MAINTENANCE OF GENETIC INFORMATION IN THE FIRST RIBOCELL

Ádám Kun[1,2,3]

[1] Parmenides Center for the Conceptual Foundations of Science, Parmenides Foundation, Kirchplatz 1, D-82049 Pullach, Germany

[2] MTA-ELTE Theoretical Biology and Evolutionary Ecology Research Group, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary

[3] Evolutionary Systems Research Group, Centre for Ecological Research, Klebelsberg Kuno u. 3., H-8237 Tihany, Hungary

1. The ribocell and the stages of the RNA world

The RNA world [1-3] is an era during the origin of life when RNA played the role of both information storage molecule and enzymes. Ribozymes are the embodiments of this last function. They not only witnessed the origin of life, they made it happen. According to Tibor Gánti [4,5], absolute life criteria are (1) inherent unity, (2) inherent stability (homeostasis), the presence of (3) a metabolism, (4) an information carrying subsystem, and (5) processes regulated and controlled by a programme. As the workhorse of metabolism, ribozymes helped to fulfil criteria (2) and (3). Information (criterium 4) are the complementary strands of ribozymes. RNA, in its dual roles as enzymes and information, regulates the processes of the ribocell. Thus, ribozymes were truly at the hearth of what it meant to be alive. The first absolute life criterium is realized by a lipid membrane, and the encapsulation of RNA enzymes and information was the first major evolutionary transition [6-8], which led to the formation of the first living cell. In modern cells, metabolism is run mostly by peptide enzymes, and information is stored in DNA, but RNAs shuttle information between DNA and peptides and they still have a rich role in regulation of cellular processes [9-14].

Questions about the origin of life can be roughly divided into structural and dynamical categories. With respect to ribozymes, structural questions mostly relate to what a certain RNA molecule can do. Dynamical questions relate to the evolution and/or the ecology of the molecules. For example, it is a structural question whether RNA can catalyse all reaction steps of a given pathway. And it is a dynamical question whether these ribozymes can act together in a common environment or whether they can all evolve within one system. Most of the experimental literature on the origin of life focuses on structural questions. Here, we will focus on dynamical questions, especially the extent of metabolic complexity a ribocell could have attained and the hurdles it needed to overcome to reach it.

I consider the RNA world to last till DNA took over as the main information-storage molecule (Fig. 1). It can be argued that after the invention of translation, it is no longer an RNA world. But the takeover of peptides cannot have happened overnight. Thus, when the first translated peptide

rolled out of a ribosome, most enzymatic functions were still fulfilled by ribozymes. The number of ribozymes might have steadily decreased as the number of peptide enzymes increased, but this, by necessity, was a gradual process. Consequently, even at this stage, RNAs acted both as an information-storage molecule(s) and as enzymes, albeit RNA's role in this last function was withdrawing. The defining feature of the RNA world was still there. Interestingly, RNA lost most of its information storage role (only RNA viruses store their genetic information in RNA), but retained some of its catalytic functions: the catalytic core of the ribosome, which is found in all living organisms, is made up of RNA [15,16]. Not as widespread as the ribosome, but there are some natural ribozymes as well, like the RNase P [17], the group I introns [18], the group II introns [19], the hammerhead ribozyme [20], the hairpin ribozyme [21], the Hepatitis Delta Virus and like ribozymes [22], the *Neurospora* Varkud Satellite Ribozyme [23], the *glmS* ribozyme [24], and the twister ribozymes [25].

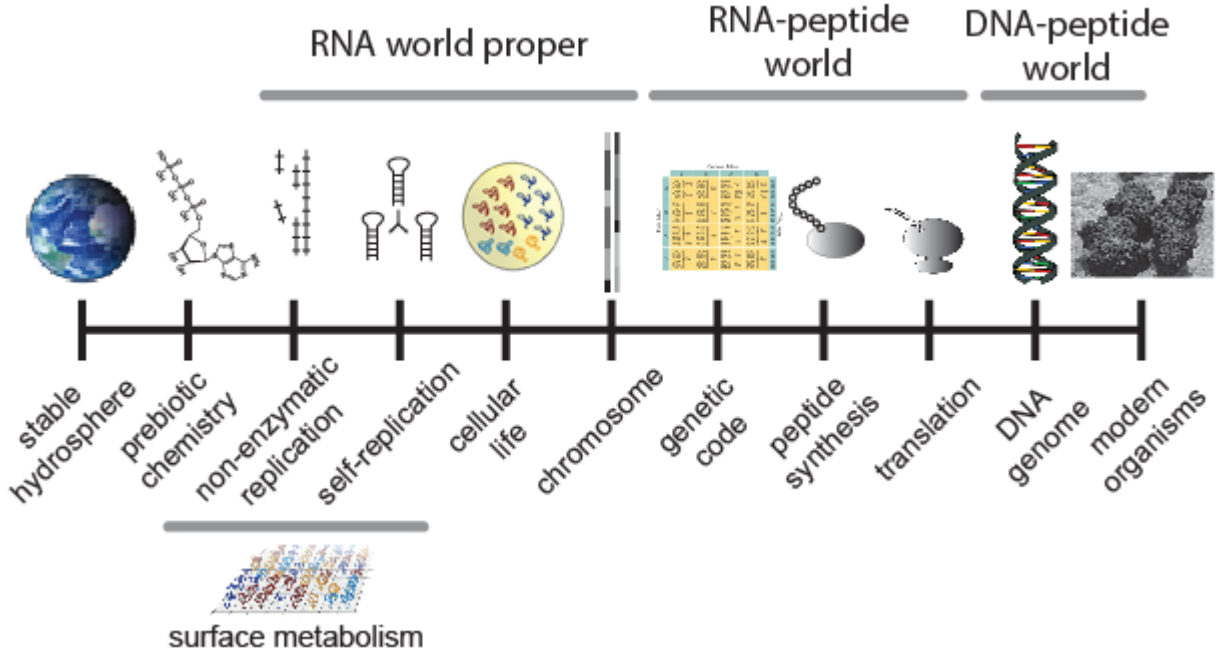

Figure 1. Stages of the RNA World. Between the two well-defined end-points of the origin of life, lies the RNA world. Non-enzymatic RNA replication could have led to the evolution of a self-replicating enzyme, which kick-started metabolism. This stage was probably still unfolded on mineral surfaces. Cellular life marks the beginning of life. The error catastrophe was overcome when the chromosome evolved. This led to complex metabolism, which could recruit amino-acids and then evolve peptide synthesis, thus leading to the RNA-peptide world stage. With the invention of DNA genome, the RNA world yielded to the now present DNA-peptide world.

1.1 Replication of the genetic information

Evolution is a powerful mechanism that can produce novelty and considerable increase in complexity, but it requires reproduction, variation and heredity. Reproduction is the capability of exponential growth achieved via autocatalysis [26,27], what chemical systems can exhibit [28-32]. However, heredity and variation are characteristics that rarely feature together in a system. The von Kiedrowski type replicators [33], for example, exhibit heredity, but the reactants cannot

be changed without destroying their ability for autocatalysis. Compositional information [34], on the other hand, have ample variability, but lacks heredity [35]. Modular polymers with complementary modules, such as RNA and DNA, have the potential for evolution. Variation in the form of mutations arise during RNA/DNA replication naturally (see below), and, via complementarity, information is passed on. The copy of a strand is yielded in two steps: first the complementary strand is produced, and by copying it, a replica of the original strand forms. However, without a mechanism for copying RNA, it also does not exhibit heredity, and thus cannot evolve. For example, RNA molecules of intermediate length can form on clay surfaces [36,37], but as they from *de novo*, and not as copies of another RNA molecule, there is no heredity. Thus, assuming RNA to be the primordial molecule that led to life also requires self-replication to be the first function on that path.

Self-replication of RNA molecules, i.e. RNA-based RNA polymerisation, it the key step leading to life. While non-enzymatic replication of RNAs [38] could have bootstrapped the evolution of ribozymes, an RNA-based RNA polymerase is a necessity. A fully functional example of this ribozyme is so far elusive. Attempts to modify a ligase ribozyme to act as a template directed polymerase first yielded an enzyme capable of adding a mere six nucleotides [39]. Later this was improved to add 14 [40], and then up to 20 nucleotides [41], which is still only about 10% of the length of the RNA molecule (Table 1). A breakthrough came in 2011 when a ribozyme capable of synthesising a strand of up to 98 nucleotides was selected [42]. A ribozyme can extend a primer with more than 200 nucleotides [43], but requires a very specific template. Unfortunately, this specific template is not its own sequence or its complementary sequence. Thus, ribozymes are capable of synthetizing long strands based on a template, but their generality is still a question.

1.2 On the metabolic complexity of ribocells

Autocatalytic replication of the informational polymer is just one of the function a living cell needs to exhibit. It should possess a membrane, which is also autocatalytically formed [44]. And it should supply all the building blocks that it cannot take up from the environment via its metabolism. Metabolism is also autocatalytic [45]. A minimal living cell (c.f. the Chemoton [46]) consists of the three subsystems of information storage, metabolism and encapsulation. Systems having only two of these subsystems are called infrabiological systems (*sensu* Szathmáry [47-49]). There are three infrabiological systems possible: an informational-metabolic system, an encapsulated informational and an encapsulated metabolic system. One of these infrabiological systems were the precursor of the first living cell. In contemporary organisms, lipid membranes encapsulate DNA information, and the system is maintained via a metabolism catalyzed by peptide enzymes. While now all subsystems are realized by chemically different entities, their chemical identity is not set into stone. Viruses, for example, are encapsulated informational systems without metabolism. Their DNA or RNA genome is encapsulated by a proteinacous capsid. Thus, encapsulation can also be done with proteins, not only with lipids. In the RNA world, one molecular species, RNA, fulfils the role of information storage and enzyme. Ribozyme based replication is an infrabiological system in itself. It has information, as it can be copied, and the information is used to produce the ribozyme which does the copying of

information. Consequently, the infrabiological system having information and metabolism is especially important for the origin of life.

The informational – metabolic stage probably flourished on some mineral surface. A prebiotic soup or warm little ponds (cf. Darwin) dilute material too much and foster hydrolysis as opposed to polymerization. Some kind of semi-compartmentalization can solve this problem. For example, negatively charged RNA molecules can be attached to mineral surfaces having positive charge, e.g. pyrite. In such system, formation of larger organic molecules could be favoured [50]. Furthermore, mineral surfaces can catalyse reactions [37,51]; and in an enantioselective manner, which may be responsible for the homochirality we observe today [52,53]. They can also protect ribozymes from the pervasive UV radiation [54] of the primordial environment.

At the stage of surface metabolism, there could have already been a budding metabolism catalysed by a small array of ribozymes. Apart from the copying of the genetic information, the assembly of nucleotides from its constituents could have been functions that enhanced the local proliferation of ribozymes. On the other end of the origin of life, in the stage shifting from RNA to DNA information storage, a small array of enzymes is not enough, a complex metabolism run by potentially hundreds if not thousands of genes is required. The metabolic pathways and characteristics common to all living organisms were already in place in the last universal common ancestor (LUCA). This entity is a very late descendant of the first cell, albeit there are some claims that it could still had an RNA genome [55]. LUCA definitely had peptide enzymes, the universality of the genetic code and the similarities of the ribosomes are testimonies of it. Translation, the process that produces polypeptides based on the information stored in a nucleic acid, requires a considerable number of enzymes. The question naturally arises, how to proceed from a system having one or a few enzymatic activities to one having several hundreds.

In this chapter, we first discuss the obstacles in the path of increasing number of ribozymes. Then the minimal ribozyme diversity of the first ribocell is estimated.

2. The error thresholds

An evolving system requires heredity, which is achieved by copying nucleic acid polymers via base-pairing. A complementary base is ligated at the end of the new polymer, resulting–at the end of the replication process–in a complementary strand. The replication of this strand yields the copy of the original strand. Consequently, by this two-stage process, genetic information can be propagated. However, there are errors during this process resulting in changed copies of the original strand. These errors are the mutations that, on the one hand, provide the variation on which selection can act, but, on the other hand, they can destroy a functional ribozyme.

2.1 Introducing the error threshold

Let us assume that there is a sequence that should be copied as it will fold into an enzyme. We will call this sequence the enzyme sequence (it might also be called the master sequence or the wild-type sequence). It replicates with a replication rate of a_{enzyme}. During replication, mutation can occur, and the resulting sequence is not the enzyme sequence but a parasite. A parasite accepts the catalytic aid of the ribozymes, but it does not contribute to the system as it does not

have any enzymatic activity or other useful role. The probability of such mutation is μ. The parasite replicates with a rate of a_{parasite}. There is a uniform, sequence-independent death rate (or wash-out rate) D. The concentrations of the enzyme (x_{enzyme}) and the parasite (x_{parasite}) are governed by the following two equations:

$$\frac{dx_{\text{enzyme}}}{dt} = a_{\text{enzyme}}(1-\mu)x_{\text{enzyme}} - Dx_{\text{enzyme}} \qquad \text{Eq.1}$$

$$\frac{dx_{\text{parasite}}}{dt} = a_{\text{enzyme}}\mu x_{\text{enzyme}} + a_{\text{parasite}}x_{\text{parasite}} - Dx_{\text{parasite}} \qquad \text{Eq.2}$$

This is a simplification of the quasispecies model of Eigen [56], which included all possible mutants of the master sequence. Here, we lump them together as proposed by Maynard Smith [57]. The persistence of the enzyme in equilibrium ($x_{\text{enzyme}} > 0$) requires that

$$(1-\mu)a_{\text{enzyme}} > a_{\text{parasite}} \qquad \text{Eq.3}$$

Thus, not only has the replication rate of the enzyme be higher than that of the parasite, it has to be considerably higher if mutation rate is high. Given the replication rate of the enzyme and the parasite, a critical mutation rate, the error threshold, can be determined. The error threshold is thus the critical mutation rate above which information cannot be maintained despite it having higher replication rate than the parasite. However, while there could be mutants that have lower replication rate than the wild-type, there are considerable number of them having lower replication rates.

Shorter sequences–usually–have faster replication rate. The mutation leading to shorter sequence is called deletions. A deletions can occur, for example, by slippage of the replicase on longer stretches of repetitions [58], which then results in either an insertion or a deletion (together called indels). In contemporary organisms, deletions are more frequent than insertions [59]. Most indels are rather short, 80%+ of them are 1-10 bp [60] or 1-5 bp long [61,62]. Small indels can destroy enzymatic activity, but longer deletions are required for considerably faster replicating parasites. Longer deletions, while rarer, are also observable. Thus, there will be shorter and faster mutants competing with the wild-type ribozyme.

We know from the pioneering work of Sol Spiegelman [63], that faster replicating, non-functional mutations of a functional RNA go to fixation. Starting from the about 3300–3600 nucleotides long RNA genome of the Qβ phage and replicating it, they have arrived–after 75 passages–at an RNA replicating fifteen times faster than the original, but being only 550 nucleotides long. This RNA was not a functional phage. In a similar experiment [64], we have replicated a modified *Neurospora* VS ribozyme with the Qβ replicase, and after some time transferring a sample to a fresh solution of NTPs and replicase. After a few transfers, the ribozyme could not be detected in the population. The population was dominated by a sequence roughly third of the length of the functional ribozyme and replicating nearly twice as fast.

Consequently, if there is only selection for replication speed, then a shorter mutant of the wild-type enzyme will outcompete it and the functional RNA (the information) will be lost. Phages can survive the high error rate of their replicases [65-69] and the competition with their faster replicating mutants by selection on function. Only functional virions can infect a new host and

replicate in it. The higher level evolutionary unit, the capsid encapsulated virus genome, the virion, allows for the apparent replication rate of the functional virus to be higher than that of its shorter mutants. Similarly, when ribozymes are encapsulated into droplets, and droplets are selected for further replication based on total enzymatic activity, then the ribozyme can be maintained despite the constant re-emergence of the shorter and faster replicating mutants [64]. Thus, a higher level evolutionary unit is required to satisfy Eq.3. This is assumed in all models of the error threshold either explicitly or implicitly. We also make this assumption here, and I will further elaborate on compartmentalization in the next section.

2.2 The fitness landscape and neutrality of mutations

The mutation rate μ in Eq. 3 is the probability of replication resulting in a sequence that is not an enzyme, but a parasite. While mutations change the genotype, they do not necessarily change the phenotype. In the original formulation of the error threshold [56], all mutations led out of the master sequence, i.e. all mutations were considered to result in a different phenotype. Even if we remain true to the original model, not all mutations to a coding DNA sequence (the genotype) result in a change of the amino-acid sequence of the coded peptide (the phenotype). The genetic code is degenerate: multiple triplets code for the same amino-acid. These synonymous mutations are neutral, and the fitness of the individual bearing the mutated sequence is the same as those harbouring the wild-type sequence. As the mutation rate μ is the rate at which the sequence changes to another having lower fitness, the actual mutation rate of the replication process can be higher. Furthermore, even if an amino-acid changes, it might not affect the activity and stability of the peptide. Experiments determining the distribution of fitness effects [70] shows that there are non-synonymous mutations with neutral effect on fitness. According to extensive mutagenesis of *Salmonella enterica*'s HisA protein, an isomerase in the L-histidine biosynthesis pathway, 2.5% of the non-synonymous mutations are neutral [71]. Furthermore, assaying 64% of all possible single mutant of the antibiotic resistance factor TEM-1 β-lactamase, 320 (32,3%) was found to have the same minimum inhibitory concentration to antibiotics as the wild-type [72]. And 4.8% of the analysed mutations of ribosomal proteins are neutral [73]. A study analysed the E3 ubiquitin ligase activity of 5153 mutants of the RING domain of BRCA1 (breast cancer 1 protein) [74]: 90 (1.7%) had a ligase activity score not differing more than 1% of the wild-type's, and 435 (8.4%) were within 5% of the value of the activity score. Thus, not all mutations lead to decrease of fitness.

At the early stages of the origin of life, coded peptides were not yet present, and the faithful replication of ribozymes were the key problem [2]. With regard to RNA, there are mutations that do not affect the secondary structure of it [75-77]. It was estimated that compared to the 4^L different sequences of length L, the number of different structures is 2.35^L [78]. Accordingly, there are considerably more sequences than structures. Usually a few (1–3) mutations do not change the secondary structure of an RNA. Thus there is a phenotypic error threshold [79,80], which is the critical error rate above which the phenotype cannot be maintained despite selection for it.

Neutral mutations are mostly substitutions, i.e. mis-incorporations of a non-canonical base-pair into the sequence. While the two base-pair system offers some protection against mutations [81,82], there are still ample possibilities for base substitutions. Both effects are deeply rooted in the chemistry of the bases. Hydrogen bonds can form between guanine (G) and cytosine (C), and between adenine (A) and uracil (U). In these canonical base pairs, there is always a larger purine derivative (G or A) facing a smaller pyrimidine derivative (C or U). The size difference allows for the easy recognition of A–G and C–U misspairs, and consequently transversions are rare [83]. Transition (A↔G and U↔C mutations) are not this easy to catch. In their normal state, the non-canonical base-pairs A–C rarely forms, but the G–U bond is quite strong, and plays an important role in RNA secondary structure. There could be base pairs whose donor-acceptor side-chains are orthogonal to each other, and non-canonical purine-pyrimidine pairs would be disfavoured. Only two such pair could exist, and it would further lessen mutation probabilities [81,82]. Resistance to mutations is important, but not the sole determinant of a base suitability, and the prebiotic environment exerted its own selective force on them [84,85].

Mutations occur mostly for chemical reasons. The current set of two base-pairs are such that by tautomerisation their hydrogen donor-acceptor characteristic change to mimic that of the other base of the same size. Consequently, the imino form of adenine pairs with cytosine, and the imino form of cytosine with adenine. Similarly, the enol forms of guanine or uracil form base-pairs with the other base. In these unfavourable states, non-canonical base-pairs form, which then could be inherited. Tautomerization is the main mechanism by which base substitutions occur [86]. Spontaneous deamination is another chemical reason via which mutations can arise. Cytosine becomes uracil, adenine hypoxanthine and guanin xanthine. These last two cannot be found in an RNA or DNA, and thus error correcting mechanisms might be able to detect them. However, uracil is natural in RNA (but not in DNA), and the ensuing transition goes undetected. While reactive oxygen species can cause oxidative deamination, which can lead to mutation, the oxygen level at the origin of life was very low, and this source of mutation was probably not important. Non-enzymatic copying of RNA has an error rate of at least 0.01 mutation/base/replication [29].

Mutation rates are often given as mutations/base/replication, but, so far, there is no length in our formulation of the error threshold. The mutation rate μ is given as mutations/sequence/replication. The two quantities can be easily exchanged. Let us assume that the enzyme to be replicated has a length of L and the per base mutation rate is denoted by u. Then

$$(1-\mu)=(1-u)^{L} \qquad \text{Eq.4}$$

We can transform Eq.3 into the more frequently seen form:

$$L<\frac{\ln s}{u} \qquad \text{Eq.5}$$

, where $s=a_{\text{enzyme}}/a_{\text{parasite}}$. If we assume that $\ln s \approx 1$ and $u=0.01$, then the length of the maintainable enzyme is smaller than 100 nucleotides. While there are ample examples of ribozymes with less than a hundred nucleotides, it is clearly not enough for the genome of a whole ribo-organism. Moreover, even the putative RNA-dependent RNA polymerase ribozymes are longer than 100 nucleotides, they are around 200 nucleotides long (Table 1). Longer enzymes

could be more accurate, but accuracy is required for a longer enzyme in the first place. Thus we arrive at the Eigen's paradox: "no large genome without enzymes, and no enzymes without a large genome" [57].

However, as discussed earlier, the formulation of the error threshold in Eq.5. does not take the possibility of neutral mutations into account. A formula for the phenotypic error threshold based on the neutrality of some of the mutations was derived from first principles [80,87]:

$$L < \frac{-\ln s}{\ln\left((1-u)+\lambda-(1-u)\lambda\right)} \qquad \text{Eq.6}$$

The fraction of neutral mutations (λ) can be estimated by the analysis of RNA secondary structures. The secondary structure of an RNA is a good proxy for its structure [88] and it can be calculated easily [89-91]. Computationally, each position can be mutated and all of the $3L$ sequences differing by only one nucleotide analysed. An average fraction of neutral mutation can then be obtained. The range of λ for a set of 305 ribozyme sequences are between 9% and 71% (median and mean are 30%) [92]. Alternatively, one can average over a sample of sequences folding into a target structure, as was done for the $tRNA^{Phe}$ structure [93]. The mean λ was found to be 0.2871±0.2489. While the average fraction of neutral mutant seems to be rather similar, there is considerable between and within sequence variation. Mutations in single stranded regions of the structure change the structure less frequently, than mutations to double stranded regions. Generalizing from the literature, we have proposed [94] that structural elements in a secondary structure can be classified into four types: neutral structure, connecting structure, forbidden structure and critical structure. Neutral structures can be freely changed, and in some cases, they can even be removed. Connection structures position the critical elements, and as long as the structure is intact, they can fulfil their role. These are the parts of the structure that give it its mutational robustness. Forbidden structures are not found in functional RNAs, as their presence abolishes the function. Critical structures harbour sites that are important for their exact chemical characteristic, often they are the catalytic site or the substrate-binding sites of the ribozyme. These sites cannot be inferred from the secondary structure alone, only wet-lab experiment can tell us about their existence. Their presence overestimates the neutrality obtained from pure secondary-structure studies.

A fitness landscape, a map from genotype to phenotype to fitness, was constructed based on mutagenesis data for the VS ribozyme and the hairpin ribozyme [79,88]. We were able to show the true extent of the difference between the genotypic and the phenotypic error threshold. The maintainable sequence length was 6-7 times as much as previously thought based on Eq.5. This considerable increase in maintainable genome length allows us to replicate know ribozymes [92], but still a magnitude lower error rate would be needed for the replication of the genome of a minimal riboorganism, and it is a long shot from the genome size of contemporary organisms (Fig. 2).

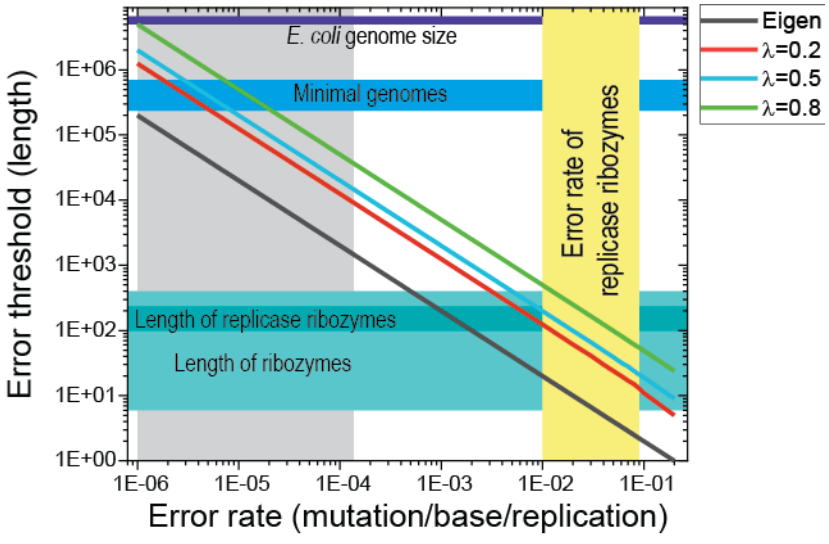

Figure 2. The error threshold for the original Eigen's formulation (Eq.6) (dark grey line), and for various fractions of neutral mutations (Eq. 7), when lns=1. Populations characterized by the parameter space below the lines are viable, while above it the sequence cannot be maintained. The vertical coloured regions represent the error rate of viral RNA-replicases (light grey) and the replicase ribozymes (yellow). The horizontal regions represent milestones in length, such as the replicase ribozymes, the minimal present-day genomes of bacteria and *Escherichia coli* (an example of a free-living organism).

The mutation rate of the RNA-dependent RNA polymerase ribozyme should be quite low. While for self-replication the error rate of 0.1% per base per replication would be low enough, mutation rates for these ribozymes are higher (Table 1). Presently, the main problem with these ribozymes are processivity and generality. They can extend a primer by a very limited number of nucleotides, far less than their size. So far, there was little effort to lower the mutation rate of the replicase ribozymes. Theory tells us [95] that if a modest increase in size can increase the fidelity of the enzyme, then genome size and replication fidelity can gradually increase. So far increase in length increased the fidelity of the replicase, albeit there are considerable variations and very few data points (Fig. 3).

Table 1. RNA-dependent RNA polymerase ribozymes

Ribozyme name	Error rate	Length	Processivity	Reference	Note
b1-233t	1.5×10^{-1}	98	6	[39]	
R18	3.3×10^{-2}	198	14	[40]	
B6.61	n.a.	193	20	[41,96]	
R18	4.1×10^{-2}	198	23–32	[97]	in ice
R18	4.26×10^{-2}	192		[42]	
tC19	2.68×10^{-2}	198	95	[42]	
Z	n.a.	221	n.a.	[42]	generality increased
tC19Z	8.8×10^{-3}	198	24	[42]	able to copy Hammerhead minizyme
Y	1.7×10^{-2}	195	30	[43]	

	5.2×10^{-2} *				
tC9Y	2.3×10^{-2}	202	206 118 *	[43]	specific template
24-3	8×10^{-2} 3.4×10^{-2}	180	61	[98]	able to copy functional RNA
$t5^{+1}$	2.6×10^{-2}	220+135	n.a.	[99]	triplet polymerase
Zcore	8.7×10^{-2}	110	n.a.	[99]	triplet polymerase
0core	9.5×10^{-2}	112	n.a.	[99]	triplet polymerase

* in ice

n.a. Data not available

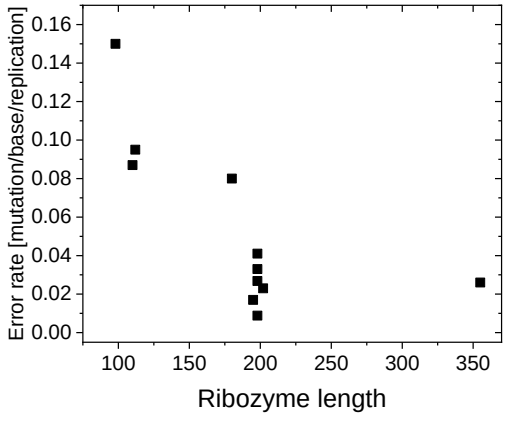

Figure 3. Error rate of the replicase ribozymes as function of their length. The 350 bases long ribozyme is a system with a 220 bases long ribozyme and a 135 bases long helper, see Table 1.

3. Compartmentalization

Splitting information into smaller pieces was seen as the solution for the error threshold by Eigen and Schuster [100]. Owing to their short length, fragments can be replicated by error-prone replicases, even though the replication of the whole genetic information in one piece (in one chromosome) is not feasible due to the error threshold. Thus, there will be good copies of fragments and consequently no information is lost because of mutations. However, the independently replicating fragments or ribozymes are in competition with each other, hence the strong replicative coupling in the hypercyclic organisation was envisioned. In a hypercycle, each of the members can catalyse the replication the next member. While short cycles can be stable, they can be destroyed by various parasites and their evolvability is very limited [101]. So, while Eigen and Schuster have identified the problem, they were not able to give a satisfactory solution to it. There should be other mechanisms to allow the coexistence of independently replicating ribozymes.

A living cell is a prime example of molecular cooperation. Replicators toil for the greater good of the whole. Individually, they would all be better off just accepting the catalytic aid or benefits the other replicators give, and not giving anything in return. That is the central problem of the evolution of cooperation: while greater benefits can be reaped if everyone co-operates compared to when no one does, the highest pay-off is obtained by exploiting others. The rather bleak message is that rational actors should not co-operate hold when interacting entities meet randomly. However, if the co-operators meet with each other more frequently than with cheaters (parasites), then co-operation could be the evolutionarily favoured outcome [102]. There should be some form of population structure or viscosity, so that the benefits of co-operation are reaped by the co-operators and not by the parasites. Encapsulation into a lipid vesicle ensures that the assistance of the molecular co-operators as well as the harm caused by parasites stay local.

3.1 Surface metabolism and transient compartmentalization

Cellular encapsulation is the ultimate form of compartmentalization, but earlier stages could also exist. Mineral surfaces, for example, limit the diffusion of macromolecules thus creating a viscous population. Small compartments in porous rocks at hydrothermal vents can house a rudimentary metabolism by providing chemical energy [103-106]. It can also offer the higher level selection required for the coexistence of functional ribozymes with parasites [107]. Not only hard rocks, but ice can also be a form of compartmentalization [97]. As RNA is quite labile, some reactions are better carried out in ice [108-110]. Irrespective of the exact nature of the surface, theory suggest that a diverse set of replicators can coexist on them [101,111-114]. Replicators can locally enhance their own replication as well as the replication of other replicators. While parasites thus can gain enzymatic boost to their replication, localities where they proliferate become less and less conductive to growth. Thus, the parasites' own replication limits their spread. Moreover the parasites that can actually coexist with the ribozymes are the ones that do not replicate much faster than the ribozymes [115]. The aggressive, fast parasites dominate their own neighbourhood fast, and then without any enzymes around they die. However, a weaker parasite would allow the enzymes to grow and thus the environment remains such that they can still grow. While these parasites, at the moment, are a drain on the resources of the system, can become useful by evolving an enzymatic function [116]. As there are no stabilizing selection on their function, as they have none, they can explore sequence space and might hit upon some useful function. Surfaces, however, cannot sustain an arbitrary diverse metabolism, as not many different types can coexist [117].

Transient compartmentalization would be the next step toward fully cellular life. In such a system, replicators are fully compartmentalized in some stages of their lifespan, but not all of it. For example, drying lipid vesicles transform them into lamellar structures releasing their content. Upon wetting the system, the newly forming vesicles take up some material, e.g. nucleic acids from their environment [118]. The drying-wetting cycles are conductive to condensation reactions as during the drying phase concentration can be relatively high, and the membranes also organize the compounds [119]. Thus, there is a prebiotically plausible way to have transient compartmentalization.

We have investigated the dynamics of a transiently compartmentalized system [64] with the help of *in vitro* compartmentalization and microfluidics [120,121]. Tiny (12 pL) aqueous droplets were loaded with the Qβ replicase, NTP, a modified *Neurospora* Varkud satellite ribozyme and a substrate. The VS ribozyme [23] is a self-cleaving ribozyme that can also be modified to be a trans-acting ribozyme [122]. The droplets can be selected based on the concentration of the cleaved substrate, which is a proxy for the number of functional ribozymes in the droplet. Selected droplets were collected, and their content pooled. New droplets were formed from this pool, and the RNAs inside the droplets were allowed to replicate again. The RNAs had only spent a portion of their life cycle compartmentalized, hence it is a transient compartmentalization. Even such incomplete compartmentalization allows the ribozymes to coexist with the parasites [64]. Droplets in which parasites proliferate and achieve high concentration will have low product concentration and are not selected. Here, coexistence rests on the number of RNA encapsulated in the droplets at the beginning of the encapsulated phase. If there is one or a few droplets at the beginning, then the ribozyme can coexist with the parasites, otherwise it cannot.

3.2 The Stochastic Corrector Model

Full compartmentalization means that cells stay intact and their content is passed on to the next generation. At the origin of cellular life, we cannot assume the cell to have full control over cell division. It was a stochastic process. Even if we would have a chromosome in the cell containing all the information, due to the stochastic nature of primordial cell division, information could be lost. Just before cell division, there would be two copies of the chromosome in the cell. Each independently would either get to one or the other daughter cell. Half of the time each daughter cells will have one chromosome, but in the other half of the time, only one of the cells will have chromosomes, the other will end up empty. Thus, information can be lost due to the stochastic nature of chromosome segregation. This is the assortment load. A way to avoid such a loss of information is to have more copies of the chromosome in the cell. The probability of having zero chromosome in a particular daughter cell after random division is $\left(\frac{1}{2}\right)^{\nu_{\max}}$, where $\nu_{\max}$ is the number of chromosomes in the parent cell before division. With increasing number of chromosomes, the probability of ending up with an empty cell is diminishing. However, because of the error threshold, chromosomes could not have evolved before a sufficiently accurate replicase evolved.

The first cell encapsulated independently replicating ribozymes. Let us assume that there was τ different types of ribozymes, each having a function indispensable for the cell. Let us further assume that cells divide when their internal concentration reach some predefined value, measured by the number of RNAs in the cell ($\nu_{\max}$). At this point the cell divides and each of the ribozymes assort randomly to daughter cells. If there are exactly two copies of each ribozymes in the cell before division, then the probability of them assorting evenly to the daughter cells is 0.5^{τ}, thus the probability of ending up with two viable daughter cells is diminishing as the number of different types increase. When there are more than one types to be maintained, both daughter cells can end up unviable, as one can lack one of the essential genes, while the other the other.

Here again, more copies of the ribozymes (redundancy) help alleviate the assortment load. There is a sharp boundary between the redundancy allowing for a certain number of ribozyme types to coexist and not permitting them to coexist. We call this the second error threshold [123], as there is a critical redundancy below which information is lost and the population is not viable. We have found that at least 100 types can coexist [123], which, as discussed in the next section, would be enough for a minimal ribocell to function. However, two problems remain: (1) internal competition between the independently replicating ribozymes can still destroy the system; and (2) parasites can still outcompete functional RNAs within a cell. Both of these problems are exacerbated when the cells are allowed to grow larger and have more RNAs.

Ribozymes are competing for the same resources: the available NTPs and the replicase ribozyme. As the limiting resources are the same, it already limits the number of RNAs that can coexist [124]. Moreover, if there are differences in their growth rates (e.g. differences in their affinities to the replicase), then, due to internal competition, the faster replicating ribozymes can dominate the population. The longer the RNAs can replicate the lower the frequency of the slower growing ones will be at the time of cell division. This again can lead to loss of information. Similar problem is caused by the appearance of parasites. Parasites, by their quicker replication, will take resources and space from the ribozymes, thereby lowering the effective redundancy in the cell. Upon cell division, the number of ribozymes is less than $\nu_{\max}$, increasing the probability that one or more of the required ribozyme types will be lost from one or both of the daughter cells.

The stochastic nature of ribozyme assortment into daughter cells can alleviate some of the problems associated with greater copy number. Just by chance, even from an imbalanced ribozyme distribution a daughter cell can end up with a favourable internal composition. For example, if parasites assort mostly to one of the daughter cells, then the other will be mostly free of them. Or if most of the faster replicating ribozymes will go to one of the cells then the internal competition will be lessened in the other for a while. This mechanism was termed the stochastic correction of internal composition, and the model framework described here is the Stochastic Corrector Model [125,126] (Fig. 4). It was shown that two types can easily coexist, even if their replication rates are different [125]. Furthermore, in an infinite population, an arbitrary number of types can coexist [127]. In an infinite population, all possible internal compositions and all possible divisions are realized, and thus selection can act on the rare but very beneficial, stochastically generated compositions. In a finite population, only a limited number of genes can coexist [128] at mutation rates comparable to that of replicase ribozymes. Compositions can also become better by primordial sex, the exchange and mixing of genetic material between cells [129,130]. While bad compositions can become better, good compositions can lose their good status by primordial sex. Sex can facilitate replicator coexistence to some extent, but it is not a universal remedy for the error catastrophes.

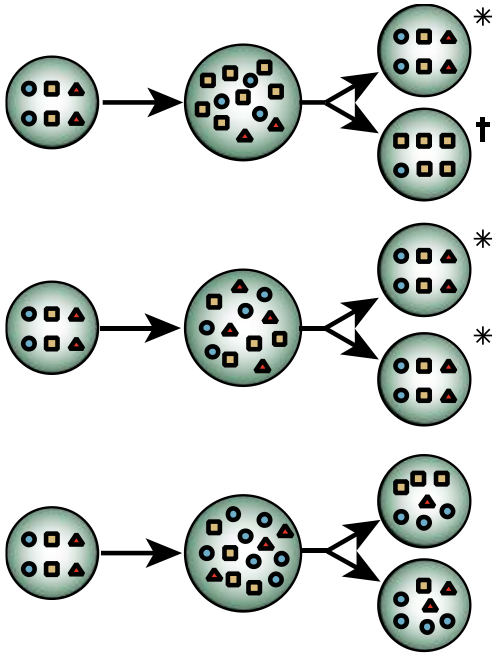

Figure 4. The Stochastic Corrector Model. Three essential, independently replicating ribozymes are in the ribocells depicted by a circle, a square and a triangle. Initially they all have the same concentration, and this uniform distribution is advantageous for the ribocell. The ribozymes multiply and the internal distribution of the constituents can become uneven. Stochastic division can restore the beneficial distribution (marked with an asterisk), however some daughter cells can end up with a missing gene, and it becomes unviable (marked with a cross).

While we do not yet know how much information can stably coexist in a compartmentalized system, it seems that the path to life is a narrow one fraught with dangers. Primordial cells need to navigate between the mythical Scylla and Charybdis of the origin of life [131,132]: one the one side, too little redundancy increase assortment load, and on the other, internal competition and parasites swamp the cells. In-between, we need enough different genes to coexist so as the cell function and serve as the basis for further evolution. It is important to note again, that once the fidelity of replication can increase, the information content of the cell can also increase, and gradually there could be more and more complex systems [95]. How many genes is needed for a minimal ribocell?

4. Minimal gene content of the first ribocell

A minimal organism has as few genes as possible. Most research in this field focuses on DNA-peptide organism, deriving the minimal gene-set based on contemporary living bacteria [48,133-135]. Present day metabolic pathways might not be the same are ones in a ribocell, albeit some vestiges of the primordial metabolism is still with us [136]. Functionally, present-day metabolism and primordial metabolism need to fulfil the same roles.

Minimal gene-sets found in contemporary organisms can be as low as 140 genes: *Tremblaya princeps* has 140 genes [137], *Nasuia deltocephalinicola* has 167 genes [138], *Hodgkinia cicadicola* has 189 genes [139], *Carsonella ruddii* has 213 genes [140], *Zinderia insecticola* has 231 genes [141], and *Sulcia muelleri* has 263 genes [141-145]. However, these symbionts of

insects are barely alive in the sense that they lack genes for membrane and cell wall synthesis, lack transporters, most of carbon metabolism [146] and some even lack some genes for DNA replication and translation. Other symbionts and intracellular parasites have around 500–600 genes (*Mysoplasma genitalium*, *Buchnera* sp. [135]). The minimalized, synthetic *Mycoplasma mycoides* JCVI- syn3.0 genome consists of 473 genes [147].

Moya and co-workers [148] have compared eight bacterial genomes to establish the minimal common set of genes found in all of them. This gave an estimate of functional minimal set of genes required for a living cell. Their estimate includes 16 genes for the replication of the genetic material, 106 genes for translation, 15 genes for enzyme folding and modification, 5 genes for cellular processes and 56 genes for energetic and intermediary metabolism, giving a grand total of 198 genes (the original estimate also included eight poorly characterized genes, which we omit here). Later they suggested 50 enzymes to be able to fulfil all minimal functionality for the intermediary metabolism [149]. This intermediary metabolism produces energy from glucose via glycolysis; assembles nucleotide-triphosphates and deoxynucleotide triphosphates from ribose, nucleobases and phosphates; forms a lipid species and produces the required co-enzymes.

An RNA organism requires considerably less genes than the above organisms and estimates, as there is no translation. One cannot emphasize enough how many genes are required for translation, which contributes to the difficulty of understanding its evolution [150]. If we subtract the genes for translation, and the genes for dNTP production from the minimal 198 genes [148], we arrive at 88 genes. The original set of genes included ones for peptide folding and salvage, and while peptides are not yet present in a ribocell, ribozymes might also require chaperons and salvage pathways. However, as discussed below, the number of genes for cell-level processes is probably underestimated. Compared to comparison-based estimates, the *Mycoplasma mycoides* JCVI- syn3.0 includes considerable number of genes for regulation (9), cell division (1) and transport (31) [147]. This is also true for an inferred minimal *Bacillus subtilis* genome [151]. Consequently, the minimal gene content of a ribo-organism might be around 100 genes.

4.1 Intermediate metabolism

The most important enzymatic function of a ribocell is the replication of the genetic material. It requires at least one enzyme, the RNA-dependent RNA polymerase (see section 1.1). However, this one enzyme might not be enough for the replication of RNA. Most of the know replicases require a primer, the synthesis of which and the initiation of the replication process is thus complicated [152]. While there is an example of RNA replicase that requires no primer [99], that system consists of two RNAs, the enzyme and a co-enzyme. Thus, even the most elementary process of a cell, the replication of the genetic material, requires two or more functional RNAs.

It has been demonstrated that replicase ribozymes are able to replicate aptamers [98,99], tRNA [98] or the Hammerhead ribozyme [42]. Thus, the replication of functional RNA is within the capabilities of the selected replicases. However, none can–at the moment–replicate itself. The problem lies with their processivity, i.e. the number of nucleotides they are able to add to the growing strand (see Table 1). In order to achieve full self-replication, the replicase had to be in many pieces, each of them replicated independently and the functional ribozyme self-assembles or it is being ligated together by ligase ribozymes [153]. Replicases can self-assemble from their

parts, but the successful copying of their parts as well as their complementary sequences have not yet been demonstrated. The replicase, which operates by the addition of triplets [99], can copy some fragments of itself and its complementary strand, but full self-replication is not achieved.

Still there is an often overlooked enzymatic activity that is required for the core function of RNA replication [152,154]: the unwinding of the resulting double-stranded RNA. Double stranded RNA is inert in the sense that the ribozyme strand cannot fulfil its catalytic role, nor can the template strand be replicated. Something has to unzip the two strands. Mostly thermal cycling/gradient [155,156] or some other oscillatory process [157] is assumed to take care of this conundrum. While short RNAs can be reliably separated in this manner, longer double-stranded RNAs are still bound too strongly and the environment conductive for separation is also damaging to RNA strands. Some helicase function is needed in the ribocell. It was suggested that the ancestor of the small subunit of the ribosome had such a function [158]. Most probably, it also adds at least one functional RNA to our list of minimal functions.

Apart from the template-based polymerisation of RNA and associated functions, the ribocell needs a constant supply of activated nucleotides (NTPs) [159]. While nucleosides can form from formamide [160-162] and other plausible prebiotic synthesis have also been proposed [163,164], and they can even be activated to some extent [165], supplies will run out quickly once ribocells begin to consume them. There should be ribozymes that contribute to the formation of activated nucleotides. Activated ribose and a nucleobase can be condensed by a ribozyme [166,167] to yield nucleosides. Phosphorylation of a single nucleoside has not yet been demonstrated, but 5'-OH of RNA [168] or 3'-OH of a DNA [169] can be triphosphorylated by a ribozyme. Generally, nucleic acid oligomers can be phosphorylated by ribozymes [170-174]. In these ribozymes, the substrate oligomer is bound to the enzyme via base-pairing. That is the simplest way of substrate binding, and it also makes the interaction specific. We know that a wide variety of RNA aptamers can bind nucleoside triphosphates [175-179] and catalyse phosphorylation, thus a ribozyme catalysing the phosphorylation of nucleosides is conceivable. With that, the replication of the genetic information would be no longer dependent on the exogenous supply of activated nucleotides. However, there should still be a supply of ribose and nucleobases. For both fatty acid and phospholipid membranes, ribose has the best permeability coefficient among aldopentoses and hexoses, and consequently it can accumulate inside a ribocell [180]. The formose reaction [181] under the right conditions [182] can supply ribose. As for the nucleobases, they were probably supplied by the environment for a long period in the RNA world. Modern biosyntheses of nucleobases are complicated and require a substantial number of enzymes, which suggests that a minimal ribocell did not had the ability the synthesize nucleobases *de novo*. If present, nucleotides can spontaneously diffuse through membranes composed of fatty-acids [183,184], and can also diffuse through membranes composed of phospholipids with 12–14 long carbon chains [185]. Consequently, ribocells could rely on exogeneous nucleobase supply.

A replicase, a helicase and some enzyme to produce activated nucleotides are still a long shot from even a minimal metabolism. The minimal intermediate metabolism proposed [149]–apart from the assembly of nucleotides–is able to harness energy, synthetize the membrane constituent, and produce the cofactors (Fig. 5). Energy (ATP) can be generated via glycolysis. For the investment of two ATPs, the cell gains four ATP from a molecule of glycose. Other sugars can be

channelled to this pathway either via the pentose-phosphate pathway or by specific isomerases able to convert sugars into sugars in that pathway. The pentose-phosphate pathway can then also be employed to produce ribose from other sugars. Whether all of these functions can be catalysed by ribozymes is still an open empirical questions, but as the repertoire of ribozymes seem to be quite diverse [3,186], we can assume that it is the case.

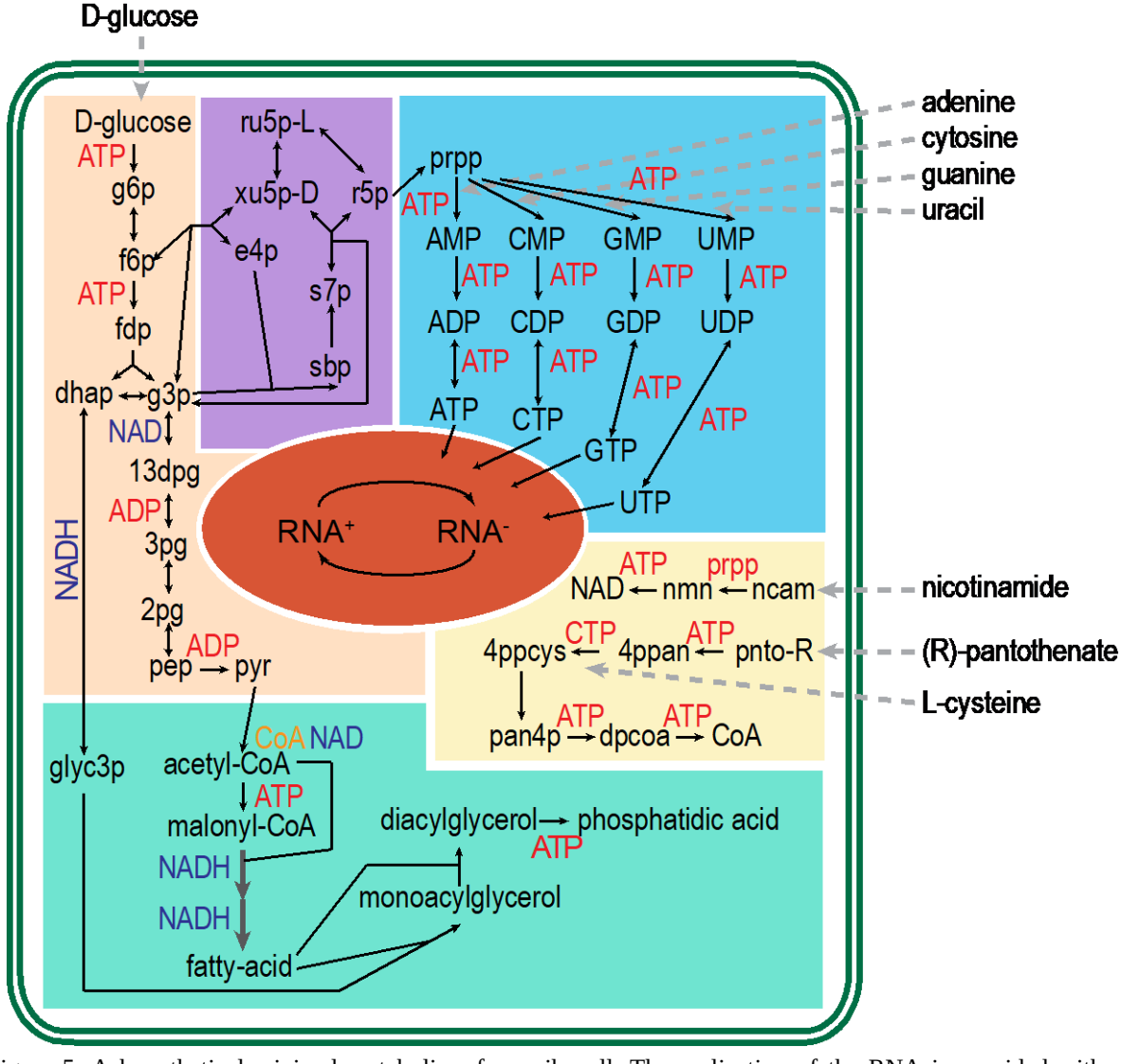

Figure 5. A hypothetical minimal metabolism for a ribocell. The replication of the RNA is provided with nucleotides and energy (where ADP is consumed ATP is produced); apart from CoA and NAD synthesis a sketch of phospholipid biosynthesis is also shown. Organic compounds to be taken up are depicted outside of the cell membrane. –Abbreviations are from the BiGG database [187]. 13dpg: 3-phospho-D-glyceroyl phosphate; 2pg: D-glycerate 2-phosphate; 3pg: 3-phospho-D-glycerate; 4ppan: D-4'-phosphopantothenate; 4ppcys: N-((R)-4-phosphopantothenoyl)-L-cysteine; dhap: dihydroxyacetone phosphate; dpcoa: dephospho-CoA; e4p: D-erythrose 4-phosphate; f6p: D-fructose 6-phosphate; fdp: D-fructose 1,6-bisphosphate; g3p: glyceraldehyde 3-phosphate; g6p: D-glucose 6-phosphate; glyc3p: glycerol 3-phosphate; ncam: nicotinamide; nmn: β-nicotinamide D-ribonucleotide; pan4p: pantetheine 4'-phosphate; pep: phosphoenolpyruvate; pnto-R: (R)-pantothenate; prpp: 5-phospho- α -D-ribose 1-diphosphate; pyr: pyruvate; r5p: α-D-ribose 5-phosphate; ru5p-L: L-ribulose 5-phosphate; s7p: sedoheptulose 7-phosphate; sbp: sedoheptulose-bisphosphatase; xu5p-D: D-xylulose 5-phosphate;

4.2 Cell level processes

Too often when contemplating metabolism, cell level processes are forgotten. We know more and more about the metabolic networks of organisms [188], and the core of this network is remarkably similar. But, transport, regulation of gene expression, control of cell division, etc. is very diverse.

For example, we have already assumed in the above minimal intermediate metabolism, that the ribocell can take up glycose or other sugar sources, nucleobases, the precursors of cofactors, some amino acids and inorganic materials (like phosphate). This requires some transporters as the membrane should not and cannot be fully permeable to these molecules. The lipid bilayer surrounding the cell not only provides the encapsulation needed for group selection to work but it also keeps the valuable materials inside the cell. Ions, for example phosphorylated compounds, can hardly cross the cell membrane [189,190] so they mostly stay inside or cannot enter the cell. Thus, the same mechanism protecting the cell from losing synthetized compounds also hinders the uptake of materials. Some kinds of transport that controls the in- and outflow of material is required. RNA can change the permeability of the membrane [191] and ribozymes can even act as membrane transporters [192] allowing control over the exchange of material with the environment. Whether RNA and the kind of membrane produced / formed by the ribocell can modulate permeability to the extent required is an empirical question. We have argued [150] that the original role for polypeptides could have been the formation of pores and channels. These polypeptides–if they existed–were not translated, just polymerized. Thus, while the full apparatus of translation was not required, a ribozyme–much like the ribosome–capable of amino-acid polymerization is needed. In this scenario, the transport or permeability modifying RNA is replaced by a polymerase, so the required number of genes is roughly the same.

One of the absolute life criteria is that processes are regulated and controlled. Genes for regulation, and cell levels processes are quite rare when homologous genes in multiple organisms are concerned. These genes are very much environment-dependent, and do not conserve well. However, they are still extremely important for the functioning of the cell. Their numbers are underestimated in minimal gene content estimates. On the other hand, present day functional RNAs are either directly connected with translation or with the regulation of gene expression [9-11], thus the regulatory role of RNA is well-preserved. But, it is not so easy to pinpoint what kind of regulation does a rudimentary riboorganism need.

Enzymatic activity can be controlled in two ways: either the enzyme itself respond to a signal, switching on and off, or the transcription–copying–of the enzyme from the template is affected by a signal. Both can be realized by RNAs. Present day functional RNAs, like small-interfering RNAs and microRNAs, mostly act post-transcriptionally [193], inhibiting translation from mRNAs. In an RNA world, such element would affect the product of transcription, the ribozymes themselves. And as such, the ribozymes then would fall into the class of allosterically controllable ribozymes, also referred to as aptazymes. Rationally designed and *in vitro* evolved aptazymes can respond to temperature change, light, small molecules or oligonucleotides [194-198]. A smaller set of RNAs can also directly control translation, inhibiting or facilitating the production of enzymes from the chromosome. Furthermore, small molecules can modulate regulatory regions in the chromosome, as in the lac-operon. Such regulatory mechanism could

have existed in the RNA world as well. As for our estimate of minimal gene content, the information to be stored is longer with the incorporation of regulatory regions of genes or the effector-recognition domains of allosterically controlled ribozymes. Especially for the shorter ribozymes, some increase in length might be possible if it conveys a fitness advantage. We should not forget, that the error threshold still looms large on the horizon.

The evolution of chromosome solves the problem stemming from the random distribution of independently-replicating genes to daughter cells. At the same time, it also allows very specialized enzymes to evolve [199]. But it also requires an array of enzymes to function [200], as ribozymes need to be transcribed from a large molecule containing all ribozymes ligated together. While cleavage seems to be an easily evolvable function, it needs to be site-specific across the chromosome, so it only cleaves between ribozymes, but not within them. This is yet another cell-level function a riboorganism requires.

A proposed function-set will pose a challenge to empirical ribozyme research: ribozymes should be able to catalyse the proposed reactions, and these ribozymes need to be able to work together in one compartment. Not only the maintenance of large set of independent replicators but their working together biochemically is a challenge. Ribozymes are usually metallozymes [201,202], and altering the prevalent metal-ion concentration can change their enzymatic activity [203]. Ribozymes are evolved by themselves, and there is scarce indication that multiple ribozymes can function at the same time. We know that two [204,205] or three [206] engineered ribozymes can work in concert. The challenge is to have 60–100 ribozymes to work together in a cell. As we are very far from realizing this number, a more accurate estimate would be superfluous.

We are a long way from solving the mystery of the first cell, but more and more of the puzzle-pieces are known. The problems, both dynamical and structural, have been identified, and for some, solutions proposed. Here we have reviewed some of the dynamical problems the first cell needed to overcome via having the right set of ribozymes cooperating with each other.

Acknowledgement

The research was funded by the National Research, Development and Innovation Office (NKFIH) under grant numbers K119347 and GINOP-2.3.2-15-2016-00057, and by the Volkswagen Stiftung initiative “Leben? – Ein neuer Blick der Naturwissenschaften auf die grundlegenden Prinzipien des Lebens” under project “A unified model of recombination in life”. This work was carried out as part of EU COST action CM1304 “Emergence and Evolution of Complex Chemical Systems.”

References

1. Yarus, M. (2011) *Life from an RNA World: The Ancestor Within*. Harvard University Press: Harvard, USA.
2. Kun, Á.; Szilágyi, A.; Könnyű, B.; Boza, G.; Zachár, I.; Szathmáry, E. (2015) The dynamics of the RNA world: Insights and challenges. *Ann. N.Y. Acad. Sci.* **1341**, 75–95.
3. Joyce, G.F. (2002) The antiquity of RNA-based evolution. *Nature* **418**(6894), 214–220.
4. Gánti, T. (2003) *The Principles of Life*. Oxford University Press: Oxford.
5. Gánti, T. (1971) *Az Élet Princípiuma*. Gondolat: Budapest.

6. Szathmáry, E. (2015) Toward major evolutionary transitions theory 2.0. *PNAS* **112**(33), 10104–10111.
7. Maynard Smith, J.; Szathmáry, E. (1995) *The Major Transition in Evolution*. W.H. Freeman: Oxford, UK.
8. Szathmáry, E.; Maynard Smith, J. (1995) The major evolutionary transitions. *Nature* **374**, 227–232.
9. Meli, M.; Albert-Fournier, B.; Maurel, M.C. (2001) Recent findings in the modern RNA world. *Int. Microbiol.* **4**(1), 5–11.
10. Spirin, A.S. (2002) Omnipotent RNA. *FEBS Lett.* **530**(1–3), 4–8.
11. Collins, L.J.; Kurland, C.G.; Biggs, P.; Penny, D. (2009) The modern RNP world of eukaryotes. *J. Hered.* **100**(5), 597–604.
12. Huang, B.; Zhang, R. (2014) Regulatory non-coding RNAs: revolutionizing the RNA world. *Mol. Biol. Rep.* **41**(6), 3915–3923.
13. Patil, V.S.; Zhou, R.; Rana, T.M. (2014) Gene regulation by non-coding RNAs. *Crit. Rev. Biochem. Mol. Biol.* **49**(1), 16–32.
14. Ghildiyal, M.; Zamore, P.D. (2009) Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* **10**(2), 94–108.
15. Moore, P.B.; Steitz, T.A. (2002) The involvement of RNA in ribosome function. *Nature* **418**(6894), 229–235.
16. Nissen, P.; Hansen, J.; Ban, N.; Moore, P.B.; Steitz, T.A. (2000) The structural basis of ribosome activity in peptide bond synthesis. *Science* **289**(5481), 920–930.
17. Guerrier-Takada, C.; Gardiner, K.; Marsh, T.; Pace, N.; Altman, S. (1983) The RNA moiety of ribonuclease P is the catalytic subunit of the enzyme. *Cell* **35**(3), 849–857.
18. Kruger, K.; Grabowski, P.; Zaug, A.J.; Sands, J.; Gottschling, D.E.; Cech, T.R. (1982) Self-splicing RNA: autoexcision and autocyclization of the ribosomal RNA intervening sequence of *Tetrahymena*. *Cell* **31**(1), 147–157.
19. Peebles, C.L.; Perlman, P.S.; Mecklenburg, K.L.; Pertillo, M.L.; Tabor, J.H.; Jarrell, K.A.; Cheng, H.-L. (1986) A self-splicing RNA excises an intron lariat. *Cell* **44**(2), 213–223.
20. Forster, A.C.; Symons, R.H. (1987) Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active site. *Cell* **49**(2), 211–220.
21. Hampel, A.; Tritz, R.R. (1989) RNA catalytic properties of the minimum (-)sTRSV sequences. *Biochemistry* **28**(12), 4929–4933.
22. Sharmeen, L.; Kuo, M.Y.P.; Dinner-Gottlieb, G.; Taylor, J. (1988) Antigenomic RNA of human hepatitis delta viruses can undergo self-cleavage. *J. Virol.* **62**(8), 2674–2679.
23. Saville, B.J.; Collins, R.A. (1990) A site-specific self-cleavage reaction performed by a novel RNA in *Neurospora* mitochondria. *Cell* **61**(4), 685–696.
24. Winkler, W.C.; Nahvi, A.; Roth, A.; Collins, J.A.; Breaker, R.R. (2004) Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**(6980), 281–286.
25. Roth, A.; Weinberg, Z.; Chen, A.G.Y.; Kim, P.B.; Ames, T.D.; Breaker, R.R. (2014) A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat. Chem. Biol.* **10**(1), 56–60.
26. Bag, B.G.; von Kiedrowski, G. (1996) Templates, autocatalysis and molecular replication. *Pure Appl. Chem.* **68**(11), 2145.
27. Szathmáry, E.; Gladkih, I. (1989) Sub-exponential growth and coexistence of non-enzymatically replicating templates. *J. Theor. Biol.* **138**(1), 55–58.
28. Zachar, I.; Kun, Á.; Fernando, C.; Szathmáry, E. (2013) Replicators: From molecules to organisms. In *Handbook of Collective Robotics: Fundamentals and Challenges*, Kernbach, S., Ed. Pan Stanford Publishing.
29. Orgel, L.E. (1992) Molecular replication. *Nature* **358**(6383), 203–209.
30. Bissette, A.J.; Fletcher, S.P. (2013) Mechanisms of autocatalysis. *Angew. Chem. Int. Ed.* **52**(49), 12800–12826.
31. Kassianidis, E.; Philp, D. (2006) Design and implementation of a highly selective minimal self-replicating system. *Angew. Chem. Int. Ed.* **45**(38), 6344–6348.
32. Patzke, V.; Von Kiedrowski, G. (2007) Self replicating systems *ARKIVOC*, 293–310.
33. von Kiedrowski, G. (1986) A self-replicating hexadeoxynucleotide. *Angew. Chem. Int. Ed.* **25**(10), 932–935.
34. Segré, D.; Ben-Eli, D.; Lancet, D. (2000) Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies. *PNAS* **97**(8), 4112–4117.
35. Vasas, V.; Szathmáry, E.; Santos, M. (2010) Lack of evolvability in self-sustaining autocatalytic networks constraints metabolism-first scenarios for the origin of life. *PNAS* **107**(4), 1470–1475.
36. Huang, W.; Ferris, J.P. (2003) Synthesis of 35-40 mers of RNA oligomers from unblocked monomers. A simple approach to the RNA world. *Chem Commun (Camb)*, 1458–1459.

37. Ferris, J.P. (2006) Montmorillonite-catalysed formation of RNA oligomers: The possible role of catalysis in the origins of life. *Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci.* **361**(1474), 1777–1786.
38. Szostak, J. (2012) The eightfold path to non-enzymatic RNA replication. *J. Sys. Chem.* **3**(1), 2.
39. Ekland, E.H.; Bartel, D.P. (1996) RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**, 373–376.
40. Johnston, W.K.; Unrau, P.J.; Lawrence, M.S.; Glasen, M.E.; Bartel, D.P. (2001) RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**(5520), 1319–1325.
41. Zaher, H.S.; Unrau, P.J. (2007) Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **13**(7), 1017–1026.
42. Wochner, A.; Attwater, J.; Coulson, A.; Holliger, P. (2011) Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332**(6026), 209–212.
43. Attwater, J.; Wochner, A.; Holliger, P. (2013) In-ice evolution of RNA polymerase ribozyme activity. *Nature Chemistry* **5**, 1011–1018.
44. Szathmáry, E. (1999) Chemes, genes, memes: A revised classification of replicators. *Lectures on Mathematics in the Life Sciences* **26**, 1–10.
45. Kun, Á.; Papp, B.; Szathmáry, E. (2008) Computational identification of obligatorily autocatalytic replicators embedded in metabolic networks. *Genome Biol.* **9**, R51.
46. Gánti, T. (2003) *Chemoton Theory*. Kluwer Academic/Plenum Publishers: New York.
47. Szathmáry, E. (2006) The origin of replicators and reproducers. *Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci.* **361**(1474), 1761–1776.
48. Szathmáry, E. (2005) Life: in search of the simplest cell. *Nature* **433**, 469–470.
49. Szathmáry, E.; Mauro, S.; Fernando, C. (2005) Evolutionary potential and requirements for minimal protocells. *Top. Curr. Chem.* **259**, 167–211.
50. Wächtershäuser, G. (1998) Origins of life in an iron-sulfur world. In *The Molecular Origins of Life*, Brack, A., Ed. Cambridge University Press: Cambridge, pp 207–218.
51. Ferris, J.P. (2002) Montmorillonite catalysis of 30–50 mer oligonucleotides: laboratory demonstration of potential steps in the origin of the RNA world. *Origins Life Evol. Biosphere* **32**(4), 311–332.
52. Joshi, P.C.; Aldersley, M.F.; Ferris, J.P. (2011) Homochiral selectivity in RNA synthesis: montmorillonite-catalyzed quaternary reactions of D, L-purine with D, L- pyrimidine nucleotides. *Origins Life Evol. Biosphere* **41**(3), 213–236.
53. Hazen, R.M.; Filley, T.R.; Goodfriend, G.A. (2001) Selective adsorption of l- and d-amino acids on calcite: Implications for biochemical homochirality. *PNAS* **98**(10), 5487–5490.
54. Biondi, E.; Branciamore, S.; Maurel, M.-C.; Gallori, E. (2007) Montmorillonite protection of an UV-irradiated hairpin ribozyme: evolution of the RNA world in a mineral environment. *BMC Evol. Biol.* **7**(suppl. 2), S2.
55. Poole, A.M.; Logan, D.T. (2005) Modern mRNA proofreading and repair: clues that the Last Universal Common Ancestor possessed an RNA genome? *Mol. Biol. Evol.* **22**(6), 1444–1455.
56. Eigen, M. (1971) Selforganization of matter and the evolution of biological macromolecules. *Naturwissenscaften* **10**, 465–523.
57. Maynard Smith, J. (1983) Models of evolution. *Proc. R. Soc. London, Ser. B* **219**(1216), 315–325.
58. Levinson, G.; Gutman, G.A. (1987) Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* **4**(3), 203–221.
59. Gregory, T.R. (2004) Insertion–deletion biases and the evolution of genome size. *Gene* **324**, 15–34.
60. Zhang, W.; Sun, X.; Yuan, H.; Araki, H.; Wang, J.; Tian, D. (2008) The pattern of insertion/deletion polymorphism in *Arabidopsis thaliana*. *Mol. Genet. Genomics* **280**(4), 351–361.
61. Bhangale, T.R.; Rieder, M.J.; Livingston, R.J.; Nickerson, D.A. (2005) Comprehensive identification and characterization of diallelic insertion–deletion polymorphisms in 330 human candidate genes. *Hum. Mol. Genet.* **14**(1), 59–69.
62. Boschiero, C.; Gheyas, A.A.; Ralph, H.K.; Eory, L.; Paton, B.; Kuo, R.; Fulton, J.; Preisinger, R.; Kaiser, P.; Burt, D.W. (2015) Detection and characterization of small insertion and deletion genetic variants in modern layer chicken genomes. *BMC Genomics* **16**(1), 562.
63. Mills, D.R.; Peterson, R.E.; Spiegelman, S. (1967) An extracellular Darwinian experiment with a self-duplicating nucleic acid molecule. *PNAS* **58**, 217–224.
64. Matsumura, S.; Kun, Á.; Ryckelynck, M.; Coldren, F.; Szilágyi, A.; Jossinet, F.; Rick, C.; Nghe, P.; Szathmáry, E.; Griffiths, A.D. (2016) Transient compartmentalization of RNA replicators prevents extinction due to parasites. *Science* **354**(6317), 1293–1296.
65. Tromas, N.; Elena, S.F. (2010) The rate and spectrum of spontaneous mutations in a plant RNA virus. *Genetics* **185**(3), 983–989.

66. Campagnola, G.; McDonald, S.; Beaucourt, S.; Vignuzzi, M.; Peersen, O.B. (2015) Structure-function relationships underlying the replication fidelity of viral RNA-dependent RNA polymerases. *J. Virol.* **89**(1), 275–286.
67. Huang, J.; Brieba, L.G.; Sousa, R. (2000) Misincorporation by wild-type and mutant T7 RNA polymerases: identification of interactions that reduce misincorporation rates by stabilizing the catalytically incompetent open conformation. *Biochemistry* **39**(38), 11571–11580.
68. Drake, J.W. (1993) Rates of spontaneous mutation among RNA viruses. *PNAS* **90**, 4171–4175.
69. Sanjuán, R.; Nebot, M.R.; Chirico, N.; Mansky, L.M.; Belshaw, R. (2010) Viral mutation rates. *J. Virol.* **84**(19), 9733–9748.
70. Eyre-Walker, A.; Keightley, P.D. (2007) The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**(8), 610–618.
71. Lundin, E.; Tang, P.-C.; Guy, L.; Näsvall, J.; Andersson, D.I. (2018) Experimental determination and prediction of the fitness effects of random point mutations in the biosynthetic enzyme HisA. *Mol. Biol. Evol.* **35**(3), 704–718.
72. Jacquier, H.; Birgy, A.; Le Nagard, H.; Mechulam, Y.; Schmitt, E.; Glodt, J.; Bercot, B.; Petit, E.; Poulain, J.; Barnaud, G., *et al.* (2013) Capturing the mutational landscape of the beta-lactamase TEM-1. *Proceedings of the National Academy of Sciences* **110**(32), 13067–13072.
73. Lind, P.A.; Berg, O.G.; Andersson, D.I. (2010) Mutational robustness of ribosomal protein genes. *Science* **330**(6005), 825–827.
74. Starita, L.M.; Young, D.L.; Islam, M.; Kitzman, J.O.; Gullingsrud, J.; Hause, R.J.; Fowler, D.M.; Parvin, J.D.; Shendure, J.; Fields, S. (2015) Massively parallel functional analysis of BRCA1 RING domain variants. *Genetics* **200**(2), 413–422.
75. Huynen, M.A.; Stadler, P.F.; Fontana, W. (1996) Smoothness within ruggedness: the role of neutrality in adaptation. *PNAS* **93**(1), 397–401.
76. Huynen, M.A. (1996) Exploring phenotype space through neutral evolution. *J. Mol. Evol.* **43**, 165–169.
77. van Nimwegen, E.; Crutchfield, J.P.; Huynen, M.A. (1999) Neutral evolution of mutational robustness. *PNAS* **96**(17), 9716–9720.
78. Haslinger, C.; Stadler, P.F. (1999) RNA structure with pseudo-knots: graph-theoretical and combinatorial properties. *Bull. Math. Biol.* **61**(3), 437–467.
79. Kun, Á.; Santos, M.; Szathmáry, E. (2005) Real ribozymes suggest a relaxed error threshold. *Nat. Genet.* **37**(9), 1008–1011.
80. Takeuchi, N.; Poorthuis, P.H.; Hogeweg, P. (2005) Phenotypic error threshold; additivity and epistasis in RNA evolution. *BMC Evol. Biol.* **5**(1), 9.
81. Szathmáry, E. (1991) Four letters in the genetic alphabet: a frozen evolutionary optimum? *Proc. R. Soc. London, Ser. B* **245**, 91–99.
82. Szathmáry, E. (1992) What is the optimum size for the genetic alphabet? *PNAS* **89**, 2614–2618.
83. Hershberg, R.; Petrov, D.A. (2010) Evidence that mutation is universally biased towards AT in Bacteria. *PLoS Genet.* **6**(9), e1001115.
84. Krishnamurthy, R. (2015) On the emergence of RNA. *Isr. J. Chem.* **55**(8), 837–850.
85. Eschenmoser, A. (1999) Chemical etiology of nucleic acid structure. *Science* **284**(5423), 2118–2124.
86. Fu, L.-Y.; Wang, G.-Z.; Ma, B.-G.; Zhang, H.-Y. (2011) Exploring the common molecular basis for the universal DNA mutation bias: Revival of Löwdin mutation model. *Biochem. Biophys. Res. Commun.* **409**(3), 367–371.
87. Schuster, P.; Stadler, P.F. (1999) Nature and evolution of early replicons. In *Origin and Evolution of Viruses*, Domingo, E.; Webster, R.G.; Holland, J., Eds. Academic Press: New York, pp 1–24.
88. Kun, Á.; Maurel, M.-C.; Santos, M.; Szathmáry, E. (2005) Fitness landscapes, error thresholds, and cofactors in aptamer evolution. In *The aptamer handbook*, Klussmann, S., Ed. WILEY-VCH Verlag GmbH & Co. KGaA: Weinheim, pp 54–92.
89. Hofacker, I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.* **31**, 3429–3431.
90. Hofacker, I.L.; Fontana, W.; Stadler, P.F.; Bonhoeffer, S.; Tacker, M.; Schuster, P. (1994) Fast folding and comparison of RNA secondary structures. *Monatchefte für Chemie* **125**, 167–188.
91. Lorenz, R.; Bernhart, S.H.; Höner zu Siederdissen, C.; Tafer, H.; Flamm, C.; Stadler, P.F.; Hofacker, I.L. (2011) ViennaRNA Package 2.0. *Algorithms for Molecular Biology* **6**(1), 26.
92. Szilágyi, A.; Kun, Á.; Szathmáry, E. (2014) Local neutral networks help maintain inaccurately replicating ribozymes. *PLoS ONE* **9**(10), e109987.
93. Reidys, C.; Forst, C.V.; Schuster, P. (2001) Replication and mutation on neutral networks. *Bull. Math. Biol.* **63**, 57–94.
94. Kun, Á.; Szathmáry, E. (2015) Fitness landscapes of functional RNAs. *Life* **5**(3), 1497–1517.

95. Scheuring, I. (2000) Avoiding Catch-22 of early evolution by stepwise increase in copying fidelity. *Selection* **1**, 13–23.
96. Wang, Q.S.; Cheng, L.K.L.; Unrau, P.J. (2011) Characterization of the B6.61 polymerase ribozyme accessory domain. *RNA* **17**(3), 469–477.
97. Attwater, J.; Wochner, A.; Pinheiro, V.B.; Coulson, A.; Holliger, P. (2010) Ice as a protocellular medium for RNA replication. *Nat. Comm.* **1**, 76.
98. Horning, D.P.; Joyce, G.F. (2016) Amplification of RNA by an RNA polymerase ribozyme. *PNAS* **113**(35), 9786–9791.
99. Attwater, J.; Raguram, A.; Morgunov, A.S.; Gianni, E.; Holliger, P. (2018) Ribozyme-catalysed RNA synthesis using triplet building blocks. *eLife* **7**, e35255.
100. Eigen, M.; Schuster, P. (1979) *The Hypercycle: A Principle of Natural Self-organization*. Springer-Verlag: Berlin.
101. Szilágyi, A.; Zachar, I.; Scheuring, I.; Kun, Á.; Könnyű, B.; Czárán, T. (2017) Ecology and evolution in the RNA World: Dynamics and stability of prebiotic replicator systems. *Life* **7**(4), 48.
102. Nowak, M.A. (2006) Five rules for the evolution of cooperation. *Science* **314**(5805), 1560–1563.
103. Koonin, E.V.; Martin, W. (2005) On the origin of genomes and cells within inorganic compartments. *Trends Genet.* **21**(12), 647–654.
104. Martin, W.; Russell, M.J. (2007) On the origin of biochemistry at an alkaline hydrothermal vent. *Philos. Trans. R. Soc. London, Ser. B* **362**(1486), 1887–1926.
105. Sleep, N.H.; Bird, D.K.; Pope, E.C. (2011) Serpentinite and the dawn of life. *Philos. Trans. R. Soc. London, Ser. B* **366**(1580), 2857–2869.
106. Russell, M.J.; Hall, A.J.; Martin, W. (2010) Serpentinization as a source of energy at the origin of life. *Geobiology* **8**(5), 355–371.
107. Branciamore, S.; Gallori, E.; Szathmáry, E.; Czárán, T. (2009) The origin of life: chemical evolution of a metabolic system in a mineral honeycomb? *J. Mol. Evol.* **69**(5), 458–469.
108. Vlassov, A.V.; Johnston, B.H.; Landweber, L.F.; Kazakov, S.A. (2004) Ligation activity of fragmented ribozymes in frozen solution: implications for the RNA world. *Nucleic Acids Res.* **32**(9), 2966–2974.
109. Trinks, H.; Schröder, W.; Biebricher, C.K. (2005) Ice and the origin of life. *Origins Life Evol. Biosphere* **35**(5), 429–445.
110. Kanavarioti, A.; Monnard, P.-A.; Deamer, D.W. (2001) Eutectic phases in ice facilitate nonenzymatic nucleic acid synthesis. *Astrobiology* **1**(3), 271–281.
111. Czárán, T.; Szathmáry, E. (2000) Coexistence of replicators in prebiotic evolution. In *The Geometry of Ecological Interactions*, Dieckmann, U.; Law, R.; Metz, J.A.J., Eds. Cambridge University Press: Cambridge, pp 116–134.
112. Czárán, T.; Könnyű, B.; Szathmáry, E. (2015) Metabolically Coupled Replicator Systems: Overview of an RNA-world model concept of prebiotic evolution on mineral surfaces. *J. Theor. Biol.* **381**, 39–54.
113. Takeuchi, N.; Hogeweg, P. (2009) Multilevel selection in models of prebiotic evolution II: A direct comparison of compartmentalization and spatial self-organization. *PLoS Comp. Biol.* **5**(10), e1000542.
114. Hogeweg, P.; Takeuchi, N. (2003) Multilevel selection in models of prebiotic evolution: Compartments and spatial self-organization. *Origins Life Evol. Biosphere* **33**(4–5), 375–403.
115. Colizzi, E.S.; Hogeweg, P. (2016) Parasites sustain and enhance RNA-like replicators through spatial self-organisation. *PLoS Comp. Biol.* **12**(4), e1004902.
116. Könnyű, B.; Czárán, T.; Szathmáry, E. (2008) Prebiotic replicase evolution in a surface-bound metabolic system: parasites as a source of adaptive evolution. *BMC Evol. Biol.* **8**, 267.
117. Könnyű, B.; Czárán, T. (2013) Spatial aspects of prebiotic replicator coexistence and community stability in a surface-bound RNA world model. *BMC Evol. Biol.* **13**, 204.
118. Deamer, D.W.; Barchfeld, G.L. (1982) Encapsulation of macromolecules by lipid vesicles under simulated prebiotic conditions. *J. Mol. Evol.* **18**(3), 203–206.
119. Rajamani, S.; Vlassov, A.; Benner, S.; Coombs, A.; Olasagasti, F.; Deamer, D. (2008) Lipid-assisted synthesis of RNA-like polymers from mononucleotides. *Origins Life Evol. Biosphere* **38**(1), 57–74.
120. Guo, M.T.; Rotem, A.; Heyman, J.A.; Weitz, D.A. (2012) Droplet microfluidics for high-throughput biological assays. *Lab on a Chip* **12**(12), 2146–2155.
121. Griffiths, A.D.; Tawfik, D.S. (2006) Miniaturising the laboratory in emulsion droplets. *Trends Biotechnol.* **24**(9), 395–402.
122. Guo, H.C.T.; Collins, R.A. (1995) Efficient *trans*-cleavage of a stem-loop RNA substrate by a ribozyme derived from *Neurospora* VS RNA. *The EMBO Journal* **14**(2), 368–376.
123. Hubai, A.G.; Kun, Á. (2016) Maximal gene number maintainable by stochastic correction – The second error threshold. *J. Theor. Biol.* **405**, 29–35.

124. Szilágyi, A.; Zachar, I.; Szathmáry, E. (2013) Gause's principle and the effect of resource partitioning on the dynamical coexistence of replicating templates. *PLoS Comp. Biol.* **9**(8), e1003193.
125. Szathmáry, E.; Demeter, L. (1987) Group selection of early replicators and the origin of life. *J. Theor. Biol.* **128**(4), 463–486.
126. Grey, D.; Hutson, V.; Szathmáry, E. (1995) A re-examination of the stochastic corrector model. *Proc. R. Soc. London, Ser. B* **262**(1363), 29–35.
127. Fontanari, J.F.; Santos, M.; Szathmáry, E. (2006) Coexistence and error propagation in pre-biotic vesicle models: A group selection approach. *J. Theor. Biol.* **239**(2), 247–256.
128. Hubai, A.G.; Kun, Á. The coexistence of independent genes is aided by multilevel selection, but only to a limited extent. In *Modelling Biological Evolution 2017: Developing Novel Approaches*, Leicester, UK, 2017.
129. Vig-Milkovics, Z.; Zachar, I.; Kun, Á.; Szilágyi, A.; Szathmáry, E. (2018) Moderate sex between protocells can balance between a decrease in assortment load and an increase in parasite spread.
130. Santos, M.; Zintzaras, E.; Szathmáry, E. (2003) Origin of sex revisited. *Origins of Life and Evolution of the Bioshpere* **33**, 405–432.
131. Niesert, U.; Harnasch, D.; Bresch, C. (1981) Origin of life between Scylla and Charybdis. *J. Mol. Evol.* **17**(6), 348–353.
132. Mizuuchi, R.; Ichihashi, N. (2018) Sustainable replication and coevolution of cooperative RNAs in an artificial cell-like system. *Nature Ecology & Evolution*, doi:10.1038/s41559-41018-40650-z.
133. Koonin, E.V. (2000) How many genes can make a cell: The Minimal-Gene-Set Concept. *Annu. Rev. Genom. Hum. Genet.* **1**(1), 99–116.
134. Fehér, T.; Papp, B.; Pál, C.; Pósfai, G. (2007) Systematic genome reductions: Theoretical and experimental approaches. *Chem. Rev.* **107**(8), 3498–3513.
135. Islas, S.; Becerra, A.; Luisi, P.L.; Lazcano, A. (2004) Comparative genomics and the gene complement of a minimal cell. *Origins Life Evol. Biosphere* **34**(1), 243–256.
136. Benner, S.A.; Ellington, A.D.; Tauer, A. (1989) Modern metabolism as a palimpsest of the RNA world. *PNAS* **86**(18), 7054–7058.
137. McCutcheon, John P.; von Dohlen, Carol D. (2011) An interdependent metabolic patchwork in the nested symbiosis of mealybugs. *Curr. Biol.* **21**(16), 1366–1372.
138. Bennett, G.M.; Moran, N.A. (2013) Small, smaller, smallest: The origins and evolution of ancient dual symbioses in a phloem-feeding insect. *Genome Biology and Evolution* **5**(9), 1675–1688.
139. McCutcheon, J.P.; McDonald, B.R.; Moran, N.A. (2009) Origin of an alternative genetic code in the extremely small and GC–rich genome of a bacterial symbiont. *PLoS Genet.* **5**(7), e1000565.
140. Tamames, J.; Gil, R.; Latorre, A.; Pereto, J.; Silva, F.; Moya, A. (2007) The frontier between cell and organelle: genome analysis of *Candidatus* Carsonella ruddii. *BMC Evol. Biol.* **7**(1), 181.
141. McCutcheon, J.P.; Moran, N.A. (2010) Functional convergence in reduced genomes of bacterial symbionts spanning 200 My of evolution. *Genome Biology and Evolution* **2**, 708–718.
142. Chang, H.-H.; Cho, S.-T.; Canale, M.C.; Mugford, S.T.; Lopes, J.R.S.; Hogenhout, S.A.; Kuo, C.-H. (2015) Complete genome sequence of "Candidatus *Sulcia muelleri*" ML, an obligate nutritional symbiont of maize leafhopper (*Dalbulus maidis*). *Genome Announcements* **3**(1).
143. McCutcheon, J.P.; Moran, N.A. (2007) Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *PNAS* **104**(49), 19392–19397.
144. Woyke, T.; Tighe, D.; Mavromatis, K.; Clum, A.; Copeland, A.; Schackwitz, W.; Lapidus, A.; Wu, D.; McCutcheon, J.P.; McDonald, B.R., *et al.* (2010) One bacterial cell, one complete genome. *PLoS ONE* **5**(4), e10314.
145. Wu, D.; Daugherty, S.C.; Van Aken, S.E.; Pai, G.H.; Watkins, K.L.; Khouri, H.; Tallon, L.J.; Zaborsky, J.M.; Dunbar, H.E.; Tran, P.L., *et al.* (2006) Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PLoS Biol.* **4**(6), e188.
146. McCutcheon, J.P.; Moran, N.A. (2011) Extreme genome reduction in symbiotic bacteria. *Nature Reviews Microbiology* **10**, 13–26.
147. Hutchison, C.A.; Chuang, R.-Y.; Noskov, V.N.; Assad-Garcia, N.; Deerinck, T.J.; Ellisman, M.H.; Gill, J.; Kannan, K.; Karas, B.J.; Ma, L., *et al.* (2016) Design and synthesis of a minimal bacterial genome. *Science* **351**(6280), aad6253.
148. Gil, R.; Silva, F.J.; Peretó, J.; Moya, A. (2004) Determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev.* **68**(3), 518–537.
149. Gabaldón, T.; Peretó, J.; Montero, F.; Gil, R.; Latorre, A.; Moya, A. (2007) Structural analyses of a hypothetical minimal metabolism. *Philos. Trans. R. Soc. Lond., Ser. B: Biol. Sci.* **362**(1486), 1761–1762.

150. Kun, Á.; Radványi, Á. (2018) The evolution of the genetic code: Impasses and challenges. *BioSyst.* **164**, 217–225.
151. Reuß, D.R.; Commichau, F.M.; Gundlach, J.; Zhu, B.; Stülke, J. (2016) The blueprint of a minimal cell: *MiniBacillus*. *Microbiol. Mol. Biol. Rev.* **80**(4), 955–987.
152. Cheng, L.K.L.; Unrau, P.J. (2010) Closing the circle: Replicating RNA with RNA. *Cold Spring Harb. Perspect. Biol.* **2**(10), a002204.
153. Mutschler, H.; Wochner, A.; Holliger, P. (2015) Freeze–thaw cycles as drivers of complex ribozyme assembly. *Nature Chemistry* **7**(6), 502–508.
154. Kováč, L.; Nosek, J.; Tomáška, L.u. (2003) An overlooked riddle of life's origins: Energy-dependent nucleic acid unzipping. *J. Mol. Evol.* **57**(1), S182–S189.
155. Kreysing, M.; Keil, L.; Lanzmich, S.; Braun, D. (2015) Heat flux across an open pore enables the continuous replication and selection of oligonucleotides towards increasing length. *Nature Chemistry* **advance online publication**.
156. Krammer, H.; Möller, F.M.; Braun, D. (2012) Thermal, autonomous replicator made from transfer RNA. *Phys. Rev. Lett.* **108**(23), 238104.
157. Ball, R.; Brindley, J. (2014) Hydrogen peroxide thermochemical oscillator as driver for primordial RNA replication. *Journal of The Royal Society Interface* **11**(95), 20131052.
158. Zenkin, N. (2012) Hypothesis: Emergence of translation as a result of RNA helicase evolution. *J. Mol. Evol.* **74**(5), 249–256.
159. Martin, L.; Unrau, P.; Müller, U. (2015) RNA synthesis by *in vitro* selected ribozymes for recreating an RNA world. *Life* **5**(1), 247.
160. Saladino, R.; Crestini, C.; Pino, S.; Costanzo, G.; Di Mauro, E. (2012) Formamide and the origin of life. *Phys. Life. Rev.* **9**(1), 84–104.
161. Saladino, R.; Crestini, C.; Costanzo, G.; Negri, R.; Di Mauro, E. (2001) A possible prebiotic synthesis of purine, adenine, cytosine, and 4(3H)-pyrimidinone from formamide implications for the origin of life. *Biorg. Med. Chem.* **9**(5), 1249–1253.
162. Pino, S.; Sponer, J.; Costanzo, G.; Saladino, R.; Mauro, E. (2015) From formamide to RNA, the path is tenuous but continuous. *Life* **5**(1), 372.
163. Powner, M.W.; Gerland, B.; Sutherland, J.D. (2009) Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242.
164. Patel, B.H.; Percivalle, C.; Ritson, D.J.; DuffyColm, D.; Sutherland, J.D. (2015) Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nature Chemistry* **7**(4), 301–307.
165. Saladino, R.; Šponer, J.; Šponer, J.; Costanzo, G.; Pino, S.; Di Mauro, E. (2018) Chemomimesis and molecular Darwinism in action: From abiotic generation of nucleobases to nucleosides and RNA. *Life* **8**(2), 24.
166. Lau, M.W.L.; Cadieux, K.E.C.; Unrau, P.J. (2004) Isolation of fast purine nucleotide synthase ribozymes. *JACS* **126**(48), 15686–15693.
167. Unrau, P.J.; Bartel, D.P. (1998) RNA-catalysed nucleotide synthesis. *Nature* **395**(6699), 260–263.
168. Moretti, J.E.; Müller, U.F. (2014) A ribozyme that triphosphorylates RNA 5′-hydroxyl groups. *Nucleic Acids Res.* **42**(7), 4767–4778.
169. Camden, A.J.; Walsh, S.M.; Suk, S.H.; Silverman, S.K. (2016) DNA oligonucleotide 3′-phosphorylation by a DNA enzyme. *Biochemistry* **55**(18), 2671–2676.
170. Biondi, E.; Maxwell, A.W.R.; Burke, D.H. (2012) A small ribozyme with dual-site kinase activity. *Nucleic Acids Res.* **40**(15), 7528–7540.
171. Li, Y.; Breaker, R.R. (1999) Phosphorylating DNA with DNA. *PNAS* **96**(6), 2746–2751.
172. Curtis, E.A.; Bartel, D.P. (2005) New catalytic structures from an existing ribozyme. *Nat. Struct. Mol. Biol.* **12**, 994–1000.
173. Lorsch, J.R.; Szostak, J.W. (1994) *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature* **371**, 31.
174. Saran, D.; Nickens, D.G.; Burke, D.H. (2005) A trans acting ribozyme that phosphorylates exogenous RNA. *Biochemistry* **44**(45), 15007–15016.
175. Jiménez, J.I.; Xulvi-Brunet, R.; Campbell, G.W.; Turk-MacLeod, R.; Chen, I.A. (2013) Comprehensive experimental fitness landscape and evolutionary network for small RNA. *PNAS* **110**(37), 14984–14989.
176. Davis, J.H.; Szostak, J.W. (2002) Isolation of high-affinity GTP aptamers from partially structured RNA libraries. *Proceedings of the National Academy of Sciences* **99**(18), 11616–11621.
177. Sassanfar, M.; Szostak, J.W. (1993) An RNA motif that binds ATP. *Nature* **364**(6437), 550–553.

178. Vu, Michael M.K.; Jameson, Nora E.; Masuda, Stuart J.; Lin, D.; Larralde-Ridaura, R.; Lupták, A. (2012) Convergent evolution of adenosine aptamers spanning bacterial, human, and random sequences revealed by structure-based bioinformatics and genomic SELEX. *Chem. Biol.* **19**(10), 1247–1254.
179. Curtis, E.A.; Liu, D.R. (2013) Discovery of widespread GTP-binding motifs in genomic DNA and RNA. *Chem. Biol.* **20**(4), 521–532.
180. Sacerdote, M.G.; Szostak, J.W. (2005) Semipermeable lipid bilayers exhibit diastereoselectivity favoring ribose. *PNAS* **102**(17), 6004–6008.
181. Breslow, R. (1959) On the mechanism of the formose reaction. *Tetrahedron Lett.* **1**(21), 22–26.
182. Ricardo, A.; Carrigan, M.A.; Olcott, A.N.; Benner, S.A. (2004) Borate minerals stabilize ribose. *Science* **303**(5655), 196–196.
183. Mansy, S.S.; Szostak, J.W. (2008) Thermostability of model protocell membranes. *PNAS* **105**(36), 13351–13355.
184. Mansy, S.S. (2010) Membrane transport in primitive cells. *Cold Spring Harb. Perspect. Biol.* **2**(8), a002188.
185. Chakrabarti, A.C.; Breaker, R.R.; Joyce, G.F.; Deamer, D.W. (1994) Production of RNA by a polymerase protein encapsulated within phospholipid vesicles. *J. Mol. Evol.* **39**(6), 555–559.
186. Chen, X.; Li, N.; Ellington, A.D. (2007) Ribozyme catalysis of metabolism in the RNA World. *Chem. Biodivers.* **4**(4), 633–655.
187. King, Z.A.; Lu, J.; Dräger, A.; Miller, P.; Federowicz, S.; Lerman, J.A.; Ebrahim, A.; Palsson, B.O.; Lewis, N.E. (2016) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* **44**(D1), D515–D522.
188. Feist, A.M.; Herrgard, M.J.; Thiele, I.; Reed, J.L.; Palsson, B.O. (2009) Reconstruction of biochemical networks in microbial organisms. *Nat. Rev. Microbiol.* **7**(2), 129–143.
189. Davis, B.D. (1958) On the importance of being ionized. *Arch. Biochem. Biophys.* **78**(2), 497–509.
190. Paula, S.; Volkov, A.G.; Van Hoek, A.N.; Haines, T.H.; Deamer, D.W. (1996) Permeation of protons, potassium ions, and small polar molecules through phospholipid bilayers as a function of membrane thickness. *Biophys. J.* **70**(1), 339–348.
191. Khvorova, A.; Kwak, Y.-G.; Tamkun, M.; Majerfeld, I.; Yarus, M. (1999) RNAs that bind and change the permeability of phospholipid membranes. *PNAS* **96**(19), 10649–10654.
192. Janas, T.; Janas, T.; Yarus, M. (2004) A membrane transporter for tryptophan composed of RNA. *RNA* **10**(10), 1541–1549.
193. Großhans, H.; Filipowicz, W. (2008) The expanding world of small RNAs. *Nature* **451**, 414–416.
194. Silverman, S.K. (2003) Rube Goldberg goes (ribo)nuclear? Molecular switches and sensors made from RNA. *RNA* **9**(4), 377-383.
195. Frommer, J.; Appel, B.; Müller, S. (2015) Ribozymes that can be regulated by external stimuli. *Curr. Opin. Biotechnol.* **31**, 35-41.
196. Kuwabara, T.; Warashina, M.; Taira, K. (2000) Allosterically controllable ribozymes with biosensor functions. *Curr. Opin. Chem. Biol.* **4**(6), 669–677.
197. Navani, N.K.; Li, Y. (2006) Nucleic acid aptamers and enzymes as sensors. *Curr. Opin. Chem. Biol.* **10**(3), 272–281.
198. Breaker, R.R. (2002) Engineered allosteric ribozymes as biosensor components. *Curr. Opin. Biotechnol.* **13**(1), 31–39.
199. Szilágyi, A.; Kun, Á.; Szathmáry, E. (2012) Early evolution of efficient enzymes and genome organization. *Biology Direct* **7**, 38.
200. Szathmáry, E.; Maynard Smith, J. (1993) The evolution of chromosomes II. Molecular mechanisms. *J. Theor. Biol.* **164**(4), 447–454.
201. Scott, W.G. (2007) Ribozymes. *Curr. Opin. Struct. Biol.* **17**(3), 280–286.
202. Musiari, A.; Rowinska-Zyrek, M.; Gallo, S.; Sigel, R.K.O. (2014) Metal ions in ribozymes and riboswitches. In *DNA in Supramolecular Chemistry and Nanotechnology*, Stulz, E.; Clever, G.H., Eds. John Wiley & Sons, Ltd.
203. Landweber, L.F.; Pokrovskaya, I.D. (1999) Emergence of a dual-catalytic RNA with metal-specific cleavage and ligase activities: the spandrels of RNA evolution. *PNAS* **96**(1), 173–178.
204. Drude, I.; Vauléon, S.; Müller, S. (2007) Twin ribozyme mediated removal of nucleotides from an internal RNA site. *Biochem. Biophys. Res. Commun.* **363**(1), 24–29.
205. Welz, R.; Bossmann, K.; Klug, C.; Schmidt, C.; Fritz, H.-J.; Müller, S. (2003) Site-directed alteration of RNA sequence mediated by an engineered twin ribozyme. *Angew. Chem. Int. Ed.* **42**(21), 2424–2427.
206. Vaidya, N.; Manapat, M.L.; Chen, I.A.; Xulvi-Brunet, R.; Hayden, E.J.; Lehman, N. (2012) Spontaneous network formation among cooperative RNA replicators. *Nature* **491**(7422), 72–77.