GU-3') hybridized with a complementary PNA strand (red, sequence amino-AC-carboxyl) in antiparallel orientation, with hydrogen bonding between complementary nucleobases depicted by dotted lines. (b) Crystal structure of duplex formed by a DNA strain (blue) hybridized with a complementary D-Lys-based chiral PNA (red) in antiparallel orientation, showing lateral and top (i.e., down the helix axis) views. The typical structure of RNA/PNA and DNA/PNA heteroduplexes is an extended double helix whose features are intermediate between those of A-form dsRNA and B-form dsDNA. Reproduced with permission from ref 656. Copyright 2003 National Academy of Sciences.

ization of aminoethylglycine into PNA or PNA-like oligomers has not been achieved yet, and the process seems hindered by intramonomer N-acyl transfer reactions.<sup>662</sup> Nevertheless, DNA-templated PNA polymerization using PNA tetramer or pentamer units has been reported, and the combination of this amplification method with a selection step has allowed the experimental evolution of PNA in the laboratory.<sup>663</sup> Also, PNA can direct the template-dependent polymerization of activated RNA monomers,<sup>664</sup> and despite its achiral nature, it shows enantiomeric cross-inhibition in the incorporation of D- and L-ribonucleotides.<sup>665</sup> The interaction and self-assembly of PNA oligomers on metal surfaces and pyrite crystals has also been investigated.<sup>666–668</sup>

In turn, modified PNA can play certain biochemical functions, as revealed by the synthesis of PNAzymes. These PNA-based enzymes carry a 2,9-dimethylphenanthroline–Cu<sup>2+</sup> complex and act as sequence-specific RNA endonucleases, by hybridizing to the target RNA and cleaving its phosphodiester backbone at an unpaired ribonucleotide.<sup>669,670</sup> Additionally, a novel PNA-related molecular family termed thioester PNA (tPNA) has been developed, which combines base-pairing interactions with the side-chain functionalities of peptides and



**Figure 26.** Secondary structure diagrams for seven natural ribozymes that catalyze cleavage and/or ligation of the RNA backbone. (a) Group I autocatalytic intron: paired elements (P1–P10, and the optional P11–P17, organized into three domains at the tertiary structure level) are shown; the central catalytic core is shaded in gray, and the internal guide sequence (IGS), which pairs with the 5' exon sequence, is highlighted. Adapted with permission from ref 696. Copyright 2005 Elsevier. (b) Group II autocatalytic intron: the core structure (which contains six domains, I–VI, with domain IV including a coding region or open reading frame, orf) and consensus nucleotides are depicted; nucleotides relevant for catalysis are boxed or shaded; tertiary interactions (two of which are marked by dashed lines) are shown by Greek symbols; exon binding sites (EBS) and intron binding sites (IBS) are highlighted. Adapted with permission from ref 697. Copyright 2001 Elsevier. (c) RNA subunit of the *E. coli* (bacterial A-type) RNase P: its structural elements are shown; solid lines represent tertiary interactions. Adapted with permission from ref 698. Copyright 2006 Elsevier. (d) Hammerhead ribozyme: its standard structure is depicted, showing the I–III stems and consensus nucleotides; the arrow indicates the cleavage site ( $H = A, C$ , or  $U$ ). Adapted with permission from ref 699. Copyright 2001 Nature Publishing Group. (e) Hairpin ribozyme: structure of the naturally occurring ribozyme (uppercase letters, nucleotides numbered 1–50, consensus positions are indicated) hybridized to an external substrate (lowercase letters, numbered –5 to +9); four helices are formed, folding into two independent domains A (helices H1 and H2) and B (helices H3 and H4); an arrow marks the cleavage site. Adapted with permission from ref 694. Copyright 2003 Wiley. (f) Hepatitis delta virus (HDV) ribozyme: the genomic strain is depicted, with continuous lines connecting different structural elements; dotted lines represent 5' or 3' additional sequences not required for self-cleavage; arrow indicates cleavage site. Adapted with permission from ref 700. Copyright 2005 Nature Publishing Group. (g) Varkud satellite (VS) ribozyme: nucleotide numbering is that of full-length VS RNA; domains I–VI are shown, and tertiary interaction between complementary trimers at domains Ib and V is indicated with a dotted line; arrow indicates cleavage site. Adapted with permission from ref 694. Copyright 2003 Wiley.

proteins. Monomers of tPNAs can self-assemble onto oligo-(dipeptide) backbones, and dynamic sequence modifications can occur in response to the available templates in solution.<sup>671</sup> Therefore, PNA and PNA-related nucleic acid analogues show physicochemical and biochemical features that make them putative achiral genetic molecules that might have existed before the advent of RNA.<sup>672,673</sup> Remarkably, the combination of amino acid and nucleobase chemistries that led to the synthesis of PNA exemplifies the success of a systems chemistry-inspired approach in the search for a prebiotic information-bearing material.

In summary, parallel lines of research have been followed to screen the structural neighborhood of natural nucleic acids for potential alternative genetic systems from which RNA could have further evolved.<sup>674</sup> If this was the case, the continuity of genetic information, when passed to RNA, would have been guaranteed by the specific base-pairing. Nevertheless, none of

the above-described nucleic acid analogues seem to exist in present-day biology, and it would be nothing but speculation to suggest any individual or coordinated contribution to putative pre-RNA worlds. Alternatively, RNA-based life might have been preceded by replicating, and eventually evolving, molecular structures without chemical resemblance to nucleic acids. If such an information-bearing system lacked nucleobase-like chemical groups, the transmission of information to RNA might have been impaired (i.e., a translationlike mechanism would have been required for conversion of the coded information to an RNA-based system) or prevented (i.e., the previous cryptic information would have been lost, with RNA being the outcome of a “genetic takeover” process). The latter possibility was already suggested by Cairns-Smith<sup>473</sup> to explain a possible evolutionary transition from a clay-based world, where the information would have been stored and replicated in the form of defects or irregularities in the crystalline structure of silicates,

to a period dominated by nucleic acid-encoded genetic information. In any case, it is reasonable to assume that different alternative genetic systems could have originated once the prebiotic synthesis of monomers yielded a heterogeneous repertoire of small molecules prone to self-assemble and polymerize, two processes likely favored by mineral surfaces (see sections 2.1 and 2.2), into information-bearing molecular structures. The interplay among them might have been driven by the availability of monomers, their ease of polymerizing into progressively longer molecules, their relative physicochemical stability, and their permeability to membrane-based compartments.

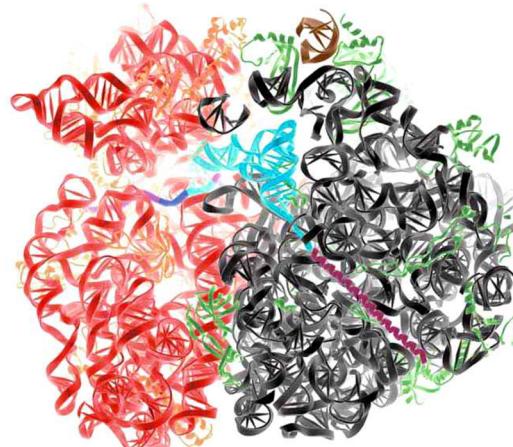
**3.3.2. The RNA World Hypothesis: Combining Genotype and Phenotype at the Molecular Level.** RNA is ubiquitous and plays different roles in nature: apart from storing genetic information in RNA viruses and viroids, this nucleic acid is involved in the control of gene expression, RNA processing and editing, catalysis of some steps of the flow of genetic information, translation, structural scaffolding of certain ribonucleoprotein aggregates, and cellular transport.<sup>675–681</sup> The chemical basis of such versatility relies on the fact that RNA is usually a single-stranded molecule, thus being able to promote intramolecular base-pairing and adopt a much greater variety of three-dimensional structural/functional motifs than double-stranded DNA.<sup>682,683</sup> RNA secondary structure (the planar depiction of the intramolecular nucleotide pairing) offers a simplified yet appropriate representation of the genotype (sequence, object of mutations) to phenotype (shape and function, object of selection) map, useful for addressing relevant evolutionary questions. Therefore, the explicit relationships between sequence, structure, and function make RNA an optimal model for the experimental and computational study of molecular evolution.<sup>684–686</sup>

In extant organisms, protein synthesis cannot be performed in the absence of RNA, and the ribonucleotides are precursors of the deoxyribonucleotides in the nucleic acid biosynthetic networks.<sup>16</sup> Thus, it was suggested that the appearance of RNA should have preceded that of proteins and DNA.<sup>575,616,617</sup> In 1970, Baltimore<sup>687</sup> and Temin and Mizutani<sup>688</sup> independently discovered the existence of protein enzymes endowed with RT activity in certain viruses, which were later called retroviruses, thus proving that the genetic information stored in RNA can be copied into DNA. The next milestone came in the early 1980s, when Cech and co-workers<sup>689,690</sup> and Altman and co-workers<sup>691</sup> independently demonstrated that RNA performs catalytic functions in current organisms. The first RNA enzyme discovered was an autocatalytic intron present in the precursor of the large subunit ribosomal RNA (rRNA) of the ciliate protozoan *Tetrahymena thermophila*, which can catalyze its self-cleavage and the further exon ligation without the intervention of proteins.<sup>689</sup> The term “ribozyme” was coined for the general concept of an RNA molecule with enzymelike activity.<sup>690</sup> The following year, catalytic activity was discovered in the RNA component of the bacterium *Escherichia coli* RNase P, a ribonucleoprotein that cuts phosphodiester bonds during the maturation of tRNA molecules.<sup>691</sup> RNase P provided the first example of a multturnover ribozyme acting on an external substrate.

These discoveries led to the hypothesis that, prior to the origin of nucleic acid-coded proteins, some kind of protometabolism based exclusively on ribozymes might have been possible, likely being assisted by metal cations, amino acids, or small molecule cofactors. This intriguing possibility

was postulated in a short paper published by Gilbert in 1986,<sup>618</sup> who claimed that before the advent of proteins the informational and catalytic properties might have been combined in RNA, and that “one can contemplate an RNA world, containing only RNA molecules that serve to catalyze the synthesis of themselves”. The idea of an RNA world alone implied that, prior to the origin of proteins, RNA could have been a multivalent catalytic molecule, despite having been outperformed by protein-based enzymes later on. Indeed, molecular vestiges of the originally abundant ribozymes can be found in the plethora of nucleotide-based coenzymes that modulate the catalytic function of contemporary protein enzymes.<sup>38,692</sup> Since the pioneering works of Cech and Altman, eight classes of ribozymes have been identified in extant organisms, including seven that catalyze cleavage or ligation of the RNA backbone (Figure 26): group I and II autocatalytic introns, RNase P, hammerhead, hairpin, hepatitis delta virus ribozyme, and VS ribozyme.<sup>693–695</sup>

The eighth type of natural ribozyme, the ribosome’s peptidyltransferase center, catalyzes the formation of a peptide bond between two amino acids previously joined to its specific tRNAs, the specificity being dictated by the genetic code.<sup>701–703</sup> Indeed, the discovery that the catalytic core of the ribosome is composed of rRNA (Figure 27) was one of the major pieces of



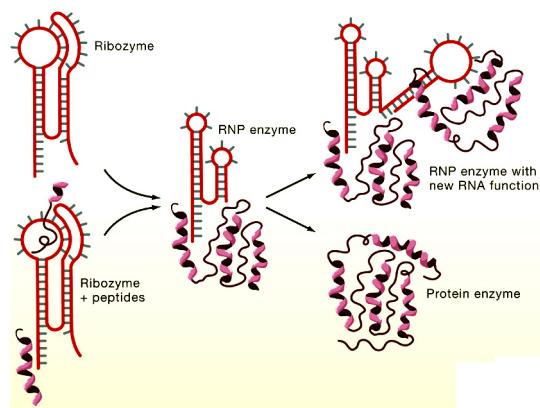
**Figure 27.** Crystal structure of *Thermus thermophilus* 70S ribosome with the small subunit (30S) on the left (16S rRNA is depicted in red and ribosomal proteins in orange), and the large subunit (50S) on the right (23S rRNA is depicted in black, SS rRNA in brown, and ribosomal proteins in green). Locations of peptidyl-tRNA (light blue) at the subunit interface, mRNA (purple—light blue—dark blue), and a modeled nascent polypeptide chain (purple) are shown. Adapted with permission from ref 702. Copyright 2001 American Association for the Advancement of Science.

evidence supporting the RNA world hypothesis. Nevertheless, the functional repertoire of natural ribozymes seems to be currently limited to catalysis of certain reactions required for the flux of genetic information within the cell.

Additionally, both artificial functional RNA molecules and improved versions of certain natural ribozymes have been obtained through in vitro selection of nucleic acids. The first experiments aimed at determining whether Darwinian evolution was possible in cell-free systems involved in vitro replication of the RNA genome of bacteriophage Q $\beta$  by the viral RNA polymerase enzyme. The phenotype under selection was the replication speed of the RNA genome, and the

mutation rate of the viral polymerase ensured genetic variation during the experiment. After several serial transfers of aliquots of the replication mixture into new reaction tubes containing fresh enzyme and ribonucleotides, the  $\text{Q}\beta$  genome evolved by deleting sequences not required for its recognition by the polymerase, thus shortening the replication time.<sup>704</sup> Two decades later, it was suggested that the production of metabolic ribozymes could be achieved by *in vitro* replication (using  $\text{Q}\beta$  replicase) and selection, through a process that would involve a large number of mutant RNA variants able to differentially bind to an stable analogue of the transition state complex of the reaction to be catalyzed, with catalytic efficiency being the criterion for selection.<sup>705</sup> In parallel, the possibility to chemically synthesize random-sequence nucleic acid pools with up to  $10^{15}$  molecules, together with the availability of the RT enzyme and PCR technology, allowed setting up a experimental protocol of *in vitro* selection of nucleic acids by successive rounds of amplification and activity selection (see section 5.3). The stepwise process of amplification-selection was termed systematic evolution of ligands by exponential enrichment or SELEX,<sup>706</sup> and the term "aptamer" (from the Latin word *aptus*, which means fitting) was coined to denote the *in vitro*-evolved, target-binding RNA.<sup>707</sup> Thanks to the optimization of *in vitro* selection systems, it is currently possible to develop RNA and DNA aptamers specific to a variety of molecular targets, as well as to engineer new ribozymes and even deoxyribozymes or DNAzymes. By use of SELEX-derived methodologies, the repertoire of ribozymes has been largely increased, some of them endowed with metabolic functions that might have operated during the RNA world. Relevant examples include ribozymes that synthesize ribonucleotides from activated ribose and pyrimidine,<sup>708</sup> and others that catalyze carbon–carbon bond formation, amide bond formation, acyl transfer, N- and S-alkylation, and Michael addition.<sup>16,677,709</sup> Despite all these advances, the possibility of complex metabolisms based exclusively on ribozymes is still purely conjectural.

Indeed, the RNA world hypothesis posits that as metabolic requirements became more sophisticated, increasing demand on different kinds of catalysis was a selective pressure leading to the transition to protein enzymes. Thus, plausible pathways for the evolution of biological catalysts, from ribozymes to ribonucleoprotein enzymes and protein-only enzymes, have been postulated (Figure 28).<sup>710</sup> Among the catalytic ribonucleoprotein complexes, the efforts to trace evolutionary pathways toward modern ribosomes are particularly relevant. Given the phylogenetic distribution of the main cellular RNAs involved in translation, it is assumed that LUCA had the same kind of translational apparatus as current cells.<sup>613,614</sup> It has been proposed that the original translation, eventually catalyzed by ribozymes with primordial peptidyltransferase activity, not only allowed the synthesis of the first RNA-coded proteins but also provided short peptides able to bind RNA, thus increasing its available structural and functional spaces.<sup>284,680</sup> Some of the first translated peptides, eventually complemented by random-sequence oligopeptides produced by nonenzymatic polymerization (see section 2.2.1), might have bound to the structural motifs present in the peptidyltransferase ribozyme. Such a process would have progressively improved the stability of the protoribosome, as well as the accuracy of primeval translation by positive feedback. This evolutionary pathway involving RNA and peptides could have paved the way to a progressive



**Figure 28.** Schematic model for evolution of biological catalysis. Primordial RNA-only ribozymes as well as complexes of ribozymes and random peptides could have acted as catalysts during the first steps of the RNA world. In a more advanced stage, upon the advent of peptidyltransferase ribozymes, the availability of RNA-coded proteins allowed the assembly of ribonucleoprotein (RNP) complexes. Some of the RNPs could have shown novel or improved catalytic activities, ultimately including the translation of mRNA on protoribosomes. Later, some RNP enzymes (upper right) evolved by adding or discarding some RNA subunits and fine-tuning their catalytic activity. In parallel, most RNP complexes (lower right) evolved to protein-only enzymes. Reproduced with permission from ref 710. Copyright 2009 Elsevier.

optimization toward the complex, accurate, and highly regulated ribosome present in LUCA and its descendants.

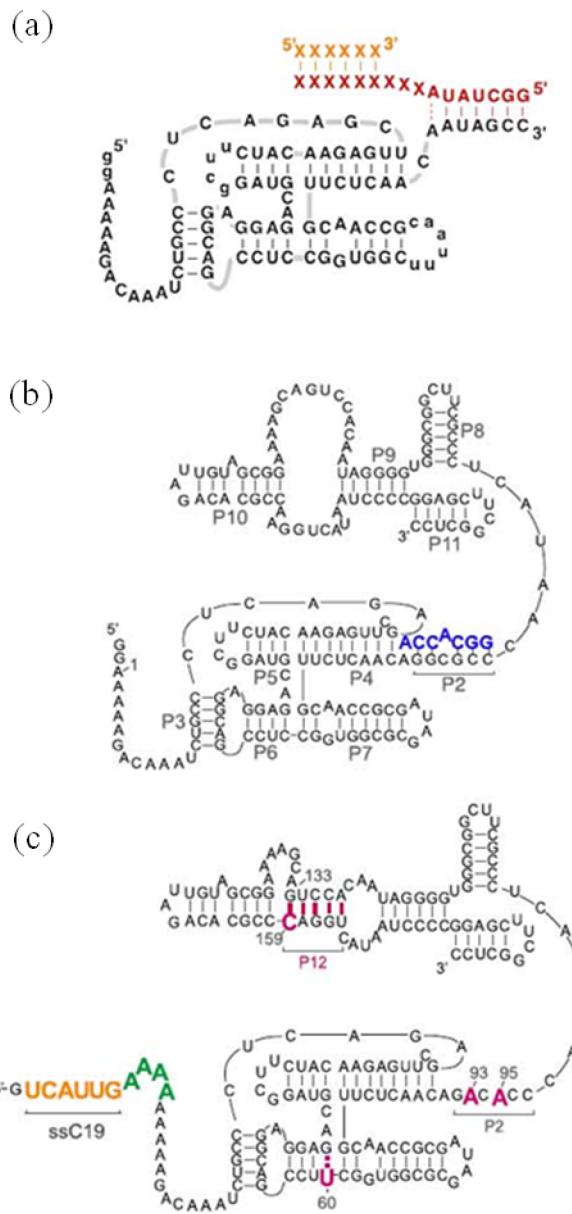
The unveiled functional plasticity of RNA has increased the plausibility of the RNA world hypothesis, since ribozymes similar to some of the natural and *in vitro*-evolved ones, as well as their complexes with peptides and proteins, might represent one of the missing links in the transition from the RNA-only world to cellular life. Nevertheless, relevant open questions remain to be solved, including the pathways eventually followed to produce progressively longer and more complex ribozymes from the relatively short RNA oligomers obtained by random polymerization on clay mineral surfaces or lipid-based systems (section 2.2.2). It has been claimed that, in the nucleotide pools generated by abiotic polymerization, ribozyme species already existed and showed different functionalities, including the ability to catalyze RNA self-replication.<sup>16,39</sup> However, abiotic polymerization of ribonucleotides usually renders oligomers of less than 50–80 nucleotides (nt) in length, while relevant template-dependent RNA ligase and RNA polymerase ribozymes are roughly 120–190 nt long and show conserved nucleotides in their catalytic cores. Therefore, the hypothesis of an early origin of complex ribozymes remains highly controversial.

Alternatively, a stepwise model of ligation-based, modular evolution of RNA has been developed *in silico*, showing two main advantages with respect to the previous hypothesis favoring direct evolution of complex ribozymes.<sup>265</sup> On one hand, short RNA modules resulting from template-independent polymerization on different abiotic microenvironments (section 2.2.2) might act as the first functional RNAs, including RNA ligase ribozymes with hairpin structure.<sup>711</sup> On the other hand, modular evolution shortens adaptation times and allows the attainment of complex structures that could not be otherwise directly selected.<sup>712</sup> This model favors the emergence of complex functional molecules by covalent binding of the RNA

subsystems generated by abiotic polymerization, which might have previously shown its own dynamics and protometabolic network of interactions. Therefore, ligation-mediated modular evolution of RNA could have bridged the gap between the outcome of prebiotic chemistry and a relatively complex template-dependent RNA polymerase ribozyme that led to a fully fledged RNA world. This kind of approaches highlight the general idea that the earliest ribozymes (e.g., the hairpin RNA ligase) must have arisen through physicochemical processes (in this case, the thermodynamics of base-pairing that dictates the abundance of hairpin structures in pools of RNA oligomers obtained by template-independent polymerization), while their catalytic activity could have been further enhanced and fine-tuned by natural selection.

The implication of ligation-based mechanisms in the origin of the RNA world has been explored experimentally. Non-enzymatic template-directed ligation experiments have shown that, with DNA oligomers as both templates and substrates, phosphodiester bond formation can be induced by cyanogen bromide. The chemical ligation occurs with high fidelity, which can be further increased by increasing temperature.<sup>267</sup> Additionally, Mg<sup>2+</sup>-dependent chemical ligation of oligoribonucleotides on RNA templates has been achieved,<sup>713</sup> showing a high 3'-5' regioselectivity in the products.<sup>268</sup> It was further documented that nonenzymatic, template-dependent RNA ligation can also occur in water in the presence of 3',5'-cyclic AMP, without requiring any added metal ion or cofactor.<sup>266,714</sup> Thus, experimental evidence suggests that once short RNA oligomers were originated by template-independent polymerization, their use as partially complementary templates and substrates by chemical ligation processes could have stepwise increased the length of RNA molecules, with the ligation fidelity later being enhanced thanks to the advent of RNA ligase ribozymes.<sup>715-717</sup> The likely relevance of RNA ligation in early evolution has been exemplified by the development of experimental systems based on ligase ribozymes that undergo self-sustained replication with exponential growth,<sup>718,719</sup> even if the ribozymes are synthesized entirely from L-ribonucleotides.<sup>720</sup>

Once the RNA world was initiated, the persistence and transmission of the heritable information encoded in any RNA molecule necessarily required its (faithful enough) replication. This process should be called "self-replication" (in this biochemical context, which is different from that discussed in section 3.2.2) only if the nucleotide sequences of the RNA template and the ribozyme (i.e., genotype and phenotype, interpreted at the molecular level) are identical. The hypothetical RNA molecule that could potentially act as both the genetic information to be copied and the RNA polymerase that can copy it would be a true "RNA replicase". It is evident that two identical molecules are required for this process: one used as the template and the other acting as the actual replicase. The search for a putative RNA replicase have been conducted following two approaches that used previously in vitro-evolved ribozymes: template-directed ligation of preformed RNA oligonucleotides<sup>583,715,716,718</sup> and template-dependent polymerization of ribonucleotides in a sequence-specific manner.<sup>721,722</sup> Noticeably, both ribozyme families are evolutionarily related, because the most efficient template-dependent RNA polymerases so far obtained were in vitro-evolved and engineered from the 119 nt-long class I ligase ribozymes (Figure 29a) as the starting point.<sup>716,723</sup> The first milestone in the way toward an RNA replicase was ribozyme R18, a 189 nt long RNA



**Figure 29.** Secondary structure model of the main in vitro-evolved ribozymes with template-dependent RNA polymerase activity. (a) A ribozyme (black strand) derived from the class I RNA ligase ribozyme (uppercase residues are those that comprise the ribozyme core), which is able to promote limited RNA polymerization: it incorporates any of the four ribonucleotides at positions indicated by an X, provided that an RNA primer (orange strand) pairs with the RNA template (red strand, which must also pair with the 3' end of the ribozyme). (b) The R18 RNA polymerase ribozyme (combined with an RNA heptamer, depicted in blue, that completes the ribozyme by forming the helix P2) is able to copy up to 14 ribonucleotides on favorable external RNA templates by extension of a complementary RNA primer. (c) The tC19Z RNA polymerase ribozyme can polymerize a wide spectrum of RNA molecules up to 95 nt in length, provided that the 3' terminus of a complementary primer pairs with the external RNA template. Mutations isolated from the selection process are depicted in magenta and orange while engineered residues are shown in green. Adapted with permission from refs 721 and 724. Copyright 2001 and 2011 American Association for the Advancement of Science.

polymerase able to add up to 14 ribonucleotides on an external RNA template (Figure 29b).<sup>721</sup> Further evolution and

engineering of R18 led to a new RNA polymerase ribozyme with improved polymerase activity and sequence generality: the 187 nt long variant termed tC19Z (Figure 29c).<sup>724</sup> This ribozyme can polymerize RNA molecules of up to 95 nt in length, on a range of different primer-template sequences. Therefore, we are already facing an RNA polymerase ribozyme that efficiently catalyzes the replication of distinct RNA molecules, including other ribozymes, whose length is up to half its own length.

The pathway to RNA polymerase ribozymes with improved performance is accompanied by a moderate increase in their polymerization fidelity. While nonenzymatic template-dependent RNA polymerization (using nonactivated ribonucleotides in lipid-based systems) renders a misincorporation rate of  $9.9 \times 10^{-2}$  per nucleotide copied,<sup>724</sup> the error rate of R18 is around  $1.1 \times 10^{-2}$  mutations per nucleotide copied,<sup>721</sup> and that of tC19Z is  $8.8 \times 10^{-3}$ .<sup>724</sup> However, such error rate values have been determined by sequencing of full-length extension products, while it cannot be assumed that the aborted sequences (which constitute the majority of the copied products) showed a similar error rate. Indeed, nucleotide mismatching is one of the main reasons for stalling and polymerization termination. These fidelity values are thus clearly overestimated, though they are useful for comparing the relative performance of such polymerization systems. If it is assumed that the error rates associated with primordial ribozyme-catalyzed RNA replication were similar to (or higher than) those of R18 or tC19Z, an extensive exploration of the available RNA sequence space would have been guaranteed. Thus, the selection of advantageous mutations in the replicated templates, eventually including the polymerase itself, could have occurred. Progressively smaller mutation rates would have permitted the emergence of ecological organization of complex populations of replicators in the RNA world.<sup>725</sup> In parallel, higher fidelity likely allowed replication of longer RNA molecules,<sup>562</sup> such as the “progenomic” RNAs previously built by means of optimized ligation<sup>540</sup> and/or recombination<sup>726</sup> mechanisms acting on preformed ribozymes and other modular assemblies. Thus, the initial RNA genomes might have propagated thanks to the availability of improved RNA polymerase ribozymes, which would have also replicated a number of molecular parasites (e.g., short RNA polymers, highly mutated sequences, or nonfunctional molecules produced by ligation or recombination) acting as alternative templates. The challenge for the RNA polymerase was therefore to faithfully replicate the genome that contained its own sequence, and to compete successfully against the molecular parasites unavoidably present in the system.

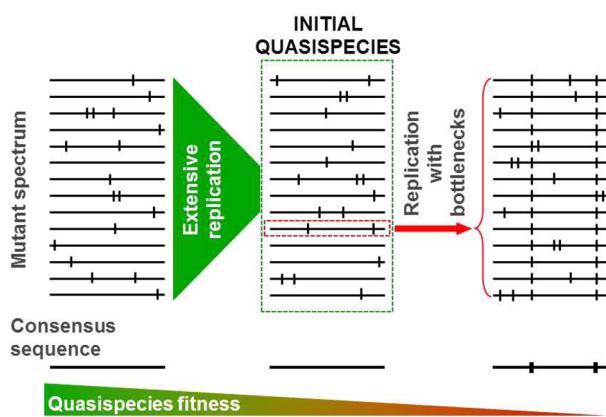
Therefore, the study of replication at intermediate error rates, between those of RNA polymerase ribozymes ( $\sim 10^{-2}$  mutations per nucleotide copied) and those of highly accurate DNA polymerases operating in current cells ( $\sim 10^{-9}$ ), is required to understand the evolutionary dynamics in a putative RNA world. Interestingly, evolution at high mutation and recombination rates can be experimentally explored by use of RNA-based systems that are currently replicating in our biosphere, as we discuss in the next section.

**3.3.3. RNA Virus Quasispecies as a Model System for the RNA World.** Viruses are obligate intracellular parasites that can persist extracellularly by enclosing their RNA or DNA genome within a protein capsid. They outnumber cellular organisms by orders of magnitude and can infect archaeal, bacterial, or eukaryotic cells in all ecosystems. Although viruses

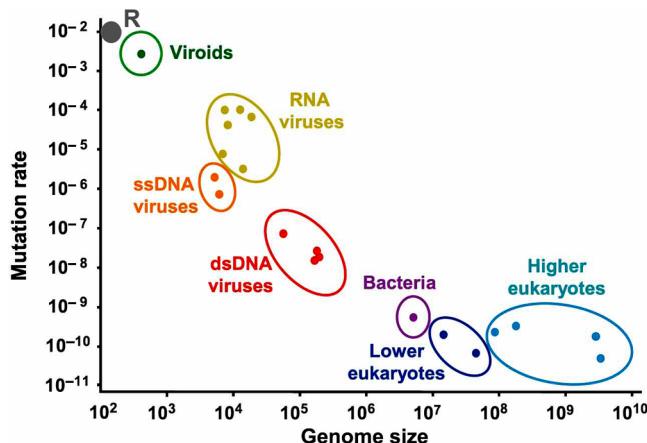
are not autonomously alive, and both their origin and antiquity remains unclear, it is evident that they have coevolved with their cellular hosts, thus promoting a continuous interchange of genetic information by a process known as horizontal gene transfer (HGT).<sup>727–729</sup> In turn, viroids are independently replicating small circular RNAs (of length 256–401 nt) that lack protein-coding genes and are capable of infecting plants. While viruses can be considered as parasites of the translation machinery of their host cell, viroids parasitize the host transcription machinery.<sup>730,731</sup> It has been claimed that RNA viruses and viroids are extant descendants of the primeval RNA world, when they could have infected RNA-containing precellular systems.<sup>732,733</sup> On the contrary, viruses might have originated once modern cells had entered the scene, their simplicity being explained as an inherent consequence of their parasitic nature and the selective pressures operating on them to keep a minimal genome able to replicate faster within the host cell.<sup>734</sup> However, the fascinating and controversial topic of the origin of viruses is beyond the scope of this review. In any case, both RNA viruses and viroids constitute the best natural model systems for studying the evolution of RNA populations in the laboratory.

RNA viruses and viroids are characterized by possessing very high mutation rates in comparison to cells, with an average of  $10^{-4}$  substitutions per nucleotide copied. This value lies between two extreme systems: it is roughly 2 orders of magnitude smaller than that of the tC19Z RNA polymerase ribozyme (see previous section) and 5 orders of magnitude higher than that of (viral or cellular) DNA-based systems. The mutation rate of RNA viruses and viroids is due to the absent or low proofreading activity of the RNA-dependent RNA polymerases and reverse transcriptases that replicate their genome.<sup>735,736</sup> As a consequence, RNA viruses and viroids propagate in their hosts as highly heterogeneous populations of closely related but nonidentical genomes, as was shown for the first time by Domingo et al. in 1978,<sup>737</sup> through analysis of the sequence heterogeneity of the Qβ bacteriophage. The complex population structure exhibited by this RNA virus was termed “viral quasispecies” (Figure 30).

Quasispecies theory had been developed by Eigen in the early 1970s<sup>562</sup> as a general model to explain the dynamics of the first hypothetical replicative molecules, in the context of the origin of information in a precellular world (see references above, in section 3.2.2). In the original quasispecies concept, the error-prone replication of an informative molecule originates a stable cloud of mutants around a master sequence. The whole mutant spectrum is called a molecular quasispecies. This pioneering work predicted that the maximum length of any heritable genetic polymer is limited by the copying fidelity associated with its replication. The inverse correlation between genome length and replication error rate has been confirmed experimentally: while cellular genomes (whose replication relies on a high-fidelity DNA polymerase enzyme) can be up to roughly  $10^9$  nt long, RNA viruses show a maximum length of  $10^5$  nt, viroids are at most  $4 \times 10^2$  nt long, and the sequence of RNA molecules replicated by in vitro-evolved RNA polymerase ribozymes spans less than  $2 \times 10^2$  nt (Figure 31). Although quasispecies were initially conceived as steady-state distributions of infinite size in equilibrium,<sup>539,562</sup> their dynamics (characterized by a continuous process of mutant generation, competition, and selection) has provided an interpretation of the great adaptive potential of current replicative entities based on RNA.<sup>738,739</sup> Indeed, quasispecies structure, high recombi-



**Figure 30.** Schematic representation of the dynamics of viral quasispecies. Individual viral genomes within the mutant spectrum are depicted as horizontal lines, and different mutations, derived from the error-prone replication catalyzed by the viral enzymes, are marked on them. The consensus sequence (also called population or average sequence) reflects the most abundant nucleotide at each position. The extensive replication of the quasispecies originates a new population with a different distribution of mutations within the mutant spectrum, which is not reflected in the consensus sequence. In turn, replication from either one or a small number of genomes introduces a population bottleneck and, as a result, the mutations present in the parental genome(s) are maintained in the progeny, together with the novel mutations introduced during their error-prone replication. Extensive replication generally results in fitness increase, whereas bottlenecking usually decreases quasispecies fitness.



**Figure 31.** Mutation rate per nucleotide vs genome size for viroids, viruses, and cellular organisms, derived from experimental data. The inverse correlation between these two parameters is clearly observed. The gray circle labeled R has been added to approximately represent the corresponding mutation rate and sequence length of template-dependent RNA polymerase ribozymes obtained through in vitro evolution. Adapted with permission from ref 736. Copyright 2009 American Association for the Advancement of Science.

nation rates, and large population sizes allow RNA viruses to quickly react to selective pressures exerted by the host environment.<sup>740</sup>

Several decades of experimental research on the genotypic and phenotypic diversity of different RNA virus families, as well as the development of ad hoc mathematical models, have shown that the target of selection is the whole viral population rather than individual viral particles.<sup>741,742</sup> Indeed, it is now evident that quasispecies behavior is influenced by the

ensemble of mutants that compose the population,<sup>743,744</sup> their evolution and fate being determined by the interactions among the individual viral genomes. In agreement with this experimental evidence, computer simulations of molecular quasispecies dynamics have shown that the adaptive behavior of evolving quasispecies cannot be explained by the features of any particular genome in the mutant spectrum, not even those of the most abundant one, but by the result of a collective search of the population as a whole.<sup>745</sup> In particular, it is currently recognized that minority genomes present in the mutant spectra of viral quasispecies, which can be monitored by ultradeep sequencing or high-throughput microarrays (see section 5.2), may play a relevant biological role. It has been experimentally shown that, in some cases, minority components of the population can be maintained as a molecular memory of the viral variants that were dominant at an earlier phase of the quasispecies evolution.<sup>746–748</sup> The maintenance of minority memory genomes within the quasispecies allows the population to quickly react to selective pressures analogous to those previously experienced. Thus, quasispecies memory provides an additional adaptive mechanism that might have modulated the evolution of heterogeneous, complex populations of genetic molecules in the precellular world.

Additionally, Diaz-Arenas and Lehman<sup>749</sup> have demonstrated the quasispecies behavior of a population of catalytic RNA molecules evolving in vitro. The experimental setup included a continuous in vitro evolution system (see section 5.3) that could keep track of hundreds of generations of class I ligase ribozyme populations evolving at a high mutation rate. The typical sequence distribution of viral quasispecies was found, with a majority sequence surrounded by a cloud of mutationally connected variants present at lower frequencies. This finding suggests that primordial RNA molecules evolving at high error rates could have benefited, as Eigen<sup>562</sup> proposed, from the robustness provided by the quasispecies structure of their populations, which eventually allowed recovery of genotypes that might have been removed, otherwise, by random drift.

#### 4. SYSTEMIC INTEGRATION APPROACHES

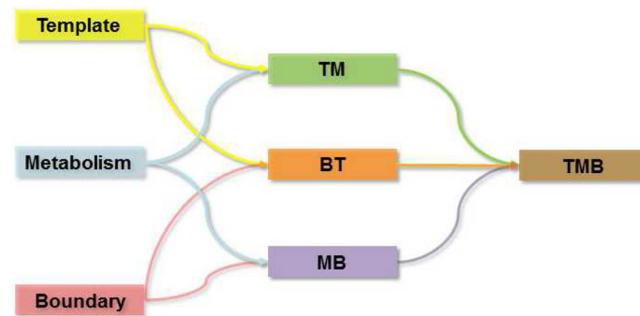
In previous sections we have reviewed various approaches toward the construction of complex chemical networks as precursors of protometabolisms, membrane structures as potential boundaries of protocells, and replicating, sequence-transferring (bio)polymers as the seed of protogenomes. These systems, even if they were essentially made of a single type of component (i.e., peptidic, amphiphilic, or nucleotidic) were shown to have quite intricate dynamics and several emergent properties, such as their ability to self-organize, self-assemble, or self-replicate. Nevertheless, in order to build up cell-like architectures with all the functions required to be considered alive, the proper integration of these three systems should be the following step. In fact, the combination of complex chemical but still infrabiological systems (*sensu* Szathmáry) into a living entity was probably the main challenge that supramolecular assemblies had to solve during the process of the origins of life. Having said that, integrative experimental approaches are also difficult in the laboratory, since many, not always compatible chemical and physical events need to be coupled in time and space, making diverse types of compounds and dynamic structures come together.

An initial problem encountered when facing the integration of the components of life into a cell-like entity relates to the terminology employed to define such entity. As Walde<sup>750</sup> has

argued, we need to differentiate among several concepts like artificial cell, synthetic cell, protocell, and minimal cell. The first two terms are synonyms and denote a man-made living cellular unit. In this respect, there are many possible degrees of “artificiality”. For example, a cellular construct might be tried by assembling biological components extracted from living cells, following reconstructive or reconstitution techniques (for an interesting review, see ref 751). There are also semisynthetic approaches, where some components of the cell come from current biology while others are synthetic ones.<sup>752</sup> From a chemical standpoint, the pursued ideal case would be the construction of a living cell from scratch, that is, by self-organization of its most basic molecular building blocks. This is actually what must have happened on the prebiotic Earth, when the first or most primitive cell-like structures appeared and developed into increasingly complex cellular organizations, without any cognitive agent (e.g., an experimental biologist) witnessing the process or interfering with it. Although the artificial bottom-up synthesis of a full-fledged cell would mean a truly deep understanding of the basic principles governing the phenomenon of life, and this should therefore be kept as an ultimate research goal, we do not know if that will ever be brought to reality. Therefore, the more modest concepts of protocell<sup>753</sup> and minimal cell<sup>754</sup> can be of much help.

A “minimal cell” can be defined as an artificial or semiartificial cell-like system containing the minimal and sufficient number of components to be considered alive. Insights into the minimal cell concept were initiated by Morowitz in the 1960s,<sup>755</sup> followed by the theoretical and/or experimental works of Deamer, Dyson, Ganti, Luisi, Maturana and Varela, Mavelli and Ruiz-Mirazo, Monnard, Moya, Pohorille, Rasmussen, Shuler, Szathmáry, Szostak, Ueda, Venter, and Yomo, among others (reviewed in refs 474, 750, 752, and 756). As we will discuss below, many theoretical approaches to minimal cellular life have been developed, but the experimental research program focused on the construction of liposome-based minimal cells is still in its infancy. A related concept is that of the “minimal genome”, coined in the top-down context of comparative genomics. It is based on the identification of the minimal gene set: “the smallest possible group of genes that would be sufficient to sustain a functioning cellular life form in the presence of a full complement of essential nutrients and in the absence of environmental stress”.<sup>757</sup> Such a minimal genome has been postulated to contain between 200 and 300 genes.<sup>752,758–760</sup>

Although this review will not be addressing all current efforts toward the construction of artificial/synthetic or minimal cells, it is now evident that investigation into the origins of life can benefit from some of the biochemical and/or computational approaches presently being developed with this aim, provided that they take into account reasonable evolutionary constraints. Otherwise, as stated by Szathmáry et al.,<sup>474</sup> the “synthesis of a living chemical system may not shed too much light on the historical process of the origination of life”. A useful way to confront the integration of protometabolism, protogenome, and protomembrane components, schematized by the previous authors,<sup>474</sup> consists of combinatorially connecting such fundamental components into doublet systems, which may then evolve into a ternary (super)system (Figure 32). As we will show below, different models of these binary, infrabiological subsystems, as well as of the ternary allegedly biological one, have already been proposed and their experimental implementation has been seriously tackled in the last years.



**Figure 32.** Different combinations of superchemical though infrabiological subsystems, inspired by Ganti's general scheme of the chemoton, based on three coupled autocatalytic cycles: template (T), metabolic (M), and boundary (B) subsystems. The TMB ternary supersystem would already meet all the requirements for life. Color code is the same as in Figure 1.

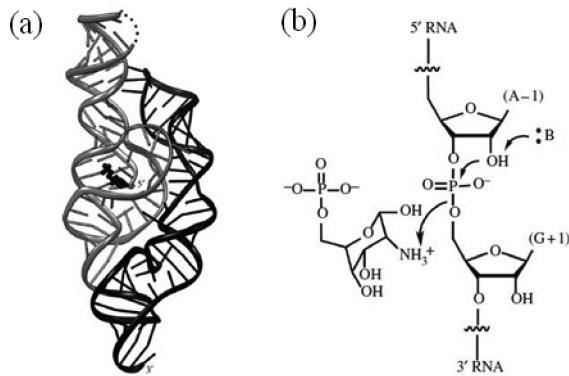
Therefore, it is becoming an achievable goal to synthesize protocells that can self-maintain, self-construct, and self-reproduce. Even if the systems reported so far are still quite rudimentary, the new techniques coming from synthetic biology and systems chemistry (see section 5) promise an exciting future that will certainly inform us about feasible pathways in protocellular evolution.

Thus, in this section we will follow the scheme of Szathmáry to separately review experimental results regarding the development of template/metabolism, boundary/template, and metabolism/boundary subsystems. The resulting ensembles, though lacking at least one of the critical functions or properties inherent to life, are all important and would be probably of relevance in the race toward a fully operating protocell. However, Darwinian evolution does not apply to individuals but to populations, and such may be the case for prebiological evolution. The notion of minimal cell cannot be defined univocally, and there were probably many options simultaneously explored in chemical space on the prebiotic Earth until informational molecules, protometabolic cycles, and membrane-based compartments were appropriately coupled, giving rise to the first population of full-fledged living organisms.

#### 4.1. Template–Metabolism Integration

As we discussed in section 3, triggering and maintaining template replication processes or protometabolic dynamics, on their own, is far from a simple task. Therefore, we cannot expect their combination to be easier. However, experimental approaches toward the integration of genetic and metabolic subsystems may also come from analysis of the functional interactions that RNA can establish with other nucleic acids, proteins, peptides, and small effector molecules. The chemical nature of RNA, together with its modular and hierarchical architecture based on secondary/tertiary interactions, defines a plethora of structural motifs that can function as interaction sites for small ligands. The RNA-binding ligands include (i) nucleotides, nucleosides and nucleobases; (ii) other aromatic and heterocyclic compounds; (iii) sugars; and (iv) amino acids.<sup>761</sup> Some of these chemical families are present in RNA-binding antibiotics, a heterogeneous group of ligands that modulate different RNA-catalyzed functions, including splicing and protein synthesis. Therefore, antibiotics have been conceived as low molecular weight partners of the RNA world,<sup>762,763</sup> and their coevolution with genetic and functional

macromolecules make them useful markers for tracing a functional phylogeny of cellular life.<sup>764</sup> With regard to other effector molecules, it has been described that, at least, one current catalytic RNA (the *glmS* ribozyme–riboswitch) employs a small molecule (glucosamine 6-phosphate, GlcN6P) as a coenzyme. In this system, binding of GlcN6P induces a structural rearrangement of the ribozyme, leading to its self-cleavage from the mRNA encoding the metabolic enzyme GlcN6P synthetase (Figure 33).<sup>765</sup> This example



**Figure 33.** (a) Crystal structure of *glmS* ribozyme–riboswitch at 1.7 Å resolution, showing core domain in gray, stabilizing domain in black, and cofactor GlcN6P as a black ball-and-stick model. (b) Catalytic mechanism of the self-cleaving ribozyme: GlcN6P binds to the active site of the prefolded though inactive ribozyme, where its amine group could function as a general acid and electrostatic catalyst (while the nucleobase of G40 might function as a general base catalyst). Cleavage leads to negative feedback regulation of the GlcN6P synthetase enzyme. Adapted with permission from ref 765. Copyright 2011 Royal Society of Chemistry.

suggests that noncovalently bound small cofactors could have expanded the functional repertoire of ribozymes and that some cofactor-dependent ribozymes might have further evolved to bind their effector molecules, including amino acids, covalently.

Turning to the interactions between RNA and polypeptides, it has been shown that, in present-day cells and viruses, RNA-binding peptides (usually <40 aa) are involved in a variety of processes, mainly related to transcriptional or translational regulation,<sup>766</sup> while RNA-binding proteins form different macromolecular complexes, including ribonucleoprotein assemblies with diverse catalytic functions.<sup>103,767</sup> Additionally, a growing number of RNA-binding peptides have been identified from combinatorial library experiments.<sup>768</sup> The high frequency at which RNA-binding peptides can be found in nature and in the laboratory suggests that peptidic binders could have been operating since the RNA world<sup>766</sup> and that, likely, the main structural features of the peptide instead of its particular sequence could have been evolutionarily selected. Indeed, due to the limited catalytic potential of natural and in vitro-evolved ribozymes (see section 3.3.2), it has been suggested that their interaction with basic peptides or other cofactors might have facilitated ribozyme–substrate contacts and, thus, catalytic processes.<sup>43,710</sup>

To date, few experimental approaches have explored the potential of RNA to evolve in vitro in peptide-containing media. In a reported example, the combination of structure-based design and in vitro RNA selection in the presence of a peptide during the SELEX process has been used to develop a ribonucleopeptide (RNPT) receptor specific for ATP, in which

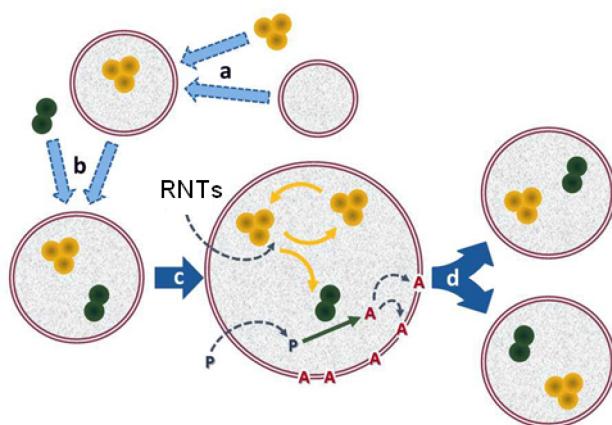
the binding domain consists of a 20 nt long RNA subsequence and a 17 aa long peptide subunit.<sup>769</sup> The recognition mode of this RNPT for the adenine ring of the ATP has been recently elucidated,<sup>770</sup> being different from those of the previously reported RNA-only aptamers specific for ATP analogues. A similar approach has allowed development of RNPT complexes specific for phosphotyrosine<sup>771</sup> or dopamine.<sup>772</sup> These results exemplify that an RNA-binding peptide can stabilize the functional structure of RNA during its exploration of the sequence space (in the form of a RNPT), thus suggesting possible scenarios of RNA–peptide coevolution during the RNA world. In turn, a ribozyme able to ligate RNA to a 22 aa long peptide through formation of a stable phosphoamide bond has been evolved in vitro,<sup>773</sup> which hints at potential mechanisms for increasing the functional diversity of both RNA molecules and covalently linked RNPT complexes.

Taking as a reference these successful experiments, the systems chemistry approach to the origins of life would clearly benefit from a research program aimed at studying in vitro RNA evolution of ribozymes or aptamers in the presence of different peptides, antibiotics, cofactors and/or other effectors that might have been available at the early stages of biochemical evolution. The outcome of those experiments in heterogeneous media, ranging from slightly doped solutions to increasingly complex mixtures of molecules, in combination with other informative approaches (e.g., in silico models about the potential for modular construction of complex ribozymes from RNA oligomers with additional protometabolic functions),<sup>265</sup> ought to shed light on possible synthetic pathways toward the integration of protogenomic and protometabolic subsystems.

#### 4.2. Boundary–Template Integration

Following the empirical pieces of evidence supporting the RNA world hypothesis, it was postulated that RNA-based protocells could have been able to self-maintain and self-reproduce before the advent of modern DNA/RNA/protein cells. The possibility of synthesizing an evolvable, RNA-based protocell was discussed by Szostak et al.<sup>37</sup> The experimental design of this minimal RNA cell or “ribocyte” would require the in vitro coupling of two different replicating systems: the RNA genome and the membrane-based compartment in which it is contained. This could be achieved by encapsulating an RNA replicase (still hypothetical; see section 3.3.2) within a reproducing vesicle (see section 3.3.1) that would grow and divide either gradually or stepwise, spontaneously or environmentally driven (i.e., externally fed). Following this approach, the simplest protocell model would consist of a self-reproducing vesicle containing an RNA replicase able to copy an identical RNA molecule, used as a self-complementary template. Nevertheless, as stated by the authors, this simple model would not show a sustained capacity for survival and reproduction, since it lacks the required interdependence between genome and membrane.

In order to enable an evolutionary optimization of the RNA replicase, a second and still unknown ribozyme able to catalyze the synthesis of the membrane lipids from their precursors would be required. In this more complete, and also more difficult to implement, model (Figure 34), the RNA replicase should drive the template copying of both molecules that put together the bipartite RNA genome of the protocell: its own sequence and that of the second ribozyme. Once synthesized by the RNA replicase, the activity of the second ribozyme would lead to a progressive increase of the vesicle size, and the growth



**Figure 34.** Proposed boundary–template integrative model for primordial synthesis of an RNA-based protocell: (a) A self-reproducing vesicle is combined with an RNA replicase (yellow). (b) This system is further combined with a second ribozyme (green) that is able to synthesize amphiphilic molecules (A) from precursor substrates (P), thus leading to an RNA protocell containing two ribozymes. (c) In such a “ribocyte”, the RNA replicase is capable of replicating itself and also making copies of the membrane-forming ribozyme, provided that ribonucleotides (RNTs) are available in the surrounding medium and can permeate the vesicle membrane. (d) Activity of the second ribozyme converts the previously internalized precursors into amphiphiles, which are further incorporated into the membrane; this leads to a progressive increase of the vesicle size and its subsequent division into two daughter vesicles, thus triggering Darwinian evolution of the whole (membrane–genome coupled) system.

of the membrane compartment would trigger the subsequent division of the parent vesicle into two. The daughter vesicles would ideally contain the same amount of “genes” of both ribozymes, whose sequences will be progressively modulated as a result of the environmental selective pressures. Thus, this improved RNA protocell system would be, as a whole, subject to Darwinian evolution.<sup>37,752</sup> Computational studies of this theoretical model are currently underway,<sup>774</sup> with the aim to simulate precisely its evolutionary dynamics and check its global coherence, as well as to anticipate possible additional difficulties that its real implementation might involve.

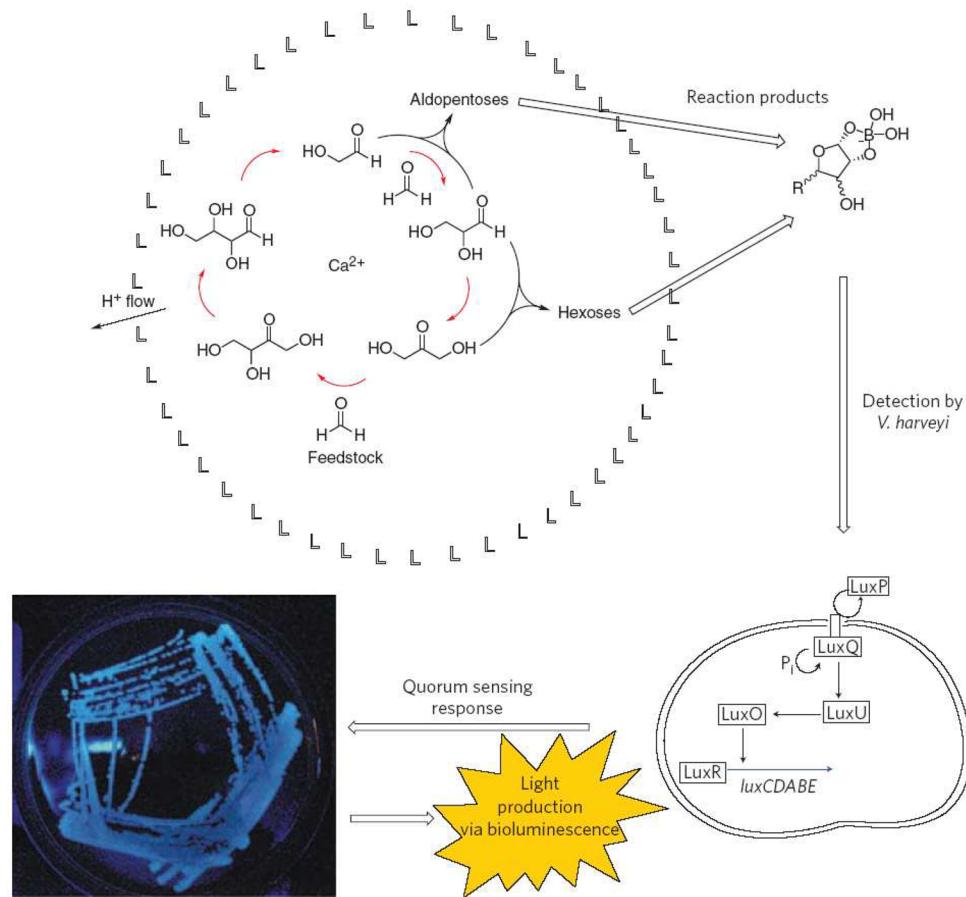
In addition to those theoretical proposals, an experimental line of research has been developed over the years for encapsulating genetic molecules within simple self-reproducing vesicles, formed by amphiphilic molecules allegedly synthesized under prebiotic conditions. Pioneering experiments showed that poly(U) and poly(C) RNA oligomers can be encapsulated within liposomes of different compositions.<sup>775</sup> Later, Luisi and co-workers<sup>776</sup> carried out one of the first empirical investigations on biopolymerization reactions within vesicles, in which ADP was condensed into poly(A), catalyzed by a polynucleotide phosphorylase enzyme (in a similar way as was reported independently by Chakrabarti et al.),<sup>536</sup> while the compartment could acquire additional lipid molecules from the external hydrolysis of a precursor compound (oleic anhydride, in this case). However, no direct coupling or synchronization was achieved between the RNA condensation reaction and the process of membrane growth and its eventual division. Despite the relevance of these experiments, which were among the first proofs of principle that biochemical reactions could be carried out within liposomes, the same limitation was apparent when RNA replication inside vesicles was achieved by making use of

the Q $\beta$  replicase.<sup>777</sup> So, as will be shown below for other approaches (see sections 4.3 and 4.4), the challenge remains in the actual coupling between different subsystems rather than in their mere combination as processes taking place in proximate space–temporal coordinates. In parallel to these encapsulation assays, investigations on the compatibility between RNA molecules and phospholipid vesicles have been more directly addressed. For instance, Yarus and co-workers<sup>778</sup> have experimentally analyzed how RNA molecules could bind to the vesicles and affect their permeability or even destabilize them.<sup>533</sup> Furthermore, the potential role of RNAs as membrane transporters,<sup>534</sup> or their effect in terms of lipid ordering within the bilayered membrane,<sup>535</sup> have also been explored by the same group.

More recently, the group of Szostak took a similar line of research, aiming at the functional integration of template-dependent polynucleotide replicators and prebiotic compartments. In particular, they have extensively studied the stability and permeability of vesicles formed by fatty acids, fatty alcohols, and fatty glycerol monoesters (see section 3.1.3). They were also able to encapsulate oligonucleotides and to study the growth and division of the nucleic acid-containing vesicles (section 3.2.1). Furthermore, they showed that the self-cleavage reaction of a hammerhead ribozyme can proceed within liposomes made of myristoleic acid and its glycerol ester.<sup>779</sup> Due to an osmotically driven competition between vesicles with different RNA loads,<sup>524</sup> the ones containing ribozyme molecules grew spontaneously. This suggests that osmotic pressure can be a factor affecting the coupling of chemical processes inside the compartment with the growth and division of vesicles.<sup>780</sup> These experiments put Szostak’s group in the right direction toward the performance of template-directed polymerization of activated nucleotides within the same type of vesicles.<sup>526</sup> The permeability of membranes composed by fatty acids and their corresponding alcohols and glycerol monoesters allowed 2-amino-2,3-dideoxyguanosine-5-phosphorimidazolides to pass through and polymerize on a DNA template (dC<sub>15</sub>).<sup>641</sup> This nonenzymatic template copying represents remarkable progress toward the construction of protocells (reviewed in refs 31 43, and 781).

Nevertheless, these experiments still lack a genome-driven connection between the reactions produced inside the vesicle lumen and the reproduction of the vesicles themselves.<sup>782</sup> According to the complete theoretical scheme envisioned,<sup>37</sup> vesicle reproduction should ideally rely on the ribozyme-catalyzed endogenous synthesis of the amphiphilic molecule, either from water-soluble precursors only or from water-soluble and -insoluble ones. The earliest attempts to achieve minimal self-producing and, potentially, self-reproducing vesicle systems focused on synthesizing the main molecular component of the membrane (e.g., lethitin) with the help of the necessary enzymes, encapsulated within the compartment.<sup>783</sup> If the complexity of that system, which involved four different enzymes (to catalyze the production of phosphatidylcholine from glycerol 3-phosphate via phosphatidic acid), could be reduced just to one or two reaction steps, preferably catalyzed by ribozymes, the dream of a good many researchers on the origins of life would be fulfilled.

Noticeably, these empirical approaches assume that, somehow, all the low molecular weight components required (ribonucleotides, energy-rich molecules, and precursors of the membrane-forming amphiphiles) are freely supplemented by the medium and can permeate the vesicle membrane. Thus, as



**Figure 35.** Schematic representation of encapsulation of the formose reaction into lipid vesicles. The reaction is initiated by an increase of pH outside the vesicles, and the produced carbohydrate–borate complexes diffuse out and are coupled with the quorum-sensing mechanism of the marine bacterium *Vibrio harveyi*. In particular, a proportional luminescent response results from a cascade of biochemical processes, which is initiated by the successful binding of those carbohydrate–borate complexes to the LuxP/LuxQ signal transduction proteins of the bacterium. Reproduced with permission from ref 794. Copyright 2009 Nature Publishing Group.

will be discussed below, the autonomous maintenance and reproduction of such template/boundary binary subsystems would require integration of the necessary metabolic reactions initially catalyzed by additional RNA-only or RNA–peptide ribozymes, as well as certain coupled protometabolic networks within the system. This, in Szathmáry's terminology,<sup>474</sup> would allow the conversion of units of evolution into proper units of life. But first we will briefly summarize how relatively simple metabolic or protometabolic reaction systems have been empirically explored in the context of lipid compartments.

#### 4.3. Metabolism–Boundary Integration

Early attempts to carry out biologically relevant chemistry (i.e., metabolic transformations) in lipid compartments can be traced back to the experimental work of Oparin<sup>784</sup> with coacervates, and even further back.<sup>504</sup> However, here we will mainly focus on the modern way of conceiving and investigating biological cellularity, through closed lipid bilayers or vesicles, and on approaches that use recent techniques and developments coming from biophysics, biochemistry, and molecular and cellular biology. In fact, the current flourishing of experimental studies on compartmentalized chemistries and so-called “bioreactors” can be explained as a result of the successful merging of liposome technologies and cell-free-extract technologies.<sup>785</sup>

Liposome technology has provided researchers with a suite of methods for vesicle preparation, from the classical film or injection methods to the more recent droplet transfer or microfluidic-assisted methods (see section 5.4), whereas cell-free-extract technology has allowed the efficient synthesis of biopolymers within them.<sup>786</sup> This involves important biotechnological challenges, and two limitations become evident: (i) cell-free extracts are black boxes, as far as their exact composition and responsibility for the success of the processes implemented are concerned; and (ii) it becomes no longer a strictly bottom-up approach, nor a search that can be classified as just a metabolism-compartment combination (since polynucleotide templates are, one way or the other, also necessarily present in the system). Further refinements, such as use of the protein synthesis using recombinant elements (PURE system)<sup>787,788</sup> have been quite helpful with regard to the first point but leave the second one equally unsatisfactory.

Significant achievements in this area include the first ribosome-catalyzed polypeptide synthesis,<sup>789</sup> the first biosynthesis of a functionally folded protein (green fluorescent protein, GFP),<sup>790</sup> and the cascade production of two well-folded proteins (an RNA polymerase that triggered the subsequent transcription and translation of GFP),<sup>791</sup> all within vesicle compartments. Nevertheless, these experiments cannot be strictly considered as the integration of two infrabiological

subsystems but as the encapsulation of a more complex mixture of compounds, partially derived from extant living organisms (*E. coli*, in most of these cases).<sup>792</sup> Therefore, we will turn our attention more specifically here to the less common experimental investigations carried out in the presence of lipid compartments, without use of macromolecular structures and extracts or kits of biological origin (enzymes, nucleotide templates, ribosomes, etc.) as the means to obtain the desired chemical/metabolic reactions.

A common goal along those lines has been to try the spontaneous, nonenzymatic polymerization of both nucleic acids and peptides within vesicles. As for the former, we already mentioned the work of Szostak and co-workers<sup>526</sup> realizing this, together with the work from Deamer and co-workers,<sup>273,274</sup> which has proved that hydration–dehydration cycles in multilayered lipid phases could also play an important role in that type of process. Regarding spontaneous peptide bond formation or amino acid condensation reactions, the contribution of liposomes has been effective in different experimental contexts: simulating a glycine-rich hydrothermal environment,<sup>197</sup> in polycondensations of NCA amino acids leading to preferentially homochiral sequences, or in the oligomerization potential of thioglutamic acid,<sup>271,793</sup> which is enhanced in contact with dimyristoylphosphatidylcholine (DMPC) lipid bilayers.<sup>272</sup> A scenario in which lipid vesicles and oligopeptides could indeed benefit from each other's presence and give rise to self-encapsulated protometabolisms has been further explored theoretically, making use of computational methods.<sup>531</sup>

Apart from biopolymer/oligomer chemistries, there have been several other attempts to implement reaction networks of prebiological interest, using simpler compounds within lipid compartments. Among these, a particularly successful approach was carried out by Davis and co-workers,<sup>794</sup> who designed a set of experiments to encapsulate the autocatalytic, sugar-synthesizing formose reaction (see section 3.1.2) within diphyanoyl-phosphatidylcholine (DPhPC) vesicles (Figure 35). They managed to do so by controlling the different conditions that trigger the process (basic pH, high enough concentration of formaldehyde, in situ production of the catalyst, concentration of calcium hydroxide, etc.) in the lumen of the vesicles. Their results corroborated the expected change in the final sugar product distribution, as compared to results in free solution, due to the strong sensitivity of the formose reaction to variations in boundary conditions, such as the actual volume of the reaction compartment. Interestingly, the authors also showed how this membrane-bounded protometabolic cycle can be engaged with the quorum-sensing mechanism of bacteria (e.g., the marine bacterium *Vibrio harveyi*), a kind of phenomenon that may have been related to the increasing complexity of metabolism in protocells or the signaling mechanisms between them. The relevance of these experiments in the field of origins of life will surely depend on the role that sugars are assumed to have played in the general problem of metabolism emergence (see section 3.1.2). In any case, the approach represents a more truly bottom-up strategy in which a protometabolic network of chemical reactions is combined with the self-assembly processes underlying vesicle stability.

On similar lines, various groups have investigated how catalytically active minerals entrapped within vesicles could help in the harvesting of energy resources required to drive internal, typically endergonic biosynthetic processes. These experiments can be classified into two groups: (i) those working with

photoactive semiconducting minerals, like CdS<sup>795</sup> or TiO<sub>2</sub>,<sup>796</sup> and (ii) those proposing a chemo-autotrophic scenario, of the “iron–sulfur-world” type, in which the production of a chemical pair (e.g., FeS/H<sub>2</sub>S) would constitute the free source of redox energy in the compartmentalized system.<sup>797</sup> In both cases, by either irradiation or precipitation of iron sulfide within the vesicles, the capacity for energy release was demonstrated as a precondition for the subsequent synthesis of complex organic molecules in the system.

The interest of these approaches involving compartments that contain reactants of lower molecular complexity (i.e., remaining closer to chemistry than to biochemistry) is quite apparent to our eyes, although they are far less popular in the field as compared to biopolymer-guided research. Indeed, they could be critical to determine some of the necessary conditions to put together an energetically autonomous system of production of components (i.e., a protometabolic cellular organization), leading to macromolecular devices of progressively higher complexity. It is still an open question whether the shortcuts being tried by semisynthetic strategies, including the ones reviewed next, will tell us anything about how nature got there.

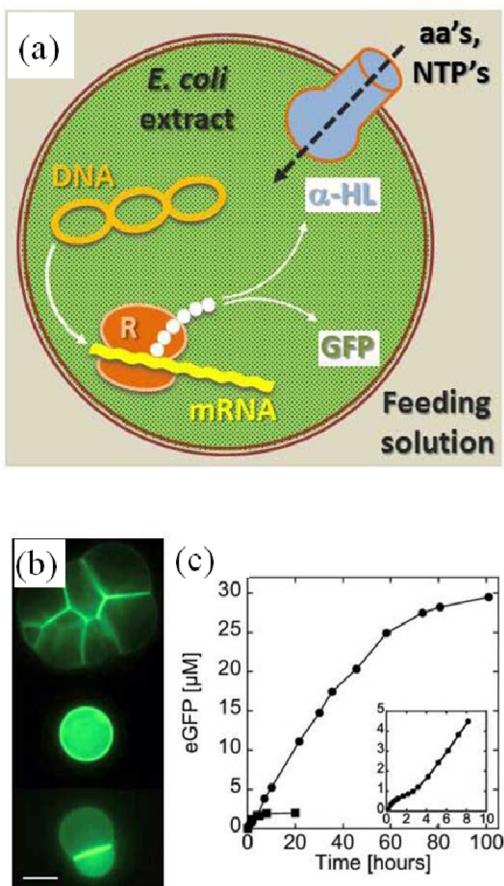
#### 4.4. Toward the Integration of Template, Metabolism, and Boundary

A number of experiments combining enzymatic activities within nucleic acid-containing vesicles and liposomes have been performed, as we already mentioned in the previous section. All of them could be interpreted, of course, as steps forward toward the construction of a minimal cell that includes all three template/metabolism/boundary (TMB) subsystems.<sup>750,752,759,785,798–800</sup> Among them, particularly relevant for the purposes of this review are those focused on nucleic acid replication within protocellular systems, including here RNA<sup>536,776</sup> and DNA<sup>801</sup> enzymatic polymerization, as well as DNA amplification by means of PCR.<sup>777</sup> Besides, the growing number of protein expression systems inside liposomes that have been developed during the past decade must also be acknowledged in this regard.<sup>789,790,802–806</sup> However, as also discussed above, these experimental systems can hardly be taken as bottom-up reconstructions of the process of origins of living systems, since they make use of components directly extracted or borrowed from them.

Given the difficulty of the challenge, another possibility has been to try a complex system that combines from scratch (i.e., through self-assembly of the parts) those three subsystems but implemented by completely alternative molecular pieces. This was the strategy taken by Rasmussen et al.,<sup>807</sup> who proposed the integration of PNAs (see section 3.3.1), a light sensitizer, and a lipid compartment, coupled together in a single chemical system with potential to grow and divide. So far, the idea has been just partly realized, in particular regarding the coupling between metabolism (i.e., photocatalytic lipid production) and vesicle self-assembly.<sup>808</sup> Although this approach could be criticized in terms of the lack of prebiotic plausibility of some of the components used, like the 8-oxo-G Ru photocatalyst, it represents a truly systemic, bottom-up attempt to achieve TMB integration, even if the result will be quite far from natural life.

The remaining option has been to assume the need to merge with top-down strategies and deal with semisynthetic hybrids. In this regard, apart from the approaches already cited or the direct handling, mimicking, and redesigning of naturally occurring microorganisms,<sup>809</sup> a very interesting advance took

place when a “cell-like bioreactor” (Figure 36) was constructed through the encapsulation in large ( $>10 \mu\text{m}$  in diameter)



**Figure 36.** (a) Schematic representation of a “cell-like bioreactor” obtained by encapsulation of *E. coli*-derived transcription from a DNA plasmid (orange) and translation from mRNA (yellow) systems within a liposome. Two proteins are produced: green fluorescent protein (GFP) and pore-forming  $\alpha$ -hemolysin ( $\alpha$ -HL), the latter allowing amino acids (aa), nucleoside triphosphates (NTP), and other low molecular weight substrates present in a feeding solution to enter the liposome. (b) The system synthesizes significant amounts of GFP and glows green inside an aggregate of liposomes (top), a single one (center) and a doublet (bottom); scale bar =  $20 \mu\text{m}$ . (c) Time course of expression of  $\alpha$ -HL/GFP (●, the inset corresponds to the first 10 h) in comparison to the expression of GFP inside a liposome without  $\alpha$ -HL under low osmotic pressure (■). Adapted with permission from ref 810. Copyright 2004 National Academy of Sciences.

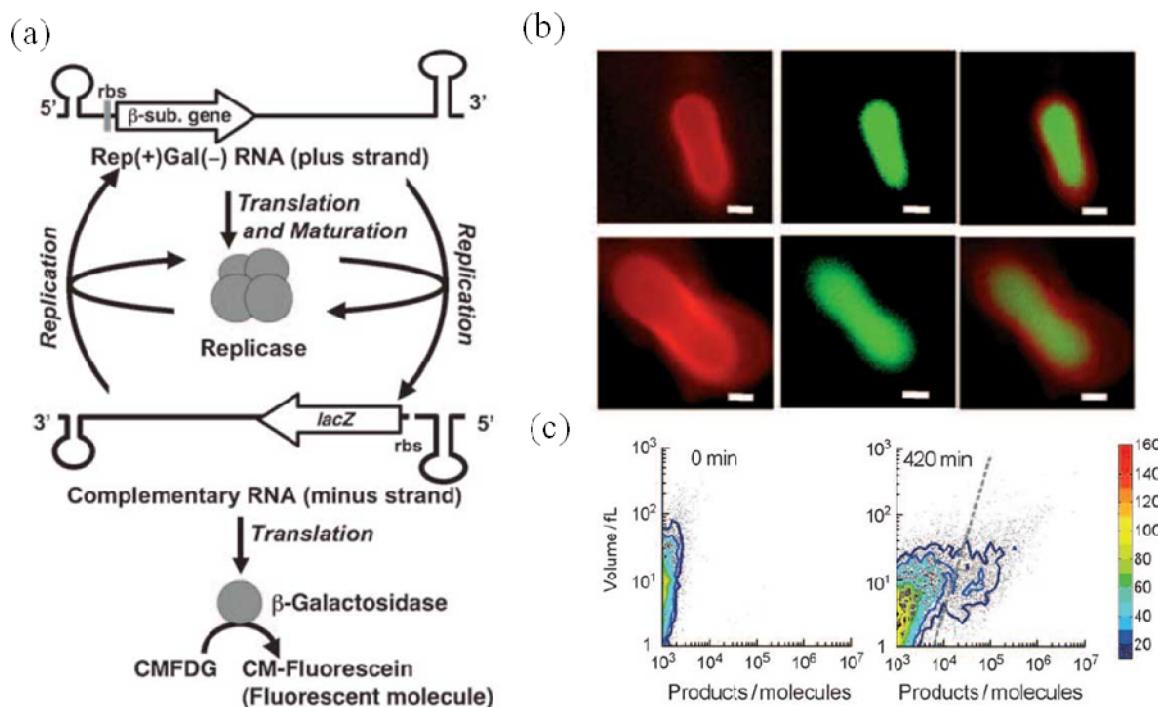
phospholipid vesicles of a DNA plasmid and an *E. coli* extract that provided the molecular machinery required for the expression (transcription and translation) of the DNA-encoded information.<sup>810</sup> Two genes had to be translated, the first one coding for the GFP used as a visual marker for protein synthesis and the second one for  $\alpha$ -hemolysin toxin, a pore-forming protein that increases the permeability of the membrane. The expression of  $\alpha$ -hemolysin and its integration as a heptamer in the bilayer to form a 1.4 nm diameter pore allowed that low molecular weight molecules (i.e., with a cutoff of  $\sim 3 \text{ kDa}$ , including ribonucleotides, amino acids, and energy-rich compounds), externally added from a feeding solution, crossed the membrane and fueled the transcription and translation reactions. This bioreactor produced proteins for up to 4 days

and was therefore considered as a relevant step toward the construction of an artificial cell.<sup>811</sup> More recently, in order to make the process more independent from the *E. coli* total extracts originally used, a cell-free version of the  $\alpha$ -hemolysin-containing system has been developed,<sup>812</sup> which allows real-time quantification of the concurrent expression and insertion of membrane proteins into lipid bilayers. Nevertheless, both variants of this system use DNA instead of RNA as the genetic molecule, and they lack the potential to self-replicate and evolve, essential requirements for moving from synthetic approaches to origins research.

A combination of some previous achievements was implemented in an alternative system that allowed RNA replication by a self-encoded RNA polymerase enzyme in liposomes composed by a mixture of two phospholipids and cholesterol.<sup>813</sup> The encapsulated mixture contained one template RNA sequence, including the gene of the catalytic subunit of  $Q\beta$  RNA polymerase protein ( $Q\beta$  replicase), as the genotype, and an *in vitro* translation system as the machinery for expressing the encoded polymerase (Figure 37). The cell-free, reconstituted reaction system consisted of a mixture of 144 gene products, including enzymes, ribosomal proteins, other proteins, rRNAs, and tRNAs, plus amino acids and other low molecular weight compounds. Once translated, the  $Q\beta$  replicase was able to copy its own template sequence, and the inherent high error rate of this enzyme generated the genetic diversity required for the evolvability of the compartmentalized system. Nevertheless, contrarily to previous approaches,<sup>810,812</sup> this “self-encoding system” contains neither transmembrane channels nor active transport machinery, and it works for a limited time (around 6 h) by consuming the encapsulated substrates. Additionally, the observed low efficiency of self-replication in the liposome seems mainly due to the competition between replication and translation machineries, together with unexpected interactions among the molecules, inactivation of enzymes, degradation of RNA, and accumulation of inhibitory products.<sup>799</sup>

In a step forward toward the effective coupling between encapsulated biochemical reactions and vesicle dynamics, it has been recently described that the activation and inhibition of enzyme-catalyzed reactions within vesicles can be triggered by exogenous addition of charged amphiphiles upon vesicle fusion.<sup>814</sup> Also, the enzymatic amplification of DNA has been linked to self-reproduction of the compartment where it is enclosed, through the interplay between the polyanionic nucleic acid and the cationic membrane of the vesicle.<sup>815</sup> In this recently reported system, the growth and spontaneous division of giant vesicles is accelerated by amplification of the enclosed DNA, and the process leads to the distribution of DNA to the daughter vesicles.

In conclusion, the relevant conceptual advances made during the past decade, together with the availability of technologies for generating and analyzing complex molecular mixtures (see section 5) have opened the way for developing more integrative approaches, aimed at synthesizing compartmentalized systems that display the key properties of life. Computational platforms that take seriously the challenge of working with minimal cell models should keep contributing to the task.<sup>816–818</sup> As a result, the synthesis of an encapsulated entity capable of self-replication and evolution, fed only by low molecular weight nutrients, is now at least foreseeable. If this goal is ever achieved, those autonomous and evolving entities would surely shed light on some of the transitions that must have occurred



**Figure 37.** Self-replication of genetic information in liposomes containing an RNA genome and a reconstituted reaction system with 144 gene products. (a) Schematic representation of the Rep(+)Gal(−) system: the catalytic subunit of the Q $\beta$  replicase enzyme is encoded on the (+)-RNA strand, while  $\beta$ -galactosidase is encoded on the complementary (−)-strand. Both strands can be replicated by the Q $\beta$  replicase. Upon the expression of  $\beta$ -galactosidase, nonfluorescent 5-chloromethylfluorescein digalactopyranoside (CMFDG) is hydrolyzed to yield green fluorescent 5-chloromethylfluorescein (CM-fluorescein). (b) Reaction within the liposomes observed under fluorescence microscopy: left, red images (membranes); middle, green images (hydrolyzed CMFDG); right, overlay of the previous images; scale bar = 1  $\mu$ m. (c) Analysis of the system, using 15000 liposomes, by fluorescence-activated cell sorting (FACS): the results in number of product molecules (horizontal) and internal aqueous volume (vertical) of each liposome (individual dots) are shown, and contour maps are overlaid (frequency depicted in color code). At 350 min (data not shown) and 420 min, the reacted liposomes are those with a substantial amount of fluorescent products (right of the dashed lines). Adapted with permission from ref 813. Copyright 2008 Wiley.

during the origins of life. For that purpose, according to the systems chemistry approach we envisage, the experimental efforts toward the recreation of primeval protocells should be conducted with mixtures of molecular components with increasing complexity. Although this field has inherited the necessarily reductionist tradition of organic chemistry and biochemistry, one should bear in mind that higher levels of heterogeneity in the system would probably increase the probabilities of success of this type of experiments, as recently pointed out by Szostak.<sup>31</sup> A solution or suspension made by mixing a small number of pure chemical components is the typical setup in current laboratories but not a realistic scenario for the appearance of life on early Earth.

## 5. CURRENT METHODOLOGICAL TOOLS FOR THE CHALLENGE

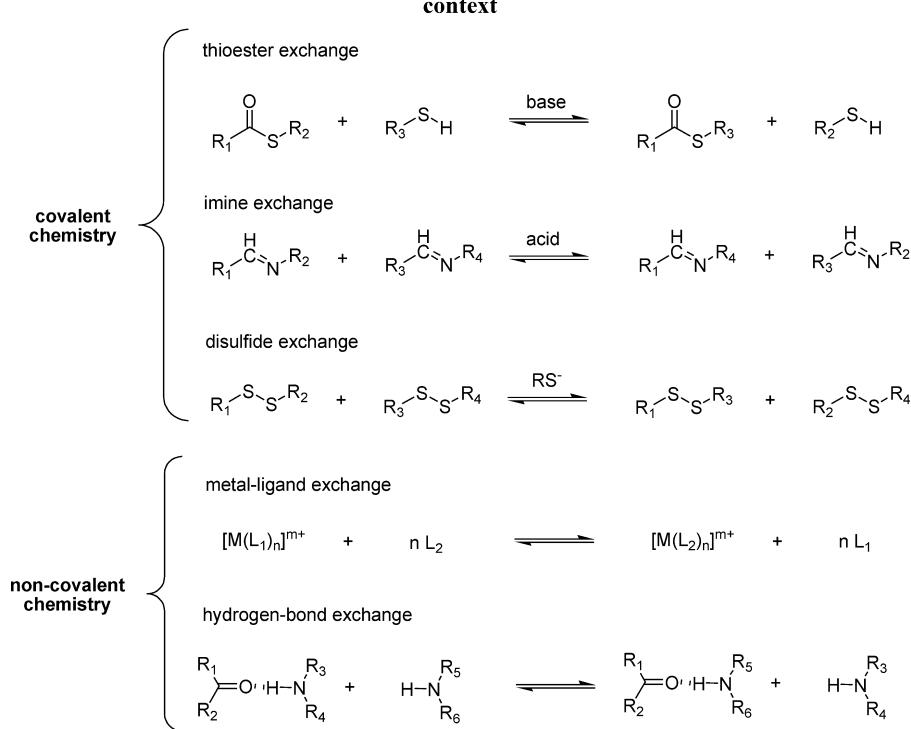
The systems chemistry approach for studying the origins of life benefits from several conceptual and technological developments incorporated into chemistry and biotechnology in recent decades, as well as novel techniques that are just coming out at present. Therefore, it is important to be updated about those methodological tools that can specifically contribute to this type of enterprise. The present section deals, in particular, with some major advances that have taken place in dynamic combinatorial methods, high-throughput techniques applied to biochemistry and molecular biology, *in vitro* evolution of biomolecules, and micro- and nanofluidics.

### 5.1. Dynamic Combinatorial Chemistry

Although the concept of systems chemistry has gotten ever more popular in recent years, it is not easy to identify which are the fundamental theoretical bases underlying this new field. In contrast to traditional synthetic chemistry, systems chemistry intends to deal with complex mixtures of molecules, inspired by the complex networks of signaling and metabolic pathways observed in biological organisms. Synthetic chemists had never faced before such an ambitious task, and at the beginning of the twentieth century they lacked a theoretical framework embracing the behavior of complex chemical networks. The emergence of dynamic combinatorial chemistry (DCC)<sup>819–821</sup> was probably the most important event toward the development of systems chemistry, since it provides a theoretical framework to understand how networks of molecules behave in equilibrium and under the action of external stimuli.

DCC is defined as combinatorial chemistry under thermodynamic control.<sup>822–827</sup> A dynamic combinatorial library (DCL) is formed by a set of interconvertible molecules through reversible bond formation. This definition implies that the distribution of library members is governed by their relative free energies. Moreover, any external stimulus, either a chemical or a physical process able to change the free energy of one or more of the DCL members, can ultimately alter the whole energy landscape as well as the products distribution of the library. Numerous examples illustrate the use of such behavior for screening and amplifying highly effective receptors of biological and nonbiological substrates.<sup>828–831</sup> Other possible

**Scheme 14.** Examples of Reversible Reactions Used in DCC That Are Relevant in a Prebiotic Context



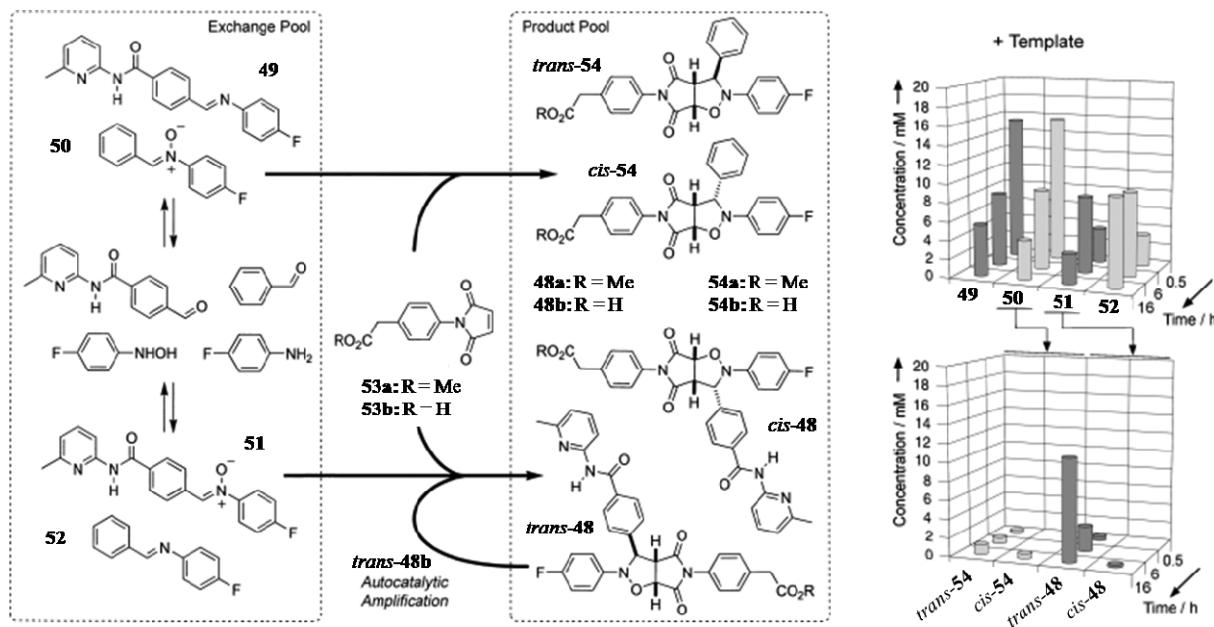
applications include the development of interlocked molecules,<sup>832–835</sup> transmembrane transporters,<sup>836–839</sup> adaptive materials,<sup>840–842</sup> sensors,<sup>843,844</sup> catalysts,<sup>845,846</sup> and self-replicators.<sup>847</sup> It is far from our aim to cover extensively the literature on DCC (for excellent reviews on the subject, see refs 822 and 824); instead, we will mainly focus on cases relevant to the origins of life (i.e., DCLs in water environments, with biomolecules or potential precursors of biomolecules), leaving aside, for example, the wide literature on DCLs in organic solvents.

From the large variety of possible reversible reactions used in DCC, some of them have certainly specific relevance from a prebiotic perspective. In particular, in water and under physiological conditions, imine exchange, disulfide exchange, and thioester exchange would be operative, as well as the entire possible noncovalent and metal–ligand interactions available (see Scheme 14). This scenario implies a rich chemistry that may be applicable to numerous prebiotic substrates, provided that the conditions in some primordial aqueous environments were moderate. Imine chemistry, which derives from carbonyl chemistry in combination with ammonia or organic amines, may have been involved in the generation of sugar diversity. Disulfide bridges are relevant for the chemistry of peptides when these contain cysteine.<sup>848</sup> Moreover, amino acids have shown to be activated as thioesters by carbonyl sulfide, which is present in volcanic gases (see section 2.1.2). Noncovalent interactions, on the other hand, lead to numerous self-assembled structures from lipid molecules, peptides, and nucleic acids. Metal–ligand coordination may have also played a role, for example, in the folding of primitive informational or functional oligomers, formed by either nucleotides or amino acids, and in catalysis performed by ribozymes or, later, protein enzymes.

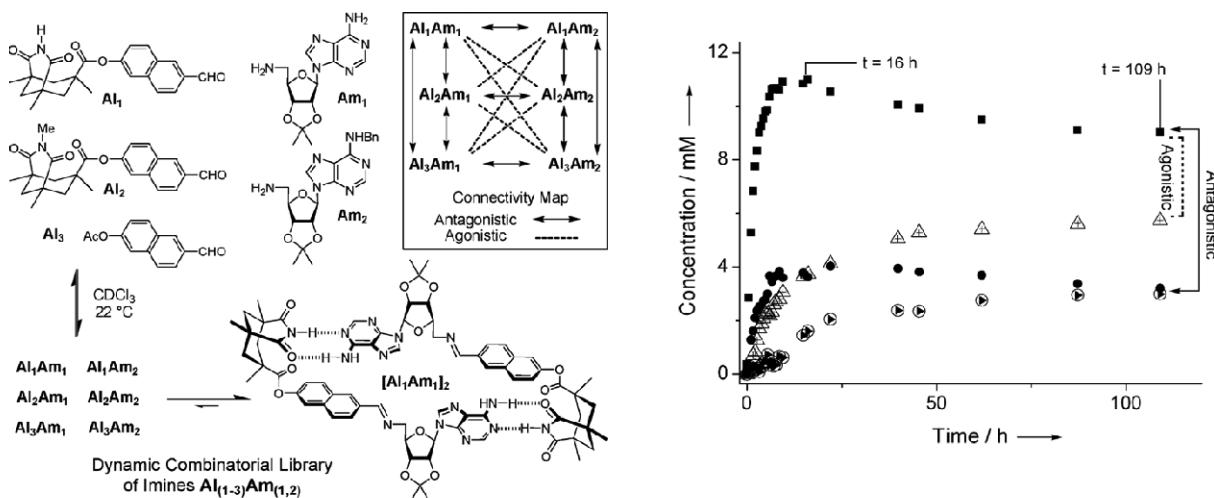
Furthermore, given the combinatorial explosion that probably characterized chemistry on the primordial Earth,

prebiotic dynamic libraries could cover a very large chemical space, determined by the ample variety of structures and functional groups available. Reversibility in such a scenario would allow the exploration of multiple reaction pathways toward functional chemical complexity. Chemical processes such as catalysis and self-replication are of interest for prebiological chemical evolution because they confer kinetic stability. In this respect, an important requisite for chemical evolution is the persistence of chemical systems in the far-from-equilibrium regime, as we recognized in the first section of this review. Some approaches related to DCC offer experimental and theoretical foundations for this idea. Multiple reversible chemistries may be simultaneously operated, or sequentially if they were orthogonal, with equilibration periods between the various conditions that activate both types of reactions.<sup>849–854</sup> The network topologies and product distributions obtained in such systems are different than those obtained from DCLs that simply converge into thermodynamic equilibrium. Combining DCLs with irreversible chemistries (e.g., catalysis, self-replication, or self-assembly events) is another way to make them kinetically driven, normally implying a significant increase in the amplification of target-selected library members.

When one considers the importance that enzymatic catalysis has in our current biosphere for the establishment of networks that support biological functions, the development of artificial enzymes becomes a remarkable goal in chemical science.<sup>855,856</sup> As previously argued, catalysis may be one of the driving forces for the transition from thermodynamically stable systems to kinetically stable ones. Catalyst development is unfortunately a difficult and challenging task. The advent of combinatorial chemistry, together with high-throughput screening techniques, proved to be a good complement to rational design and certainly allows exploration of vast regions of chemical space.<sup>857</sup> A further step in the search for efficient artificial catalysts with novel catalytic functions is the use of DCC (for reviews see refs



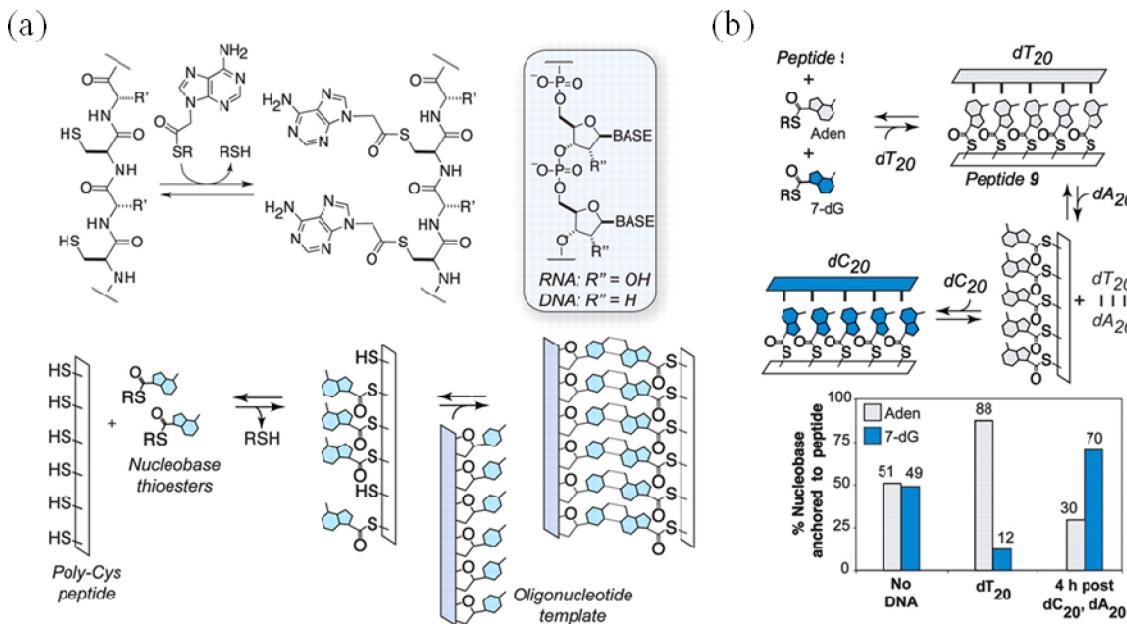
**Figure 38.** Dynamic combinatorial library based on imine/nitroxide chemistry, in which the self-replicator (*trans*-48b) emerges from an initial set of exchanging library products (49–52) after the addition of maleimides (53), while its nonreplicating counterparts (*cis*-48, *trans*-54, and *cis*-54) do not get amplified at all. The products distribution before (top) and after (bottom) the addition of 53 (in the presence of *trans*-48b template at the start of the experiment) is shown on the right. Adapted with permission from ref 860. Copyright 2008 Wiley–VCH.



**Figure 39.** (Left) Dynamic combinatorial library based on imine chemistry, in which one of the six imines generated from three aldehyde and two amine derivatives has the ability to dimerize through hydrogen bonds and template its own production. (Right) Evolution over time of the concentration of six library imines is shown, revealing agonistic and antagonistic connections between them. Adapted with permission from ref 861. Copyright 2008 American Chemical Society.

822, 846, and 858). Synthetic efforts are reduced in this approach to the preparation of building blocks, which are then able to generate molecular diversity by the process of self-assembly. Subsequently, self-selection of the best self-assembled catalysts, by association with the transition-state analogue (TSA) of a target reaction, relieves the limitation of rational design, leading to the discovery of reaction mechanisms impossible to predict *a priori*. In reality, the strategy to follow is a bit more complicated, since recognition of both substrate and TSA is necessary in order to achieve selectivity and activity, as very well exemplified by enzymes.<sup>859</sup> Alternatively, *in vitro* evolution of proteins is a biochemically based method to produce novel catalytic proteins (see section 5.3)

Within the context of dynamic catalyst discovery, the potential emergence of self-replicators (see section 3.2.2) would represent a complete change of scenario. The development of “smart complex mixtures” with autocatalytic properties is still nothing but a dream, although two recent examples have shed some light on this possibility. Sadownik and Philp<sup>860</sup> (Figure 38) and Giuseppone and co-workers<sup>861,862</sup> (Figure 39) have reported independently the combination of reversible dynamic chemistry with self-replication, showing that as soon as the self-replicator appears on the scene it is strongly amplified. Although self-replication in such systems is thermodynamically driven, and therefore does not fully satisfy the kinetic requirements for life’s self-replicating entities, these findings



**Figure 40.** (a) Schematic representation of the reversible linkage of nucleobase thioesters on a peptide backbone bearing Cys residues in every second position. The process is templated by DNA oligonucleotides. (b) Schematic representation of a competition experiment: incubation of adenine and 7-deazaguanine thioesters with the peptide backbone in the presence of dT<sub>20</sub> determines the coverage of the peptide with mainly adenine moieties. The successive addition of dA<sub>20</sub> (to remove dT<sub>20</sub> in the form of a duplex) and dC<sub>20</sub> triggers a significant rearrangement of nucleobases on the peptide backbone. Adapted with permission from ref 671. Copyright 2009 American Association for the Advancement of Science.

somehow approach the behavior of nucleic acid and peptide replicators

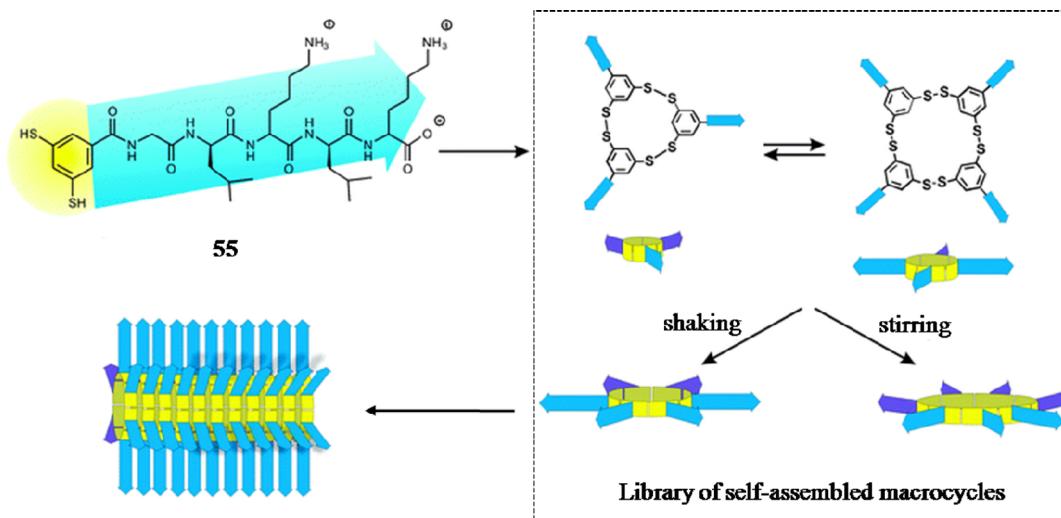
Dynamic combinatorial approaches have actually been applied to systems based on peptides and nucleic acids, mainly for the discovery of new ligands that can bind to such biopolymers, eventually useful as drugs with biotechnological or clinical applications. The search for efficient peptide–ligand interactions, for instance, faces several problems such as the lack of precise knowledge on the folding mechanisms of many proteins. DCLs offer possible solutions to this problem, since they respond to subtle interactions without the need for a previous understanding of those interactions. Moreover, the selection process has been performed in some case in combination with high-throughput screening, for example, with imine libraries potentially containing over  $4 \times 10^4$  members derived from neuraminidase inhibitors.<sup>863,864</sup> Only two members were amplified in selection experiments against the neuraminidase enzyme, one of them shown to be a good inhibitor of its catalytic activity. DCC has also allowed researchers to find good ligands for duplex and quadruplex DNA structures,<sup>865–868</sup> and this approach has been used for enhancing nucleic acid–nucleic acid recognition events.<sup>869</sup> This body of work, which might have important implications for the chemistry of primordial RNA and peptide oligomers, has been extensively reviewed in the literature.<sup>822–825</sup>

With peptides, DCLs have been explored even further, aiming at understanding the outcome of external stimuli on libraries where reversible chemistry is coupled to self-assembly processes. Several examples have been reported involving hydrazone, disulfide, and thioester exchanges, as well as metal–ligand coordination complexes.<sup>870</sup> In turn, there are several self-assembly motifs typical from peptides that might be useful coupled with the aforementioned reversible reactions: for example,  $\beta$ -hairpins, coiled-coils,  $\pi$ – $\beta$  assemblies, and  $\beta$ -sheets. One of the main objectives of this research area is the

development of dynamic and complex systems with unique emergent functions derived from their self-assembled structure.<sup>871</sup> Bio(nano)materials of this type would potentially adapt their properties to external stimuli, just as autonomous systems change their function in response to their environment.

As an example, disulfide redox chemistry based on the natural amino acid Cys can be easily employed to connect and disconnect self-templating peptide fragments, as Krishnan-Ghosh and Balasubramanian<sup>872</sup> have proved. In a library formed by short (4 aa) and long (10 aa) peptides, both able to form  $\beta$ -hairpins given that their sequences are based on Leu-Lys repeating units, the presence of a Cys residue at their respective N-terminal positions determined the formation of different self-assemblies depending on the redox conditions. This kind of processes, with bond formation being a result of disulfide bridges or other reversible reactions, may have been responsible for the ligation of primitive peptide oligomers, potentially leading to a great deal of dynamic diversity. Since the aforementioned discovery, the same principle has been used by other authors to probe, and in some case even switch,<sup>873,874</sup> the conformation of peptide assemblies with coiled-coil structures, using either disulfide<sup>873,874</sup> or thioester<sup>875–879</sup> chemistry.

Ghadiri and co-workers,<sup>671</sup> in turn, described an impressive example of reversible chemistry and self-assembly cooperating to produce sequence-adapting PNAs (see section 3.3.1) which evolve under an external stimulus. Their system is based on a PNA molecule bearing Cys in every second position. The Cys residues allow reversible introduction of different nucleobases by means of thioester bonds. In a series of experiments they demonstrated that addition of a DNA template results in the reorganization of nucleobases in the dynamic peptide analogue (Figure 40); that is, the system is able to rearrange Watson–Crick base pairs in order to adapt to a complementary template present in solution. Such adaptability somehow resembles one



**Figure 41.** Schematic representation of the process by which the peptide-functionalized dithiol (**55**) gives rise to a library of four self-assembled macrocycles. Although the trimer and tetramer form initially as major products, the ability of the hexamer and heptamer to assemble into columnar stacks held by  $\beta$ -sheet interactions determines their complete formation after a few days. The type of mechanical agitation is responsible for the major formation of hexamers or heptamers. Adapted with permission from ref 880. Copyright 2010 American Association for the Advancement of Science.

of the features of living systems: the selection of a nucleotide mutation in response to environmental pressure.

Another example of emergent properties arising from the combination of a reversible reaction and self-assembly has been described by the group of Otto and co-workers<sup>880</sup> (Figure 41). In this case, a library of macrocycles of various sizes is assembled from a peptide building block possessing thiol functionalities. Initially, the establishment of disulfide bridges leads to cyclic trimers and tetramers, but after several days, hexamers and heptamers raise their concentrations and become predominant, their relative proportion also depending on the mechanical stimuli applied to the library. The reason for such behavior is the ability of cyclic hexamers and heptamers to form  $\beta$ -sheet-mediated amyloid-like fibers, which in turn accelerate the synthesis of their own constituents. Therefore, the interplay of two reversible processes does not necessarily make a chemical system tend to equilibrium. Instead, it is possible for the system to escape from a thermodynamically controlled regime, for instance, by the emergence of kinetic self-replicating assemblies.

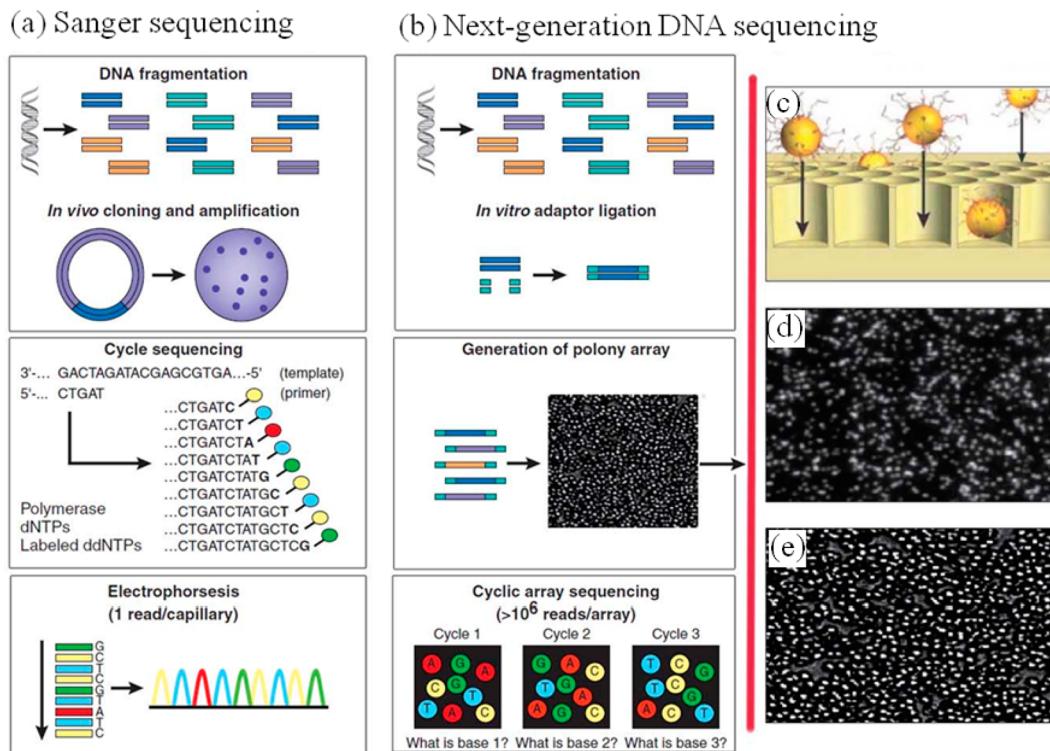
Finally, dynamic combinatorial self-reproducing micellar systems have also been described. By reaction of a lipophilic aldehyde with several distinct hydrophilic amines, Giuseppone and co-workers<sup>862,881</sup> were able to build up micelles composed of mixtures of the corresponding imine derivatives. By virtue of the reversibility of the imine bond, it was observed that the amphiphile leading to the most stable micelles was self-selected. Moreover, exponential micellar growth was observed due to speeding up of the lipophilic aldehyde incorporation into micelles from the water phase. On the basis of these results, it is reasonable to think of prebiotic micelles and vesicles (i.e., protocells) as dynamic combinatorial mixtures able to self-reproduce and select their own constituents.

## 5.2. High-Throughput Biochemical Techniques: Sequencing Technologies and Microarrays

As has been summarized in the sections above, the problem of the origins of life is related to the characterization of plausible synthetic pathways to biological monomers and their subsequent nonenzymatic polymerization. Mixtures of

oligomers of peptides and nucleic acids, or their analogous molecules, would be the result of these processes. A crucial methodological aspect of research on origins is, thus, the analysis of such complex mixtures. Additionally, highly diverse populations of nucleic acids, such as molecular quasispecies, generated by the error-prone replication mechanisms that were eventually operating during the RNA world (see sections 3.3.2 and 3.3.3) must be characterized. The technological revolution that has occurred in recent decades in the field of molecular biology can help tackle these issues.

The advent of novel techniques in biochemistry and molecular biology since the 1950s made possible the determination of the amino acid sequence of proteins, as well as the nucleotide sequence of fragments of cellular DNA genomes and those of their DNA- or RNA-based subcellular parasites. The first protein to be sequenced was bovine insulin, thanks to the pioneering work of Sanger and Thompson in 1951–1953.<sup>882</sup> After six decades of evolution, the two major direct methods for protein sequencing are the classical Edman degradation reaction and, currently, mass spectrometry (MS)-based techniques. In Edman's method, the amino acid placed at the amino terminus of a peptide is labeled, cleaved from the peptide without disrupting the bonds between the remaining amino acids, extracted into an organic solvent, and further identified by means of chromatography or electrophoresis.<sup>883</sup> Peptides up to 60 residues long can be sequenced, and the process can be automated<sup>884,885</sup> and combined with high-performance analytical technologies (reviewed in ref 886). Protein and peptide sequencing via MS can be achieved by means of two alternative ionization methods: electrospray ionization or matrix-assisted laser desorption/ionization (MALDI).<sup>887,888</sup> During the last two decades, the outstanding performance of MS-based techniques for protein characterization and for studying protein–protein interactions has been fundamental in the field of proteomics.<sup>889–891</sup> Nevertheless, for most current applications the sequence of a given protein is determined indirectly from the DNA or RNA sequence that encodes it, thus benefiting from the high-throughput nucleic acid sequencing techniques (see below). Regarding RNA, the



**Figure 42.** Workflow of (a) conventional, high-throughput shotgun Sanger sequencing technology versus (b) next-generation sequencing (NGS). In NGS, *in vitro* adaptor ligation and PCR-based amplification results in an array of millions of spatially immobilized PCR colonies or “polonies”, each of them consisting of many copies of a single shotgun library fragment. Such a polony array can be generated by different methods. (c) In pyrosequencing, the sequencing features are produced by emulsion PCR, with amplicons captured to the surface of 28- $\mu\text{m}$  beads. (d) In the sequencing-by-synthesis technique, a dense array of clonally amplified sequencing features is generated directly onto a surface by bridge PCR. (e) In the sequencing-by-ligation method, clonally amplified 1- $\mu\text{m}$  beads are used to generate a dense, disordered array of sequencing features. Later, the three NGS methodologies use successive iterations of enzymatic interrogation, by either a DNA polymerase or a DNA ligase, and imaging to build up a contiguous sequencing read for each array feature. Further details are given in the text. Adapted with permission from ref 901. Copyright 2008 Nature Publishing Group.

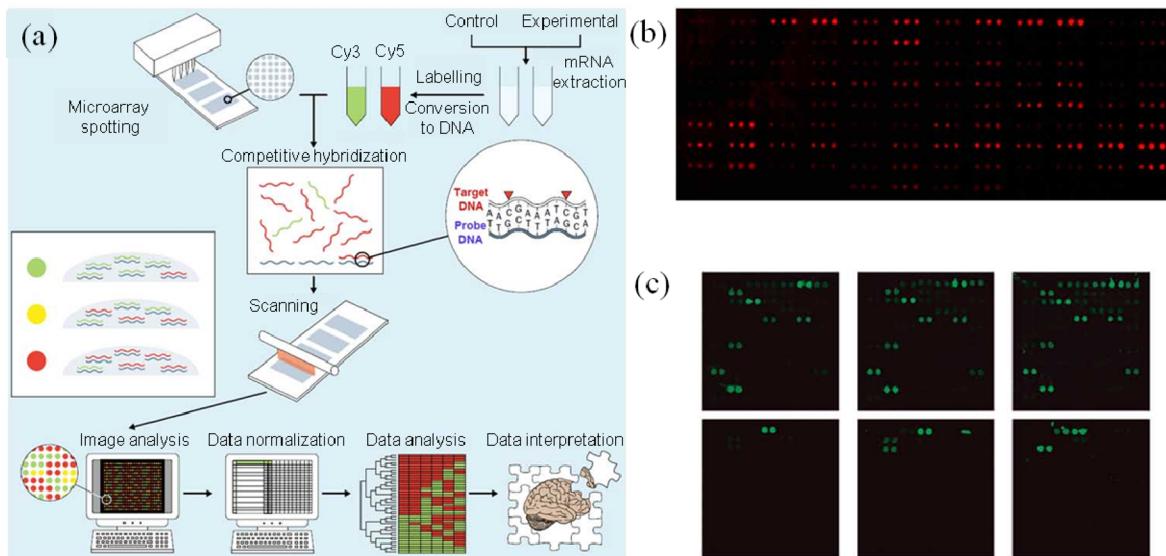
first sequenced molecule was the 77 ribonucleotide long alanine-tRNA from *Saccharomyces cerevisiae* in 1965.<sup>892</sup> Since then, RNA sequencing has run in parallel to the progressively automated DNA sequencing technology.

The first efficient DNA sequencing method was also developed by Sanger and Coulson in the 1970s.<sup>893</sup> It was based on the abortive synthesis of DNA strands complementary to the queried template, thanks to the use of 2',3'-dideoxynucleoside 5'-triphosphates or “dideoxes” as chain terminators, with the resulting fragments being resolved by gel electrophoresis (Figure 42a). An alternative system was published two years later by Maxam and Gilbert,<sup>894</sup> which relied on controlled chemical breakage of the DNA backbone at one or two of its four nucleotides. The first sequenced gene was that of the bacteriophage MS2 coat protein,<sup>895</sup> an RNA virus whose genome was also the first to be sequenced in its entirety.<sup>896</sup>

In the 1990s, the chain termination method for DNA sequencing gave rise to automatic DNA sequencers, based on fluorescent labeling and capillary electrophoresis, allowing read-lengths of up to 1000 nt. This high-throughput technology inaugurated the era of genome sequencing projects and, thus, the field of genomics.<sup>897</sup> The first cellular genome to be completed was that of the bacterium *Haemophilus influenzae* in 1995.<sup>898</sup> Currently, more than 8000 complete viral and cellular genomes have been already sequenced or are in progress. Advances in DNA sequencing and bioinformatics have also

allowed unveiling the “metagenome” of complex communities containing different cellular<sup>899</sup> or viral<sup>900</sup> species.

Since 2005, thanks to advances in microfluidics (see section 5.4) and nanotechnology, a number of “next-generation” DNA sequencing (NGS, also called second-generation or ultradeep) techniques have been developed, allowing an enormous increase in sequencing speed together with a drastic reduction in the cost per nucleotide (reviewed in refs 901 and 902). In all of them, an initial DNA library is prepared by random fragmentation of the molecule to be sequenced, followed by ligation of common adaptor sequences and enzymatic generation of clonally clustered amplicons by different variants of the PCR technique (Figure 42b). The first NGS platform was based on the so-called “pyrosequencing” method. It uses emulsion PCR and captures the amplicons (currently up to 800 nt long) onto the surface of microbeads, being able to sequence tens of millions nucleotides in a single run.<sup>903</sup> In turn, the “sequencing-by-synthesis” technique, released in 2006, relies on a bridge PCR primed by oligonucleotides tethered to a solid substrate. It can produce billions of nucleotides in a single run, with read-lengths up to 36 nt.<sup>904,905</sup> The third ultradeep method developed so far is known as “sequencing-by-ligation” because the reaction is driven by a DNA ligase rather than a polymerase. This method queries twice each sequenced nucleotide, thus reducing the error rate in each of the 35 nt long sequence reads.<sup>906</sup>



**Figure 43.** (a) Workflow of a typical microarray experiment designed to study differential gene expression profiles in control (labeled with the fluorochrome Cy3) and experimental (labeled with Cy5) cells. Adapted with permission from ref 923, copyright 2004 Elsevier. (b) Sample image obtained by use of DNA microarrays for characterization of native RNA structure in a viral genome. Adapted with permission from ref 925, copyright 2011 Elsevier. (c) Image of protein microarrays designed for identification and quantification of protein–peptide interactions. Adapted with permission from ref 924, copyright 2010 Nature Publishing Group.

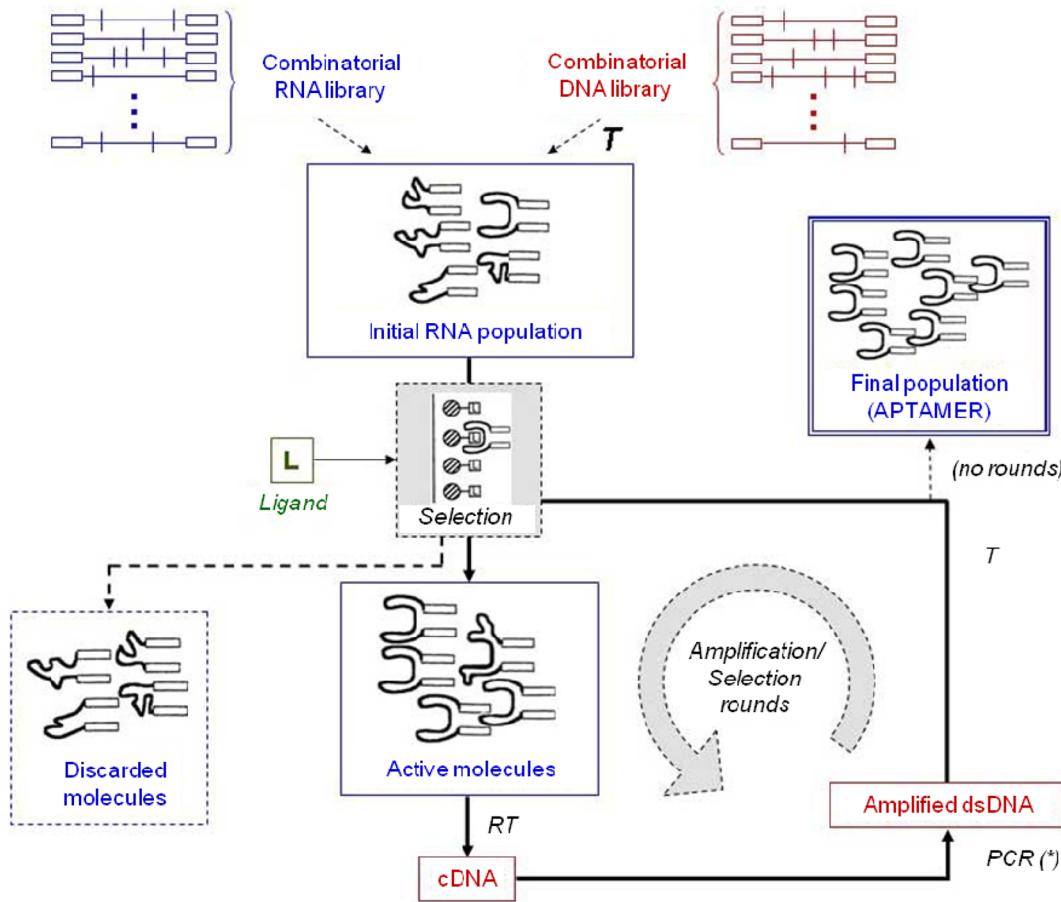
Next-generation sequencing is increasingly applied to different experimental approaches related to origins of life research. As an example, Deamer and co-workers<sup>274</sup> have used this technique for analysis of the molecular diversity generated by lipid-assisted, template-dependent RNA polymerization from nonactivated ribonucleotides under enzyme-free prebiotically plausible conditions. Additionally, NGS in combination with computational analysis has been applied to construction of the empirical fitness landscape for the *in vitro*-selected class II RNA ligase ribozyme, thanks to the genotype–phenotype mapping of  $\sim 10^7$  unique RNA genotypes.<sup>907</sup> Also relevant for this review, NGS is proving to be especially useful for identifying low-frequency genome variants in complex mixtures such as molecular or viral quasispecies (see section 3.3.3). As an example, pyrosequencing has been successfully applied to the detection of minority genomes in the human immunodeficiency virus type 1 (HIV-1) *pol* gene at proportions as low as 5%,<sup>908</sup> 2%,<sup>909</sup> or even 0.1%<sup>910</sup> of the quasispecies. In turn, parallel allele-specific sequencing has allowed the detection of HIV-1 variants present at levels ranging from 0.1% to 0.01% of the sample.<sup>911</sup> The performance of three NGS platforms has been recently analyzed by comparing the quality of their respective sequence reads and single-nucleotide polymorphism (SNP) detection performance.<sup>912</sup>

Currently, a novel family of “third-generation” techniques is being developed, allowing fast and reliable sequencing of individual nucleic acid molecules. Single DNA sequencing can be performed by electrophoretically driving molecules in solution through a nanoscale pore<sup>913,914</sup> as well as by stepwise enzymatic polymerization of individual DNA molecules onto surface-immobilized primer–template duplexes.<sup>915</sup> The growing family of third-generation methods will likely shape the future of sequencing because single-molecule approaches maximize cost effectiveness and do not require the nucleic acid to be previously amplified.<sup>916,917</sup> Therefore, advances in sequencing techniques offer an unprecedented opportunity to

characterize the diversity within any complex, heterogeneous population of nucleic acids.

In parallel with the revolution in sequencing, a rapid advance has also occurred in the technologies aimed at studying nucleic acid hybridization and protein–ligand interactions. The hybridization of two single-stranded molecules of DNA, RNA, or artificial nucleic acid analogues (see section 3.3.1) requires specific hydrogen-bonding interactions between their complementary nucleobases. In turn, protein–ligand bioaffinity (with the ligand being another protein, nucleic acid molecule, metabolite, or any other chemical species) is based on noncovalent interactions of different types: electrostatic interactions, hydrophobic interactions, van der Waals forces,  $\pi$ – $\pi$  stacking interactions, and H-bonds. Typical examples of protein–ligand interactions relevant in biology include enzyme–substrate binding and antibody–antigen recognition.

These kinds of biomolecular interactions were initially monitored by means of membrane-based blot methods,<sup>918,919</sup> useful to identify the main parameters involved in the biorecognition processes: temperature, ionic strength, concentration of divalent cations, pH, and presence of denaturing agents. This heritage, together with the advances in chemistry and nanotechnology aimed at immobilizing different kinds of molecules onto solid surfaces, allowed the development of high-performance biosensors. A biosensor is an analytical device that contains a biological receptor or probe, which interacts with the target molecule, and a signal transducer that detects the molecular interaction event and translates it into a measurable signal.<sup>920</sup> The two main families of biosensors currently used are based on bioaffinity and biocatalytic processes, involving different kinds of probes: proteins, natural or artificial nucleic acids, carbohydrates, combinations of these biomolecules, macromolecular assemblies, and even whole cells or fragments of tissues.<sup>921</sup> Regarding transduction systems, current biosensors benefit from the sensitivity and specificity offered by distinct optical, electrochemical, electrical, mechanical, acoustic, or thermal methods.<sup>922</sup>



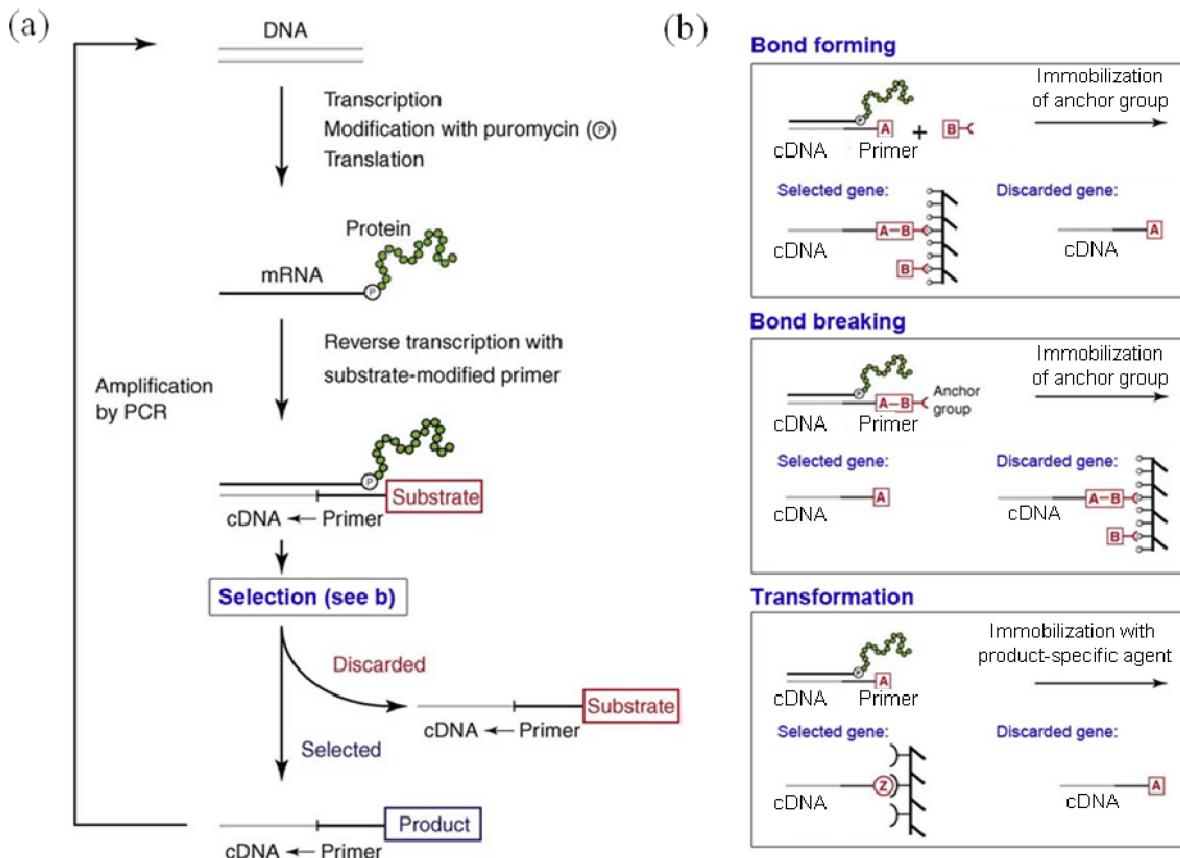
**Figure 44.** Workflow of systematic evolution of ligands by exponential enrichment (SELEX) applied, in this example, to the selection of RNA aptamers specific for a target or ligand molecule (*L*). Two alternative starting RNA and DNA combinatorial libraries are shown, and the primer-binding sites are depicted by boxes in every molecule. The enzymatic steps of each of the *n* amplification/selection rounds (no. rounds) are reverse transcription (RT), polymerase chain reaction (PCR) amplification, and in vitro transcription (T). The asterisk denotes that the PCR (as well as, eventually, other steps) is mutagenic when in vitro evolution rather than in vitro selection is performed. Further details are given in the text.

In the field of biosensors, a rapid advance has been produced in the so-called “microarrays” or “biochips”, which are based on the possibility of covalently immobilizing thousands of probe molecules (e.g., nucleic acids, proteins, etc.) onto a solid substrate (chemically modified glass, silicon, gold, etc) (Figure 43).<sup>923–925</sup> The probe molecules are arranged in miniaturized bidimensional arrays of dots, typically 10–150  $\mu\text{m}$  in diameter. The sample to be analyzed is fluorescently labeled and hybridized to the microarray, and the specific target–probe interactions are detected by means of a high-resolution scanner. Microarray techniques provided the possibility to perform high-throughput analyses, dramatically increasing the speed and performance of experimental work in genomics and proteomics.<sup>926,927</sup>

Nucleic acid microarray technology was initiated in the 1990s<sup>928</sup> and allows the production of biochips following two alternative strategies: *in situ* synthesis of up to millions of short oligonucleotide probes by photolithographic technology, or mechanical deposition of hundreds or thousands of probe molecules onto a solid support by printing of presynthesized nucleic acid molecules.<sup>929,930</sup> DNA microarrays are currently used in different fields of genomics, including (i) genotyping and detection of single-nucleotide polymorphisms (SNPs) in certain genes,<sup>931,932</sup> (ii) “transcriptomics” or study of gene expression profiles in an organism<sup>933,934</sup> or in complex

environmental samples,<sup>935</sup> (iii) DNA sequencing,<sup>936</sup> (iv) phylogenetic analysis of microbial communities,<sup>937</sup> (v) subtyping of microorganisms,<sup>937–939</sup> and (vi) determination of genomic RNA structure in viral genomes.<sup>925,940</sup> RNA microarrays have also been developed to monitor the interactions of RNA probes with different RNA-binding molecules.<sup>941</sup> In turn, microarrays containing oligomeric PNA (see section 3.3.1) probes have been constructed, taking advantage of the superior hybridization features and improved chemical and enzymatic stability of PNA relative to natural nucleic acids.<sup>942,943</sup> Thus, microarrays provide a high-throughput platform for investigation of possible molecular interactions between nucleic acids and/or analogue polymers in a prebiotic scenario. Regarding the genotyping of complex molecular populations such as viral quasispecies, microarray-based technology allows the detection of low-frequency genome variants within the mutant spectrum at frequencies as low as 1%,<sup>944</sup> or even 0.1% when microarray technology is combined with MS.<sup>945,946</sup>

In turn, protein microarrays appeared once the printing technology was developed for DNA biochips and have become a relevant technology in the field of proteomics. They are currently applied to the study of the interactions of proteins with other proteins, nucleic acids, ligands, and inhibitors.<sup>947–949</sup> Among the most specific and sensitive bioaffinity-based protein



**Figure 45.** Workflow of enzyme selection by mRNA display. (a) A DNA library, either chemically synthesized or of genomic origin, is transcribed into mRNA. Then a DNA oligonucleotide containing the antibiotic puromycin is attached to the 3'-end of each mRNA. During the subsequent in vitro translation, ribosome synthesizes the mRNA-encoded protein until it stalls at the RNA–DNA junction. Before the ribosome dissociates from the complex, it covalently joins the puromycin to the protein chain, thereby generating mRNA–protein fusions. This library is reverse-transcribed with a substrate-modified primer, thus attaching the substrate to the cDNA/RNA/protein complex. Proteins that catalyze the reaction of the substrate modify their encoding complementary DNA (cDNA) with the product. Selected cDNA sequences are amplified by PCR and used as input for further rounds of selection (or evolution, if mutagenic PCR is used), while the remaining genes are discarded. (b) Three examples of selection strategies followed to isolate novel enzymes by mRNA display. Adapted with permission from ref 965. Copyright 2010 Elsevier.

microarrays are those that use antibodies as capture probes to detect the presence of an antigenic molecule in a sample, a methodology with relevant applications in biotechnology and biomedicine.<sup>950,951</sup>

Other technological achievements in the field of microarray technology include the use of different chemically modified substrates with improved performance,<sup>929</sup> combination of microarrays with microfluidic technologies (see section 5.4),<sup>952</sup> setup of hydrogel three-dimensional microarrays on plastic substrates,<sup>953</sup> functionalization of microarray surfaces with dendrimeric structures,<sup>954</sup> and development of “nanoarrays” with spot sizes of 100 nm spaced 1 μm apart, a range of dimensions that allows densities of up to 100 million spots/cm<sup>2</sup>.<sup>955,956</sup> Nevertheless, despite their broad usefulness in biotechnology, microarray-based methods face some technical limitations, mainly related to the need for fluorescent labeling of the sample to be analyzed. Therefore, during the past decade several families of alternative, nanotechnology-inspired detection techniques have been developed in order to produce label-free biosensors. They are either sensitive to mass increases, such as those occurring upon nucleic acid hybridization or protein–ligand interaction, or to certain chemical signatures of specific molecular groups present in the target molecule (comprehensive reviews can be found in refs 957–959).

Thus, current nucleic acid- and protein-based microarrays and other high-throughput biosensors are well-suited to monitor the interactions among biomolecules in a systems chemistry-based investigation of precellular or acellular worlds.

### 5.3. In Vitro Evolution of Nucleic Acids and Other Biomolecules

In vitro evolution of nucleic acids allows the screening of large combinatorial libraries of RNA or DNA molecules for a specific function, such as catalysis or target-binding.<sup>706,707</sup> These cell-free evolution experiments have led to the selection of novel ribozymes (see section 3.3.2), artificial DNA enzymes (called deoxyribozymes or DNAzymes), and target-binding nucleic acid molecules (i.e., aptamers). The iterative in vitro process involves serial cycles of amplification and selection of the functional nucleic acid until the desired activity is reached. The outcome of directed evolution can be monitored through analysis of the genotype and phenotype of the evolved molecules.

The starting point of a stepwise in vitro evolution experiment is a large and complex pool of nucleic acid molecules (in general, 10<sup>10</sup>–10<sup>16</sup>) with a central region of random or mutagenized sequence (usually 15–100 nt long) flanked by two primer-binding sites of defined sequence required for the amplification steps. In vitro evolution of RNA involves three

enzyme-catalyzed steps: reverse transcription (i.e., conversion of RNA into DNA), DNA amplification by PCR, and transcription of a new RNA pool (Figure 44). In vitro evolution of DNA is experimentally easier to perform, since amplification requires only one step of PCR. The successive amplification-selection rounds increase the ratio of active versus inactive nucleic acid, and the pool finally becomes dominated by molecules with the desired phenotype. The whole cell-free process occurs in vitro, and all the experimental variables can be controlled and fine-tuned. In vitro evolution is distinguished from in vitro selection in that the former includes continuous introduction of genetic variation in the molecular population, through mutation (usually by error-prone PCR) and/or recombination.<sup>709,960</sup> A further step in experimental directed evolution was the development of a system for continuous in vitro evolution of nucleic acids, which, in contrast to the stepwise protocol, combines the processes of selection, amplification, and genetic diversification within the same reaction mixture.<sup>961,962</sup> This is particularly relevant here, because continuous in vitro evolution allows mimicking some of the evolutionary processes and transitions that might have operated in replicating RNA populations in a hypothetical precellular world.<sup>749</sup>

Related methods have been optimized for the selection and evolution of other biomolecules, such as peptides and proteins. The technology known as mRNA display relies on the covalent binding of proteins to their encoding mRNA, thus allowing the selection of functional proteins from an in vitro translated protein library of at least  $10^{12}$  distinct sequences (Figure 45).<sup>963</sup> In a relevant example, this technology has been used for the isolation of novel RNA ligase enzymes from a partially randomized noncatalytic scaffold, thus exemplifying the possibility of creating de novo new enzymatic activities following biochemical approaches alternative to the use of DCC (see section 5.1).<sup>964,965</sup> Also, mRNA display has allowed the in vitro selection of cyclic peptides that contain a majority of unnatural amino acids.<sup>966</sup> Recently, a different experimental system called phage-assisted continuous evolution (PACE) has been described that enables the continuous directed evolution of gene-encoded proteins, which are cell-to-cell transferred (the host being the bacterium *E. coli*) through a bacteriophage whose life cycle is dependent on the activity of interest.<sup>967</sup> These experimental approaches are very relevant in the field of (bio)molecular evolution and can be applied to the study of otherwise intractable questions related to the evolution of catalytic functions of peptides or proteins in a precellular context.

In vitro selection of nucleic acids has been fundamental in the field of the origins and early evolution of life, in particular for supporting the RNA world hypothesis, thanks to the selection of novel ribozyme activities (section 3.3.2). Additionally, this methodology has produced several families of aptamers with different biotechnological applications. Targets specifically recognized by RNA or DNA aptamers cover a wide range of size and molecular complexity, including simple ions, small molecules (e.g., synthetic organic molecules, amino acids, nucleotides, monosaccharides, lipids, or metabolites), polysaccharides, peptides, proteins, nucleic acids, macromolecular assemblies, micelles and vesicles, viruses, organelles, or even whole cells.<sup>968–971</sup> In general, after 6–15 rounds of in vitro selection or evolution, the sensitivity and specificity of an aptamer for its target molecule (present in a homogeneous solution, in heterogeneous mixtures, or even in complex

biological samples) rival those of a monoclonal antibody for its antigen. These reasons, together with their cost effectiveness, make aptamers very useful tools in biotechnology with increasing applications in biosensing, diagnostics, and therapy.<sup>972</sup> Although aptamers are artificial molecules, riboswitches are considered as “natural aptamers” embedded in mRNAs, since they sense small effector molecules and regulate the expression of the RNA-encoded gene.<sup>973–975</sup> Recently, the unveiled ability of the nucleic acid analogues PNA,<sup>663</sup> TNA,<sup>409</sup> and HNA<sup>651</sup> (see section 3.3.1) to fold into three-dimensional structures allowed the design of an in vitro evolution protocol of XNA aptamers, some of which bind to their respective ligands with an affinity similar to that of previously evolved RNA aptamers.

In addition to nucleic acid aptamers, peptide aptamers have been obtained from randomized peptide libraries via different modifications of in vitro selection procedures<sup>976,977</sup> and show relevant applications in analytical chemistry.<sup>978</sup> Allosteric nucleic acid enzymes that fuse one ribozyme or DNAzyme and one aptamer are known as “aptazymes”, whose applications in biotechnology are increasingly recognized.<sup>979</sup> Recently, the SELEX methodology has been employed to improve the selection capability of DCLs (see section 5.1). The combination of DCC and SELEX allows introduction of an amplification step after a binding ligand is selected from the library.<sup>980</sup> Nucleic acids, and in particular DNA, can alternatively be utilized to encode small molecules, in the so-called encoded self-assembling chemical libraries (ESAC),<sup>981–985</sup> a methodology that may help to deal with large DCLs in the near future. Microfluidic chip technology has also been incorporated into SELEX-based methods, in order to automate the system and to monitor in real time the population size, composition, and complexity, as well as the performance of the evolved molecules.<sup>986,987</sup>

The applications reviewed in the paragraphs above show the potentiality of molecular evolution in a noncellular environment. An analogous type of evolution probably shaped how nucleic acids and peptides interacted with each other in the prebiotic world, as well as with other small biomolecule precursors present in the heterogeneous medium. All the novel biochemical technologies and in vitro evolution strategies described will certainly help to unravel the meaning and limits of chemical evolution, but many of them are still experimentally complicated. In this respect, big efforts are being devoted to automate and miniaturize them by means of micro- and nanofluidic approaches.

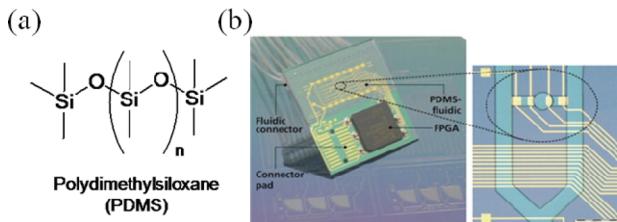
#### 5.4. Microfluidic and Nanofluidic Approaches

The ability to precisely and reproducibly manipulate (bio)molecules, together with the availability of methods to actuate fluids at the microscale, fueled the development of the field of microfluidics.<sup>988</sup> Over the last two decades, microfluidics has become a very useful tool for chemists and biologists, since it allows replication of laboratory benchtop technology on a miniature chip-scale device. Integrated microfluidic systems provide many advantages in analytical chemistry and biotechnology, including extremely low reagent consumption (i.e., in the subnanoliter range), precise spatial and temporal control of processes, and fine-tuning of all the experimental variables involved. Microfluidic technology currently provides a wide range of platforms for high-throughput combinatorial drug screening, novel biomaterials synthesis, biosensing, microscale

genomic and proteomic analyses, as well as single-molecule manipulation.

In the early 1990s, the first microfabrication techniques were focused on the setup of capillary electrophoresis on a chip by glass etching, a useful technology with a number of relevant applications (reviewed in ref 989). In particular, high-throughput DNA sequencing and other genotyping methods based on microchannel electrophoretic separations<sup>990</sup> or flow chambers that include a microwell-containing fiber optic slide were developed.<sup>993</sup> This promoted the complete transformation of sequencing system platforms to a microfabricated chip format. The availability of micrometric channels and micro-reactor compartments allowed the construction of different lab-on-a-chip devices for rapid and parallel testing of chemical and/or biological samples.<sup>991–993</sup>

The next milestone in this field was the development of soft lithography methods thanks to the use of poly(dimethylsiloxane) (PDMS), a polymeric silicon compound with unique physicochemical properties (Figure 46a).<sup>994</sup> These



**Figure 46.** (a) Chemical structure of PDMS, a silicon-based polymer widely used in photolithography for the fabrication of micrometer-size patterns. (b) Image of a PDMS-based microfluidic device with compartmentalized, electronically programmable membranes, which allow control of the transport and concentration of DNA and other biomolecules within a network of channels and microreactors. Adapted with permission from ref 1000. Copyright 2008 Elsevier.

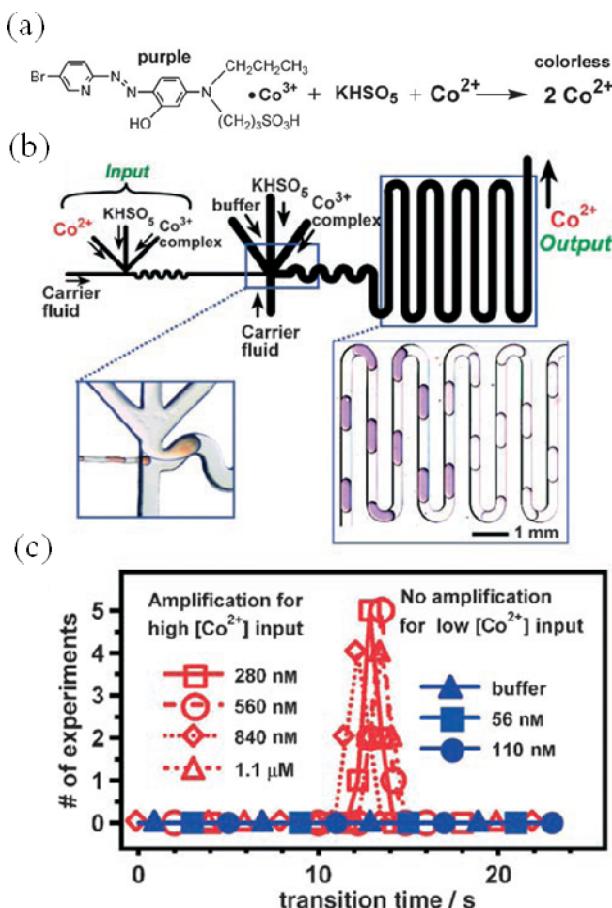
methods start with the design and photolithographic fabrication onto a silicon wafer of a prototype containing the desired micrometer-size patterns. Then PDMS is poured over the wafer and left to harden. When removed, all the micrometric details of the pattern stay imprinted in the PDMS, thus creating an elastic shape or “stamp”. Upon hydrophilic surface modification of the PDMS block, a piece of glass slide, or alternatively a solid polymeric surface, is placed on the activated, imprint-containing side of the PDMS stamp. Once the bonds relax to their normal state, the solid surface becomes permanently sealed to the PDMS, thus creating a waterproof network of channels with the desired geometry.<sup>995,996</sup>

The typical microchannel diameter of PDMS-based devices is 10–100  $\mu\text{m}$ , thus fitting the size of bacterial (i.e., 1–2  $\mu\text{m}$ ) or eukaryotic (i.e., 10–20  $\mu\text{m}$ ) cells and allowing control of the cellular microenvironment at relevant length and time scales. This is the main reason why microfluidics has gradually moved from the fields of chemistry and materials science<sup>997</sup> to those of biotechnology, biomedicine,<sup>998</sup> and even microbial ecology.<sup>999</sup> Especially relevant for the systems approach embraced in this review, PDMS-based microfluidics allows the fabrication of compartments containing electronically programmable membranes that can be used to control the transport and concentration of DNA and other biomolecules within a network of channels and microreactors (Figure 46b).<sup>1000</sup> The techniques to generate populations of stable vesicles with well-defined, monodisperse sizes in microfluidic devices are in fact

being optimized, by use of the double-emulsion method<sup>1001</sup> or controlled solvent extraction processes.<sup>1002</sup> In this context, for instance, the behavior of replicating combinatorial molecules in a ternary fluid of amphiphiles, hydrocarbons, and water has been investigated with the purpose of determining eventual transitions from independent replicators toward cell-like collective structures.<sup>1003</sup> As another example, the cell-free expression of GFP has been demonstrated within microfluidically produced lipid vesicles.<sup>1002</sup>

In parallel, advances in the understanding of self-organization and self-assembly of molecules, as well as the bottom-up rationale that led to the advent of nanotechnology, allowed the construction of the first nanofluidic systems in the decade of the 2000s.<sup>1004</sup> The nanometric dimensions of the channels and reactors within these devices are comparable to the size of nucleic acids, proteins, and other macromolecules, thus showing broad applicability in different fields of chemistry, molecular evolution, and biotechnology, including chemical analysis and nanopore-mediated separation<sup>1005</sup> and DNA manipulation and sorting.<sup>1006</sup> Noticeably, nanofluidics has been used for the investigation of synthetic or semisynthetic enclosed lipid bilayer structures analogous to biological nanocompartments. As an example, surface-immobilized vesicles with controlled geometry, membrane composition, and interior contents were produced, and fluid-state lipid bilayer networks of nanotubes have been used to control the materials exchange between conjugated containers.<sup>1007</sup> It has also been shown that lipid vesicles with diameters of tens of nanometers, as well as other nanometer-sized objects such as polymer beads and gold nanoparticles, can be trapped and manipulated without external intervention thanks to the electrostatic properties of nanofluidic systems.<sup>1008</sup> This paved the way to contact-free confinement of single biomolecules by methods such as ion concentration polarization.<sup>1009</sup> Therefore, these methodologies are suitable for monitoring chemical reactions in artificial biomimetic compartments, with applicability in the fields of analytical chemistry, biotechnology and synthetic biology, as well as in experimental research with protocells.<sup>1010</sup>

Finally, the past decade has seen the emergence of droplet-based microfluidic systems, which allow miniaturization of reactions through their compartmentalization inside plugs (i.e., droplets large enough to block the channel) of femtoliter to microliter volumes.<sup>1011,1012</sup> These plugs are produced by appropriately injecting streams of two immiscible liquids into a central microchannel. Moreover, there are several methods to load the plugs with reagents. Then, it is in principle possible to control multiple chemical reactions on the millisecond time scale by creating a network of convergent and divergent channels, and by adapting flow velocities to adjust mixing times.<sup>1013</sup> In this respect, microfluidic droplets provide an opportunity to compartmentalize reaction networks, so each reaction occurs at the right place and the right time, and to maintain them away from equilibrium through the supply of reagents and removal of products. An example of a chemical network showing reaction rate amplification with a threshold response is presented in Figure 47.<sup>1014</sup> The system is based on a two-stage process autocatalytic in  $\text{Co}^{2+}$ , the key being that the output of the first stage is the input of the second stage. The reagents are encapsulated in aqueous plugs carried by the water-immiscible perfluorodecaline, and only when the  $\text{Co}^{2+}$  concentration in the plugs exceeds a threshold value is the reaction rate amplified. Extending this kind of studies to more



**Figure 47.** Droplet-based microfluidic system with rate amplification of a reaction network. (a) Autocatalytic decomposition of a  $\text{Co}^{3+}$  complex, in the presence of  $\text{KHSO}_5$ , into the autocatalyst  $\text{Co}^{2+}$ . (b) Schematic representation of microfluidic system for two-stage amplification. The first and second stages of the reaction occur in the thinner and thicker channels, respectively. Left micrograph: merging junction of thin and thick channels. Right micrograph: abrupt change in color of the plugs containing the autocatalytic reaction when the purple  $\text{Co}^{3+}$  complex converts into colorless  $\text{Co}^{2+}$ . (c) Transition times for various concentration inputs of  $\text{Co}^{2+}$ . Below a threshold initial concentration (110 nM), there is not enough time for the formation of  $\text{Co}^{2+}$  in the plugs along the thin channels (i.e., in the first stage), so no response is observed in the second stage (blue symbols). On the contrary, above a threshold initial concentration of  $\text{Co}^{2+}$  (280 nM), enough  $\text{Co}^{2+}$  ions are produced in the first stage, and therefore the reaction can be triggered in the second stage (response shown with red symbols). Reproduced with permission from ref 1014. Copyright 2004 American Chemical Society.

complex chemical and biochemical networks may help understand important biological functions relevant to the origins of life. Droplet-based microfluidic approaches have also been developed to perform cell studies, PCR, enzyme kinetics, protein crystallizations, hemostasis models, small-molecule and polymer syntheses, and gel particle preparations.<sup>1011,1012</sup>

In summary, micro- and nanofluidics provide very valuable tools to investigate the behavior of protocellular compartments, biomolecules, and reaction networks in a systems chemistry context.

## 6. CONCLUDING REMARKS AND OPEN QUESTIONS

This review shows the wide variety of perspectives from which we can look at such a controversial topic as the origins of life.

Although finding consensus about the nature and definition of life is a very difficult issue, and will remain as a subject of debate probably for a long time, there is nowadays relatively widespread agreement on which features should be shared by the simplest living systems. They must possess a genetic apparatus able to store and transmit information to their progeny, some sort of metabolism for gathering nutrients and energy from the environment, and a selectively permeable boundary that separates and distinguishes them from that environment. Hence, in order to explain how the first organisms might have appeared on Earth, or elsewhere, it is necessary to develop chemistries that enable the synthesis of information-bearing polymers, protometabolic networks, and protocellular compartments under compatible prebiotic conditions. Moreover, there is the need for finding thermodynamically and kinetically plausible pathways to integrate the three kinds of subsystems into far-from-equilibrium, autonomous agents with open-ended evolution capacities.

In this direction, the first step would be the synthesis of the monomeric units that constitute peptides, nucleic acids or their analogue polymers, and vesicle membranes. Some of these building blocks could have been delivered to Earth from extraterrestrial sources, while their endogenous synthesis on our planet is another feasible option. Traditional synthetic approaches have been shown to be valuable for obtaining amino acids, nucleobases, lipidlike molecules, sugars, and certain sources of inorganic phosphate. Such synthetic routes still face, however, numerous problems. Indeed, most of them are thermodynamically uphill, giving rise to low yields and poor selectivity for many transformations of interest.

In contrast to this view, novel systemic approaches have recently allowed researchers to obtain some of the building blocks of life in much more efficient ways. Catalysis, in particular organocatalysis and organometallic catalysis, could have enlarged and diversified the set of compounds potentially present in different prebiotic scenarios. It is now known that amino acids can assist the synthesis of sugars; aldehydes catalyze the formation of amino acids and drive the synthesis of some nucleobases; formamide is crucial in the linkage of nucleobases with sugars to produce nucleosides as well as in nucleoside phosphorylation; etc. Inorganic species such as phosphates, silicates and borates, on the other hand, also play a relevant role by stabilizing molecular intermediates, preventing their degradation and even catalyzing other important reactions. This kind of processes, and many others still waiting for discovery, might have been operating at the same time and in the same place, if we assume the plausible existence of complex molecular mixtures on the early Earth.

The same would apply for polymerization reactions and for the assembly of protocellular compartments. It has been demonstrated that polymerization of amino acids and nucleotides can be assisted by mineral (e.g., montmorillonite clay) or metal surfaces, lipid membranes, eutectic phases in water–ice, and small cofactor molecules. Vesicles composed of mixtures of simple amphiphilic molecules and surfactants, on the other hand, have shown the required stability and permeability for encapsulation of nucleic acids and peptides. In this scenario of high molecular heterogeneity, some authors even suggest peptide–lipid, nucleolipid, and nucleopeptide worlds as ancient chemical relics of the current, extremely intricate relationships between peptide, nucleic acid, and lipid biochemistries.

In all these synthetic processes, an important question is whether the current homochirality of biomolecules could have been prebiotically originated or, on the contrary, could be the result of biological evolution. In favor of the former hypothesis, there exists a wide variety of possible physicochemical phenomena able to generate chiral molecules from nonchiral matter. The fundamental question is then determining the mechanism or set of mechanisms that are ultimately responsible for the emergence of homochirality. Among the various options, small chiral biases can be generated through deterministic effects, such as the action of circularly polarized light, the parity violation principle, the Earth's rotation, and the asymmetry of quartz crystals. The amplification of chirality from such minor enantiomeric excesses is, however, a difficult issue that cannot be explained by taking into account only deterministic argumentations. In contrast, there are a number of theories based on stochastic phenomena that can also lead to mirror symmetry breaking. All theories within this group imply systems that are far from equilibrium and are therefore more interesting for the purpose of studying living entities, provided that flows of matter and/or energy are continuously present. Energy flows are required as the driving force for self-organization processes (i.e., crystallization, polymerization, supramolecular self-assembly, etc.) where chiral dissipative structures are formed. If one of these structures catalyzes its own formation and inhibits the formation of the structure with opposite handedness, any stochastic chiral fluctuation occurring within the system would be amplified. This behavior is especially efficient in systems where autocatalysis is embodied in a complex network of reactions.

As a result of the presence of autocatalytic reaction networks, some chemical processes known as oscillating reactions are able to develop spatial and temporal patterns of order. In biology, oscillatory behaviors are generic mechanisms behind complex functions such as metabolic and cell division cycles, growth of bacterial colonies, morphogenesis in multicellular organisms, neuronal signal transmission, and circadian rhythms. It then seems reasonable to think that oscillating reactions may have been behind the generation of functional features of protocellular organisms, although they have not been found yet for molecules of prebiotic relevance. In any case, their nature as autocatalytic reaction networks also points out the relevance of these two concepts (autocatalysis and reaction network) for increasing complexity in chemical systems.

Autocatalysis, in particular, is considered crucial in all scenarios for the origins of life, regardless of whether these are based on gene-first or metabolism-first premises. In addition to its probable role in the origin of homochirality and in oscillatory reactions, autocatalysis is a fundamental concept from the point of view of any system that ought to maintain, grow, and produce more of itself and more of its kind: that is, to account for some of the main properties of life. Only when the genetic, metabolic, and boundary components of protocells acquired the coordinated capacity to self-sustain, multiply in number, and transmit their structural and functional features to an offspring could Darwinian evolution have started operating. It therefore seems that autocatalysis, in its different versions (i.e., related to metabolic networks, vesicles, information-bearing polymers, or to their collective coupling) probably paved the way for the transition from chemistry to biology.

One of these versions is cyclic autocatalysis, in which there is net production of at least one intermediate species within a closed cycle of chemical transformations (e.g., in the formose

reaction or in metabolic pathways). This kind of autocatalysis may be intimately related to the establishment of proto-metabolic networks: it does not necessarily imply some form of catalysis for each and every step of the cycle but achieves self-sustenance by stoichiometric means. Most probably, autocatalytic networks of this kind may have been assisted, at the very beginning, by mineral or organic catalysts. Then, only after evolutionary mechanisms entered the scene would primitive enzymes have overtaken the catalytic role. Much experimental and computational research is, however, needed to explore these issues in the coming years.

The reproduction of vesicles made of fatty acid derivatives and other simple, potentially prebiotic amphiphiles has also been demonstrated, although it is not a trivial issue. Besides mechanical extrusion, this process generally occurs through growth and subsequent division, when vesicles reach sizes at which they become unstable. Vesicle growth can be induced by the addition of more amphiphilic molecules, in the form of micelles, or by generating the amphiphile *in situ* from a precursor. Interestingly, this type of vesicle growth has been shown to be autocatalytic under certain conditions, when the incorporation of new amphiphile molecules is accelerated by the presence of preformed vesicles. Moreover, increasingly detailed studies are being carried out on the kinetics of vesicle growth and eventual self-reproduction, as well as on the influence of physical parameters (e.g., osmotic tension, initial size, fluidity, bilayer lipid uptake/release coefficients composition, lamellarity, overall charge, permeability), or additional properties, like the ability to generate pH gradients. The results of these studies open new avenues of research concerning the self-reproduction of vesicles that could contain oligonucleotides, oligopeptides, small cofactor molecules, or mixtures of these, either in their lumen or in contact with their membrane.

The third type of autocatalysis relevant to the origins of life is template-directed molecular autocatalysis, where one of the products of a reaction catalyzes its own formation. This notion, which lies at the heart of genetic mechanisms for processing and transmitting information, has been proved in minimal systems with oligonucleotides, oligopeptides, and small synthetic molecules. In slightly more complex sets of molecules, composed of a few oligonucleotide or oligopeptide replicators, it has been shown that the occurrence of cross-catalytic and cross-inhibition events leads to network behaviors. These results are of paramount importance to support and tinge the RNA world hypothesis, which is a very reasonable assumption as an intermediate stage in the process bridging prebiotic chemistry and DNA/RNA/protein-based life. However, such an RNA-based world could hardly have relied exclusively on the interaction dynamics of a population of macromolecular replicators in an (otherwise homogeneous) aqueous solution.

Although the RNA world hypothesis, mainly based on the capacity of RNA to play both roles of information storage and catalysis, is supported by much theoretical and experimental data (including the ongoing development of *in vitro*-evolved RNA polymerase ribozymes), it is unlikely that the first self-replicating ribozymes could have survived and evolved in bulk solution, given the heterogeneous chemical map of the prebiotic Earth. That is why the role of mineral surfaces, membrane compartments, small cofactor molecules, and metabolic networks is being considered as a new, key element of this hypothesis. Studies addressing the evolution of RNA virus quasispecies are also very important to understand the dynamics of RNA populations under conditions eventually

similar to those operating on our planet before the advent of cellular life.

Several approaches have tried to explore more complex systems *in vitro*, combining various of these fundamental components and their collective and dynamic properties: either nucleic acid templates in interaction with cofactors plus other metabolic/catalytic precursors, protocellular compartments encapsulating such templates, protometabolic cycles taking place within those compartments, or even all three infrabiological subsystems interacting with each other. The difficulty of carrying out this task from strict bottom-up assumptions has led many researchers to adopt semisynthetic strategies, in which part of the compounds and processes implemented are borrowed from extant living organisms (e.g., *E. coli* extracts). Those experiments are relevant as proofs of principle that complex chemistries can be carried out under simpler than *in vivo* boundary conditions. Nevertheless, it becomes very difficult to tell, under such work premises, their actual significance in evolutionary terms; that is, their relative order or position within the sequence of transitions that lead from physicochemical to biological systems. Working out this general evolutionary pathway is probably the most ambitious enterprise in the hands of the coming generation of systems chemists, who will hopefully be able to tackle it with the help of new technologies developed in the field.

In fact, as can be inferred from the increasing complexity described in this review, especially regarding the integration of all key components into minimal or model protocells, the ability to deal experimentally with such heterogeneous chemical and biochemical systems represents an outstanding challenge in itself. Scientists in this emerging field are, however, starting to foresee the maturity of a number of conceptual frameworks and technological developments of which prebiotic systems chemistry could certainly take advantage. First, since DCC deals with large libraries of interconvertible compounds that respond to external stimuli, this subfield provides a body of knowledge to understand and tame the combinatorial explosion that probably characterized chemistry on the prebiotic Earth. Second, high-throughput biochemical techniques such as sequencing technologies and microarrays are nowadays necessary to understand complex biochemical networks, and these techniques may assist investigations on the origins of life at the stage where populations of peptides and RNA oligomers have been prebiotically synthesized. In a further step, *in vitro* evolution of nucleic acids and other biomolecules is a very useful methodology to study the principles underlying evolution and adaptation at the molecular level (e.g., the functional evolution of information-bearing molecules). Finally, microfluidics has become an excellent tool for chemists and biologists, since it allows researchers to deal with extremely low amounts of material with high precision. Additionally, micro- and nanofluidics represent ideal methodologies to work with protocells and molecular assemblies in the relevant size ranges.

These new, increasingly integrative techniques, and others soon to appear, are opening up the field of chemistry to the study of complex systems and networks of molecules in interaction. However, such a move does not imply, by any means, forgetting about molecular details: chemistry will remain “the science of molecules and their transformations”, so it will be crucial to have precise knowledge about the properties of individual molecular entities and their one-to-one interactions. Much in the same way as systems biology is presently stepping forward, systems chemistry will need to cut

across different levels of description of the phenomenon under scrutiny, and use different experimental and theoretical methodologies to construct a coherent body of knowledge about that phenomenon.

On the basis of all these considerations, the present and future of research on the origins of life seems seeded with many new avenues to be explored. It is impressive to turn our view back to the first scientific ideas of Darwin on the origin of living beings “in some warm little pond”, to the hypotheses of Oparin and Haldane, or to the first experimental approaches of Miller and Urey, and realize how far this scientific community has come. The field of the origins of life is not surrounded anymore by a halo of religious or metaphysical speculations, as it used to be not so long ago, even among scientists. It is true that some questions about the topic are still deeply intriguing, and a number of them could remain unanswered forever, because they represent historical events that cannot be reproduced in the laboratory. However, many reasonable theories have already been put forward and, what is more important, these theories are subject to experimental verification through an increasing variety of approaches and techniques. Some of those hypotheses, which in the past seemed to be irreconcilable antagonists, have started to merge during the past decade, allowing us to become fully aware of the impressive mosaic of interconnected physical and chemical events that were probably behind the origins of life. Although chemistry operating on the prebiotic Earth must have been extraordinarily complex and heterogeneous, we believe it is not impossible to understand. A number of concepts and methodologies, developed over the past 30 years, are now mature enough to ensure a brilliant future for such an old and challenging endeavor of human beings: getting to know about their ancient origins from inert chemical matter.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail andres.delaescosura@uam.es.

### Author Contributions

<sup>†</sup>K.R.-M. and C.B. contributed equally to this paper.

### Notes

The authors declare no competing financial interest.

### Biographies



Kepa Ruiz-Mirazo, permanent researcher (former Ramón y Cajal fellow) of the University of the Basque Country (UPV/EHU), has a first degree in physics (solid-state physics), a master's degree in science education, and a Ph.D. in sciences (complex systems interdisciplinary

program). At present, he shares his academic time between the Department of Logic and Philosophy of Science, in Donostia–San Sebastian, and the Biophysics Unit (CSIC-UPV/EHU), in Leioa (Bilbao), developing both experimental and theoretical protocell models, in the wider context of a constant philosophical reflection on the problems of definition, origins, and evolution of life.



Carlos Briones is a staff scientist of the Spanish National Research Council (CSIC) in the Department of Molecular Evolution at the Centro de Astrobiología (CSIC-INTA, associated to the NASA Astrobiology Institute). He received his Ph.D. in chemistry (biochemistry and molecular biology) in 1997 at the Universidad Autónoma de Madrid. His research interests are mainly focused on the origins and early evolution of life, including experimental and theoretical approaches to the RNA world hypothesis, in vitro evolution of nucleic acids, and genetics of RNA viruses. He is also involved in the development of microarrays and bionanotechnology-inspired bio-sensors.



Andrés de la Escosura received his Ph.D. in chemistry in 2005 at the Universidad Autónoma de Madrid, working on molecular and supramolecular materials within the group of Tomás Torres. In 2006, he joined the group of Roeland J. M. Nolte and Jeroen J. L. M. Cornelissen, at the Radboud University Nijmegen, where he held a Marie Curie fellowship and explored the use of protein cages to build biohybrid materials. In 2009, he moved back to the Universidad Autónoma de Madrid, where he has held an Assistant Professor position and, since February 2012, a Ramón y Cajal fellowship. His main research interests are at the interface of chemistry and biology. He is currently working on the construction and study of biohybrid nanomaterials.

## ACKNOWLEDGMENTS

K.R.-M., Ramón y Cajal Research Fellow during the preparation of this paper, acknowledges support from the

Basque Government (Grant IT 590-13), Spanish Ministry of Science (MEC Grant FFI2011-25665) and COST Action CM0703 (Systems Chemistry). The work of C.B. at Centro de Astrobiología was supported by MICINN (Grants BIO2007-67523, EUI2008-00158, and BIO2010-20696), CSIC (Grant 200920I040) and FEDER Funds from the EU. A.d.l.E., Ramón y Cajal Research Fellow, received support from the Spanish MICINN and MEC (CTQ2011-24187/BQU and Consolider Ingenio 2010, CSD2007-00010 Nanociencia Molecular), and the Comunidad de Madrid (MADRISOLAR-2, S2009/PPQ/1533). We are also very grateful to Gonen Ashkenasy, Juli Peretó, Eörs Szathmáry, and Peter Walde, as well as to the editor of the journal and three anonymous reviewers, for reading the manuscript and providing useful comments and suggestions for improvement. This review is dedicated to the memory of our friend and colleague Dr. G. Christian Claessens, who regrettably passed away last June 2012.

## REFERENCES

- (1) Although there are more radical views on life, which argue for the possibility of living systems not necessarily based on aqueous chemistry (e.g., those coming from “strong” artificial life; see Langton C. G. In *Artificial Life I (Proceedings of the First Conference on Artificial Life)*; Langton, C. G., Ed.; Addison-Wesley: Redwood City, CA, 1989) we will take the more widespread position that life has an inherent chemical nature, which is obviously true for all life we know on Earth.
- (2) *Prebiotic Chemistry: From Simple Amphiphiles to Model Protocells*; Walde, P., Ed.; Topics in Current Chemistry, Vol. 259; Springer-Verlag: Berlin/Heidelberg, 2005.
- (3) Ludlow, R. F.; Otto, S. *Chem. Soc. Rev.* **2008**, *37*, 101.
- (4) The term “systems chemistry” was first used by von Kiedrowski to describe the chemical origins of biological organization. See Stankiewicz, J.; Eckardt, L. H. *Angew. Chem., Int. Ed.* **2006**, *45*, 342.
- (5) Woese, C. R.; Fox, G. E. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 5088.
- (6) Staley, J. T. *Curr. Opin. Biotechnol.* **2003**, *14*, 347.
- (7) Blumberg, B. S. *Philos. Trans.: Math. Phys. Eng. Sci.* **2011**, *369*, 508.
- (8) Brack, A. *Astrobiology* **2012**, *12*, 370.
- (9) Gayon, J.; Malaterre, C.; Morange, M.; Raulin-Cerceau, M.; Tirard, S. *Origins Life Evol. Biospheres* **2010**, *40* (2), 119.
- (10) Bedau, M. A.; Cleland, C. *The Nature of Life*; Cambridge University Press: Cambridge, U.K., 2010.
- (11) Ganti, T. *The Principles of Life*; Oxford University Press: Oxford, U.K., 2003.
- (12) Ruiz-Mirazo, K.; Pereto, J.; Moreno, A. *Origins Life Evol. Biospheres* **2004**, *34*, 323.
- (13) This palaeontological evidence comes from the analysis of microfossils (Schopf, J. W.; Packer, B. M. *Science* **1987**, *237*, 70.) and stromatolites (Allwood, A.C.; Walter, M.R.; Kamber, B.S.; Marshall, C.P.; Burch, I.W. *Nature* **2006**, *441*, 714.) dated roughly 3.5 billion years old.
- (14) Orgel, L. E. *PLoS Biol.* **2008**, *6*, e200.
- (15) Shapiro, R. Q. *Rev. Biol.* **2006**, *81*, 105.
- (16) Joyce, G. F.; Orgel, L. E. In *The RNA World*, 3rd ed.; Gesteland, R. F., Cech, T. R., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2006; Chapt. 2, p 23.
- (17) Kauffman, S. A. *Investigations*; Oxford University Press: Oxford, U.K., 2000.
- (18) Segre, D.; Ben-Eli, D.; Deamer, D. W.; Lancet, D. *Origins Life Evol. Biospheres* **2001**, *31*, 119.
- (19) Morowitz, H. J. *Energy Flows in Biology*; Academic Press: New York, 1968.
- (20) Morowitz, H. J. *Beginnings of Cellular Life*; Yale University Press: New York, 1992.

- (21) Nicolis, G.; Prigogine, I. *Self-organization in Non-equilibrium Systems*; Wiley: New York, 1977.
- (22) The terms “self-organizing” and “self-assembling” will be used in this paper to refer to “phenomena in which a large number of molecules spontaneously generate a global, irreducible correlation that brings and holds them together”. See Ruiz-Mirazo, K. In *Encyclopedia of Systems Biology*; Dubitzky, W., Wolkenhauer, O., Yokota, H., Cho, K. H., Eds.; Springer: Heidelberg, Germany, 2012. The difference between the two lies in the thermodynamic conditions under which that correlation is generated and maintained (far-from-equilibrium, dissipative conditions for self-organization; close-to-equilibrium, quasi-conservative ones for self-assembly).
- (23) Eschenmoser, A. *Origins Life Evol. Biospheres* **2007**, *37*, 309.
- (24) Stano, P.; Luisi, P. L. *Origins Life Evol. Biospheres* **2007**, *37*, 303.
- (25) Pross, A. *J. Phys. Org. Chem.* **2008**, *21*, 724.
- (26) Pross, A. *Chem.—Eur. J.* **2009**, *15*, 8374.
- (27) Wagner, N.; Pross, A. *Entropy* **2011**, *13*, 518.
- (28) Wagner, N.; Pross, A.; Tannenbaum, E. *BioSystems* **2010**, *99*, 126.
- (29) Wachtershauser, G. *Microbiol. Rev.* **1988**, *52*, 452.
- (30) Budin, I.; Szostak, J. W. *Annu. Rev. Biophys.* **2010**, *39*, 245.
- (31) Szostak, J. W. *Philos. Trans. R. Soc., B* **2011**, *366*, 2894.
- (32) Mann, S. *Angew. Chem., Int. Ed.* **2013**, *52*, 155.
- (33) Lazcano, A.; Miller, S. L. *Cell* **1996**, *85*, 793.
- (34) Wachtershauser, G. *J. Theor. Biol.* **1997**, *187*, 483.
- (35) Lahav, N. *Biogenesis. Theories of Life's Origins*; Oxford University Press: Oxford, U.K., 1999.
- (36) Fry, I. *The Emergence of Life on Earth*; Rutgers University Press: New Brunswick, NJ, 2000.
- (37) Szostak, J. W.; Bartel, D. P.; Luisi, P. L. *Nature* **2001**, *409*, 387.
- (38) Joyce, G. F. *Nature* **2002**, *418*, 214.
- (39) Orgel, L. E. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99.
- (40) Eschenmoser, A. *Tetrahedron* **2007**, *63*, 12821.
- (41) *Protocells: Bridging Nonliving and Living Matter*; Rasmussen, S., Bedau, M. A., Chen, L., Deamer, D., Krakauer, D. C., Packard, N. H., Stadler, P. F., Eds.; MIT Press: Cambridge, MA, 2008.
- (42) Lazcano, A. *Cold Spring Harbor Perspect. Biol.* **2010**, *2*, No. a002089.
- (43) Schrum, J. P.; Zhu, T. F.; Szostak, J. W. *Cold Spring Harbor Perspect. Biol.* **2010**, *2*, No. a002212.
- (44) In the second half of 2012, special issues have been published [*Acc. Chem. Res.* **45**, 12 and *Chem. Soc. Rev.* **41**, 16] about the origins of life and chemical evolution, which deal with many aspects of what is being called herein prebiotic systems chemistry.
- (45) Cleaves, H. J.; Scott, A. M.; Hill, F. C.; Leszczynski, J.; Sahai, N.; Hazen, R. *Chem. Soc. Rev.* **2012**, *41*, 5502.
- (46) Research on the origins of life constitutes a very broad area, with contributions from many different fields. This is even more so after the advent of astrobiology and synthetic biology, as mentioned before. Although we recognize the importance of any approach that tries to help in this interdisciplinary effort, covering all of them is beyond the scope of this review. Thus, our survey will mainly refer to work done on the study of physicochemical processes that could shed light on the transition from inanimate to animate matter, leaving out other more conceptual aspects as well as other more strictly biological issues. Two major current enterprises of astrobiology, the study of interstellar chemistry and the search for extraterrestrial life, are not reviewed, either. Despite the eventual relevance of these two issues to the question addressed here, we would like to focus on processes that could have led to the emergence of life on an Earth-like planet or satellite. We apologize in advance to anyone whose favourite subject is glossed over or ignored. Nevertheless, we hope that our review article will be of interest to any scientific mind willing to get an insight into the rather intricate and intriguing subject of the transition from nonliving to living matter.
- (47) Miller, S. L. *Science* **1953**, *117*, 528.
- (48) Berry, S. *J. Mol. Evol.* **2002**, *54*, 595.
- (49) Shen, L.; Ji, H.-F. *PLoS One* **2011**, *6*, No. e22494.
- (50) Khiem Van, N.; Burrows, C. *J. Acc. Chem. Res.* **2012**, *45*, 2151.
- (51) Saladino, R.; Crestini, C.; Costanzo, G.; DiMauro, E. In *Prebiotic Chemistry: From Simple Amphiphiles to Protocell Models*; Walde, P., Ed.; Topics in Current Chemistry, Vol. 259; Springer: Berlin/Heidelberg, 2005; Chapt. 2, p 29.
- (52) Dalko, P. I.; Moisan, L. *Angew. Chem., Int. Ed.* **2004**, *43*, 5138.
- (53) Chyba, C.; Sagan, C. *Nature* **1992**, *355*, 125.
- (54) Bernstein, M. P.; Sandford, S. A.; Allamandola, L. J.; Gillette, J. S.; Clemett, S. J.; Zare, R. N. *Science* **1999**, *283*, 1135.
- (55) Botta, O.; Bada, J. L. *Surv. Geophys.* **2002**, *23*, 411.
- (56) Pizzarello, S. *Acc. Chem. Res.* **2006**, *39*, 231.
- (57) Pizzarello, S.; Shock, E. *Cold Spring Harbor Perspect. Biol.* **2010**, *2*, No. a002105.
- (58) Burton, A. S.; Stern, J. C.; Elsila, J. E.; Glavin, D. P.; Dworkin, J. P. *Chem. Soc. Rev.* **2012**, *41*, 5459.
- (59) Schmitt-Kopplin, P.; Gabelica, Z.; Gougeon, R. D.; Fekete, A.; Kanawati, B.; Harir, M.; Gebefuegi, I.; Eckel, G.; Hertkorn, N. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 2763.
- (60) Hargreaves, W. R.; Mulvihill, S. J.; Deamer, D. W. *Nature* **1977**, *266*, 78.
- (61) Monnard, P. A.; Deamer, D. W. *Anat. Rec.* **2002**, *268*, 196.
- (62) Pohorille, A.; Deamer, D. *Res. Microbiol.* **2009**, *160*, 449.
- (63) Ourisson, G.; Nakatani, Y. In *Lectures in Astrobiology*; Gargaud, M., Martin, H., Reisse, J., Eds.; Springer: Berlin/Heidelberg, 2005; Vol. 1.
- (64) Lombard, J.; Lopez-Garcia, P.; Moreira, D. *Nat. Rev. Microbiol.* **2012**, *10*, 507.
- (65) Deamer, D. W. In *The Molecular Origins of Life*; Brack, A., Ed.; Cambridge University Press: Cambridge, U.K., 1998; Chapt. 8, p 189.
- (66) Ourisson, G.; Nakatani, Y. *Tetrahedron* **1999**, *55*, 3183.
- (67) Thomas, J. A.; Rana, F. R. *Origins Life Evol. Biospheres* **2007**, *37*, 267.
- (68) Cronin, J. R. In *The Molecular Origins of Life*; Brack, A., Ed.; Cambridge University Press: Cambridge, U.K., 1998; Chapt. 6, p 119.
- (69) Bada, J. L. *Earth Planet. Sci. Lett.* **2004**, *226*, 1.
- (70) Hanczyc, M. M.; Szostak, J. W. *Curr. Opin. Chem. Biol.* **2004**, *8*, 660.
- (71) Deamer, D. W. *Nature* **1985**, *317*, 792.
- (72) Deamer, D. W.; Pashley, R. M. *Origins Life Evol. Biospheres* **1989**, *19*, 21.
- (73) Dworkin, L. P.; Deamer, D. W.; Sandford, S. A.; Allamandola, L. J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 815.
- (74) Deamer, D. W.; Mahon, E. H.; Bosco, G. In *Early Life on Earth, Nobel Symposium*; Bengtson, S., Ed.; Columbia University Press: New York, 1994; Vol. 84.
- (75) Nooner, D. W.; Gibert, J. M.; Gelpi, E.; Oro, J. *Geochim. Cosmochim. Acta* **1976**, *40*, 915.
- (76) McCollom, T. M.; Ritter, G.; Simoneit, B. R. T. *Origins Life Evol. Biospheres* **1999**, *29*, 153.
- (77) Rushdi, A. I.; Simoneit, B. R. T. *Origins Life Evol. Biospheres* **2001**, *31*, 103.
- (78) Simoneit, B. R. T.; Rushdi, A. I.; Deamer, D. W. *Adv. Space Res.* **2007**, *40*, 1649.
- (79) Volkov, A. G.; Gugeshashvili, M. I.; Deamer, D. W. *Electrochim. Acta* **1995**, *40*, 2849.
- (80) Ariga, K.; Yuki, H.; Kikuchi, J.; Dannemuller, O.; Albrecht-Gary, A. M.; Nakatani, Y.; Ourisson, G. *Langmuir* **2005**, *21*, 4578.
- (81) Summons, R. E.; Albrecht, P.; McDonald, G.; Moldowan, J. M. *Space Sci. Rev.* **2008**, *135*, 133.
- (82) Apel, C. L.; Deamer, D. W.; Mautner, M. N. *Biochim. Biophys. Acta* **2002**, *1559*, 1.
- (83) Maurer, S. E.; Deamer, D. W.; Boncella, J. M.; Monnard, P. A. *Astrobiology* **2009**, *9*, 979.
- (84) Rushdi, A. I.; Simoneit, B. R. T. *Origins Life Evol. Biospheres* **2006**, *36*, 93.
- (85) Eichberg, J.; Sherwood, E.; Epps, D. E.; Oro, J. *J. Mol. Evol.* **1977**, *10*, 221.
- (86) Epps, D. E.; Sherwood, E.; Eichberg, J.; Oro, J. *J. Mol. Evol.* **1978**, *11*, 279.
- (87) Rao, M.; Eichberg, J.; Oro, J. *J. Mol. Evol.* **1982**, *18*, 196.

- (88) Gissot, A.; Camplo, M.; Grinstaff, M. W.; Barthelemy, P. *Org. Biomol. Chem.* **2008**, *6*, 1324.
- (89) Yanagawa, H.; Ogawa, Y.; Furuta, H.; Tsuno, K. *J. Am. Chem. Soc.* **1989**, *111*, 4567.
- (90) Bonaccio, S.; Wessicken, M.; Berti, D.; Walde, P.; Luisi, P. L. *Langmuir* **1996**, *12*, 4976.
- (91) Rosemeyer, H. *Chem. Biodiversity* **2005**, *2*, 977.
- (92) Johnson, A. P.; Cleaves, H. J.; Dworkin, J. P.; Glavin, D. P.; Lazcano, A.; Bada, J. L. *Science* **2008**, *322*, 404.
- (93) Pascal, R.; Boiteau, L.; Commeiras, A. In *Prebiotic Chemistry: From Simple Amphiphiles to Protocell Models*; Walde, P., Ed.; Topics in Current Chemistry, Vol. 259; Springer-Verlag: Berlin/Heidelberg, 2005; Chapt. 3, p 69.
- (94) Cordova, A.; Ibrahim, I.; Casas, J.; Sundén, H.; Engqvist, M.; Reyes, E. *Chem.—Eur. J.* **2005**, *11*, 4772.
- (95) Cordova, A.; Zou, W.; Dziedzic, R.; Ibrahim, I.; Reyes, E.; Xu, Y. *Chem.—Eur. J.* **2006**, *12*, 5175.
- (96) *Modified Amino Acids, Organocatalysis and Enzymes*; Hughes, A. B., Ed.; Amino Acids, Peptides and Proteins in Organic Chemistry, Vol. 2; Wiley—VCH: Weinheim, Germany, 2009.
- (97) Weber, A. L.; Pizzarello, S. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12713.
- (98) Pizzarello, S.; Weber, A. L. *Origins Life Evol. Biospheres* **2010**, *40*, 3.
- (99) Pizzarello, S.; Weber, A. L. *Science* **2004**, *303*, 1151.
- (100) Leman, L. J.; Orgel, L. E.; Ghadiri, M. R. *J. Am. Chem. Soc.* **2006**, *128*, 20.
- (101) Borsenberger, V.; Crowe, M. A.; Lehbauer, J.; Raftery, J.; Hellwell, M.; Bhutia, K.; Cox, T.; Sutherland, J. D. *Chem. Biodiversity* **2004**, *1*, 203.
- (102) Smith, J. M.; Borsenberger, V.; Raftery, J.; Sutherland, J. D. *Chem. Biodiversity* **2004**, *1*, 1418.
- (103) Cech, T. R. *Cold Spring Harbor Symp. Quant. Biol.* **2009**, *74*, 11.
- (104) Di Giulio, M. *J. Mol. Evol.* **1997**, *45*, 571.
- (105) Cavalier-Smith, T. *J. Mol. Evol.* **2001**, *53*, 555.
- (106) Ramakrishnan, V. *Cold Spring Harbor Symp. Quant. Biol.* **2009**, *74*, 25.
- (107) Zaia, D. A. M.; Zaia, C.; De Santana, H. *Origins Life Evol. Biospheres* **2008**, *38*, 469.
- (108) Plankensteiner, K.; Reiner, H.; Rode, B. M. *Mol. Diversity* **2006**, *10*, 3.
- (109) Cleaves, H. J.; Chalmers, J. H.; Lazcano, A.; Miller, S. L.; Bada, J. L. *Origins Life Evol. Biospheres* **2008**, *38*, 105.
- (110) Urey, H. C. *Proc. Natl. Acad. Sci. U.S.A.* **1952**, *38*, 351.
- (111) Tian, F.; Toon, O. B.; Pavlov, A. A.; De Sterck, H. *Science* **2005**, *308*, 1014.
- (112) Parker, E. T.; Cleaves, H. J.; Dworkin, J. P.; Glavin, D. P.; Callahan, M.; Aubrey, A.; Lazcano, A.; Bada, J. L. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 5526.
- (113) Parker, E. T.; Cleaves, H. J.; Callahan, M. P.; Dworkin, J. P.; Glavin, D. P.; Lazcano, A.; Bada, J. L. *Origins Life Evol. Biospheres* **2011**, *41*, 201.
- (114) Huber, C.; Eisenreich, W.; Wachtershauser, G. *Tetrahedron Lett.* **2010**, *51*, 1069.
- (115) Fitz, D.; Reiner, H.; Rode, B. M. *Pure Appl. Chem.* **2007**, *79*, 2101.
- (116) *Origins and Synthesis of Amino Acids*; Hughes, A. B., Ed.; Amino Acids, Peptides and Proteins in Organic Chemistry, Vol. 1; Wiley—VCH: Weinheim, Germany, 2009.
- (117) Strecker, A. *Liebigs Ann. Chem.* **1850**, *75*, 27.
- (118) Taillades, J.; Beuzelin, I.; Garrel, L.; Tabacik, V.; Bied, C.; Commeiras, A. *Origins Life Evol. Biospheres* **1998**, *28*, 61.
- (119) Pitsch, S.; Eschenmoser, A.; Gedulin, B.; Hui, S.; Arrhenius, G. *Origins Life Evol. Biospheres* **1995**, *25*, 297.
- (120) Oro, J. *Biochem. Biophys. Res. Commun.* **1960**, *2*, 407.
- (121) Sanchez, R. A.; Ferris, J. P.; Orgel, L. E. *Science* **1966**, *154*, 784.
- (122) Robertson, M. P.; Miller, S. L. *Nature* **1995**, *377*, 257.
- (123) Robertson, M. P.; Miller, S. L. *Nature* **1995**, *375*, 772.
- (124) Fuller, W. D.; Sanchez, R. A.; Orgel, L. E. *J. Mol. Biol.* **1972**, *67*, 25.
- (125) Powner, M. W.; Sutherland, J. D. *Philos. Trans. R. Soc., B* **2011**, *366*, 2870.
- (126) Powner, M. W.; Gerland, B.; Sutherland, J. D. *Nature* **2009**, *459*, 239.
- (127) Szostak, J. W. *Nature* **2009**, *459*, 171.
- (128) Miyakawa, S.; Cleaves, H. J.; Miller, S. L. *Origins Life Evol. Biospheres* **2002**, *32*, 195.
- (129) Miyakawa, S.; Cleaves, H. J.; Miller, S. L. *Origins Life Evol. Biospheres* **2002**, *32*, 209.
- (130) Shapiro, R. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4396.
- (131) Nelson, K. E.; Robertson, M. P.; Levy, M.; Miller, S. L. *Origins Life Evol. Biospheres* **2001**, *31*, 221.
- (132) Orgel, L. E. *Origins Life Evol. Biospheres* **2004**, *34*, 361.
- (133) Cleaves, H. J.; Nelson, K. E.; Miller, S. L. *Naturwissenschaften* **2006**, *93*, 228.
- (134) Menor-Salvan, C.; Ruiz-Bermejo, M.; Osuna-Esteban, S.; Veintemillas-Verdaguer, S. *Origins Life Evol. Biospheres* **2009**, *39*, 250.
- (135) Menor-Salvan, C.; Ruiz-Bermejo, D. M.; Guzman, M. I.; Osuna-Esteban, S.; Veintemillas-Verdaguer, S. *Chem.—Eur. J.* **2009**, *15*, 4411.
- (136) Menor-Salvan, C.; Marin-Yaseli, M. R. *Chem. Soc. Rev.* **2012**, *41*, 5404.
- (137) Donaldson, D. J.; Tervahattu, H.; Tuck, A. F.; Vaida, V. *Origins Life Evol. Biospheres* **2004**, *34*, 57.
- (138) Ruiz-Bermejo, M.; Menor-Salvan, C.; Osuna-Esteban, S.; Veintemillas-Verdaguer, S. *Origins Life Evol. Biospheres* **2007**, *37*, 123.
- (139) Hazen, R. M.; Sverjensky, D. A. *Cold Spring Harbor Perspect. Biol.* **2010**, *2*, No. a002162.
- (140) Saladino, R.; Crestini, C.; Ciciriello, F.; Costanzo, G.; Di Mauro, E. *Origins Life Evol. Biospheres* **2006**, *36*, 523.
- (141) Saladino, R.; Botta, G.; Pino, S.; Costanzo, G.; Di Mauro, E. *Chem. Soc. Rev.* **2012**, *41*, 5526.
- (142) Saladino, R.; Crestini, C.; Ciciriello, F.; Costanzo, G.; Di Mauro, E. *Chem. Biodiversity* **2007**, *4*, 694.
- (143) Saladino, R.; Crestini, C.; Ciciriello, F.; Pino, S.; Costanzo, G.; Di Mauro, E. *Res. Microbiol.* **2009**, *160*, 441.
- (144) Saladino, R.; Crestini, C.; Pino, S.; Costanzo, G.; Di Mauro, E. *Phys. Life Rev.* **2012**, *9*, 84.
- (145) Barks, H. L.; Buckley, R.; Grieves, G. A.; Di Mauro, E.; Hud, N. V.; Orlando, T. M. *ChemBioChem* **2010**, *11*, 1240.
- (146) Saladino, R.; Crestini, C.; Costanzo, G.; Negri, R.; Di Mauro, E. *Bior. Med. Chem.* **2001**, *9*, 1249.
- (147) Saladino, R.; Ciambecchini, U.; Crestini, C.; Costanzo, G.; Negri, R.; Di Mauro, E. *ChemBioChem* **2003**, *4*, 514.
- (148) Saladino, R.; Crestini, C.; Ciambecchini, U.; Ciciriello, F.; Costanzo, G.; Di Mauro, E. *ChemBioChem* **2004**, *5*, 1558.
- (149) Saladino, R.; Crestini, C.; Neri, V.; Brucato, J. R.; Colangeli, L.; Ciciriello, F.; Di Mauro, E.; Costanzo, G. *ChemBioChem* **2005**, *6*, 1368.
- (150) Saladino, R.; Neri, V.; Crestini, C.; Costanzo, G.; Graciotti, M.; Di Mauro, E. *J. Am. Chem. Soc.* **2008**, *130*, 15512.
- (151) Saladino, R.; Neri, V.; Crestini, C.; Costanzo, G.; Graciotti, M.; Di Mauro, E. *J. Mol. Evol.* **2010**, *71*, 100.
- (152) Shanker, U.; Bhushan, B.; Bhattacharjee, G.; Kamaluddin. *Astrobiology* **2011**, *11*, 225.
- (153) Hanczyc, M. M.; Fujikawa, S. M.; Szostak, J. W. *Science* **2003**, *302*, 618.
- (154) Bean, H. D.; Sheng, Y. H.; Collins, J. P.; Anet, F. A. L.; Leszczynski, J.; Hud, N. V. *J. Am. Chem. Soc.* **2007**, *129*, 9556.
- (155) Sheng, Y. H.; Bean, H. D.; Mamajanov, I.; Hud, N. V.; Leszczynski, J. *J. Am. Chem. Soc.* **2009**, *131*, 16088.
- (156) Benner, S. A. *Acc. Chem. Res.* **2004**, *37*, 784.
- (157) Sutherland, J. D. *Angew. Chem., Int. Ed.* **2007**, *46*, 2354.
- (158) Mittapalli, G. K.; Reddy, K. R.; Xiong, H.; Munoz, O.; Han, B.; De Riccardis, F.; Krishnamurthy, R.; Eschenmoser, A. *Angew. Chem., Int. Ed.* **2007**, *46*, 2470.

- (159) Mittapalli, G. K.; Osornio, Y. M.; Guerrero, M. A.; Reddy, K. R.; Krishnamurthy, R.; Eschenmoser, A. *Angew. Chem., Int. Ed.* **2007**, *46*, 2478.
- (160) Krishnamurthy, R. *Acc. Chem. Res.* **2012**, *45*, 2035.
- (161) Levy, M.; Miller, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 7933.
- (162) Weber, A. L. *Origins Life Evol. Biospheres* **2008**, *38*, 279.
- (163) Sacerdote, M. G.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6004.
- (164) Ricardo, A.; Carrigan, M. A.; Olcott, A. N.; Benner, S. A. *Science* **2004**, *303*, 196.
- (165) Lambert, J. B.; Gurusamy-Thangavelu, S. A.; Ma, K. *Science* **2010**, *327*, 984.
- (166) Kim, H.-J.; Benner, S. A. *Science* **2010**, *329*, 902 (comment on ref 165).
- (167) Lambert, J. B.; Gurusamy-Thangavelu, S. A.; Ma, K. *Science* **2010**, *329*, 902 (response to comment on ref 165).
- (168) Kim, H.-J.; Ricardo, A.; Illangkoon, H. I.; Kim, M. J.; Carrigan, M. A.; Frye, F.; Benner, S. A. *J. Am. Chem. Soc.* **2011**, *133*, 9457.
- (169) Costanzo, G.; Saladino, R.; Crestini, C.; Ciciriello, F.; Di Mauro, E. *J. Biol. Chem.* **2007**, *282*, 16729.
- (170) Crowe, M. A.; Sutherland, J. D. *ChemBioChem* **2006**, *7*, 951.
- (171) Anastasi, C.; Crowe, M. A.; Pownier, M. W.; Sutherland, J. D. *Angew. Chem., Int. Ed.* **2006**, *45*, 6176.
- (172) Anastasi, C.; Crowe, M. A.; Sutherland, J. D. *J. Am. Chem. Soc.* **2007**, *129*, 24.
- (173) Anastasi, C.; Buchet, F. E.; Crowe, M. A.; Hellwell, M.; Raftery, J.; Sutherland, J. D. *Chem.—Eur. J.* **2008**, *14*, 2375.
- (174) Pownier, M. W.; Sutherland, J. D. *Angew. Chem., Int. Ed.* **2010**, *49*, 4641.
- (175) Pownier, M. W.; Sutherland, J. D.; Szostak, J. W. *J. Am. Chem. Soc.* **2010**, *132*, 16677.
- (176) Sutherland, J. D. *Cold Spring Harbor Perspect. Biol.* **2010**, *2*, No. a005439.
- (177) Pownier, M. W.; Zheng, S.-L.; Szostak, J. W. *J. Am. Chem. Soc.* **2012**, *134*, 13889.
- (178) Ritson, D.; Sutherland, J. D. *Nat. Chem.* **2012**, *4*, 895.
- (179) Ritson, D.; Sutherland, J. D. *Angew. Chem., Int. Ed.* **2013**, *52*, 5845.
- (180) Drobner, E.; Huber, H.; Wachtershauser, G.; Rose, D.; Stetter, K. O. *Nature* **1990**, *346*, 742.
- (181) Blochl, E.; Keller, M.; Wächtershäuser, G.; Stetter, K. O. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 8117.
- (182) Yamagata, Y.; Watanabe, H.; Saitoh, M.; Namba, T. *Nature* **1991**, *352*, 516.
- (183) De Graaf, R. M.; Schwartz, A. W. *Origins Life Evol. Biospheres* **2000**, *30*, 405.
- (184) Schwartz, A. W. *Philos. Trans. R. Soc., B* **2006**, *361*, 1743.
- (185) Pasek, M. A.; Kee, T. P.; Bryant, D. E.; Pavlov, A. A.; Lunine, J. I. *Angew. Chem., Int. Ed.* **2008**, *47*, 7918.
- (186) Szostak, J. W. *J. Syst. Chem.* **2012**, *3*, 2.
- (187) Martin, R. B. *Biopolymers* **1998**, *45*, 351.
- (188) Radzicka, A.; Wolfenden, R. *J. Am. Chem. Soc.* **1996**, *118*, 6105.
- (189) Smith, R. M.; Hansen, D. E. *J. Am. Chem. Soc.* **1998**, *120*, 8910.
- (190) White, R. H. *Nature* **1984**, *310*, 430.
- (191) Danger, G.; Plasson, R.; Pascal, R. *Chem. Soc. Rev.* **2012**, *41*, 5416.
- (192) Fox, S. W.; Harada, K. *Science* **1958**, *128*, 1214.
- (193) Napier, J.; Yin, J. *Peptides* **2006**, *27*, 607.
- (194) Huber, C.; Wachtershauser, G. *Science* **1998**, *281*, 670.
- (195) Huber, C.; Eisenreich, W.; Hecht, S.; Wachtershauser, G. *Science* **2003**, *301*, 938.
- (196) Imai, E.; Honda, H.; Hatori, K.; Brack, A.; Matsuno, K. *Science* **1999**, *283*, 831.
- (197) Tsukahara, H.; Imai, E. I.; Honda, H.; Hatori, K.; Matsuno, K. *Origins Life Evol. Biospheres* **2002**, *32*, 13.
- (198) Islam, M. N.; Kaneko, T.; Kobayashi, K. *Bull. Chem. Soc. Jpn.* **2003**, *76*, 1171.
- (199) Cleaves, H. J.; Aubrey, A. D.; Bada, J. L. *Origins Life Evol. Biospheres* **2009**, *39*, 109.
- (200) Lahav, N.; White, D.; Chang, S. *Science* **1978**, *201*, 67.
- (201) Bujdak, J.; LeSon, H.; Yongyai, Y.; Rode, B. M. *Catal. Lett.* **1996**, *37*, 267.
- (202) Bujdak, J.; Rode, B. M. *Catal. Lett.* **2003**, *91*, 149.
- (203) Bujdak, J.; Rode, B. M. *Origins Life Evol. Biospheres* **1999**, *29*, 451.
- (204) Rode, B. M. *Peptides* **1999**, *20*, 773.
- (205) Rode, B. M.; Son, H. L.; Suwannachot, Y.; Bujdak, J. *Origins Life Evol. Biospheres* **1999**, *29*, 273.
- (206) Plankenstein, K.; Reiner, H.; Rode, B. M. *Peptides* **2005**, *26*, 535.
- (207) Fitz, D.; Jakschitz, T.; Rode, B. M. *J. Inorg. Biochem.* **2008**, *102*, 2097.
- (208) Li, F.; Fitz, D.; Fraser, D. G.; Rode, B. M. *Amino Acids* **2010**, *38*, 287.
- (209) Li, F.; Fitz, D.; Fraser, D. G.; Rode, B. M. *Amino Acids* **2010**, *39*, 579.
- (210) Rode, B. M.; Fitz, D.; Jakschitz, T. *Chem. Biodiversity* **2007**, *4*, 2674.
- (211) Rode, B. M. *Amino Acids* **2009**, *37*, 2.
- (212) Bada, J. L.; Lazcano, A. *Science* **2002**, *296*, 1982.
- (213) Griffith, E. C.; Vaida, V. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 15697.
- (214) Liu, R. H.; Orgel, L. E. *Origins Life Evol. Biospheres* **1998**, *28*, 47.
- (215) Orgel, L. E. *Origins Life. Evol. Biospheres* **1998**, *28*, 227.
- (216) Hill, A. R.; Bohler, C.; Orgel, L. E. *Origins Life Evol. Biospheres* **1998**, *28*, 235.
- (217) Liu, R.; Orgel, L. E. *Origins Life Evol. Biospheres* **1998**, *28*, 245.
- (218) Naka, K.; Tampo, Y.; Chujo, Y. *J. Organomet. Chem.* **2007**, *692*, 436.
- (219) Duvernay, F.; Chiavassa, T.; Borget, F.; Aycard, J. P. *J. Am. Chem. Soc.* **2004**, *126*, 7772.
- (220) Danger, G.; Michaut, A.; Bucchi, M.; Boiteau, L.; Canal, J.; Plasson, R.; Pascal, R. *Angew. Chem., Int. Ed.* **2013**, *52*, 611.
- (221) Leuchs, H. *Ber. Dtsch. Chem. Ges.* **1906**, *39*, 857.
- (222) Ferris, J. P.; Hill, A. R.; Liu, R. H.; Orgel, L. E. *Nature* **1996**, *381*, 59.
- (223) Kricheldorf, H. R. *Angew. Chem., Int. Ed.* **2006**, *45*, 5752.
- (224) Ehler, K. W.; Orgel, L. E. *Biochim. Biophys. Acta* **1976**, *434*, 233.
- (225) Boiteau, L.; Collet, H.; Lagrille, O.; Taillades, J.; Vayaboury, W.; Giani, O.; Schue, F.; Commeyras, A. *Polym. Int.* **2002**, *51*, 1037.
- (226) Lagrille, O.; Taillades, J.; Boiteau, L.; Commeyras, A. *J. Phys. Org. Chem.* **2007**, *20*, 271.
- (227) Leman, L.; Orgel, L.; Ghadiri, M. R. *Science* **2004**, *306*, 283.
- (228) Chen, F.; Yang, D. *Origins Life Evol. Biospheres* **2007**, *37*, 47.
- (229) De Duve, C. In *Frontiers of Life*; Tran Thanh Van, J., Tran Thanh Van, K., Mounolou, J.-C., Schneider, J., McKay, C., Eds.; Editions Frontières: Gif-sur-Yvette, France, 1992.
- (230) Liu, R. H.; Orgel, L. E. *Nature* **1997**, *389*, 52.
- (231) Commeyras, A.; Taillades, J.; Collet, H.; Boiteau, L.; Vandeneabeele-Trambouze, O.; Pascal, R.; Rousset, A.; Garrel, L.; Rossi, J. C.; Biron, J. P.; Lagrille, O.; Plasson, R.; Souaid, E.; Danger, G.; Selsis, F.; Dobrijevic, M.; Martin, H. *Origins Life Evol. Biospheres* **2004**, *34*, 35.
- (232) Commeyras, A.; Collet, H.; Boiteau, L.; Taillades, J.; Vandeneabeele-Trambouze, O.; Cottet, H.; Biron, J. P.; Plasson, R.; Mion, L.; Lagrille, O.; Martin, H.; Selsis, F.; Dobrijevic, M. *Polym. Int.* **2002**, *51*, 661.
- (233) Plasson, R.; Bersini, H.; Cornmeyras, A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 16733.
- (234) Gleiser, M.; Walker, S. I. *Origins Life Evol. Biospheres* **2009**, *39*, 479.
- (235) Dickson, K. S.; Burns, C. M.; Richardson, J. P. *J. Biol. Chem.* **2000**, *275*, 15828.

- (236) Given the preminent role that RNA has over DNA in the field of origins of life, mainly due to the well-recognized plausibility of the RNA world (in its more traditional or systems versions), here we will refer to the polymerization of ribonucleotides, unless otherwise stated.
- (237) Moravec, J. *Tetrahedron Lett.* **1967**, *8*, 1707.
- (238) Kawamura, K. *Anal. Sci.* **2002**, *18*, 715.
- (239) Ogasawara, H.; Yoshida, A.; Imai, E. I.; Honda, H.; Hatori, K.; Matsuno, K. *Origins Life Evol. Biospheres* **2000**, *30*, 519.
- (240) Kawamura, K. *Int. J. Astrobiol.* **2004**, *3*, 301.
- (241) Fox, S. W.; Harada, K. *Science* **1961**, *133*, 1923.
- (242) Kochetkov, N. K.; Domkin, V. D.; Budowsky, E. I.; Khromovborisov, N. N. *Biochim. Biophys. Acta* **1964**, *80*, 145.
- (243) Jacob, T. M.; Khorana, H. G. *J. Am. Chem. Soc.* **1964**, *86*, 1630.
- (244) Gottsch, B. P.; Slutsky, O. I. *Biochim. Biophys. Acta* **1964**, *87*, 163.
- (245) Inoue, T.; Orgel, L. E. *Science* **1983**, *219*, 859.
- (246) Huang, W. H.; Ferris, J. P. *J. Am. Chem. Soc.* **2006**, *128*, 8914.
- (247) Prabahar, K. J.; Cole, T. D.; Ferris, J. P. *J. Am. Chem. Soc.* **1994**, *116*, 10914.
- (248) Prabahar, K. J.; Ferris, J. P. *J. Am. Chem. Soc.* **1997**, *119*, 4330.
- (249) Monnard, P. A.; Kanavarioti, A.; Deamer, D. W. *J. Am. Chem. Soc.* **2003**, *125*, 13734.
- (250) Huang, W. H.; Ferris, J. P. *Chem. Commun.* **2003**, *12*, 1458.
- (251) Miyakawa, S.; Ferris, J. P. *J. Am. Chem. Soc.* **2003**, *125*, 8202.
- (252) Ferris, J. P. *Philos. Trans. R. Soc., B* **2006**, *361*, 1777.
- (253) Ertem, G.; Snellinger-O'Brien, A. M.; Ertem, M. C.; Rogoff, D. A.; Dworkin, J. P.; Johnston, M. V.; Hazen, R. M. *Int. J. Astrobiol.* **2008**, *7*, 1.
- (254) Ertem, G. *Origins Life Evol. Biospheres* **2004**, *34*, 549.
- (255) Thyveetil, M. A.; Coveney, P. V.; Greenwell, H. C.; Suter, J. L. *J. Am. Chem. Soc.* **2008**, *130*, 4742.
- (256) Franchi, M.; Ferris, J. P.; Gallori, E. *Origins Life Evol. Biospheres* **2003**, *33*, 1.
- (257) Kawamura, K.; Ferris, J. P. *Origins Life Evol. Biospheres* **1999**, *29*, 563.
- (258) Joshi, P. C.; Aldersley, M. F.; Delano, J. W.; Ferris, J. P. *J. Am. Chem. Soc.* **2009**, *131*, 13369.
- (259) Mathew, D. C.; Luthey-Schulten, Z. *Origins Life Evol. Biospheres* **2010**, *40*, 303.
- (260) Ertem, G.; Ferris, J. P. *Origins Life Evol. Biospheres* **2000**, *30*, 411.
- (261) Joshi, P. C.; Pitsch, S.; Ferris, J. P. *Chem. Commun.* **2000**, *24*, 2497.
- (262) Joshi, P. C.; Pitsch, S.; Ferris, J. P. *Origins Life Evol. Biospheres* **2007**, *37*, 3.
- (263) Joshi, P. C.; Aldersley, M. F.; Ferris, J. P. *Origins Life Evol. Biospheres* **2011**, *41*, 213.
- (264) Costanzo, G.; Pino, S.; Ciciriello, F.; Di Mauro, E. *J. Biol. Chem.* **2009**, *284*, 33206.
- (265) Briones, C.; Stich, M.; Manrubia, S. C. *RNA* **2009**, *15*, 743.
- (266) Pino, S.; Costanzo, G.; Giorgi, A.; Di Mauro, E. *Biochemistry* **2011**, *50*, 2994.
- (267) James, K. D.; Ellington, A. D. *Chem. Biol.* **1997**, *4*, 595.
- (268) Rohatgi, R.; Bartel, D. P.; Szostak, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 3340.
- (269) Horowitz, E. D.; Engelhart, A. E.; Chen, M. C.; Quarles, K. A.; Smith, M. W.; Lynn, D. G.; Hud, N. V. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 5288.
- (270) Blocher, M.; Liu, D. J.; Walde, P.; Luisi, P. L. *Macromolecules* **1999**, *32*, 7332.
- (271) Hitz, T.; Blocher, M.; Walde, P.; Luisi, P. L. *Macromolecules* **2001**, *34*, 2443.
- (272) Zepik, H. H.; Rajamani, S.; Maurel, M. C.; Deamer, D. *Origins Life Evol. Biospheres* **2007**, *37*, 495.
- (273) Rajamani, S.; Vlassov, A.; Benner, S.; Coombs, A.; Olasagasti, F.; Deamer, D. *Origins Life Evol. Biospheres* **2008**, *38*, 57.
- (274) Olasagasti, F.; Kim, H. J.; Pourmand, N.; Deamer, D. W. *Biochimie* **2011**, *93*, 556.
- (275) Griffith, R. W. *Origins Life Evol. Biospheres* **2009**, *39*, 517.
- (276) Mullen, L. B.; Sutherland, J. D. *Angew. Chem., Int. Ed.* **2007**, *46*, 8063.
- (277) Schimmel, P.; Henderson, B. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 11283.
- (278) Wieczorek, R.; Dorr, M.; Chotera, A.; Luisi, P. L.; Monnard, P. A. *ChemBioChem* **2013**, *14*, 217.
- (279) Di Giulio, M. *Origins Life Evol. Biospheres* **2003**, *33*, 479.
- (280) Terenzi, S.; Biala, E.; Nguyen-Trung, N. Q.; Strazewski, P. *Angew. Chem., Int. Ed.* **2003**, *42*, 2909.
- (281) Yarus, M. *New Biol.* **1991**, *3*, 183.
- (282) Yarus, M.; Widmann, J. J.; Knight, R. *J. Mol. Evol.* **2009**, *69*, 406.
- (283) Turk-MacLeod, R. M.; Puthenvedu, D.; Majerfeld, I.; Yarus, M. *J. Mol. Evol.* **2012**, *74*, 217.
- (284) Yarus, M.; Caporaso, J. G.; Knight, R. *Annu. Rev. Biochem.* **2005**, *74*, 179.
- (285) Calvin, M. *Chemical Evolution*; Oxford University Press: London, 1969.
- (286) Bonner, W. A. *Origins Life Evol. Biospheres* **1994**, *24*, 63.
- (287) Cintas, P. *Angew. Chem., Int. Ed.* **2002**, *41*, 1139.
- (288) Feringa, B. L.; van Delden, R. A. *Angew. Chem., Int. Ed.* **1999**, *38*, 3419.
- (289) Weissbuch, I.; Leiserowitz, L.; Lahav, M. In *Prebiotic Chemistry: From Simple Amphiphiles to Protocell Models*; Walde, P., Ed.; Topics in Current Chemistry, Vol. 259; Springer: Berlin/Heidelberg, 2005; Chapt. 4, p 123.
- (290) Plasmon, R.; Kondepudi, D. K.; Bersini, H.; Commeyras, A.; Asakura, K. *Chirality* **2007**, *19*, 589.
- (291) Guijarro, A.; Yus, M. *The Origin of Chirality in the Molecules of Life*; RSC Publishing: Cambridge, U.K., 2009.
- (292) Blackmond, D. G. *Cold Spring Harbor Perspect. Biol.* **2010**, *2*, No. a002147.
- (293) Blackmond, D. G. *Philos. Trans. R. Soc., B* **2011**, *366*, 2878.
- (294) Evans, A. C.; Meinert, C.; Giri, C.; Goesmann, F.; Meierhenrich, U. J. *Chem. Soc. Rev.* **2012**, *41*, 5447.
- (295) Peyralans, J. J. P.; Otto, S. *Curr. Opin. Chem. Biol.* **2009**, *13*, 705.
- (296) Bada, J. L.; Miller, S. L. *BioSystems* **1987**, *20*, 21.
- (297) Joyce, G. F.; Visser, G. M.; van Boekel, C. A.; van Boom, J. H.; Orgel, L. E.; van Westrenen, J. *Nature* **1984**, *310*, 602.
- (298) Avetisov, V.; Goldanski, V. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 11435.
- (299) Kuhn, H. *Curr. Opin. Colloid Interface Sci.* **2008**, *13*, 3.
- (300) Mason, S. F.; Tranter, G. E. *Proc. R. Soc. London, A* **1985**, *397*, 45.
- (301) Quack, M. *Angew. Chem., Int. Ed.* **2002**, *41*, 4618.
- (302) Bonner, W. A. *Chirality* **2000**, *12*, 114.
- (303) Avalos, M.; Babiano, R.; Cintas, P.; Jiménez, J. L.; Palacios, J. C. *Tetrahedron: Asymmetry* **2000**, *11*, 2845.
- (304) Meierhenrich, U. J.; Nahon, L.; Alcaraz, C.; Bredehoft, J. H.; Hoffmann, S. V.; Barbier, B.; Brack, A. *Angew. Chem., Int. Ed.* **2005**, *44*, 5630.
- (305) de Marcellus, P.; Meinert, C.; Nuevo, M.; Filippi, J.-J.; Danger, G.; Deboffe, D.; Nahon, L.; d'Hendecourt, L. L. S.; Meierhenrich, U. J. *Astrophys. J. Lett.* **2011**, *727*, No. L27.
- (306) Cronin, J. R.; Pizzarello, S. *Science* **1997**, *275*, 951.
- (307) Kawasaki, T.; Hatase, K.; Fujii, Y.; Jo, K.; Soai, K.; Pizzarello, S. *Geochim. Cosmochim. Acta* **2006**, *70*, 5395.
- (308) Pizzarello, S.; Groy, T. L. *Geochim. Cosmochim. Acta* **2011**, *75*, 645.
- (309) Bailey, J. *Origins Life Evol. Biospheres* **2001**, *31*, 167.
- (310) Kovacs, K. L.; Keszthelyi, L.; Goldanski, V. *Origins Life* **1981**, *11*, 93.
- (311) Klabunovskii, E. I. *Astrobiology* **2001**, *1*, 127.
- (312) Arteaga, O.; Canillas, A.; Crusats, J.; El-Hachemi, Z.; Jellison, G. E., Jr.; Llorca, J.; Ribo, J. M. *Origins Life Evol. Biospheres* **2010**, *40*, 27.
- (313) Siegel, J. S. *Chirality* **1998**, *10*, 24.

- (314) Faglioni, F.; Passalacqua, A.; Lazzeretti, P. *Origins Life Evol. Biospheres* **2005**, *35*, 461.
- (315) Ball, P. *Chem. World* **2007**, *4*, 30.
- (316) Ribo, J. M.; Hochberg, D. *Phys. Lett. A* **2008**, *373*, 111.
- (317) Crusats, J.; Hochberg, D.; Moyano, A.; Ribo, J. M. *ChemPhysChem* **2009**, *10*, 2123.
- (318) Blanco, C.; Stich, M.; Hochberg, D. *Chem. Phys. Lett.* **2011**, *505*, 140.
- (319) Frank, F. C. *Biochim. Biophys. Acta* **1953**, *11*, 459.
- (320) Decker, P. *J. Mol. Evol.* **1974**, *4*, 49.
- (321) Hochstim, A. R. *Origins Life* **1975**, *6*, 317.
- (322) Kondepudi, D. K.; Nelson, G. W. *Phys. Rev. Lett.* **1983**, *50*, 1023.
- (323) Cintas, P. *ChemPhysChem* **2001**, *2*, 409.
- (324) Blackmond, D. G. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 5732.
- (325) Shnerb, N. M.; Louzoun, Y.; Bettelheim, E.; Solomon, S. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10322.
- (326) Saito, Y.; Hyuga, H. *J. Phys. Soc. Jpn.* **2004**, *73*, 33.
- (327) Lombardo, T. G.; Stillinger, F. H.; Debenedetti, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 15131.
- (328) Klussmann, M.; Iwamura, H.; Mathew, S. P.; Wells, D. H.; Pandya, U.; Armstrong, A.; Blackmond, D. G. *Nature* **2006**, *441*, 621.
- (329) Breslow, R.; Levine, M. S. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 12979.
- (330) Breslow, R.; Cheng, Z. L. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 9144.
- (331) Morowitz, H. J. *J. Theor. Biol.* **1969**, *25*, 491.
- (332) Klussmann, M.; White, A. J. R.; Armstrong, A.; Blackmond, D. G. *Angew. Chem., Int. Ed.* **2006**, *45*, 7985.
- (333) Klussmann, M.; Izumi, T.; White, A. J. P.; Armstrong, A.; Blackmond, D. G. *J. Am. Chem. Soc.* **2007**, *129*, 7657.
- (334) Perry, R. H.; Wu, C. P.; Neffliu, M.; Cooks, R. G. *Chem. Commun.* **2007**, *10*, 1071.
- (335) Fletcher, S. P.; Jagt, R. B. C.; Feringa, B. L. *Chem. Commun.* **2007**, *25*, 2578.
- (336) Blackmond, D. G.; Klussmann, M. *Chem. Commun.* **2007**, *39*, 3990.
- (337) Viedma, C.; Ortiz, J. E.; de Torres, T.; Cintas, P. *Chem. Commun.* **2012**, *48*, 3623.
- (338) Asakura, K.; Nagasaka, Y.; Osanai, S.; Kondepudi, D. K. *J. Phys. Chem. B* **2005**, *109*, 1586.
- (339) Kondepudi, D. K.; Kaufman, R. J.; Singh, N. *Science* **1990**, *250*, 975.
- (340) Blanco, C.; Hochberg, D. *Phys. Chem. Chem. Phys.* **2011**, *13*, 12920.
- (341) Viedma, C. *Astrobiology* **2007**, *7*, 312.
- (342) Viedma, C. *Phys. Rev. Lett.* **2005**, *94*, 5504.
- (343) Kaptein, B.; Noorduin, W. L.; Meekes, H.; van Enckevort, W. J. P.; Kellogg, R. M.; Vlieg, E. *Angew. Chem., Int. Ed.* **2008**, *47*, 7226.
- (344) Noorduin, W. L.; Meekes, H.; van Enckevort, W. J. P.; Millemaggi, A.; Leeman, M.; Kaptein, B.; Kellogg, R. M.; Vlieg, E. *Angew. Chem., Int. Ed.* **2008**, *47*, 6445.
- (345) Noorduin, W. L.; Izumi, T.; Millemaggi, A.; Leeman, M.; Meekes, H.; Van Enckevort, W. J. P.; Kellogg, R. M.; Kaptein, B.; Vlieg, E.; Blackmond, D. G. *J. Am. Chem. Soc.* **2008**, *130*, 1158.
- (346) Viedma, C.; Noorduin, W. L.; Ortiz, J. E.; de Torres, T.; Cintas, P. *Chem. Commun.* **2011**, *47*, 671.
- (347) Viedma, C.; Ortiz, J. E.; de Torres, T.; Izumi, T.; Blackmond, D. G. *J. Am. Chem. Soc.* **2008**, *130*, 15274.
- (348) Viedma, C.; Verkuij, B. J. V.; Ortiz, J. E.; de Torres, T.; Kellogg, R. M.; Blackmond, D. G. *Chem.—Eur. J.* **2010**, *16*, 4932.
- (349) Noorduin, W. L.; van der Asdonk, P.; Meekes, H.; van Enckevort, W. J. P.; Kaptein, B.; Leeman, M.; Kellogg, R. M.; Vlieg, E. *Angew. Chem., Int. Ed.* **2009**, *48*, 3278.
- (350) Hein, J. E.; Cao, B. H.; Viedma, C.; Kellogg, R. M.; Blackmond, D. G. *J. Am. Chem. Soc.* **2012**, *134*, 12629.
- (351) See ref 289.
- (352) Ribo, J. M.; Crusats, J.; Sagues, F.; Claret, J.; Rubires, R. *Science* **2001**, *292*, 2063.
- (353) Escudero, C.; Crusats, J.; Diez-Perez, I.; El-Hachemi, Z.; Ribo, J. M. *Angew. Chem., Int. Ed.* **2006**, *45*, 8032.
- (354) Crusats, J.; El-Hachemi, Z.; Ribo, J. M. *Chem. Soc. Rev.* **2010**, *39*, 569.
- (355) Micali, N.; Engelkamp, H.; van Rhee, P. G.; Christianen, P. C. M.; Scolaro, L. M.; Maan, J. C. *Nat. Chem.* **2012**, *4*, 201.
- (356) Takats, Z.; Nanita, S. C.; Cooks, R. G. *Angew. Chem., Int. Ed.* **2003**, *42*, 3521.
- (357) Nanita, S. C.; Cooks, R. G. *Angew. Chem., Int. Ed.* **2006**, *45*, 554.
- (358) Julian, R. R.; Myung, S.; Clemmer, D. E. *J. Am. Chem. Soc.* **2004**, *126*, 4110.
- (359) Julian, R. R.; Myung, S.; Clemmer, D. E. *J. Phys. Chem. B* **2005**, *109*, 440.
- (360) Soai, K.; Shibata, T.; Morioka, H.; Choji, K. *Nature* **1995**, *378*, 767.
- (361) Shibata, T.; Yonekubo, S.; Soai, K. *Angew. Chem., Int. Ed.* **1999**, *38*, 659.
- (362) Shibata, T.; Yamamoto, J.; Matsumoto, N.; Yonekubo, S.; Osanai, S.; Soai, K. *J. Am. Chem. Soc.* **1998**, *120*, 12157.
- (363) Sato, I.; Sugie, R.; Matsueda, Y.; Furumura, Y.; Soai, K. *Angew. Chem., Int. Ed.* **2004**, *43*, 4490.
- (364) Kawasaki, T.; Sasagawa, T.; Shiozawa, K.; Uchida, M.; Suzuki, K.; Soai, K. *Org. Lett.* **2011**, *13*, 2361.
- (365) Mineki, H.; Hanasaki, T.; Matsumoto, A.; Kawasaki, T.; Soai, K. *Chem. Commun.* **2012**, *48*, 10538.
- (366) Kawasaki, T.; Matsumura, Y.; Tsutsumi, T.; Suzuki, K.; Ito, M.; Soai, K. *Science* **2009**, *324*, 492.
- (367) Blackmond, D. G.; McMillan, C. R.; Ramdeehul, S.; Schorm, A.; Brown, J. M. *J. Am. Chem. Soc.* **2001**, *123*, 10103.
- (368) Blackmond, D. G. *Tetrahedron: Asymmetry* **2006**, *17*, 584.
- (369) Gehring, T.; Quaranta, M.; Odell, B.; Blackmond, D. G.; Brown, J. M. *Angew. Chem., Int. Ed.* **2012**, *51*, 9539.
- (370) Girard, C.; Kagan, H. B. *Angew. Chem., Int. Ed.* **1998**, *37*, 2923.
- (371) Mauksch, M.; Tsogoeva, S. B.; Martynova, I. M.; Wei, S. W. *Angew. Chem., Int. Ed.* **2007**, *46*, 393.
- (372) Mauksch, M.; Tsogoeva, S. B.; Wei, S. W.; Martynova, I. M. *Chirality* **2007**, *19*, 816.
- (373) Tsogoeva, S. B. *Chem. Commun.* **2010**, *46*, 7662.
- (374) Mauksch, M.; Wei, S. W.; Freund, M.; Zamfir, A.; Tsogoeva, S. B. *Origins Life Evol. Biospheres* **2010**, *40*, 79.
- (375) Blackmond, D. G. *Chirality* **2009**, *21*, 359.
- (376) Tsogoeva, S. B.; Wei, S.; Freund, M.; Mauksch, M. *Angew. Chem., Int. Ed.* **2009**, *48*, 590.
- (377) Green, M. M.; Park, J. W.; Sato, T.; Teramoto, A.; Lifson, S.; Selinger, R. L. B.; Selinger, J. V. *Angew. Chem., Int. Ed.* **1999**, *38*, 3139.
- (378) Schmidt, P. *Origins Life Evol. Biospheres* **2006**, *36*, 391.
- (379) Brandenburg, A.; Andersen, A. C.; Hofner, S. H. R.; Nilsson, M. *Origins Life Evol. Biospheres* **2005**, *35*, 225.
- (380) Gleiser, M.; Walker, S. I. *Origins Life Evol. Biospheres* **2008**, *38*, 293.
- (381) Hochberg, D. *Chem. Phys. Lett.* **2010**, *491*, 237.
- (382) Blanco, C.; Hochberg, D. *Phys. Chem. Chem. Phys.* **2011**, *13*, 839.
- (383) Blair, N. E.; Bonner, W. A. *Origins Life Evol. Biospheres* **1981**, *11*, 331.
- (384) Green, M. M.; Peterson, N. C.; Sato, T.; Teramoto, A.; Cook, R.; Lifson, S. *Science* **1995**, *268*, 1860.
- (385) Brack, A.; Horneck, G.; Cockell, C. S.; Berces, A.; Belisheva, N. K.; Eiroa, C.; Henning, T.; Herbst, T.; Kaltenegger, L.; Leger, A.; Liseau, R.; Lammer, H.; Selsis, F.; Beichman, C.; Danchi, W.; Fridlund, M.; Lunine, J.; Paresce, F.; Penny, A.; Quirrenbach, A.; Rottgering, H.; Schneider, J.; Stam, D.; Tinetti, G.; White, G. *J. Astrobiology* **2010**, *10*, 69.
- (386) Hitz, T.; Luisi, P. L. *Helv. Chim. Acta* **2003**, *86*, 1423.
- (387) Hitz, T. H.; Luisi, P. L. *Origins Life Evol. Biospheres* **2004**, *34*, 93.

- (388) Weissbuch, I.; Bolbach, G.; Zepik, H.; Shavit, E.; Tang, M.; Frey, J.; Jensen, T. R.; Kjaer, K.; Leiserowitz, L.; Lahav, M. *J. Am. Chem. Soc.* **2002**, *124*, 9093.
- (389) Nery, J. G.; Bolbach, G.; Weissbuch, I.; Lahav, M. *Angew. Chem., Int. Ed.* **2003**, *42*, 2157.
- (390) Nery, J. G.; Bolbach, G.; Weissbuch, I.; Lahav, M. *Chem.—Eur. J.* **2005**, *11*, 3039.
- (391) Eliash, R.; Nery, J. G.; Rubinstein, I.; Clodic, G.; Bolbach, G.; Weissbuch, I.; Lahav, M. *Chem.—Eur. J.* **2007**, *13*, 10140.
- (392) Rubinstein, I.; Eliash, R.; Bolbach, G.; Weissbuch, I.; Lahav, M. *Angew. Chem., Int. Ed.* **2007**, *46*, 3710.
- (393) Rubinstein, I.; Clodic, G.; Bolbach, G.; Weissbuch, I.; Lahav, M. *Chem.—Eur. J.* **2008**, *14*, 10999.
- (394) Weissbuch, I.; Illos, R. A.; Bolbach, G.; Lahav, M. *Acc. Chem. Res.* **2009**, *42*, 1128.
- (395) Illos, R. A.; Clodic, G.; Bolbach, G.; Weissbuch, I.; Lahav, M. *Origins Life Evol. Biospheres* **2010**, *40*, 51.
- (396) Hitz, T.; Luisi, P. L. *Helv. Chim. Acta* **2002**, *85*, 3975.
- (397) Weissbuch, I.; Rubinstein, I.; Weygand, M. J.; Kjaer, K.; Leiserowitz, L.; Lahav, M. *Helv. Chim. Acta* **2003**, *86*, 3867.
- (398) Rubinstein, I.; Bolbach, G.; Weygand, M. J.; Kjaer, K.; Weissbuch, I.; Lahav, M. *Helv. Chim. Acta* **2003**, *86*, 3851.
- (399) Hitz, T.; Luisi, P. L. *Biopolymers* **2000**, *55*, 381.
- (400) Morigaki, K.; Dallavalle, S.; Walde, P.; Colonna, S.; Luisi, P. L. *J. Am. Chem. Soc.* **1997**, *119*, 292.
- (401) Lee, D. H.; Granja, J. R.; Martinez, J. A.; Severin, K.; Ghadiri, M. R. *Nature* **1996**, *382*, 525.
- (402) Lee, D. H.; Severin, K.; Yokobayashi, Y.; Ghadiri, M. R. *Nature* **1997**, *390*, 591.
- (403) Saghatelyan, A.; Yokobayashi, Y.; Soltani, K.; Ghadiri, M. R. *Nature* **2001**, *409*, 797.
- (404) Breslow, R. *Tetrahedron Lett.* **2011**, *52*, 2028.
- (405) Breslow, R. *Tetrahedron Lett.* **2011**, *52*, 4228.
- (406) Breslow, R.; Cheng, Z. L. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 5723.
- (407) Hein, J. E.; Tse, E.; Blackmond, D. G. *Nat. Chem.* **2011**, *3*, 704.
- (408) Lehn, J. M. *Supramolecular Chemistry: Concepts and Perspectives*; Wiley—VCH: New York, 1995.
- (409) Yu, H.; Zhang, S.; Chaput, J. C. *Nat. Chem.* **2012**, *4*, 183.
- (410) Karsenti, E. *Nat. Rev. Mol. Cell. Biol.* **2008**, *9*, 255.
- (411) Goldbeter, A. *Biochemical Oscillations and Cellular Rhythms*; Cambridge University Press: Cambridge, U.K., 1996.
- (412) Tyson, J. J. In *Computational Cell Biology*; Fall, C. P., Marland, E. S., Wagner, J. M., Tyson, J. J., Eds.; Springer-Verlag: New York, 2002.
- (413) McKane, A. J.; Nagy, J. D.; Newman, T. J.; Stefanini, M. O. *J. Stat. Phys.* **2007**, *128*, 165.
- (414) Ishiura, M.; Kutsuna, S.; Aoki, S.; Iwasaki, H.; Andersson, C. R.; Tanabe, A.; Golden, S. S.; Johnson, C. H.; Kondo, T. *Science* **1998**, *281*, 1519.
- (415) Hoffmann, A.; Levchenko, A.; Scott, M. L.; Baltimore, D. *Science* **2002**, *298*, 1241.
- (416) Nelson, D. E.; Ihekweaba, A. E. C.; Elliott, M.; Johnson, J. R.; Gibney, C. A.; Foreman, B. E.; Nelson, G.; See, V.; Horton, C. A.; Spiller, D. G.; Edwards, S. W.; McDowell, H. P.; Unitt, J. F.; Sullivan, E.; Grimley, R.; Benson, N.; Broomhead, D.; Kell, D. B.; White, M. R. *H. Science* **2004**, *306*, 704.
- (417) Lahav, G.; Rosenfeld, N.; Sigal, A.; Geva-Zatorsky, N.; Levine, A. J.; Elowitz, M. B.; Alon, U. *Nat. Genet.* **2004**, *36*, 147.
- (418) Hirata, H.; Yoshiura, S.; Ohtsuka, T.; Bessho, Y.; Harada, T.; Yoshikawa, K.; Kageyama, R. *Science* **2002**, *298*, 840.
- (419) Plotsky, P. M.; Vale, W. *Science* **1985**, *230*, 461.
- (420) Zeitler, P.; Tannenbaum, G. S.; Clifton, D. K.; Steiner, R. A. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 8920.
- (421) Tsai, T. Y. C.; Choi, Y. S.; Ma, W. Z.; Pomerening, J. R.; Tang, C.; Ferrell, J. E. *Science* **2008**, *321*, 126.
- (422) Keener, J.; Sneyd, J. *Mathematical Physiology*; Springer-Verlag: New York, 1998.
- (423) Harmer, S. L.; Hogenesch, L. B.; Straume, M.; Chang, H. S.; Han, B.; Zhu, T.; Wang, X.; Kreps, J. A.; Kay, S. A. *Science* **2000**, *290*, 2110.
- (424) Ripperger, J. A.; Schibler, U. *Curr. Opin. Cell Biol.* **2001**, *13*, 357.
- (425) Panda, S.; Antoch, M. P.; Miller, B. H.; Su, A. I.; Schook, A. B.; Straume, M.; Schultz, P. G.; Kay, S. A.; Takahashi, J. S.; Hogenesch, J. B. *Cell* **2002**, *109*, 307.
- (426) Tu, B. P.; Kudlicki, A.; Rowicka, M.; McKnight, S. L. *Science* **2005**, *310*, 1152.
- (427) Tu, B. P.; McKnight, S. L. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 696.
- (428) Mihalcescu, I.; Hsing, W. H.; Leibler, S. *Nature* **2004**, *430*, 81.
- (429) Lakin-Thomas, P. L.; Brody, S. *Annu. Rev. Microbiol.* **2004**, *58*, 489.
- (430) Liebermeister, W. *J. Theor. Biol.* **2005**, *234*, 423.
- (431) Parrondo, J. M. R.; Harmer, G. P.; Abbott, D. *Phys. Rev. Lett.* **2000**, *85*, 5226.
- (432) McClintock, P. V. E. *Nature* **1999**, *401*, 23.
- (433) Osipovitch, D. C.; Barratt, C.; Schwartz, P. M. *New J. Chem.* **2009**, *33*, 2022.
- (434) Barratt, C. *Int. J. Chem.* **2010**, *2*, 19.
- (435) Epstein, I. R.; Showalter, K. *J. Phys. Chem.* **1996**, *100*, 13132.
- (436) Epstein, I. R.; Pojman, J. A.; Steinbock, O. *Chaos* **2006**, *16*, No. 037101.
- (437) Biosa, G.; Bastianoni, S.; Rustici, M. *Chem.—Eur. J.* **2006**, *12*, 3430.
- (438) Epstein, I. R.; Vanag, V. K.; Balazs, A. C.; Kuksenok, O.; Dayal, P.; Bhattacharya, A. *Acc. Chem. Res.* **2012**, *45*, 2160.
- (439) Belousov, B. P. *Collection of Abstracts on Radiation Medicine; Medgiz: Moscow*, 1958.
- (440) Zhabotinsky, A. M. *Biophysik* **1964**, *9*, 329.
- (441) Zaikin, A. N.; Zhabotinsky, A. M. *Nature* **1970**, *225*, 535.
- (442) Noyes, R. M.; Field, R. J.; Koros, E. *J. Am. Chem. Soc.* **1972**, *94*, 1394.
- (443) Field, R. J.; Noyes, R. M.; Koros, E. *J. Am. Chem. Soc.* **1972**, *94*, 8649.
- (444) Dekepper, P.; Epstein, I. R.; Kustin, K. *J. Am. Chem. Soc.* **1981**, *103*, 2133.
- (445) Orban, M.; Dekepper, P.; Epstein, I. R.; Kustin, K. *Nature* **1981**, *292*, 816.
- (446) Rustici, M.; Branca, M.; Caravati, C.; Marchettini, N. *Chem. Phys. Lett.* **1996**, *263*, 429.
- (447) Biosa, G.; Masia, M.; Marchettini, N.; Rustici, M. *Chem. Phys.* **2005**, *308*, 7.
- (448) Vanag, V. K.; Epstein, I. R. *Phys. Rev. Lett.* **2001**, *87*, 8301.
- (449) Vanag, V. K.; Epstein, I. R. *Science* **2001**, *294*, 835.
- (450) Epstein, I. R.; Vanag, V. K. *Chaos* **2005**, *15*, 7510.
- (451) Bansagi, T.; Vanag, V. K.; Epstein, I. R. *Science* **2011**, *331*, 1309.
- (452) Castets, V.; Dulos, E.; Boissonade, J.; De kepper, P. *Phys. Rev. Lett.* **1990**, *64*, 2953.
- (453) Lee, K. J.; McCormick, W. D.; Ouyang, Q.; Swinney, H. L. *Science* **1993**, *261*, 192.
- (454) Ouyang, Q.; Swinney, H. L. *Nature* **1991**, *352*, 610.
- (455) Turing, A. M. *Philos. Trans. R. Soc., B* **1952**, *237*, 37.
- (456) Edelstein-Keshet, L. *Mathematical Models in Biology*; McGraw-Hill: New York, 1988.
- (457) Toth, R.; Taylor, A. F.; Tinsley, M. R. *J. Phys. Chem. B* **2006**, *110*, 10170.
- (458) Taylor, A. F.; Tinsley, M. R.; Wang, F.; Huang, Z. Y.; Showalter, K. *Science* **2009**, *323*, 614.
- (459) Taylor, A. F.; Tinsley, M. R.; Wang, F.; Showalter, K. *Angew. Chem., Int. Ed.* **2011**, *50*, 10161.
- (460) Toiya, M.; Gonzalez-Ochoa, H. O.; Vanag, V. K.; Fraden, S.; Epstein, I. R. *J. Phys. Chem. Lett.* **2010**, *1*, 1241.
- (461) Beutel, K. M.; Peacock-López, E. *J. Chem. Phys.* **2006**, *125*, 4908.
- (462) Strier, D. E.; Dawson, S. P. *PLoS One* **2007**, *2*, No. e1053.
- (463) Elowitz, M. B.; Leibler, S. *Nature* **2000**, *403*, 335.

- (464) Fung, E.; Wong, W. W.; Suen, J. K.; Bulter, T.; Lee, S. G.; Liao, J. C. *Nature* **2005**, *435*, 118.
- (465) Dadon, Z.; Wagner, N.; Ashkenasy, G. *Angew. Chem., Int. Ed.* **2008**, *47*, 6128.
- (466) Kauffman, S. J. *Theor. Biol.* **1986**, *119*, 1.
- (467) Ashkenasy, G.; Jagasia, R.; Yadav, M.; Ghadiri, M. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10872.
- (468) Vaidya, N.; Manapat, M. L.; Chen, I. A.; Xulvi-Brunet, R.; Hayden, E. J.; Lehman, N. *Nature* **2012**, *491*, 72.
- (469) Pereto, J. *Chem. Soc. Rev.* **2012**, *41*, 5394.
- (470) Letelier, J. C.; Soto-Andrade, J.; Abarzua, F. G.; Cornish-Bowden, A.; Cárdenas, M. L. *J. Theor. Biol.* **2006**, *238*, 949.
- (471) Cornish-Bowden, A.; Cárdenas, M. L. *J. Theor. Biol.* **2008**, *252*, 411.
- (472) Meléndez-Hevia, E.; Montero-Gómez, N.; Montero, F. J. *Theor. Biol.* **2008**, *252*, 505.
- (473) Cairns-Smith, A. G. *Genetic Takeover and the Mineral Origins of Life*; Cambridge University Press: Cambridge, U.K., 1982.
- (474) Szathmáry, E.; Santos, M.; Fernando, C. In *Prebiotic Chemistry: From Simple Amphiphiles to Protocell Models*; Walde, P., Ed.; Topics in Current Chemistry, Vol. 259; Springer-Verlag: Berlin/Heidelberg, 2005; Chapt. 5, p 167.
- (475) Schuster, P. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7678.
- (476) de Duve, C.; Miller, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **1991**, *88*, 10014.
- (477) de Duve, C. *Blueprint for a Cell: The Nature and Origin of Life*; Neil Patterson Publishers: Burlington, NC, 1991.
- (478) Shapiro, R. *Sci. Am.* **2007**, *296*, 46.
- (479) Wächtershäuser, G. *Microbiol. Rev.* **1988**, *52*, 452.
- (480) Morowitz, H. J.; Kostelnik, J. D.; Yang, J.; Cody, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7704.
- (481) Smith, E.; Morowitz, H. J. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 13168.
- (482) Orgel, L. E. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12503.
- (483) Vasas, V.; Szathmáry, E.; Santos, M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 1470.
- (484) Shapiro, R. *A Skeptic's Guide to the Creation of Life on Earth*; Submit Books: New York, 1986.
- (485) Benko, G.; Flamm, C.; Stadler, P. F. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1085.
- (486) Benko, G.; Flamm, C.; Stadler, P. F. *Lect. Notes Comput. Sci.* **2003**, *2801*, 10.
- (487) Huber, C.; Wächtershäuser, G. *Science* **2006**, *314*, 630.
- (488) Weber, A. L. *Origins Life Evol. Biospheres* **2002**, *32*, 333.
- (489) Weber, A. L. *Origins Life Evol. Biospheres* **2000**, *30*, 33.
- (490) Weber, A. L. *Origins Life Evol. Biospheres* **2004**, *34*, 473.
- (491) Weber, A. L. *Origins Life Evol. Biospheres* **1998**, *28*, 259.
- (492) Deamer, D.; Weber, A. L. *Cold Spring Harbor Perspect. Biol.* **2010**, *2*, No. a004929.
- (493) Weber, A. L. *Origins Life Evol. Biospheres* **2005**, *35*, 523.
- (494) Weber, A. L. *Origins Life Evol. Biospheres* **2007**, *37*, 105.
- (495) Gottschalk, G. *Bacterial Metabolism*; Springer: Berlin, Heidelberg, New York, 1986.
- (496) Heinrich, R.; Schuster, S. *The Regulation of Cellular Systems*; Chapman & Hall: New York, 1996.
- (497) Meléndez-Hevia, E.; Montero-Gómez, N.; Montero, F. J. *Theor. Biol.* **2008**, *252*, 505.
- (498) Pereto, J.; Lopez-Garcia, P.; Moreira, D. *Trends Biochem. Sci.* **2004**, *29*, 469.
- (499) Lombard, J.; Moreira, D. *Mol. Biol. Evol.* **2011**, *28*, 87.
- (500) Mouritsen, O. *Life - As a Matter of Fat: The Emerging Science of Lipidomics*; Springer: New York, 2005.
- (501) Morigaki, K.; Walde, P. *Curr. Opin. Colloid Interface Sci.* **2007**, *12*, 75.
- (502) de Clue, M. S.; Monnard, P. A.; Bailey, J. A.; Maurer, S. E.; Collis, G. E.; Ziock, H. J.; Rasmussen, S.; Boncella, J. M. *J. Am. Chem. Soc.* **2009**, *131*, 931.
- (503) Zhu, T. F.; Szostak, J. W. *J. Am. Chem. Soc.* **2009**, *131*, 5705.
- (504) Hanczyc, M. M. In *Protocells: Bridging Nonliving and Living Matter*; Rasmussen, S., Bedau, M. A., Chen, L., Deamer, D., Krakauer, D. C., Packard, N. H., Stadler, P. F., Eds.; MIT Press: Cambridge, MA, 2009.
- (505) Which has become a field of research in itself since Bangham's pioneering work in the 1960s: (a) Bangham, A. D.; Horne, R. W. *J. Mol. Biol.* **1964**, *8*, 660. (b) Johnson, S. M.; Bangham, A. D. *Biochim. Biophys. Acta* **1969**, *193*, 82.
- (506) Ourisson, G.; Nakatani, Y. *Chem. Biol.* **1995**, *2*, 631.
- (507) Deamer, D. W.; Dworkin, J. P. In *Prebiotic Chemistry: From Simple Amphiphiles to Protocell Models*; Walde, P., Ed.; Topics in Current Chemistry, Vol. 259; Springer-Verlag: Berlin/Heidelberg, 2005; Chapt. 1, p 1.
- (508) Wachtershäuser, G. *Mol. Microbiol.* **2003**, *47*, 13.
- (509) Koga, Y. *J. Mol. Evol.* **2011**, *72*, 274.
- (510) Gebicki, J. M.; Hicks, M. *Nature* **1973**, *243*, 232.
- (511) Hargreaves, W. R.; Deamer, D. W. *Biochemistry* **1978**, *17*, 3759.
- (512) Pozzi, G.; Birault, V.; Werner, B.; Dannenmuller, O.; Nakatani, Y.; Ourisson, G.; Terakawa, S. *Angew. Chem., Int. Ed.* **1996**, *35*, 177.
- (513) Gotoh, M.; Sugawara, A.; Akiyoshi, K.; Matsumoto, I.; Ourisson, G.; Nakatani, Y. *Chem. Biodiversity* **2007**, *4*, 837.
- (514) The critical aggregation concentration (cac) refers to the threshold of concentration values required for supramolecular structures to start forming once the aqueous solution becomes saturated with the amphiphilic monomer.
- (515) Cistola, D. P.; Atkinson, D.; Hamilton, J. A.; Small, D. M. *Biochemistry* **1986**, *25*, 2804.
- (516) Cistola, D. P.; Hamilton, J. A.; Jackson, D.; Small, D. M. *Biochemistry* **1988**, *27*, 1881.
- (517) Haines, T. H. *Proc. Natl. Acad. Sci. U.S.A.* **1983**, *80*, 160.
- (518) Dejanovic, B.; Noethig-Laslo, V.; Sentjurc, M.; Walde, P. *Chem. Phys. Lipids* **2011**, *164*, 83.
- (519) Monnard, P. A.; Apel, C. L.; Kanavarioti, A.; Deamer, D. W. *Astrobiology* **2002**, *2*, 139.
- (520) Rendon, A.; Gil Carton, D.; Sot, J.; Garcia-Pacios, M.; Montes, R.; Valle, M.; Arrondo, J.-L. R.; Goñi, F. M.; Ruiz-Mirazo, K. *Biophys. J.* **2012**, *102*, 278.
- (521) Namani, T.; Deamer, D. W. *Origins Life Evol. Biospheres* **2008**, *38*, 329.
- (522) Ehrenfreund, P.; Rasmussen, S.; Cleaves, J.; Chen, L. H. *Astrobiology* **2006**, *6*, 490.
- (523) Groen, J.; Deamer, D. W.; Kros, A.; Ehrenfreund, P. *Origins Life Evol. Biospheres* **2012**, *42*, 295.
- (524) Chen, I. A.; Roberts, R. W.; Szostak, J. W. *Science* **2004**, *305*, 1474.
- (525) Mansy, S. S.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13351.
- (526) Mansy, S. S.; Schrum, J. P.; Krishnamurthy, M.; Tobe, S.; Treco, D. A.; Szostak, J. W. *Nature* **2008**, *454*, 122.
- (527) Deamer, D. W. *Nature* **2008**, *454*, 37.
- (528) Apel, C. L.; Deamer, D. W. *Origins Life Evol. Biospheres* **2005**, *35*, 323.
- (529) Budin, I.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 5249.
- (530) Ruiz-Mirazo, K.; Mavelli, F. In *Advances in Artificial Life*; Almeida e Costa, F., Mateus-Rocha, L., Costa, E., Harvey, I., Coutinho, A., Eds.; Springer: Heidelberg, Germany, 2007.
- (531) Ruiz-Mirazo, K.; Mavelli, F. *BioSystems* **2008**, *91*, 374.
- (532) Kaiser, H. J.; Orlowski, A.; Rog, T.; Nyholm, T. K. M.; Chai, W. G.; Feizi, T.; Lingwood, D.; Vattulainen, I.; Simons, K. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 16628.
- (533) Vlassov, A.; Khvorova, A.; Yarus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 7706.
- (534) Janas, T.; Janas, T.; Yarus, M. *RNA* **2004**, *10*, 1541.
- (535) Janas, T.; Janas, T.; Yarus, M. *Nucleic Acids Res.* **2006**, *34*, 2128.
- (536) Chakrabarti, A. C.; Breaker, R. R.; Joyce, G. F.; Deamer, D. W. *J. Mol. Evol.* **1994**, *39*, 555.
- (537) Given its importance in the field of origins of life, the RNA world hypothesis will be dealt with in a section of its own (section

- 3.3), although we are aware that some of the issues might partially overlap.
- (538) Dyson, F. J. *Origins of Life*; Cambridge University Press: Cambridge, U.K., 1985.
- (539) Eigen, M.; Schuster, P. *The Hypercycle. A Principle of Natural Self-Organization*; Springer: Berlin, 1979.
- (540) Maynard-Smith, J.; Szathmary, E. *J. Theor. Biol.* **1993**, *164*, 437.
- (541) Szathmary, E. *Philos. Trans. R. Soc., B* **2006**, *361*, 1761.
- (542) Bozic, B.; Svetina, S. *Eur. Biophys. J.* **2004**, *33*, 565.
- (543) Svetina, S. *ChemPhysChem* **2009**, *10*, 2769.
- (544) Walde, P. In *Encyclopedia of Nanoscience and Nanotechnology*; Nalwa, H. S., Ed.; American Scientific Publishers: Valencia, CA, 2004; Vol. 9.
- (545) Toyota, T.; Takakura, K.; Kageyama, Y.; Kurihara, K.; Maru, N.; Ohnuma, K.; Kaneko, K.; Sugawara, T. *Langmuir* **2008**, *24*, 3037.
- (546) Takakura, K.; Toyota, T.; Sugawara, T. *J. Am. Chem. Soc.* **2003**, *125*, 8134.
- (547) Wick, R.; Walde, P.; Luisi, P. L. *J. Am. Chem. Soc.* **1995**, *117*, 1435.
- (548) Stano, P.; Luisi, P. L. *Chem. Commun.* **2010**, *46*, 3639.
- (549) Bachmann, P. A.; Walde, P.; Luisi, P. L.; Lang, J. *J. Am. Chem. Soc.* **1991**, *113*, 8204.
- (550) Bachmann, P. A.; Luisi, P. L.; Lang, J. *Nature* **1992**, *357*, 57.
- (551) Walde, P.; Wick, R.; Fresta, M.; Mangone, A.; Luisi, P. L. *J. Am. Chem. Soc.* **1994**, *116*, 11649.
- (552) Blochlinger, E.; Blocher, M.; Walde, P.; Luisi, P. L. *J. Phys. Chem. B* **1998**, *102*, 10383.
- (553) Chen, I. A.; Szostak, J. W. *Biophys. J.* **2004**, *87*, 988.
- (554) Berclaz, N.; Muller, M.; Walde, P.; Luisi, P. L. *J. Phys. Chem. B* **2001**, *105*, 1056.
- (555) Stano, P.; Wehrli, E.; Luisi, P. L. *J. Phys.: Condens. Matter* **2006**, *18*, S2231.
- (556) Luisi, P. L. *The Emergence of Life: From Chemical Origins to Synthetic Biology*; Cambridge University Press: Cambridge, U.K., 2006.
- (557) Budin, I.; Debnath, A.; Szostak, J. W. *J. Am. Chem. Soc.* **2012**, *134*, 20812.
- (558) Zhu, T. F.; Adamala, K.; Zhang, N.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 9828.
- (559) Cheng, Z. L.; Luisi, P. L. *J. Phys. Chem. B* **2003**, *107*, 10940.
- (560) Chen, I. A.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7965.
- (561) Segre, D.; Ben-Eli, D.; Lancet, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 4112.
- (562) Eigen, M. *Naturwissenschaften* **1971**, *58*, 465.
- (563) Orgel, L. E. *Proc. R. Soc. London, B* **1979**, *205*, 435.
- (564) Spiegelman, S. Q. *Rev. Biophys.* **1971**, *4*, 213.
- (565) Maynard-Smith, J. *Nature* **1979**, *280*, 445.
- (566) Szathmary, E.; Demeter, L. *J. Theor. Biol.* **1987**, *128*, 463.
- (567) Boerlijst, M. C.; Hogeweg, P. *Physica D* **1991**, *48*, 17.
- (568) Oehlenschläger, F.; Eigen, M. *Origins Life Evol. Biospheres* **1997**, *27*, 437.
- (569) Robertson, A.; Sinclair, A. J.; Philp, D. *Chem. Soc. Rev.* **2000**, *29*, 141.
- (570) Paul, N.; Joyce, G. F. *Curr. Opin. Chem. Biol.* **2004**, *8*, 634.
- (571) von Kiedrowski, G. *Bioorg. Chem. Front.* **1993**, *3*, 113.
- (572) Lee, D. H.; Severin, K.; Ghadiri, M. R. *Curr. Opin. Chem. Biol.* **1997**, *1*, 491.
- (573) Ghosh, I.; Chrnielewski, J. *Curr. Opin. Chem. Biol.* **2004**, *8*, 640.
- (574) Schneider-Bernloehr, H.; Lohrmann, R.; Sulston, J.; Weimann, B. J.; Orgel, L. E.; Miles, H. T. *J. Mol. Biol.* **1968**, *37*, 151.
- (575) Orgel, L. E. *J. Mol. Biol.* **1968**, *38*, 381.
- (576) Inoue, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1981**, *103*, 7666.
- (577) Orgel, L. E. *Nature* **1992**, *358*, 203.
- (578) von Kiedrowski, G. *Angew. Chem., Int. Ed.* **1986**, *25*, 932.
- (579) von Kiedrowski, G.; Wlotzka, B.; Helbing, J.; Matzen, M.; Jordan, S. *Angew. Chem., Int. Ed.* **1991**, *30*, 423.
- (580) Patzke, V.; von Kiedrowski, G. *ARKIVOC* **2007**, *5*, 293.
- (581) Zielinski, W. S.; Orgel, L. E. *Nature* **1987**, *327*, 346.
- (582) von Kiedrowski, G.; Wlotzka, B.; Helbing, J. *Angew. Chem., Int. Ed.* **1989**, *28*, 1235.
- (583) Paul, N.; Joyce, G. F. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12733.
- (584) Tjivikua, T.; Ballester, P.; Rebek, J. *J. Am. Chem. Soc.* **1990**, *112*, 1249.
- (585) Nowick, J. S.; Feng, Q.; Tjivikua, T.; Ballester, P.; Rebek, J. *J. Am. Chem. Soc.* **1991**, *113*, 8831.
- (586) Wintner, E. A.; Conn, M. M.; Rebek, J. *J. Am. Chem. Soc.* **1994**, *116*, 8877.
- (587) Menger, F. M.; Eliseev, A. V.; Khanjin, N. A.; Sherrod, M. J. *J. Org. Chem.* **1995**, *60*, 2870.
- (588) Reinhoudt, D. N.; Rudkevich, D. M.; deJong, F. *J. Am. Chem. Soc.* **1996**, *118*, 6880.
- (589) Kassianidis, E.; Philp, D. *Angew. Chem., Int. Ed.* **2006**, *45*, 6344.
- (590) Kassianidis, E.; Philp, D. *Chem. Commun.* **2006**, *39*, 4072.
- (591) Issac, R.; Chmielewski, J. *J. Am. Chem. Soc.* **2002**, *124*, 6808.
- (592) Li, X. Q.; Chmielewski, J. *J. Am. Chem. Soc.* **2003**, *125*, 11820.
- (593) Rubinov, B.; Wagner, N.; Rapaport, H.; Ashkenasy, G. *Angew. Chem., Int. Ed.* **2009**, *48*, 6683.
- (594) Maury, C. P. J. *Origins Life Evol. Biospheres* **2009**, *39*, 141.
- (595) Carny, O.; Gazit, E. *Origins Life Evol. Biospheres* **2011**, *41*, 121.
- (596) Hammarström, P.; Ali, M. M.; Mishra, M.; Salagic, B.; Svensson, S.; Tengvall, P.; Lundström, I. *Origins Life Evol. Biospheres* **2011**, *41*, 373.
- (597) Wagner, N.; Ashkenasy, G. *J. Chem. Phys.* **2009**, *130*, No. 164907.
- (598) Sievers, D.; von Kiedrowski, G. *Nature* **1994**, *369*, 221.
- (599) Sievers, D.; von Kiedrowski, G. *Chem.—Eur. J.* **1998**, *4*, 629.
- (600) Achilles, T.; von Kiedrowski, G. *Angew. Chem., Int. Ed.* **1993**, *32*, 1198.
- (601) Hayden, E. J.; von Kiedrowski, G.; Lehman, N. *Angew. Chem., Int. Ed.* **2008**, *47*, 8424.
- (602) Hordijk, W.; Steel, M. J. *Syst. Chem.* **2013**, DOI: 10.1186/1759-2208-4-3.
- (603) Hong, J. I.; Feng, Q.; Rotello, V.; Rebek, J. *Science* **1992**, *255*, 848.
- (604) Pieters, R. J.; Huc, I.; Rebek, J. *Angew. Chem., Int. Ed.* **1994**, *33*, 1579.
- (605) Severin, K.; Lee, D. H.; Martinez, J. A.; Vieth, M.; Ghadiri, M. R. *Angew. Chem., Int. Ed.* **1998**, *37*, 126.
- (606) Ashkenasy, G.; Ghadiri, M. R. *J. Am. Chem. Soc.* **2004**, *126*, 11140.
- (607) Yao, S.; Ghosh, I.; Zutshi, R.; Chmielewski, J. *Nature* **1998**, *396*, 447.
- (608) Santoro, S. W.; Joyce, G. F. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 4262.
- (609) Levy, M.; Ellington, A. D. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 6416.
- (610) Levy, M.; Ellington, A. D. *J. Mol. Evol.* **2003**, *56*, 607.
- (611) Chen, J.; Korner, S.; Craig, S. L.; Rudkevich, D. M.; Rebek, J. *Nature* **2002**, *415*, 385.
- (612) Chen, J.; Korner, S.; Craig, S. L.; Lin, S.; Rudkevich, D. M.; Rebek, J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 2593.
- (613) Koonin, E. V. *Nat. Rev. Microbiol.* **2003**, *1*, 127.
- (614) Becerra, A.; Delaey, L.; Islas, S.; Lazcano, A. In *Annual Review of Ecology Evolution and Systematics*; Annual Reviews: Palo Alto, CA, 2007; Vol. 38.
- (615) Flint, S.; Enquist, L.; Racaniello, V.; Skalka, A. *Principles of Virology*, 3rd ed.; ASM Press: Washington, DC, 2009.
- (616) Woese, C. R. *Proc. Natl. Acad. Sci. U.S.A.* **1965**, *54*, 1546.
- (617) Crick, F. H. C. *J. Mol. Biol.* **1968**, *38*, 367.
- (618) Gilbert, W. *Nature* **1986**, *319*, 618.
- (619) Motorin, Y.; Helm, M. *Biochemistry* **2010**, *49*, 4934.
- (620) Piccirilli, J. A.; Krauch, T.; Moroney, S. E.; Benner, S. A. *Nature* **1990**, *343*, 33.
- (621) Watson, J. D.; Crick, F. H. C. *Nature* **1953**, *171*, 737.
- (622) Benner, S. A.; Battersby, T. R.; Eschgfaller, B.; Hutter, D.; Kodra, J. T.; Lutz, S.; Arslan, T.; Baschlin, D. K.; Blattler, M.; Egli, M;

- Hammer, C.; Held, H. A.; Horlacher, J.; Huang, Z.; Hyrup, B.; Jenny, T. F.; Jurczyk, S. C.; Konig, M.; von Krosigk, U.; Lutz, M. J.; MacPherson, L. J.; Moroney, S. E.; Muller, E.; Nambiar, K. P.; Piccirilli, J. A.; Switzer, C. Y.; Vogel, J. J.; Richert, C.; Roughton, A. L.; Schmidt, J.; Schneider, K. C.; Stackhouse, J. *Pure Appl. Chem.* **1998**, *70*, 263.
- (623) Geyer, C. R.; Battersby, T. R.; Benner, S. A. *Structure* **2003**, *11*, 1485.
- (624) Yang, Z. Y.; Sismour, A. M.; Sheng, P. P.; Puskar, N. L.; Benner, S. A. *Nucleic Acids Res.* **2007**, *35*, 4238.
- (625) Callahan, M. P.; Smith, K. E.; Cleaves, H. J.; Ruzicka, J.; Stern, J. C.; Glavin, D. P.; House, C. H.; Dworkin, J. P. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 13995.
- (626) Eschenmoser, A. *Science* **1999**, *284*, 2118.
- (627) Schneider, K. C.; Benner, S. A. *J. Am. Chem. Soc.* **1990**, *112*, 453.
- (628) Chaput, J. C.; Switzer, C. *J. Mol. Evol.* **2000**, *51*, 464.
- (629) Chen, J. J.; Tsai, C. H.; Cai, X.; Horhota, A. T.; McLaughlin, L. W.; Szostak, J. W. *PLoS One* **2009**, *4*, No. e4949.
- (630) Chen, J. J.; Cai, X.; Szostak, J. W. *J. Am. Chem. Soc.* **2009**, *131*, 2119.
- (631) Schoning, K. U.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. *Science* **2000**, *290*, 1347.
- (632) Chaput, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 9274.
- (633) Orgel, L. *Science* **2000**, *290*, 1306.
- (634) Eschenmoser, A. *Origins Life Evol. Biospheres* **2004**, *34*, 277.
- (635) Pitsch, S.; Wendeborn, S.; Jaun, B.; Eschenmoser, A. *Helv. Chim. Acta* **1993**, *76*, 2161.
- (636) Micura, R.; Kudick, R.; Pitsch, S.; Eschenmoser, A. *Angew. Chem., Int. Ed.* **1999**, *38*, 680.
- (637) Micura, R.; Bolli, M.; Windhab, N.; Eschenmoser, A. *Angew. Chem., Int. Ed.* **1997**, *36*, 870.
- (638) Bolli, M.; Micura, R.; Eschenmoser, A. *Chem. Biol.* **1997**, *4*, 309.
- (639) Bolli, M.; Micura, R.; Pitsch, S.; Eschenmoser, A. *Helv. Chim. Acta* **2004**, *80*, 1901.
- (640) Lescrinier, E.; Froeyen, M.; Herdewijn, P. *Nucleic Acids Res.* **2003**, *31*, 2975.
- (641) Schrum, J. P.; Ricardo, A.; Krishnamurthy, M.; Blain, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2009**, *131*, 14560.
- (642) Zhang, S.; Zhang, N.; Blain, J. C.; Szostak, J. W. *J. Am. Chem. Soc.* **2013**, *135*, 924.
- (643) Tarkoy, M.; Bolli, M.; Schweizer, B.; Leumann, C. *Helv. Chim. Acta* **1993**, *76*, 481.
- (644) Ahn, D. R.; Egger, A.; Lehmann, C.; Pitsch, S.; Leumann, C. J. *Chem.—Eur. J.* **2002**, *8*, 5312.
- (645) Obika, S.; Nanbu, D.; Hari, Y.; Morio, K.; In, Y.; Ishida, T.; Imanishi, T. *Tetrahedron Lett.* **1997**, *38*, 8735.
- (646) Obika, S.; Nanbu, D.; Hari, Y.; Andoh, J.; Morio, K.; Doi, T.; Imanishi, T. *Tetrahedron Lett.* **1998**, *39*, 5401.
- (647) Koshkin, A. A.; Singh, S. K.; Nielsen, P.; Rajwanshi, V. K.; Kumar, R.; Meldgaard, M.; Olsen, C. E.; Wengel, J. *Tetrahedron* **1998**, *54*, 3607.
- (648) Braasch, D. A.; Corey, D. R. *Chem. Biol.* **2001**, *8*, 1.
- (649) Petersen, M.; Wengel, J. *Trends Biotechnol.* **2003**, *21*, 74.
- (650) Doessing, H.; Vester, B. *Molecules* **2011**, *16*, 4511.
- (651) Pinheiro, V. P.; Taylor, A. I.; Cozens, C.; Abramov, M.; Renders, M.; Zhang, S.; Chaput, J. C.; Wengel, J.; Peak-Chew, S. Y.; McLaughlin, S. H.; Herdewijn, P.; Holliger, P. *Science* **2012**, *336*, 341.
- (652) Pinheiro, V. P.; Holliger, P. *Curr. Opin. Chem. Biol.* **2012**, *16*, 245.
- (653) Schneider, K. C.; Benner, S. A. *Tetrahedron Lett.* **1990**, *31*, 335.
- (654) Nielsen, P. E.; Egholm, M.; Berg, R. H.; Buchardt, O. *Science* **1991**, *254*, 1497.
- (655) Egholm, M.; Buchardt, O.; Christensen, L.; Behrens, C.; Freier, S. M.; Driver, D. A.; Berg, R. H.; Kim, S. K.; Norden, B.; Nielsen, P. E. *Nature* **1993**, *365*, 566.
- (656) Menchise, V.; De Simone, G.; Tedeschi, T.; Corradini, R.; Sforza, S.; Marchelli, R.; Capasso, D.; Saviano, M.; Pedone, C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12021.
- (657) Wittung, P.; Nielsen, P. E.; Buchardt, O.; Egholm, M.; Norden, B. *Nature* **1994**, *368*, 561.
- (658) Eriksson, M.; Nielsen, P. E. *Nat. Struct. Biol.* **1996**, *3*, 410.
- (659) Ratilainen, T.; Holmen, A.; Tuite, E.; Haaima, G.; Christensen, L.; Nielsen, P. E.; Norden, B. *Biochemistry* **1998**, *37*, 12331.
- (660) Nelson, K. E.; Levy, M.; Miller, S. L. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 3868.
- (661) Meierhenrich, U. J.; Caro, G. M. M.; Bredehoft, J. H.; Jessberger, E. K.; Thiemann, W. H. P. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9182.
- (662) Eriksson, M.; Christensen, L.; Schmidt, J.; Haaima, G.; Orgel, L.; Nielsen, P. E. *New J. Chem.* **1998**, *22*, 1055.
- (663) Brudno, Y.; Birnbaum, M. E.; Kleiner, R. E.; Liu, D. R. *Nat. Chem. Biol.* **2010**, *6*, 148.
- (664) Schmidt, J. G.; Nielsen, P. E.; Orgel, L. E. *Nucleic Acids Res.* **1997**, *25*, 4797.
- (665) Schmidt, J. G.; Christensen, L.; Nielsen, P. E.; Orgel, L. E. *Nucleic Acids Res.* **1997**, *25*, 4792.
- (666) Matysiak, S.; Reuthner, F.; Hoheisel, J. D. *Biotechniques* **2001**, *31*, 896.
- (667) Briones, C.; Mateo-Martí, E.; Gomez-Navarro, C.; Parro, V.; Roman, E.; Martin-Gago, J. A. *Phys. Rev. Lett.* **2004**, *93*, No. 208103.
- (668) Mateo-Martí, E.; Briones, C.; Rogero, C.; Gomez-Navarro, C.; Methivier, C.; Pradier, C. M.; Martin-Gago, J. A. *Chem. Phys.* **2008**, *352*, 11.
- (669) Murtola, M.; Stromberg, R. *Org. Biomol. Chem.* **2008**, *6*, 3837.
- (670) Murtola, M.; Wenska, M.; Stromberg, R. *J. Am. Chem. Soc.* **2010**, *132*, 8984.
- (671) Ura, Y.; Beierle, J. M.; Leman, L. J.; Orgel, L. E.; Ghadiri, M. R. *Science* **2009**, *325*, 73.
- (672) Nielsen, P. E. *Chem. Biodiversity* **2007**, *4*, 1996.
- (673) Nielsen, P. E. *Origins Life Evol. Biospheres* **2007**, *37*, 323.
- (674) Engelhart, A. E.; Hud, N. V. *Cold Spring Harbor Perspect. Biol.* **2010**, *2*, No. a002196.
- (675) Zappulla, D. C.; Cech, T. R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10024.
- (676) Tucker, B. J.; Breaker, R. R. *Curr. Opin. Struct. Biol.* **2005**, *15*, 342.
- (677) Strobel, S. A.; Cochrane, J. C. *Curr. Opin. Chem. Biol.* **2007**, *11*, 636.
- (678) Filbin, M. E.; Kieft, J. S. *Curr. Opin. Struct. Biol.* **2009**, *19*, 267.
- (679) Khraiwesh, B.; Arif, M. A.; Seumel, G. I.; Ossowski, S.; Weigel, D.; Reski, R.; Frank, W. *Cell* **2010**, *140*, 111.
- (680) Noller, H. F. *Cold Spring Harbor Perspect. Biol.* **2010**, *4*, No. a003681.
- (681) Djuranovic, S.; Nahvi, A.; Green, R. *Science* **2011**, *331*, 550.
- (682) Hendrix, D. K.; Brenner, S. E.; Holbrook, S. R. Q. *Rev. Biophys.* **2005**, *38*, 221.
- (683) Cruz, J. A.; Westhof, E. *Cell* **2009**, *136*, 604.
- (684) Huynen, M. A.; Stadler, P. F.; Fontana, W. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 397.
- (685) Ancel, L. W.; Fontana, W. *J. Exp. Zool.* **2000**, *288*, 242.
- (686) Schuster, P. *Rep. Prog. Phys.* **2006**, *69*, 1419.
- (687) Baltimore, D. *Nature* **1970**, *226*, 1209.
- (688) Temin, H. M.; Mizutani, S. *Nature* **1970**, *226*, 1211.
- (689) Cech, T. R.; Zaugg, A. J.; Grabowski, P. J. *Cell* **1981**, *27*, 487.
- (690) Kruger, K.; Grabowski, P. J.; Zaugg, A. J.; Sands, J.; Gottschling, D. E.; Cech, T. R. *Cell* **1982**, *31*, 147.
- (691) Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. *Cell* **1983**, *35*, 849.
- (692) Hill, H. B. *J. Mol. Evol.* **1976**, *7*, 101.
- (693) Doherty, E. A.; Doudna, J. A. *Annu. Rev. Biochem.* **2000**, *69*, 597.
- (694) Puerta-Fernandez, E.; Romero-Lopez, C.; Barroso-delJesus, A.; Berzal-Herranz, A. *FEMS Microbiol. Rev.* **2003**, *27*, 75.
- (695) Lilley, D. M. J. *Curr. Opin. Struct. Biol.* **2005**, *15*, 313.

- (696) Haugen, P.; Simon, D. M.; Bhattacharya, D. *Trends Genet.* **2005**, *21*, 111.
- (697) Bonen, L.; Vogel, J. *Trends Genet.* **2001**, *17*, 322.
- (698) Evans, D.; Marquez, S. M.; Pace, N. R. *Trends Biochem. Sci.* **2006**, *31*, 333.
- (699) Salehi-Ashtiani, K.; Szostak, J. W. *Nature* **2001**, *414*, 82.
- (700) Das, S. R.; Piccirilli, J. A. *Nat. Chem. Biol.* **2005**, *1*, 45.
- (701) Nissen, P.; Hansen, J.; Ban, N.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 920.
- (702) Yusupov, M. M.; Yusupova, G. Z.; Baucom, A.; Lieberman, K.; Earnest, T. N.; Cate, J. H. D.; Noller, H. F. *Science* **2001**, *292*, 883.
- (703) Beringer, M.; Rodnina, M. V. *Mol. Cell* **2007**, *26*, 311.
- (704) Mills, D. R.; Peterson, R. L.; Spiegelman, S. *Proc. Natl. Acad. Sci. U.S.A.* **1967**, *58*, 217.
- (705) Szathmáry, E. *Nature* **1990**, *344*, 115.
- (706) Tuerk, C.; Gold, L. *Science* **1990**, *249*, 505.
- (707) Ellington, A. D.; Szostak, J. W. *Nature* **1990**, *346*, 818.
- (708) Unrau, P. J.; Bartel, D. P. *Nature* **1998**, *395*, 260.
- (709) Joyce, G. F. *Annu. Rev. Biochem.* **2004**, *73*, 791.
- (710) Cech, T. R. *Cell* **2009**, *136*, 599.
- (711) Stich, M.; Briones, C.; Manrubia, S. C. *J. Theor. Biol.* **2008**, *252*, 750.
- (712) Manrubia, S. C.; Briones, C. *RNA* **2007**, *13*, 97.
- (713) Rohatgi, R.; Bartel, D. P.; Szostak, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 3332.
- (714) Pino, S.; Ciciriello, F.; Costanzo, G.; Di Mauro, E. *J. Biol. Chem.* **2008**, *283*, 36494.
- (715) James, K. D.; Ellington, A. D. *Origins Life Evol. Biospheres* **1999**, *29*, 375.
- (716) Bartel, D. P.; Szostak, J. W. *Science* **1993**, *261*, 1411.
- (717) Meyer, A. J.; Ellefson, J. W.; Ellington, A. D. *Acc. Chem. Res.* **2012**, *45*, 2097.
- (718) Lincoln, T. A.; Joyce, G. F. *Science* **2009**, *323*, 1229.
- (719) Ferretti, A. C.; Joyce, G. F. *Biochemistry* **2013**, *S2*, 1227.
- (720) Olea, C., Jr.; Horning, D. P.; Joyce, G. F. *J. Am. Chem. Soc.* **2012**, *134*, 8050.
- (721) Johnston, W. K.; Unrau, P. J.; Lawrence, M. S.; Glasner, M. E.; Bartel, D. P. *Science* **2001**, *292*, 1319.
- (722) McGinness, K. E.; Joyce, G. F. *Chem. Biol.* **2002**, *9*, 297.
- (723) Shechner, D. M.; Bartel, D. P. *Nat. Struct. Mol. Biol.* **2011**, *18*, 1036.
- (724) Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. *Science* **2011**, *332*, 209.
- (725) Takeuchi, N.; Hogeweg, P. *Biol. Direct* **2008**, *3*, 11.
- (726) Lehman, N. *J. Mol. Evol.* **2003**, *56*, 770.
- (727) Hambly, E.; Suttle, C. A. *Curr. Opin. Microbiol.* **2005**, *8*, 444.
- (728) Domingo, E. In *Fields Virology*; Knipe, D., Howley, P., Eds.; Lippincott Williams & Wilkins: Philadelphia, PA, 2007.
- (729) Siefert, J. L. *Methods Mol. Biol.* **2009**, *532*, 13.
- (730) Diener, T. O. *Nat. Rev. Microbiol.* **2003**, *1*, 75.
- (731) Flores, R.; Delgado, S.; Gas, M. E.; Carbonell, A.; Molina, D.; Gago, S.; De la Pena, M. *FEBS Lett.* **2004**, *567*, 42.
- (732) Diener, T. O. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 9370.
- (733) Forterre, P.; Prangishvili, D. *Res. Microbiol.* **2009**, *160*, 466.
- (734) Moreira, D.; Lopez-Garcia, P. *Nat. Rev. Microbiol.* **2009**, *7*, 306.
- (735) Drake, J. W.; Holland, J. J. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 13910.
- (736) Gago, S.; Elena, S. F.; Flores, R.; Sanjuan, R. *Science* **2009**, *323*, 1308.
- (737) Domingo, E.; Sabo, D.; Taniguchi, T.; Weissmann, C. *Cell* **1978**, *13*, 735.
- (738) Eigen, M. *Trends Microbiol.* **1996**, *4*, 216.
- (739) Domingo, E.; Holland, J. J. *Annu. Rev. Microbiol.* **1997**, *51*, 151.
- (740) Domingo, E.; Biebricher, C. K.; Holland, J. J.; Eigen, M. *Quasispecies and RNA Virus Evolution: Principles and Consequences*; Landes Bioscience: Austin, TX, 2001.
- (741) Eigen, M.; Biebricher, C. K. In *RNA Genetics*; Domingo, E., Ahlquist, P., Holland, J. J., Eds.; CRC Press: Boca Raton, FL, 1988; Vol. 3.
- (742) Manrubia, S. C. *Curr. Opin. Virol.* **2012**, *2*, 531.
- (743) Vignuzzi, M.; Stone, J. K.; Arnold, J. J.; Cameron, C. E.; Andino, R. *Nature* **2006**, *439*, 344.
- (744) Lauring, A. S.; Andino, R. *PLoS Pathog.* **2010**, *6*, e1001005.
- (745) Stich, M.; Briones, C.; Manrubia, S. C. *BMC Evol. Biol.* **2007**, *7*, 110.
- (746) Ruiz-Jarabo, C. M.; Arias, A.; Baranowski, E.; Escarmis, C.; Domingo, E. *J. Virol.* **2000**, *74*, 3543.
- (747) Ruiz-Jarabo, C. M.; Arias, A.; Molina-Paris, C.; Briones, C.; Baranowski, E.; Escarmis, C.; Domingo, E. *J. Mol. Biol.* **2002**, *315*, 285.
- (748) Briones, C.; Domingo, E.; Molina-Paris, C. *J. Mol. Biol.* **2003**, *331*, 213.
- (749) Diaz-Arenas, C.; Lehman, N. *BMC Evol. Biol.* **2010**, *10*, 80.
- (750) Walde, P. *Bioessays* **2010**, *32*, 296.
- (751) Liu, A. P.; Fletcher, D. A. *Nat. Rev. Mol. Cell. Biol.* **2009**, *10*, 644.
- (752) Luisi, P. L.; Ferri, F.; Stano, P. *Naturwissenschaften* **2006**, *93*, 1.
- (753) A protocell can be defined as “any experimental or theoretical model that involves a self-assembled compartment (typically a supramolecular structure, like a lipid vesicle) linked to chemical processes taking place around or within it, aimed at explaining how more complex biological cells or alternative forms of cellular organization may come about”. See Ruiz-Mirazo K. In *Encyclopedia of Astrobiology*; Gargaud M., Amils R., Cernicharo Quintanilla J., Cleaves H. J., Irvine W. M., Pinti D., Viso M., Eds.; Springer: Heidelberg, Germany, 2011.
- (754) *The Minimal Cell. The Biophysics of Cell Compartment and the Origin of Cell Functionality*; Luisi, P. L., Stano, P., Eds.; Springer: Dordrecht, The Netherlands, 2011.
- (755) Morowitz, H. J. *Prog. Theor. Biol.* **1967**, *1*, 35.
- (756) Sole, R. V.; Munteanu, A.; Rodriguez-Caso, C.; Macia, J. *Philos. Trans. R. Soc., B* **2007**, *362*, 1727.
- (757) Koonin, E. V. *Annu. Rev. Genomics Hum. Genet.* **2000**, *1*, 99.
- (758) Gil, R.; Silva, F. J.; Pereto, J.; Moya, A. *Microbiol. Mol. Biol. Rev.* **2004**, *68*, 518.
- (759) Forster, A. C.; Church, G. M. *Mol. Syst. Biol.* **2006**, *2*, 45.
- (760) Moya, A.; Gil, R.; Latorre, A.; Pereto, J.; Pilar Garcillan-Barcia, M.; de la Cruz, F. *FEMS Microbiol. Rev.* **2009**, *33*, 225.
- (761) Kondo, J.; Westhof, E. *J. Mol. Recognit.* **2010**, *23*, 241.
- (762) Davies, J.; von Ahsen, U.; Schroeder, R. In *The RNA World*, 1st ed.; Gesteland, R. F., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 1993; Chapt. 8, p 185.
- (763) Schroeder, R.; Waldsch, C.; Wank, H. *EMBO J.* **2000**, *19*, 1.
- (764) Briones, C.; Manrubia, S. C.; Lazaro, E.; Lazcano, A.; Amils, R. *Mol. Phylogenet. Evol.* **2005**, *34*, 371.
- (765) Ferre-D'Amare, A. R. *Philos. Trans. R. Soc., B* **2011**, *366*, 2942.
- (766) Frankel, A. D. *Curr. Opin. Struct. Biol.* **2000**, *10*, 332.
- (767) Keene, J. D.; Tenenbaum, S. A. *Mol. Cell* **2002**, *9*, 1161.
- (768) Das, C.; Frankel, A. D. *Biopolymers* **2003**, *70*, 80.
- (769) Hagihara, M.; Hasegawa, T.; Sato, S.; Yoshikawa, S.; Ohkubo, K.; Morii, T. *Biopolymers* **2004**, *76*, 66.
- (770) Nakano, S.; Mashima, T.; Matsugami, A.; Inoue, M.; Katahira, M.; Morii, T. *J. Am. Chem. Soc.* **2011**, *133*, 4567.
- (771) Nakano, S.; Hasegawa, T.; Fukuda, M.; Fujieda, N.; Tainaka, K.; Morii, T. *Nucleic Acids Symp. Ser.* **2008**, *52*, 199.
- (772) Liew, F. F.; Hasegawa, T.; Fukuda, M.; Nakata, E.; Morii, T. *Biorg. Med. Chem.* **2011**, *19*, 4473.
- (773) Baskerville, S.; Bartel, D. P. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 9154.
- (774) Mavelli, F. In *The Minimal Cell. The Biophysics of Cell Compartment and the Origin of Cell Functionality*; Luisi, P. L., Stano, P., Eds.; Springer: Dordrecht, The Netherlands, 2011; Chapt. 14, p 255.
- (775) Baeza, I.; Ibanez, M.; Santiago, J. C.; Wong, C.; Lazcano, A.; Oro, J. *Adv. Space Res.* **1986**, *6*, 39.
- (776) Walde, P.; Goto, A.; Monnard, P. A.; Wessicken, M.; Luisi, P. *L. J. Am. Chem. Soc.* **1994**, *116*, 7541.
- (777) Oberholzer, T.; Albrizio, M.; Luisi, P. L. *Chem. Biol.* **1995**, *2*, 677.

- (778) Khvorova, A.; Kwak, Y. G.; Tamkun, M.; Majerfeld, I.; Yarus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 10649.
- (779) Chen, I. A.; Salehi-Ashtiani, K.; Szostak, J. W. *J. Am. Chem. Soc.* **2005**, *127*, 13213.
- (780) Mavelli, F.; Ruiz-Mirazo, K. *Integr. Biol.* **2013**, *5*, 324.
- (781) Chen, I. A.; Walde, P. *Cold Spring Harbor Perspect. Biol.* **2010**, *2*, a002170.
- (782) Zepik, H. H.; Walde, P. *ChemBioChem* **2008**, *9*, 2771.
- (783) Schmidli, P. K.; Schurtenberger, P.; Luisi, P. L. *J. Am. Chem. Soc.* **1991**, *113*, 8127.
- (784) Oparin, A. I. In *The Origins of Prebiological Systems and of Their Molecular Matrices*; Fox, S. W., Ed.; Academic Press: New York, 1965.
- (785) Stano, P.; Carrara, P.; Kuruma, Y.; Pereira de Souza, T.; Luisi, P. L. *J. Mater. Chem.* **2011**, *21*, 18887.
- (786) *Cell-Free Protein Synthesis. Methods and Protocols*; Spirin, A. S., Swartz, J. R., Eds.; Wiley: Weinheim, Germany, 2008.
- (787) Shimizu, Y.; Inoue, A.; Tomari, Y.; Suzuki, T.; Yokogawa, T.; Nishikawa, K.; Ueda, T. *Nat. Biotechnol.* **2001**, *19*, 751.
- (788) Shimizu, Y.; Kanamori, T.; Ueda, T. *Methods* **2005**, *36*, 299.
- (789) Oberholzer, T.; Nierhaus, K. H.; Luisi, P. L. *Biochem. Biophys. Res. Commun.* **1999**, *261*, 238.
- (790) Yu, W.; Sato, K.; Wakabayashi, M.; Nakaishi, T.; Ko-Mitamura, E. P.; Shima, Y.; Urabe, I.; Yomo, T. *J. Biosci. Bioeng.* **2001**, *92*, 590.
- (791) Ishikawa, K.; Sato, K.; Shima, Y.; Urabe, I.; Yomo, T. *FEBS Lett.* **2004**, *576*, 387.
- (792) Although in this section we are mainly referring to work on protein synthesis as a paramount example of metabolic process, to be fair this limitation is also present in those experimental systems that focus on template (DNA or RNA)-dependent replication processes but require protein enzymes, like the Q $\beta$  replicase, or a kit of specific reagents, like those involved in the PCR amplification of DNA, for their effective accomplishment.
- (793) Blocher, M.; Hitz, T.; Luisi, P. L. *Helv. Chim. Acta* **2001**, *84*, 842.
- (794) Gardner, P. M.; Winzer, K.; Davis, B. G. *Nat. Chem.* **2009**, *1*, 377.
- (795) Vassiltsova, O. V.; Chuvalin, A. L.; Parmon, V. N. *J. Photochem. Photobiol. A* **1999**, *125*, 127.
- (796) Summers, D. P.; Noveron, J.; Basa, R. C. B. *Origins Life Evol. Biospheres* **2009**, *39*, 127.
- (797) Alpermann, T.; Rudel, K.; Ruger, R.; Steiniger, F.; Nietzsche, S.; Filiz, V.; Forster, S.; Fahr, A.; Weigand, W. *Origins Life Evol. Biospheres* **2011**, *41*, 103.
- (798) Chen, I. A.; Hanczyk, M. M.; Sazani, P. L.; Szostak, J. W. In *The RNA World*, 3rd ed.; Gesteland, R. F., Cech, T. R., Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2006; Chapt. 3, p 57.
- (799) Ichihashi, N.; Matsuura, T.; Kita, H.; Sunami, T.; Suzuki, H.; Yomo, T. *Cold Spring Harbor Perspect. Biol.* **2010**, *2*, No. a004945.
- (800) Nourian, Z.; Roelofsen, W.; Danelon, C. *Angew. Chem., Int. Ed.* **2012**, *51*, 3114.
- (801) Shohda, K.; Sugawara, T. *Soft Matter* **2006**, *2*, 402.
- (802) Nomura, S.; Tsumoto, K.; Hamada, T.; Akiyoshi, K.; Nakatani, Y.; Yoshikawa, K. *ChemBioChem* **2003**, *4*, 1172.
- (803) Murtas, G.; Kuruma, Y.; Bianchini, P.; Diaspro, A.; Luisi, P. L. *Biochem. Biophys. Res. Commun.* **2007**, *363*, 12.
- (804) Saito, H.; Kato, Y.; Le Berre, M.; Yamada, A.; Inoue, T.; Yosikawa, K.; Baigl, D. *ChemBioChem* **2009**, *10*, 1640.
- (805) Sunami, T.; Matsuura, T.; Suzuki, H.; Yomo, T. *Methods Mol. Biol.* **2010**, *607*, 243.
- (806) Stano, P.; Kuruma, Y.; Souza, T. P.; Luisi, P. L. *Methods Mol. Biol.* **2010**, *606*, 127.
- (807) Rasmussen, S.; Chen, L. H.; Deamer, D.; Krakauer, D. C.; Packard, N. H.; Stadler, P. F.; Bedau, M. A. *Science* **2004**, *303*, 963.
- (808) DeClue, M. S.; Monnard, P. A.; Bailey, J. A.; Maurer, S. E.; Collis, G. E.; Ziock, H. J.; Rasmussen, S.; Boncella, J. M. *J. Am. Chem. Soc.* **2009**, *131*, 931.
- (809) Gibson, D. G.; Glass, J. I.; Lartigue, C.; Noskov, V. N.; Chuang, R. Y.; Algire, M. A.; Benders, G. A.; Montague, M. G.; Ma, L.; Moodie, M. M.; Merryman, C.; Vashee, S.; Krishnakumar, R.; Assad-Garcia, N.; Andrews-Pfannkoch, C.; Denisova, E. A.; Young, L.; Qi, Z. Q.; Segall-Shapiro, T. H.; Calvey, C. H.; Parmar, P. P.; Hutchison, C. A.; Smith, H. O.; Venter, J. C. *Science* **2010**, *329*, 52.
- (810) Noireaux, V.; Libchaber, A. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17669.
- (811) Deamer, D. *Trends Biotechnol.* **2005**, *23*, 336.
- (812) Chalmeau, J.; Monina, N.; Shin, J.; Vieu, C.; Noireaux, V. *Biochim. Biophys. Acta* **2011**, *1808*, 271.
- (813) Kita, H.; Matsuura, T.; Sunami, T.; Hosoda, K.; Ichihashi, N.; Tsukada, K.; Urabe, I.; Yomo, T. *ChemBioChem* **2008**, *9*, 2403.
- (814) Caschera, F.; Sunami, T.; Matsuura, T.; Suzuki, H.; Hanczyk, M. M.; Yomo, T. *Langmuir* **2011**, *27*, 13082.
- (815) Kurihara, K.; Tamura, M.; Shohda, K.-i.; Toyota, T.; Suzuki, K.; Sugawara, T. *Nat. Chem.* **2011**, *3*, 775.
- (816) Castellanos, M.; Wilson, D. B.; Shuler, M. L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 6681.
- (817) Castellanos, M.; Kushiro, K.; Lai, S. K.; Shuler, M. L. *Biochem. Biophys. Res. Commun.* **2007**, *97*, 397.
- (818) Mavelli, F.; Ruiz-Mirazo, K. *Phys. Biol.* **2010**, *7*, No. 036002.
- (819) Lehn, J. M. *Chem.—Eur. J.* **1999**, *5*, 2455.
- (820) Cousins, G. R. L.; Poulsen, S. A.; Sanders, J. K. M. *Curr. Opin. Chem. Biol.* **2000**, *4*, 270.
- (821) Lehn, J. M.; Eliseev, A. V. *Science* **2001**, *291*, 2331.
- (822) Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J. L.; Sanders, J. K. M.; Otto, S. *Chem. Rev.* **2006**, *106*, 3652.
- (823) Ladame, S. *Org. Biomol. Chem.* **2008**, *6*, 219.
- (824) Reek, J. H. R.; Otto, S. *Dynamic Combinatorial Chemistry*; Wiley—VCH: Weinheim, Germany, 2010.
- (825) Hunta, R. A. R.; Otto, S. *Chem. Commun.* **2011**, *47*, 847.
- (826) Cougnon, F. B. L.; Sanders, J. K. M. *Acc. Chem. Res.* **2012**, *45*, 2211.
- (827) Otto, S. *Acc. Chem. Res.* **2012**, *45*, 2200.
- (828) Otto, S.; Furlan, R. L. E.; Sanders, J. K. M. *Science* **2002**, *297*, 590.
- (829) Herrmann, A. *Org. Biomol. Chem.* **2009**, *7*, 3195.
- (830) Caraballo, R.; Dong, H.; Ribeiro, J. P.; Jimenez-Barbero, J.; Ramstrom, O. *Angew. Chem., Int. Ed.* **2010**, *49*, 589.
- (831) Bhat, V. T.; Caniard, A. M.; Luksch, T.; Brenk, R.; Campopiano, D. J.; Greaney, M. F. *Nat. Chem.* **2010**, *2*, 490.
- (832) Lam, R. T. S.; Belenguer, A.; Roberts, S. L.; Naumann, C.; Jarrosson, T.; Otto, S.; Sanders, J. K. M. *Science* **2005**, *308*, 667.
- (833) Au-Yeung, H. Y.; Pantos, G. D.; Sanders, J. K. M. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 10466.
- (834) Au-Yeung, H. Y.; Pantos, G. D.; Sanders, J. K. M. *Angew. Chem., Int. Ed.* **2010**, *49*, 5331.
- (835) Cougnon, F. B. L.; Au-Yeung, H. Y.; Pantos, G. D.; Sanders, J. K. M. *J. Am. Chem. Soc.* **2011**, *133*, 3198.
- (836) Perez-Fernandez, R.; Pittelkow, M.; Belenguer, A. M.; Sanders, J. K. M. *Chem. Commun.* **2008**, *15*, 1738.
- (837) Saggiomo, V.; Luning, U. *Chem. Commun.* **2009**, *25*, 3711.
- (838) Perez-Fernandez, R.; Pittelkow, M.; Belenguer, A. M.; Lane, L. A.; Robinson, C. V.; Sanders, J. K. M. *Chem. Commun.* **2009**, *25*, 3708.
- (839) Saggiomo, V.; Goesch, C.; Herges, R.; Quesada, R.; Luning, U. *Eur. J. Org. Chem.* **2010**, *12*, 2337.
- (840) Lehn, J. M. *Prog. Polym. Sci.* **2005**, *30*, 814.
- (841) Otto, S. *Nat. Nanotechnol.* **2009**, *4*, 13.
- (842) Giuseppone, N. *Acc. Chem. Res.* **2012**, *45*, 2178.
- (843) Buryak, A.; Severin, K. *Angew. Chem., Int. Ed.* **2005**, *44*, 7935.
- (844) Severin, K. *Curr. Opin. Chem. Biol.* **2010**, *14*, 737.
- (845) Brisig, B.; Sanders, J. K. M.; Otto, S. *Angew. Chem., Int. Ed.* **2003**, *42*, 1270.
- (846) Gasparini, G.; Dal Molin, M.; Prins, L. J. *Eur. J. Org. Chem.* **2010**, *2429*.
- (847) del Amo, V.; Philp, D. *Chem.—Eur. J.* **2010**, *16*, 13304.
- (848) Unfortunately, sulfur-containing amino acids have so far not been detected in carbonaceous chondrites (see section 2), probably due to their relative instabilities (see Pizzarello, S.; Shock, E. *Cold Spring Harbor Perspect. Biol.* **2010**, *2*, a002105). Quite recently, some

light into this issue was shed by the analysis of original extracts from a nonpublished experiment by the late Stanley L. Miller (see Parker, E. T.; Cleaves, H. J.; Callahan, M. P.; Dworkin, J. P.; Glavin, D. P.; Lazcano, A.; Bada, J. L. *Origins Life Evol. Biospheres* **2011**, *41*, 201). The authors found significant amounts of methionine in these old samples, as well as degradation products from methionine and cysteine. Their results suggest that the route towards sulfur-containing organic compounds may have been robust under certain reducing conditions.

- (849) Goral, V.; Nelen, M. I.; Eliseev, A. V.; Lehn, J. M. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 1347.
- (850) Leclaire, J.; Vial, L.; Otto, S.; Sanders, J. K. M. *Chem. Commun.* **2005**, *15*, 1959.
- (851) Sarma, R. J.; Otto, S.; Nitschke, J. R. *Chem.—Eur. J.* **2007**, *13*, 9542.
- (852) Orrillo, A. G.; Escalante, A. M.; Furlan, R. L. E. *Chem. Commun.* **2008**, *42*, 5298.
- (853) Sarma, R. J.; Nitschke, J. R. *Angew. Chem., Int. Ed.* **2008**, *47*, 377.
- (854) Escalante, A. M.; Orrillo, A. G.; Furlan, R. L. E. *J. Comb. Chem.* **2010**, *12*, 410.
- (855) Breslow, R. L. *Artificial Enzymes*; Wiley—VCH: Weinheim, Germany, 2005.
- (856) van Leeuwen, P. W. N. M. *Supramolecular Catalysis*; Wiley—VCH: Weinheim, Germany, 2008.
- (857) Reetz, M. T. *Angew. Chem., Int. Ed.* **2001**, *40*, 284.
- (858) Prins, L. J.; Scrimin, P. *Angew. Chem., Int. Ed.* **2009**, *48*, 2288.
- (859) Kirby, A. J. *Angew. Chem., Int. Ed.* **1996**, *35*, 707.
- (860) Sadownik, J. W.; Philp, D. *Angew. Chem., Int. Ed.* **2008**, *47*, 9965.
- (861) Xu, S.; Giuseppone, N. *J. Am. Chem. Soc.* **2008**, *130*, 1826.
- (862) Nguyen, R.; Allouche, L.; Buhler, E.; Giuseppone, N. *Angew. Chem., Int. Ed.* **2009**, *48*, 1093.
- (863) Hochgurtel, M.; Kroth, H.; Piecha, D.; Hofmann, M. W.; Nicolau, C.; Krause, S.; Schaaf, O.; Sonnenmoser, G.; Eliseev, A. V. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 3382.
- (864) Hochgurtel, M.; Biesinger, R.; Kroth, H.; Piecha, D.; Hofmann, M. W.; Krause, S.; Schaaf, O.; Nicolau, C.; Eliseev, A. V. *J. Med. Chem.* **2003**, *46*, 356.
- (865) Karan, C.; Miller, B. L. *J. Am. Chem. Soc.* **2001**, *123*, 7455.
- (866) Whitney, A. M.; Ladame, S.; Balasubramanian, S. *Angew. Chem., Int. Ed.* **2004**, *43*, 1143.
- (867) Bugaut, A.; Jantos, K.; Wietor, J. L.; Rodriguez, R.; Sanders, J. K. M.; Balasubramanian, S. *Angew. Chem., Int. Ed.* **2008**, *47*, 2677.
- (868) Nielsen, M. C.; Ulven, T. *Chem.—Eur. J.* **2008**, *14*, 9487.
- (869) Bugaut, A.; Toulme, J. J.; Rayner, B. *Angew. Chem., Int. Ed.* **2004**, *43*, 3144.
- (870) Sadownik, J. W.; Ulijn, R. V. *Curr. Opin. Biotechnol.* **2010**, *21*, 401.
- (871) Cavalli, S.; Albericio, F.; Kros, A. *Chem. Soc. Rev.* **2010**, *39*, 241.
- (872) Krishnan-Ghosh, Y.; Balasubramanian, S. *Angew. Chem., Int. Ed.* **2003**, *42*, 2171.
- (873) Pandya, M. J.; Cerasoli, E.; Joseph, A.; Stoneman, R. G.; Waite, E.; Woolfson, D. N. *J. Am. Chem. Soc.* **2004**, *126*, 17016.
- (874) Schnarr, N. A.; Kennan, A. J. *Org. Lett.* **2005**, *7*, 395.
- (875) Woll, M. G.; Gellman, S. H. *J. Am. Chem. Soc.* **2004**, *126*, 11172.
- (876) Hadley, E. B.; Gellman, S. H. *J. Am. Chem. Soc.* **2006**, *128*, 16444.
- (877) Steinkruger, J. D.; Woolfson, D. N.; Gellman, S. H. *J. Am. Chem. Soc.* **2010**, *132*, 7586.
- (878) Price, J. L.; Hadley, E. B.; Steinkruger, J. D.; Gellman, S. H. *Angew. Chem., Int. Ed.* **2010**, *49*, 368.
- (879) Samiappan, M.; Dadon, Z.; Ashkenasy, G. *Chem. Commun.* **2011**, *47*, 710.
- (880) Carnall, J. M. A.; Waudby, C. A.; Belenguer, A. M.; Stuart, M. C. A.; Peyralans, J. J. P.; Otto, S. *Science* **2010**, *327*, 1502.
- (881) Nguyen, R.; Buhler, E.; Giuseppone, N. *Macromolecules* **2009**, *42*, 5913.

- (882) Sanger, F.; Thompson, E. O. P. *Biochem. J.* **1953**, *53*, 366.
- (883) Edman, P. *Acta Chem. Scand.* **1950**, *4*, 283.
- (884) Niall, H. D. *Methods Enzymol.* **1973**, *27*, 942.
- (885) Gevaert, K.; Houthaeve, T.; Vandekerckhove, J. *EXS* **2000**, *88*, 29.
- (886) Shively, J. E. *EXS* **2000**, *88*, 99.
- (887) Seidler, J.; Zinn, N.; Boehm, M. E.; Lehmann, W. D. *Proteomics* **2010**, *10*, 634.
- (888) Armirotti, A.; Damonte, G. *Proteomics* **2010**, *10*, 3566.
- (889) Lovric, J. *Introducing Proteomics: From Concepts to Sample Separation, Mass Spectrometry and Data Analysis*; Wiley—Blackwell: Hoboken, NJ, 2011.
- (890) Gavin, A. C.; Maeda, K.; Kuhner, S. *Curr. Opin. Biotechnol.* **2011**, *22*, 42.
- (891) Krug, K.; Nahnsen, S.; Macek, B. *Mol. Biosyst.* **2011**, *7*, 284.
- (892) Holley, R. W.; Apgar, J.; Everett, G. A.; Madison, J. T.; Marquise, M.; Merrill, S. H.; Penswick, J. R.; Zamir, A. *Science* **1965**, *147*, 1462.
- (893) Sanger, F.; Coulson, A. R. *J. Mol. Biol.* **1975**, *94*, 441.
- (894) Maxam, A. M.; Gilbert, W. *Proc. Natl. Acad. Sci. U.S.A.* **1977**, *74*, 560.
- (895) Min Jou, W.; Ysebaert, M.; Fiers, W.; Haegeman, G. *Nature* **1972**, *237*, 82.
- (896) Fiers, W.; Contreras, R.; Duerinck, F.; Haegeman, G.; Iserentant, D.; Merregaert, J.; Minjou, W.; Molemans, F.; Raeymaekers, A.; Vandenberghe, A.; Volckaert, G.; Ysebaert, M. *Nature* **1976**, *260*, 500.
- (897) Snustad, D. P.; Simmons, M. J. *Principles of Genetics*, 5th ed.; John Wiley & Sons: Hoboken, NJ, 2010.
- (898) Fleischmann, R. D.; Adams, M. D.; White, O.; Clayton, R. A.; Kirkness, E. F.; Kerlavage, A. R.; Bult, C. J.; Tomb, J. F.; Dougherty, B. A.; Merrick, J. M.; McKenney, K.; Sutton, G.; Fitzhugh, W.; Fields, C.; Gocayne, J. D.; Scott, J.; Shirley, R.; Liu, L. I.; Glodek, A.; Kelley, J. M.; Weidman, J. F.; Phillips, C. A.; Spriggs, T.; Hedblom, E.; Cotton, M. D.; Utterback, T. R.; Hanna, M. C.; Nguyen, D. T.; Saudek, D. M.; Brandon, R. C.; Fine, L. D.; Fritchman, J. L.; Fuhrmann, J. L.; Geoghegan, N. S. M.; Gnehm, C. L.; McDonald, L. A.; Small, K. V.; Fraser, C. M.; Smith, H. O.; Venter, J. C. *Science* **1995**, *269*, 496.
- (899) Venter, J. C.; Remington, K.; Heidelberg, J. F.; Halpern, A. L.; Rusch, D.; Eisen, J. A.; Wu, D. Y.; Paulsen, I.; Nelson, K. E.; Nelson, W.; Fouts, D. E.; Levy, S.; Knap, A. H.; Lomas, M. W.; Nealson, K.; White, O.; Peterson, J.; Hoffman, J.; Parsons, R.; Baden-Tillson, H.; Pfannkoch, C.; Rogers, Y. H.; Smith, H. O. *Science* **2004**, *304*, 66.
- (900) Edwards, R. A.; Rohwer, F. *Nat. Rev. Microbiol.* **2005**, *3*, 504.
- (901) Shendure, J.; Ji, H. L. *Nat. Biotechnol.* **2008**, *26*, 1135.
- (902) Shendure, J.; Aiden, E. L. *Nat. Biotechnol.* **2012**, *30*, 1084.
- (903) Margulies, M.; Egholm, M.; Altman, W. E.; Attiya, S.; Bader, J. S.; Bemben, L. A.; Berka, J.; Braverman, M. S.; Chen, Y. J.; Chen, Z. T.; Dewell, S. B.; Du, L.; Fierro, J. M.; Gomes, X. V.; Godwin, B. C.; He, W.; Helgesen, S.; Ho, C. H.; Irzyk, G. P.; Jando, S. C.; Alenquer, M. L. I.; Jarvie, T. P.; Jurage, K. B.; Kim, J. B.; Knight, J. R.; Lanza, J. R.; Leamon, J. H.; Lefkowitz, S. M.; Lei, M.; Li, J.; Lohman, K. L.; Lu, H.; Makhijani, V. B.; McDade, K. E.; McKenna, M. P.; Myers, E. W.; Nickerson, E.; Nobile, J. R.; Plant, R.; Puc, B. P.; Ronan, M. T.; Roth, G. T.; Sarkis, G. J.; Simons, J. F.; Simpson, J. W.; Srinivasan, M.; Tartaro, K. R.; Tomasz, A.; Vogt, K. A.; Volkmer, G. A.; Wang, S. H.; Wang, Y.; Weiner, M. P.; Yu, P. G.; Begley, R. F.; Rothberg, J. M. *Nature* **2005**, *437*, 376.
- (904) Fedurco, M.; Romieu, A.; Williams, S.; Lawrence, I.; Turcatti, G. *Nucleic Acids Res.* **2006**, *34* (3), No. e22.
- (905) Turcatti, G.; Romieu, A.; Fedurco, M.; Tairi, A. P. *Nucleic Acids Res.* **2008**, *36* (4), No. e25.
- (906) Shendure, J.; Porreca, G. J.; Reppas, N. B.; Lin, X. X.; McCutcheon, J. P.; Rosenbaum, A. M.; Wang, M. D.; Zhang, K.; Mitra, R. D.; Church, G. M. *Science* **2005**, *309*, 1728.
- (907) Pitt, J. N.; Ferre-D'Amare, A. R. *Science* **2010**, *330*, 376.
- (908) Hoffmann, C.; Minkah, N.; Leipzig, J.; Wang, G.; Arens, M. Q.; Tebas, P.; Bushman, F. D. *Nucleic Acids Res.* **2007**, *35* (13), No. e91.

- (909) Wang, C. L.; Mitsuya, Y.; Gharizadeh, B.; Ronaghi, M.; Shafer, R. W. *Genome Res.* **2007**, *17*, 1195.
- (910) Zagordi, O.; Klein, R.; Daumer, M.; Beerenswinkel, N. *Nucleic Acids Res.* **2010**, *38*, 7400.
- (911) Cai, F. P.; Chen, H. F.; Hicks, C. B.; Bartlett, J. A.; Zhu, J.; Gao, F. *Nat. Methods* **2007**, *4*, 123.
- (912) Suzuki, S.; Ono, N.; Furusawa, C.; Ying, B. W.; Yomo, T. *PLoS One* **2011**, *6*, No. e19534.
- (913) Branton, D.; Deamer, D. W.; Marziali, A.; Bayley, H.; Benner, S. A.; Butler, T.; Di Ventra, M.; Garaj, S.; Hibbs, A.; Huang, X. H.; Jovanovich, S. B.; Krstic, P. S.; Lindsay, S.; Ling, X. S. S.; Mastrangelo, C. H.; Meller, A.; Oliver, J. S.; Pershin, Y. V.; Ramsey, J. M.; Riehn, R.; Soni, G. V.; Tabard-Cossa, V.; Wanunu, M.; Wiggin, M.; Schloss, J. A. *Nat. Biotechnol.* **2008**, *26*, 1146.
- (914) Maitra, R. D.; Kim, J.; Dunbar, W. B. *Electrophoresis* **2012**, *33*, 3418.
- (915) Harris, T. D.; Buzby, P. R.; Babcock, H.; Beer, E.; Bowers, J.; Braslavsky, I.; Causey, M.; Colonell, J.; Dimeo, J.; Efcaivitch, J. W.; Giladi, E.; Gill, J.; Healy, J.; Jarosz, M.; Lapen, D.; Moulton, K.; Quake, S. R.; Steinmann, K.; Thayer, E.; Tyurina, A.; Ward, R.; Weiss, H.; Xie, Z. *Science* **2008**, *320*, 106.
- (916) Braslavsky, I.; Hebert, B.; Kartalov, E.; Quake, S. R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 3960.
- (917) Merriman, B.; Rothberg, J. M. *Electrophoresis* **2012**, *33*, 3397.
- (918) Southern, E. M. *J. Mol. Biol.* **1975**, *98*, 503.
- (919) Burnette, W. N. *Anal. Biochem.* **1981**, *112*, 195.
- (920) *Handbook of Biosensors and Biochips*; Marks, R. S., Lowe, C. R., Cullen, D. C., Weetall, H. H., Karube, I., Eds.; Wiley-VCH: Weinheim, Germany, 2007.
- (921) *Recognition Receptors in Biosensors*; Zourob, M., Ed.; Springer: Heidelberg, Germany, 2010.
- (922) *Biosensors and Biodetection: Methods and Protocols*; Rasooly, A., Herold, K. E., Eds.; Humana Press-Springer: New York, 2009; Vol. 1 & 2.
- (923) Bryant, P. A.; Venter, D.; Robins-Browne, R.; Curtis, N. *Lancet Infect. Dis.* **2004**, *4*, 100.
- (924) Kaushansky, A.; Allen, J. E.; Gordus, A.; Stiffler, M. A.; Karp, E. S.; Chang, B. H.; MacBeath, G. *Nat. Protoc.* **2010**, *5*, 773.
- (925) Fernandez, N.; Garcia-Sacristan, A.; Ramajo, J.; Briones, C.; Martinez-Salas, E. *Virology* **2011**, *409*, 251.
- (926) Southern, E. M. *Methods Mol. Biol.* **2001**, *170*, 1.
- (927) Stoughton, R. B. *Annu. Rev. Biochem.* **2005**, *74*, 53.
- (928) Schena, M.; Shalon, D.; Davis, R. W.; Brown, P. O. *Science* **1995**, *270*, 467.
- (929) Dufva, M. *Methods Mol. Biol.* **2009**, *529*, 1.
- (930) Juskowiak, B. *Anal. Bioanal. Chem.* **2011**, *399*, 3157.
- (931) Kozal, M. J.; Shah, N.; Shen, N. P.; Yang, R.; Fucini, R.; Merigan, T. C.; Richman, D. D.; Morris, D.; Hubbell, E. R.; Chee, M.; Gingeras, T. R. *Nat. Med.* **1996**, *2*, 753.
- (932) Poulsen, L.; Soe, M. J.; Moller, L. B.; Dufva, M. *PLoS One* **2011**, *6*, No. e14777.
- (933) Duggan, D. J.; Bittner, M.; Chen, Y. D.; Meltzer, P.; Trent, J. M. *Nat. Genet.* **1999**, *21*, 10.
- (934) Pariet, L.; Chillemi, G.; Bongiorni, S.; Spica, V. R.; Volentini, A. *New Biotechnol.* **2009**, *25*, 272.
- (935) Kammenga, J. E.; Herman, M. A.; Ouborg, N. J.; Johnson, L.; Breitling, R. *Trends Ecol. Evol.* **2007**, *22*, 273.
- (936) Gresham, D.; Dunham, M. J.; Botstein, D. *Nat. Rev. Genet.* **2008**, *9*, 291.
- (937) Loy, A.; Pester, M.; Steger, D. *Methods Mol. Biol.* **2011**, *688*, 187.
- (938) Garrido, P.; Blanco, M.; Moreno-Paz, M.; Briones, C.; Dahbi, G.; Blanco, J.; Blanco, J.; Parro, V. *Clin. Chem.* **2006**, *52*, 192.
- (939) Han, X. Q.; Lin, X. M.; Liu, B. H.; Hou, Y. H.; Huang, J. H.; Wu, S. Q.; Liu, J.; Mei, L.; Jia, G. L.; Zhu, Q. Y. *J. Virol. Methods* **2008**, *152*, 117.
- (940) Ooms, M.; Verhoef, K.; Southern, E.; Huthoff, H.; Berkhouit, B. *Nucleic Acids Res.* **2004**, *32*, 819.
- (941) Mikheikin, A. L.; Surzhikov, S. A.; Zasedateleva, O. A.; Vasiliskov, V. A.; Pan'kov, S. V.; Grechisnikova, I. V.; Kisselev, L. L.; Zaseclatelev, A. S. *BioTechniques* **2008**, *44*, 77.
- (942) Brandt, O.; Hoheisel, J. D. *Trends Biotechnol.* **2004**, *22*, 617.
- (943) Debaene, F.; Winssinger, N. *Methods Mol. Biol.* **2009**, *570*, 299.
- (944) Gerry, N. P.; Witowski, N. E.; Day, J.; Hammer, R. P.; Barany, G.; Barany, F. *J. Mol. Biol.* **1999**, *292*, 251.
- (945) Gilles, P. N.; Wu, D. J.; Foster, C. B.; Dillon, P. J.; Chanock, S. *J. Nat. Biotechnol.* **1999**, *17*, 365.
- (946) Tillib, S. V.; Mirzabekov, A. *Curr. Opin. Biotechnol.* **2001**, *12*, 53.
- (947) Weller, M. G. In *Protein Microarrays*; Schena, M., Ed.; Jones and Bartlett Publishers: Sudbury, MA, 2005.
- (948) LaBaer, J.; Ramachandran, N. *Curr. Opin. Chem. Biol.* **2005**, *9*, 14.
- (949) Petschnigg, J.; Snider, J.; Stagliar, I. *Curr. Opin. Biotechnol.* **2011**, *22*, 50.
- (950) Zhu, H.; Bilgin, M.; Bangham, R.; Hall, D.; Casamayor, A.; Bertone, P.; Lan, N.; Jansen, R.; Bidlingmaier, S.; Houfek, T.; Mitchell, T.; Miller, P.; Dean, R. A.; Gerstein, M.; Snyder, M. *Science* **2001**, *293*, 2101.
- (951) Lechuga, L. M. *Anal. Bioanal. Chem.* **2006**, *384*, 44.
- (952) Situma, C.; Hashimoto, M.; Soper, S. A. *Biomol. Eng.* **2006**, *23*, 213.
- (953) Moschallski, M.; Baader, J.; Prucker, O.; Ruhe, J. *Anal. Chim. Acta* **2010**, *671*, 92.
- (954) Rozkiewicz, D. I.; Brugman, W.; Kerkhoven, R. M.; Ravoo, B. J.; Reinhoudt, D. N. *J. Am. Chem. Soc.* **2007**, *129*, 11593.
- (955) Blattler, T.; Huwiler, C.; Ochsner, M.; Stadler, B.; Solak, H.; Voros, J.; Grandin, H. M. *J. Nanosci. Nanotechnol.* **2006**, *6*, 2237.
- (956) Erickson, D.; Mandal, S.; Yang, A. H. J.; Cordovez, B. *Microfluid. Nanofluid.* **2008**, *4*, 33.
- (957) Bodrossy, L.; Sessitsch, A. *Curr. Opin. Microbiol.* **2004**, *7*, 245.
- (958) *Label-Free Biosensors: Techniques and Applications*; Cooper, M. A., Ed.; Cambridge University Press: Cambridge, U.K., 2009.
- (959) Ray, S.; Mehta, G.; Srivastava, S. *Proteomics* **2010**, *10*, 731.
- (960) Wilson, D. S.; Szostak, J. W. *Annu. Rev. Biochem.* **1999**, *68*, 611.
- (961) Wright, M. C.; Joyce, G. F. *Science* **1997**, *276*, 614.
- (962) Diaz-Arenas, C.; Lehman, N. *Curr. Protoc. Nucleic Acid Chem.* **2010**, *9*, 1.
- (963) Roberts, R. W.; Szostak, J. W. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 12297.
- (964) Seelig, B.; Szostak, J. W. *Nature* **2007**, *448*, 828.
- (965) Golynskiy, M. V.; Seelig, B. *Trends Biotechnol.* **2010**, *28*, 340.
- (966) Guillen Schlippe, Y. V.; Hartman, M. C. T.; Josephson, K.; Szostak, J. W. *J. Am. Chem. Soc.* **2012**, *134*, 10469.
- (967) Esvelt, K. M.; Carlson, J. C.; Liu, D. R. *Nature* **2011**, *472*, 499.
- (968) *The Aptamer Handbook*; Klussmann, S., Ed.; Wiley-VCH: Weinheim, Germany, 2006.
- (969) Stoltenburg, R.; Reinemann, C.; Strehlitz, B. *Biomol. Eng.* **2007**, *24*, 381.
- (970) Janas, T.; Janas, T. *Cell. Mol. Biol. Lett.* **2011**, *16*, 25.
- (971) Recent reviews show the applicability and clinical usefulness of aptamers against either selected cellular components (Moreno, M.; Gonzalez, V. M. *Curr. Med. Chem.* **2011**, *18*, 5003.) or whole living cells (Sun, W.; Du, L. P.; Li, M. Y. *Curr. Pharm. Des.* **2011**, *17*, 80.).
- (972) Bunka, D. H. J.; Stockley, P. G. *Nat. Rev. Microbiol.* **2006**, *4*, 588.
- (973) Breaker, R. R. In *The RNA World*, 3rd ed.; Gesteland, R. F.; Cech, T. R.; Atkins, J. F., Eds.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, 2006; Chapt. 4, p 89.
- (974) Montange, R. K.; Batey, R. T. *Annu. Rev. Biophys.* **2008**, *37*, 117.
- (975) Zhang, J.; Lau, M. W.; Ferre-D'Amare, A. R. *Biochemistry* **2010**, *49*, 9123.
- (976) Hoppe-Seyler, F.; Crnkovic-Mertens, I.; Tomai, E.; Butz, K. *Curr. Mol. Med.* **2004**, *4*, S29.
- (977) Baines, I. C.; Colas, P. *Drug Discovery Today* **2006**, *11*, 334.

- (978) Johnson, S.; Evans, D.; Laurenson, S.; Paul, D.; Davies, A. G.; Ferrigno, P. K.; Walti, C. *Anal. Chem.* **2008**, *80*, 978.
- (979) Lam, B. J.; Joyce, G. F. *Nat. Biotechnol.* **2009**, *27*, 288.
- (980) Bugaut, A.; Toulme, J. J.; Rayner, B. *Org. Biomol. Chem.* **2006**, *4*, 4082.
- (981) Melkko, S.; Scheuermann, J.; Dumelin, C. E.; Neri, D. *Nat. Biotechnol.* **2004**, *22*, 568.
- (982) Melkko, S.; Zhang, Y.; Dumelin, C. E.; Scheuermann, J.; Neri, D. *Angew. Chem., Int. Ed.* **2007**, *46*, 4671.
- (983) Mannocci, L.; Zhang, Y. X.; Scheuermann, J.; Leimbacher, M.; De Bellis, G.; Rizzi, E.; Dumelin, C.; Melkko, S.; Neri, D. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17670.
- (984) Melkko, S.; Mannocci, L.; Dumelin, C. E.; Villa, A.; Sommavilla, R.; Zhang, Y. X.; Grutter, M. G.; Keller, N.; Jermutus, L.; Jackson, R. H.; Scheuermann, J.; Neri, D. *ChemMedChem* **2010**, *5*, 584.
- (985) Mannocci, L.; Melkko, S.; Buller, F.; Molnar, I.; Bianke, J. P.; Dumelin, C. E.; Scheuermann, J.; Neri, D. *Bioconjugate Chem.* **2010**, *21*, 1836.
- (986) Paegel, B. M.; Joyce, G. F. *PLoS Biol.* **2008**, *6*, 900.
- (987) Xu, Y. H.; Yang, X. R.; Wang, E. K. *Anal. Chim. Acta* **2010**, *683*, 12.
- (988) Atencia, J.; Beebe, D. J. *Nature* **2005**, *437*, 648.
- (989) Wu, D. P.; Qin, J. H.; Lin, B. C. *J. Chromatogr. A* **2008**, *1184*, 542.
- (990) Kan, C. W.; Fredlake, C. P.; Doherty, E. A. S.; Barron, A. E. *Electrophoresis* **2004**, *25*, 3564.
- (991) Auroux, P. A.; Koc, Y.; deMello, A.; Manz, A.; Day, P. J. R. *Lab Chip* **2004**, *4*, 534.
- (992) Melin, J.; Quake, S. R. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 213.
- (993) Mark, D.; Haeberle, S.; Roth, G.; von Stetten, F.; Zengerle, R. *Chem. Soc. Rev.* **2010**, *39*, 1153.
- (994) McDonald, J. C.; Duffy, D. C.; Anderson, J. R.; Chiu, D. T.; Wu, H. K.; Schueller, O. J. A.; Whitesides, G. M. *Electrophoresis* **2000**, *21*, 27.
- (995) Chovan, T.; Guttman, A. *Trends Biotechnol.* **2002**, *20*, 116.
- (996) Wolfe, D. B.; Qin, D.; Whitesides, G. M. *Methods Mol. Biol.* **2010**, *583*, 81.
- (997) Marre, S.; Jensen, K. F. *Chem. Soc. Rev.* **2010**, *39*, 1183.
- (998) Young, E. W. K.; Beebe, D. J. *Chem. Soc. Rev.* **2010**, *39*, 1036.
- (999) Austin, R. H.; Tung, C. K.; Lambert, G.; Liao, D.; Gong, X. Q. *Chem. Soc. Rev.* **2010**, *39*, 1049.
- (1000) Chemnitz, S.; Tangen, U.; Wagler, P. F.; Maeke, T.; McCaskill, J. S. *Chem. Eng. J.* **2008**, *135*, S276.
- (1001) Shum, H. C.; Lee, D.; Yoon, I.; Kodger, T.; Weitz, D. A. *Langmuir* **2008**, *24*, 7651.
- (1002) Teh, S.-H.; Khnouf, R.; Fan, H.; Lee, A. P. *Biomicrofluidics* **2011**, *5*, No. 044113.
- (1003) McCaskill, J. S.; Packard, N. H.; Rasmussen, S.; Bedau, M. A. *Philos. Trans. R. Soc., B* **2007**, *362*, 1763.
- (1004) van den Berg, A.; Craighead, H. G.; Yang, P. D. *Chem. Soc. Rev.* **2010**, *39*, 899.
- (1005) Piruska, A.; Gong, M.; Sweedler, J. V.; Bohn, P. W. *Chem. Soc. Rev.* **2010**, *39*, 1060.
- (1006) Levy, S. L.; Craighead, H. G. *Chem. Soc. Rev.* **2010**, *39*, 1133.
- (1007) Karlsson, M.; Davidson, M.; Karlsson, R.; Karlsson, A.; Bergenholz, J.; Konkoli, Z.; Jesorka, A.; Lobovkina, T.; Hurtig, J.; Voinova, M.; Orwar, O. *Annu. Rev. Phys. Chem.* **2004**, *55*, 613.
- (1008) Krishnan, M.; Mojarrad, N.; Kukura, P.; Sandoghdar, V. *Nature* **2010**, *467*, 692.
- (1009) Kim, S. J.; Song, Y. A.; Han, J. *Chem. Soc. Rev.* **2010**, *39*, 912.
- (1010) Bedau, M. A.; McCaskill, J. S.; Packard, N. H.; Rasmussen, S. *Artif. Life* **2010**, *16*, 89.
- (1011) Song, H.; Chen, D. L.; Ismagilov, R. F. *Angew. Chem., Int. Ed.* **2006**, *45*, 7336.
- (1012) Theberge, A. B.; Courtois, F.; Schaerli, Y.; Fischlechner, M.; Abell, C.; Hollfelder, F.; Huck, W. T. S. *Angew. Chem., Int. Ed.* **2010**, *49*, 5846.
- (1013) Song, H.; Tice, J. D.; Ismagilov, R. F. *Angew. Chem., Int. Ed.* **2003**, *42*, 768.
- (1014) Gerdts, C. J.; Sharoyan, D. E.; Ismagilov, R. F. *J. Am. Chem. Soc.* **2004**, *126*, 6327.