QRB QUARTERLY REVIEWS OF BIOPHYSICS
journals.cambridge.org/qrb

Nucleic acids: function and potential for abiogenesis

Falk Wachowius, James Attwater and Philipp Holliger*

MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Francis Crick Avenue, Cambridge CB2 0QH, UK

Quarterly Reviews of Biophysics (2017), **50**, e4, page 1 of 37 doi:10.1017/S0033583517000038

Abstract. The emergence of functional cooperation between the three main classes of biomolecules – nucleic acids, peptides and lipids – defines life at the molecular level. However, how such mutually interdependent molecular systems emerged from prebiotic chemistry remains a mystery. A key hypothesis, formulated by Crick, Orgel and Woese over 40 year ago, posits that early life must have been simpler. Specifically, it proposed that an early primordial biology lacked proteins and DNA but instead relied on RNA as the key biopolymer responsible not just for genetic information storage and propagation, but also for catalysis, i.e. metabolism. Indeed, there is compelling evidence for such an 'RNA world', notably in the structure of the ribosome as a likely molecular fossil from that time. Nevertheless, one might justifiably ask whether RNA alone would be up to the task. From a purely chemical perspective, RNA is a molecule of rather uniform composition with all four bases comprising organic heterocycles of similar size and comparable polarity and pK_a values. Thus, RNA molecules cover a much narrower range of steric, electronic and physicochemical properties than, e.g. the 20 amino acid side-chains of proteins. Herein we will examine the functional potential of RNA (and other nucleic acids) with respect to self-replication, catalysis and assembly into simple protocellular entities.

* Author for correspondence: Philipp Holliger, MRC Laboratory of Molecular Biology, Cambridge Biomedical Campus, Francis Crick Avenue, Cambridge CB2 0QH, UK. Tel.: 0044 1223 267092; Fax: 0044 1223 268300; Email: ph1@mrc-lmb.cam.ac.uk

1. Introduction

Life depends on the intricate interplay of myriads of different biomolecules, but the interactions of two classes of biopolymer, nucleic acids and polypeptides (proteins), are of fundamental importance. In current biology, these biopolymers are mutually interdependent: nucleic acids (DNA and RNA) are required for protein synthesis (at all levels) and proteins in turn are required to synthesize both DNA and RNA and replicate the genome. The emergence of such a molecular symbiosis and its genetic fixation in the genome has been the focus of intense enquiry. An attractive, if speculative solution to this 'chicken and egg' problem is the so-called RNA world hypothesis, which proposes a simpler, primordial biology preceding our own, in which RNA played a central role not only as the informational polymer but also as a catalyst in early metabolic pathways (Gesteland *et al.* 2005; Pressman *et al.* 2015).

The central role of RNA in protein translation and RNA splicing, together with a diverse array of different functional RNAs such as ribozymes, riboswitches, tRNA, mRNA, ncRNAs and other regulatory RNAs found to different extents in all domains of life, provide compelling support for a central role of RNA in early biology (Atkins *et al.* 2011). However, one might ask, if RNA really is the only conceivable solution driven by overwhelming functional constraints or if it is rather a reflection of life's chemical history - a 'frozen accident' - imposed by prebiotic chemistry (Sutherland, 2016). To paraphrase Monod, is the chemistry of life's genetic system based on 'chance or necessity'? One potential approach to this key question lies in a thorough exploration of the functional potential of RNA. A large body of work in the last 30 years has begun to map the functional space for RNA (and nucleic acids in general). Repertoire selection experiments (SELEX) (Ellington & Szostak, 1990; Robertson & Joyce, 1990; Tuerk & Gold, 1990) have explored the catalytic and binding potential of RNA and have generated a wide variety of RNA aptamers, sensors and catalysts attesting to an astonishing functional versatility. Similar *in vitro* evolution approaches have also uncovered a comparable functional potential in other genetic polymers such as DNA and xeno-nucleic acid (XNA) polymers not found in nature (Pinheiro *et al.* 2013; Silverman, 2016).

However, a potential weakness of these experiments with regard to nucleic acid function at the origin of life is that they have largely ignored the prebiotic molecular context. The environmental and molecular diversity of the early Earth is likely to have critically impacted on the function and evolution of early genetic polymers whatever their chemistry. Indeed, the emergence of the earliest life-like entities likely involved mutually reinforcing mechanisms of interaction and adaptation of the primordial genetic material with both the molecular environment – including peptides and molecules from simple metabolic networks – as well as their physicochemical environment. The latter might have involved for example interactions with mineral, ice or other surfaces as well as encapsulation into macromolecular compartments or demixing into colloidal or coacervate phases all of which might alter the functional potential of a given genetic polymer. Thus, investigating complex environments and compositional heterogeneity – moving beyond the paradigm of controlled monomer reactions to more realistic dynamic multi-substrate systems – may reveal novel emergent properties through complex interactions that are not evident in homogenous systems. Indeed, such 'systems chemistry' approaches have been critical for recent progress in the unified prebiotic synthesis of the building blocks for RNA, peptides and lipids (Jauker *et al.* 2015; Patel *et al.* 2015; Sutherland, 2016). Consideration of early Earth environments also includes potentially relevant cofactors, e.g. Fe^{2+} (Hsiao *et al.* 2013), phenotypes (ice-evolved polymerase ribozymes; Attwater *et al.*, 2013b) and physicochemical conditions (Budin & Szostak, 2010).

Herein we will describe recent progress in exploring these questions both with the 'classical' homogenous systems as well as novel approaches, including (controlled) degrees of chemical and compositional heterogeneity.

2. Nucleic acids as information-coding entities

The key feature that sets nucleic acids apart from other biopolymers is their remarkable capacity for stable yet accessible information storage and propagation through semi-conservative replication. Furthermore, nucleic acid molecules are not simple strings of information, but they can fold into intricate three-dimensional (3D) shapes to form specific ligands, sensors and catalysts. They unite within the same molecule the genetic information, the genotype (i.e. the sequence of nucleobases) and the phenotype (the function encoded by said sequence) (Fig. 1) and this makes them amenable to direct evolution. Thus, they represent a true molecular incarnation of information, a code that at some point in time acquired the ability to write and copy itself and evolve (Adami & LaBar, 2015). Therefore, the origin of biological information is the foundation for the origin of life.

One might start by considering, which molecular functions and processes might be required for the emergence of such a code, considered by some to resemble a physical phase transition, i.e. an abrupt change in the capacity of a chemical system to store and utilize information (Cronin & Walker, 2016). This notion is also captured in NASA's widely postulated simple definition of life as a 'chemical system capable of self-replication and evolution'. Thus, the search for the molecular embodiments of the

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

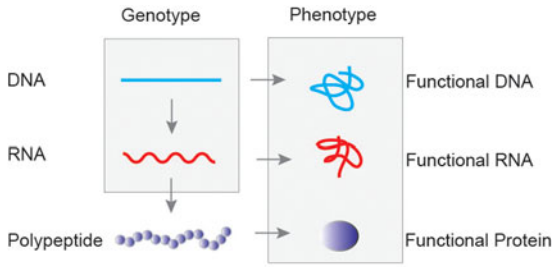

Fig. 1. Genotypes and Phenotypes. Biological information (genotype) is exclusively encoded in nucleic acids (DNA and RNA) and the flow of information is unidirectional as proposed by the central dogma from DNA via RNA to proteins. Both nucleic acids and proteins can express functional phenotypes.

transition from inanimate matter to living systems, from chemistry to early biology, simplifies to the search for chemical components that can encode and propagate information, that are capable of self-replication and ultimately evolution.

Are nucleic acids the only molecular systems capable of information storage and propagation? Various alternatives have been proposed. Cairns–Smith postulated a primary origin of information imprinted in inorganic clay crystals, based on the inherent self-organizing principles of matter, with the later 'take-over' of heritable function by organic macromolecules (Cairns-Smith, 1966). Higher level information storage and capacity for heritable change and evolution has been proposed for networks of autocatalytic metabolic reactions (so-called autocatalytic sets) (Kauffman, 1996) or as a form of compositional memory (Segre & Lancet, 2000). The first concept proposes that networks of self-sustaining chemical reactions can spontaneously self-organize and that their cooperativity and connectivity constitutes a form of distributed memory, i.e. a genotype that can evolve – at least in computer simulations (Vasas *et al.* 2012) – while a compositional memory captures the finding that preferential self-organization in some molecular systems favours a compositional or stereochemical bias, which can to some degree be propagated i.e. inherited. The validity of such concepts outside theoretical considerations has been questioned (Orgel, 2008), but the expanding toolbox of systems chemistry should bring experimental evaluation within reach. Indeed, examples of simple chemical (compositional) genotypes have recently been described (Gutierrez *et al.* 2014). However, information density of such systems is likely to be low and information propagation, mutation and evolution remains to be demonstrated.

Therefore, despite experimental progress in exploring the above concepts, there is, as yet, no compelling alternative to nucleic acids for chemical information storage. If we accept that the emergence of an ability to store, replicate and propagate information as a molecular memory to record and preserve successful phenotypes for future cycles of selection was a key event in the origin of life, then nucleic acids should be considered the prime candidate for such molecular memory for reasons of both functionality and analogy with extant biology.

2.1 Self-replication as a molecular property

Self-replication (at the genetic, cellular and organismal levels) is a defining hallmark of life. However, its beginnings are currently unknown. But self-replication as a system-level property is widespread beyond biology not just in the digital realm, e.g. in the form of computer viruses but in macromolecular and colloidal chemistry. Examples include crystal seeding, as well as colloidal self-organizing systems such as lipidic vesicles, which can display both autocatalytic growth and self-replication (Hanczyc & Szostak, 2004; Oberholzer *et al.* 1995a, b).

Autocatalytic chemical systems capable of self-replication have also been designed based on various components, including small molecules and peptides (Bissette & Fletcher, 2013; Conn *et al.* 1994; Lee *et al.* 1996). However, these systems differ from genetic systems in several crucial aspects. Key differences include the unique ability of nucleic acids (DNA, RNA and XNA) to store information both redundantly (on both strands) and at exceptionally high density (Church *et al.* 2012) using an exclusive double-sided recognition code based on non-covalent interactions by hydrogen bonding. Furthermore, and possibly even more importantly, replication in the autocatalytic chemical systems is by necessity perfect, and a 'mistake', i.e. side-reactions, etc. simply dissipate the self-replication cycle and are non-heritable. In contrast, information transfer in nucleic acid replication – while accurate – is imperfect, enabling both faithful transmission of the genetic information to the next generation, as well as generating low-level sequence diversity (i.e. mutations), which is a prerequisite for evolution.

Some autocatalytic systems have been built from synthetic nucleic acid components. These include systems involving palindromic trinucleotide ligations using carbodiimide (EDC) chemistry (Sievers & von Kiedrowski, 1994) either in solution or on longer (24-mer) duplex palindromic polypurine/polypyrimidine DNA (Li & Nicolaou, 1994). A common problem of such

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

approaches is product inhibition, which can be overcome by surface tethering and thermocycling to liberate the daughter strands from the template (Luther *et al.* 1998).

Joyce and co-worker repurposed the R3 RNA ligase ribozyme for self-ligation (Paul & Joyce, 2002) and faced the same problem but overcame product inhibition through an elegant cross-catalytic system, which allowed self-assembly of the two R3 variants from their constituent parts with true exponential growth kinetics (Lincoln & Joyce, 2009). This system has also been optimized for the sensing of ligands (Lam & Joyce, 2009) as well as for impressive speed (Robertson & Joyce, 2014). Similarly, although with much slower growth kinetics, split variants of the *Azoarcus* self-splicing intron (SSI) can self-assemble both *in cis* and *in trans* into active complexes and can form cross-catalytic assembly networks (Hayden *et al.* 2008; Vaidya *et al.* 2012). However, although both the cross-catalytic ligase and *Azoarcus* SSI can form new variants through recombination and network growth, the need to provide pre-fabricated RNA oligomer-building blocks with substantial homology to the ribozyme/SSI core constrains their ability to evolve freely.

2.2 Physicochemical properties and information storage capacity

A strong case can be made that nucleic acids are singularly suited for information storage and transmission (Benner, 2004). Beyond the specific base-pairing and redundant double-helical information encoding famously recognized by Watson & Crick, a key feature of the chemistry of nucleic acids is that information content and physicochemical properties are effectively decoupled due to the dominant influence of the polyanionic phosphodiester backbone. In contrast to the behaviour of proteins, where single mutations can have dramatic consequences on folding, structure or solubility, most nucleic acid sequences display identical physicochemical properties. Indeed, without this feature much of recombinant DNA technology, microarrays and sequencing would be technically impossible. Other features include the charge repulsion along the backbone favouring an extended conformation facilitating information readout. Finally, there are the unusual chemical properties of phosphodiester bonds combining thermodynamic instability with an unusual kinetic stability as famously pointed out by Westheimer (1987). The kinetic stability of phosphodiesters is in sharp contrast to other esters, including the chemically closely related arsenodiester linkage, which undergoes rapid hydrolysis in aqueous solution due to inefficient charge shielding of the larger arsenic atom (Fekry *et al.* 2011). In addition, the restricted number of sugar ring conformations provide a stable scaffold for the nucleobases and is essential for duplex formation, stability and the restriction of conformational polymorphism to just two main double-helical structures, A- and B-forms, under physiological conditions (Saenger & Egli, 1984).

Despite this seemingly ideal 'Goldilocks' chemistry, it should be noted that recent work has shown that these fundamental principles are stable to considerable variation in both the canonical sugar and nucleobase chemistry, which in turn give rise to a wide range of structural variation (Anosova *et al.* 2016). Building on earlier work from Orgel and Eschenmoser (Eschenmoser, 1999; Kozlov *et al.* 1999a, b; Schoning *et al.* 2000) nucleic acids in which the canonical (deoxi)ribo-furanose of DNA and RNA is replaced by ring congeners not found in nature, including HNA (1,5 anhydrohexitol nucleic acid), CeNA (cyclohexenyl nucleic acids), LNA (2′ O, 4′-C-methylene-β-D-ribonucleic acids; locked nucleic acids), ANA (arabinonucleic acids), FANA (2′-fluoro-arabinonucleic acid) and TNA (α-L-threofuranosylnucleic acids, based on a tetrose sugar) are capable of genetic information storage and propagation (Pinheiro *et al.* 2012). Furthermore, these XNAs support a replication cycle progressing through a DNA intermediate (conceptually similar to retroviral replication) enabling the *in vitro* evolution of XNA aptamers (Pinheiro *et al.* 2012) and catalysts (Taylor *et al.* 2015). So far, no prebiotic synthesis of XNAs has been described, though this argument in itself is insufficient to argue against their inherent plausibility (as prebiotic syntheses of XNAs have not been actively sought).

Similarly, there might also exist alternative patterns of information encoding. Indeed genetic information storage and transfer have been demonstrated for a range of artificial base-pair designs. These expand the genetic alphabet and can be based on alternative hydrogen-bonding patterns, hydrophobic and/or geometric compatibility or even metal ion chelation. Some of these expanded genetic alphabets have also enabled evolution of superior aptamer ligands to protein or cell-surface targets incorporating one or more bases or base-pairs (Benner, 2004; Hirao *et al.* 2012) and have even been integrated into a plasmid in a living organism (Malyshev *et al.* 2014). Importantly, both unnatural base-pairs as well as a number of XNA backbones retain their molecular memory function despite deviations from canonical helical conformations (Georgiadis *et al.* 2015; Lescrinier *et al.* 2000; Nauwelaerts *et al.* 2007) and planar base-stacking (Betz *et al.* 2013).

In contrast to the comparable tolerance to different sugar/nucleobase chemistries, the design of alternatives to the canonical phosphodiester backbone chemistry that can also support genetic information storage and propagation and allow cross-talk (i.e. helix-formation with natural nucleic acids) has proven challenging (Micklefield, 2001; Nielsen, 1995). The only successful designs fulfilling all of the above criteria are isosteric and largely isoelectronic modifications such as phosphorothioates (Eckstein, 2014) and boranophosphates (Li *et al.* 2007) (in which the non-bridging oxygen is replaced by sulphur or

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

borano-trihydride substituents, respectively). More radical departures from the canonical backbone chemistry such as peptide nucleic acids (PNAs) (Sharma & Awasthi, 2016), in which the ribofuranose-phosphate backbone of DNA/RNA is replaced by N-(2-aminoethyl)-glycine or morpholino nucleic acids (PMO), in which the sugar–phosphate linkage is substituted by a morpholino ring–phosphorodiamidate linkage are among the few exceptions. Both PNAs and PMOs show specific hybridization to target sequences, but currently cannot be replicated enzymatically and hence are not amenable to laboratory evolution. Nevertheless, using reductive amination chemistry (Li *et al.* 2002) PNA can be used in information transfer from a DNA template (Brudno *et al.* 2010; Rosenbaum & Liu, 2003) and indeed it has been proposed that PNA may have been involved in pre-biotic evolution (Nielsen, 2007; Ura *et al.* 2009).

3. The catalytic potential of nucleic acids

DNA and RNA (and XNAs) are not just repositories of genetic information, but can fold up into intricate 3D structures with specific ligand-binding activities [aptamers (Famulok & Mayer, 2014; Pfeiffer & Mayer, 2016; Sullenger & Nair, 2016)], allosteric conformational properties [riboswitches (Breaker, 2012; Peselis & Serganov, 2014; Serganov & Nudler, 2013)] and catalysts (ribozymes and deoxyribozymes) (see below). The specific and programmable hybridization properties of nucleic acids can also be exploited in the construction of intricate nano-objects and devices built from DNA (Chen *et al.* 2015; Zhang *et al.* 2014), RNA (Grabow & Jaeger, 2014; Guo, 2010) or XNA (Taylor *et al.* 2016).

In the context of an early origin of life scenario, catalysis would arguably be the most distinctive ability of nucleic acids. As storage and propagation of information is an essential property of a molecule at the dawn of life (see above), catalysis would be the key emergent property, resulting in a dual functional molecular trait. Accordingly, the relative catalytic potentials of RNA, DNA and XNAs merit some discussion.

Nucleic acids with only four different functional groups appear seemingly inferior to proteins with 20 different amino acids bearing diverse chemical functionalities with a wide range of properties, shapes and pK_a values. For example, histidine with its $pK_a \sim 6$ is well suited for acid–base catalysis and proton transfer at neutral pH. In contrast, nucleotide bases present pK_a values >9·1 and <4·3 (for nucleotides free in solution) with pK_a's closest to neutrality for the N1 nitrogen of the purine bases and the N3 nitrogen of the pyrimidine bases, and no functional groups of nucleic acids are positively charged at neutral pH (Blackburn *et al.* 2006; Ferre-D'Amare & Scott, 2010). Nevertheless nucleobase pK_a values, as amino acid pK_a values, can be modulated when protected from bulk solvent (Harris & Turner, 2002; Wilcox & Bevilacqua, 2013). Furthermore, uniquely in RNA a proximally positioned intramolecular nucleophile – the vicinal 2′ OH – allows for rapid strand cleavage and recombination/exchange (transesterification) reactions via a 2,3′ cyclic phosphate intermediate, which may have been important in early RNA oligomer pools.

3.1 RNA catalysis

The first examples of RNA catalysis were discovered by Cech and Altman, in the SSI of *Tetrahymena* (Kruger *et al.* 1982) and the RNA component of RNAse P (Guerrier-Takada *et al.* 1983) and were followed by the discovery of a wide range of self-cleaving ribozymes in viruses as well as an ever-expanding number of RNA catalysts generated by *in vitro* selection technologies. Finally and most fundamentally, RNA catalysis was found to be at the heart of both the spliceosome and the peptidyl-transferase activity of the ribosome. The landmark discovery of RNA catalysis also set the starting point for the exploration of the essential regulatory function of RNA *in vivo* (Cech & Steitz, 2014). Ribozyme catalysis is based on distinct 3D structures, with stacking, base-pairing and tertiary contacts all contributing to the complex folding of the ribozyme/substrate complex. Ribozyme and more generally RNA folding and dynamics occur in hierarchical order with structural elements forming on timescales ranging from picoseconds to seconds (Mustoe *et al.* 2014). The folding is generally facilitated by metal ions, due to the highly polyanionic character of the sugar phosphate backbone (Denesyuk & Thirumalai, 2015). Nevertheless RNA folding *in vitro* (as it has mostly been studied) is often different from the much more crowded natural *in vivo* conditions (Leamy *et al.* 2016).

RNA catalysis *in vivo* can be either solely performed by RNA, as for the small nucleolytic ribozymes, the Hammerhead (HHR), Hairpin (HP), Varkud satellite (VS), Hepatitis delta (HDV), twister and the *glmS* ribozyme (Lilley, 2011; Wilson *et al.* 2016b) or aided by proteins forming ribonucleoprotein (RNP) complexes, as for the group II intron (Pyle, 2016), RNaseP (Mondragon, 2013), the ribosome (Voorhees & Ramakrishnan, 2013) and the spliceosome (Wahl *et al.* 2009), with the RNA component responsible for catalysis and the protein component mainly acting as a scaffold and/or counterion. The principal mechanisms of naturally occurring ribozymes are either based on general acid–base catalysis as for the small nucleolytic ribozymes or on two metal ion catalysis as for group I, group II introns, RNase P and the spliceosome.

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

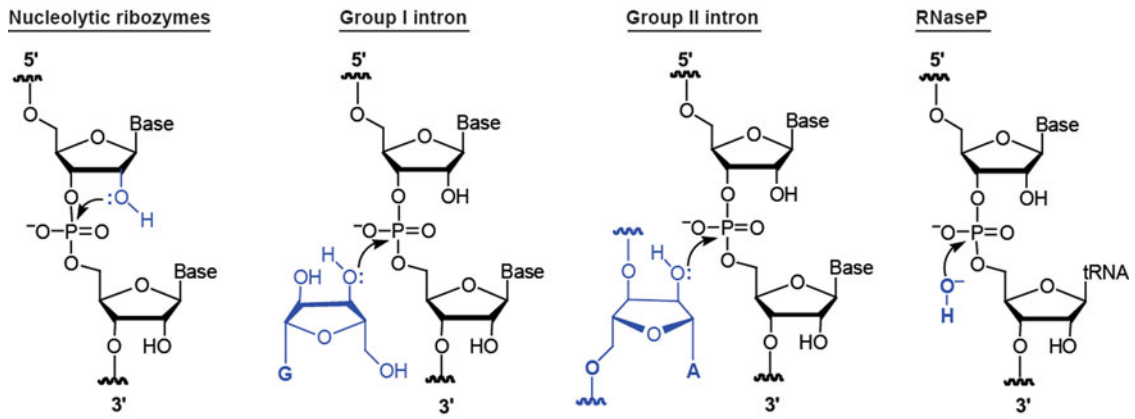

Fig. 2. First step of phosphoryl transfer reactions of natural occurring ribozymes. The nucleophile (in blue) attacks the phosphorus of the RNA phosphodiester bond.

All natural occurring ribozymes, with the notable exception of the ribosome (which performs peptidyl transfer), catalyse phosphoryl transfer reactions. This is initiated by nucleophilic attack on the phosphate by the adjacent 2′-oxygen (as for the nucleolytic ribozymes), the 3′-oxygen of an exogenous guanosine (group I intron), the 2′ oxygen of an internal adenosine (group II intron and the spliceosome) or water (RNase P) (Lilley & Eckstein, 2008) (Fig. 2).

This rather limited chemical reactivity spectrum raises the question of whether the many diverse chemical transformations necessary to support a putative RNA world could have been performed by RNA alone. It may be that there are more RNA-world molecular fossils (with more diverse chemical capabilities) waiting to be discovered, in particular considering that still only a small section of the 'RNAome' of the biosphere has been explored.

There is a strong discrepancy between the occurrence and significance of different ribozymes in the tree of life. The essential reactions catalysed by the more complex RNP structures such as the peptidyl-transferase activity of the ribosome, the RNase P catalysed tRNA maturation and RNA splicing by the spliceosome (or its simpler forerunnner the group II intron) are distinctive and found across all branches of life. On the other hand, the simpler nucleolytic ribozymes are rather sparsely distributed in biology (with the VS ribozyme only found once) and with a narrow biological function only fully explored in viruses. Nevertheless, biochemical experiments and bioinformatic search algorithms identified HHR, HDV and HP sequences in all domains of life, with their precise functions in most cases still to be explored (Jimenez *et al.* 2015; Salehi-Ashtiani *et al.* 2006; Webb *et al.* 2009). This ubiquitous presence of the small nucleolytic ribozymes suggests that either they too might be leftovers from an ancient RNA world (as well as actively participating in modern nucleic acid metabolism, and hence being part of the 'modern RNA World') (Cech, 2012) or alternatively, that this distribution might be simply a consequence of their comparative structural and functional simplicity. Indeed, the HHR fold, which is particularly ubiquitous (Hammann *et al.* 2012), is also the most likely motif for RNA cleavage identified by *in vitro* selections (Salehi-Ashtiani & Szostak, 2001), presumably due to its small size and relaxed sequence requirements, i.e the 'tyranny of the small motif'. On the other hand, evolutionary pressure has clearly also led to different outcomes for the same reaction and seemingly to alternative structural and catalytic solutions such as the HDV, Twister, etc. ribozymes (see below). In general, the nucleolytic ribozymes reveal a high sequence specificity and catalytic efficiency with their essential information content encoding catalytic function lower than that suggested by the length of the ribozyme. RNA sequences capable of catalysis, in particular RNA cleavage, are therefore rather common in sequence space. Hence, even a rather modest repertoire of random RNAs should already contain a number of active folds indicating how they could have contributed to the emergence of RNA catalysis from the pools of short RNA oligomers provided by prebiotic chemistry

The direct involvement of divalent metal ions in RNA catalysis (inner sphere coordination) by the small nucleolytic ribozymes has been largely excluded (Murray *et al.* 1998), but outer sphere coordinated divalent metal ions are likely involved in HDV catalysis (Ke *et al.* 2004), and might also play a direct role in HHR catalysis (Mir & Golden, 2016). Apart from their involvement in catalysis, metal ions fulfill a prominent role in the folding process and stabilization of the 3D structure of ribozymes (Lipfert *et al.* 2014; Sigel *et al.* 2012). From an origins perspective, metal ions were abundantly present on the early earth, making them the most likely early interacting partner for RNA, with divalent cations (such as Mg^{2+}) more efficiently decreasing the electrostatic repulsion upon folding of the RNA molecule compared with monovalent cations (such as Na^{+} and K^{+}). However, there is a fundamental functional trade-off between the essential functions of divalent metal ions in ribozyme

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

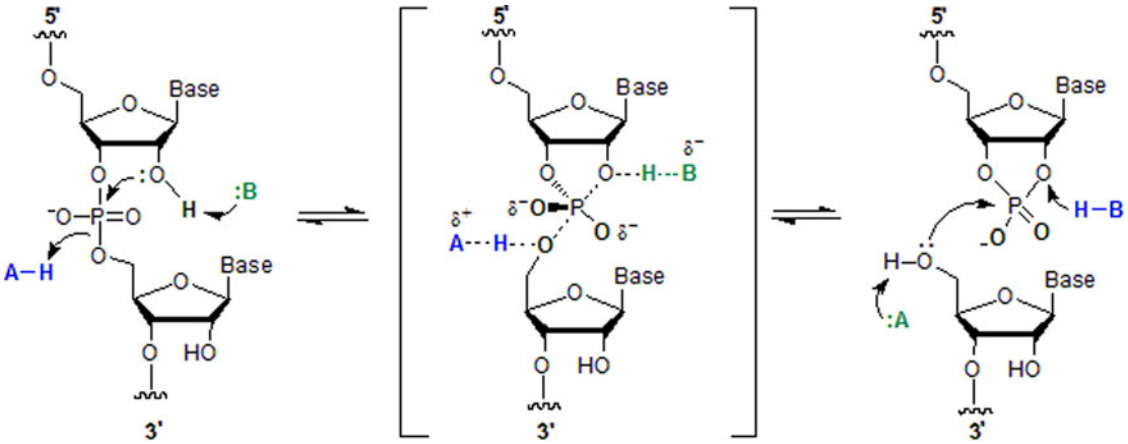

Fig. 3. Mechanism of general acid–base catalysis as performed by the small nucleolytic ribozymes. The general base (in green) is attracting a proton from the 2′-hydroxyl in the cleavage reaction or from the 5′-hydroxyl in the reversed ligation reaction. The general acid (in blue) is protonating the 5′-oxyanion leaving group for cleavage or the 2′oxyanion for ligation. The proposed trigonal bipyramidal phosphorane transition state is shown in the centre.

folding and catalysis, and the increased degradation rate of RNA in their presence. This trade-off has to be considered as a major evolutionary driving force both towards the assembly of folded RNA structures – as double-stranded RNA (dsRNA) is much more robust against degradation compared with single-stranded RNA (ssRNA) – and towards the replacement of structural metal ions by peptidic or proteinaceous counterions (see Section 6).

High-resolution structures of examples of all the natural classes of ribozymes are now available, including at least 20 different structures for the HHR and HP ribozymes. Starting with the first crystal structure of an HHR variant (Scott *et al.* 1995), the crystal structures of the HDV (Ferre-D'Amare *et al.* 1998) the HP (Rupert & Ferre-D'Amare, 2001), the *glmS* (Klein & Ferre-D'Amare, 2006) and finally also the VS ribozyme (Suslov *et al.* 2015) were solved over the following 20 years. Similarly, high-resolution structures of the more complex RNA structures and RNP complexes were obtained for the group I intron (Adams *et al.* 2004), the group II intron (Toor *et al.* 2008), RNase P (Kazantsev *et al.* 2005), the ribosome (Ban *et al.* 2000) and very recently also the spliceosome (Yan *et al.* 2015). Recent technical breakthroughs in CryoEM (cryo-electron-microscopy) techniques (Nogales & Scheres, 2015; Vinothkumar & Henderson, 2016) revolutionized structural biology of large RNP complexes such as the ribosome (Frank, 2016) and the spliceosome (Nguyen *et al.* 2016) resulting in unprecedented and detailed pictures of RNA catalysis by these complex molecular machines. While RNA catalysis at the heart of the ribosome had been suspected some time ago (Noller *et al.* 1992) to be confirmed by the structure of the peptidyl-transferase site (Nissen *et al.* 2000), the conjectured ribozyme catalysis of the spliceosome could only recently be ascertained by a combination of biochemical and structural studies (Fica *et al.* 2013; Nguyen *et al.* 2015; Wan *et al.* 2016), identifying the U2–U6 snRNA as the catalytic complex and showing, likely ancestral similarities to group II intron two metal ion catalysis.

Mechanistically, RNA undergoes non-enzymatic degradation by an internal transesterification reaction, through nucleophilic attack of the 2′-oxygen on the adjacent 3′-phosphodiester forming a 2′,3′-cyclic phosphate and 5′-hydroxyl. The reaction is catalysed by the deprotonation of the 2′-hydroxyl and is therefore increased at higher pH values. This transesterification proceeds through a concerted S_N2 mechanism, with the 2′-oxygen, the 5′-oxygen and the phosphorus in an in-line geometry. However, the main contribution to cleavage rates is believed to arise from deprotonation events (by a factor of 10^5–10^6) with the optimal orientation, i.e. in-line geometry less important and contributing only a factor of around 10^2 to the observed rate enhancement (Emilsson *et al.* 2003; Lilley, 2005) (measured for ribozyme catalysed cleavage reactions but likely similar for the non-enzymatic reaction). The non-enzymatic degradation of RNA phosphodiesters is about 10^4-fold faster than that of DNA at neutral pH and even more accelerated at basic pH (though slower at acidic pH). This stability divergence is likely one of the functional drivers for the switch from RNA to DNA for information storage in living systems as genomes became larger.

The 'classical' (HHR, HP, VS, HDV, *glmS*) small nucleolytic ribozymes all catalyse phosphodiester cleavage of RNA by general acid–base catalysis along the mechanistic trajectory described above (Fig. 3). The active structures of the HHR, HP and VS ribozyme are formed by multihelix junctions and all three bind their substrate RNA by Watson–Crick base-pairing on both sides of the cleavage site, therefore the reverse ligation reactions are possible according to the principle of microscopic

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

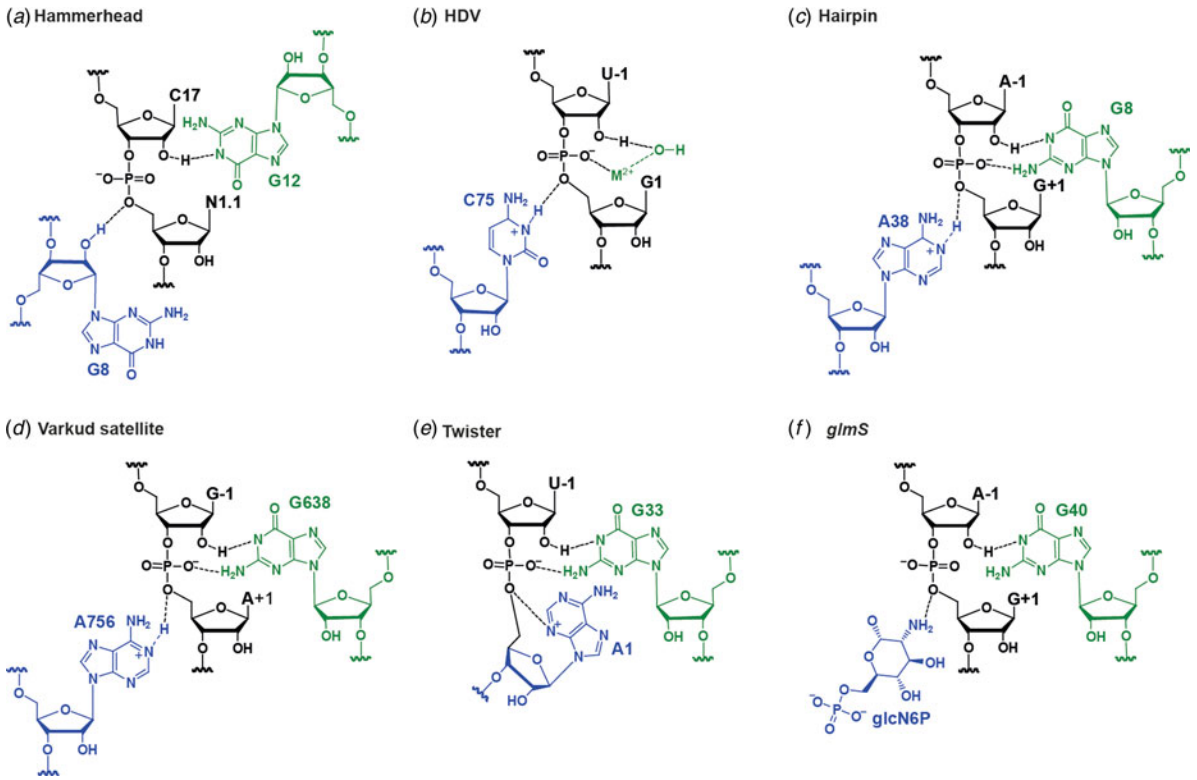

Fig. 4. Proposed cleavage mechanism of the small nucleolytic ribozymes, based on general acid–base catalysis. The general acid is displayed in blue and the general base in green. The general acid is in all cases G, with the exception of the Varkud satellite, where a hydrated metal ion acts as general base.

reversibility. The HP ribozyme applies the N1 of G8 and N1 of A39, as general base and acid, respectively (reversed in the ligation reaction) (Kath-Schorr *et al.* 2012). Similary, the VS ribozyme uses the N1 of G638 as general base and the N1 of A756 as general acid (Suslov *et al.* 2015). In HHR catalysis the N1 of G12 attracts the proton from the 2′-oxygen nucleophile, acting as general base and the 2′-hydroxyl of G8 is positioned near the 5′-oxygen leaving group, fulfilling the role of the general acid (Martick & Scott, 2006) (Fig. 4). In contrast, the HDV and *glmS* ribozymes, whose active structure is formed by pseudoknots, only basepair with their substrates 3′ to the cleavage site; therefore the intermolecular reverse ligation reaction is (akin to RNase A) excluded under standard reaction conditions. In the HDV ribozyme, the pK_a shifted N3 imine proton of the catalytic C75 acts as a general acid and a hydrated Mg^{2+} ion as general base (Das & Piccirilli, 2005; Nakano *et al.* 2000), with mainly C75 contributing to the observed rate enhancement. The *glmS* catalytic riboswitch has an absolute requirement for G40, with the N1 of G40 acting as general base and with the amino group of the glucosamine-6-phosphate substrate in close proximity to the 5′-oxygen leaving group, consistent with its function as general acid (Jansen *et al.* 2006; Klein *et al.* 2007) (Fig. 4).

The small nucleolytic ribozymes are the favourite study objects for RNA catalysis, related to their small size and the fact that they provide different structural and mechanistic solutions. They may also embody independent evolutionary trajectories towards the same chemical problem, therefore representing an example of convergent evolution at the molecular level. In principle, RNA cleavage by the different nucleolytic ribozymes could have been based on the same active site nucleotides arranged on different structural scaffolds. However, detailed biochemical, structural and biophysical methods have elucidated not only different structural arrangements, but also unique constellations of functional groups, pH and metal ions inside the framework of general acid–base catalysis within this group of ribozymes. Furthermore, even different constructs of the same ribozyme can have different structural folds and catalytic rates, as was shown for the HHR, in which the full-length variant (Martick & Scott, 2006) was found to adopt a different structural arrangement compared with a previously crystallized minimal variant (Scott *et al.* 1995). This shows that seemingly irrelevant residues distal to the catalytic core can lead to major structural changes, impact catalytic turnover and influence metal ion requirements and overall stability through non-Watson–Crick long-range tertiary interactions. An interesting recent finding in this context was the identification of a minimal HHR variant with a strong increase in catalytic activity, based solely on the interaction of a single AU Hoogsteen

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

base pair, formed by an A residing in the loop region of stem 2 of the HHR and an unpaired U from the 3′-end of the substrate RNA (O'Rourke *et al.* 2015).

Recent additions to the above-mentioned nucleolytic ribozymes are the Twister ribozyme (Roth *et al.* 2014) (Fig. 4) and related variants (Twister sister, Pistol and Hatchet) (Harris *et al.* 2015; Li *et al.* 2015; Weinberg *et al.* 2015) that were identified by sequence- and structure-based bioinformatics algorithms. The Twister motif was identified in all domains of life, but its exact biological functions remain to be explored. The Twister ribozyme forms a double pseudoknot structure with its catalytic mechanism recently elucidated by a combination of structural (Eiler *et al.* 2014; Liu *et al.* 2014; Ren *et al.* 2014), biochemical (Wilson *et al.* 2016a) and modelling (Gaines & York, 2016) studies, using A and G as general acid and base, respectively (similar to the HP and VS ribozymes). The crystal structures were obtained from different Twister variants, (*O. sativa*) (Huang *et al.* 2014), an environmental variant (env) (Eiler *et al.* 2014) and a minimized variant thereof (env22) (Ren *et al.* 2014), showing the same overall ribozyme fold but with a partially different arrangement at the catalytic site.

As a significant difference to the HP and VS, which are using the N1 of A, the Twister applies the more acidic proton of the N3 of the conserved catalytic A (A1, adjacent to the cleavage site) for protonation of the 5′-oxygen (Fig. 4). This can only be achieved by a specific electrostatic environment causing a strong rise in pK_a towards neutrality (Kosutic *et al.* 2015). Similarly, a perturbed pK_a of A in the catalytic centre of the lead-dependent ribozyme was previously identified by NMR (Legault & Pardi, 1997). This not only adds a new mechanism to the repertoire of natural RNA catalysis, but also demonstrates how ribozymes can transcend their limited chemical functionalities, by forming micro-environments resulting in dramatically altered pK_a's of specified functional groups and thereby exploring a much broader array of catalytic strategies. Nevertheless, even though these new ribozyme variants comprise a divergent structural scaffold and a new catalytic mechanism, they all represent variations on the theme of RNA transesterification chemistry.

The advent of deep sequencing technology has not only revolutionized genomics (Koboldt *et al.* 2013), but also provided a much more detailed picture of the fitness landscape of functional RNAs such as RNA aptamers (Jimenez *et al.* 2013) and short ribozymes (Ameta *et al.* 2014; Petrie & Joyce, 2014; Pitt & Ferre-D'Amare, 2010). A recently introduced mutation analysis method for ribozymes also relies on an in-depth deep sequencing analysis (Kobori *et al.* 2015). For this approach, the starting sequence compromises 97% of the wild-type bases, doped with 1% of each of the remaining nucleobases, and after the ribozyme catalysed reaction the active and inactive variants are separated and analysed by deep sequencing. Such detailed mutational analyses presents an ideal complement to the previously developed combinatorial NAIM (nucleotide analogue interference mapping) approaches that introduced base or sugar-modified nucleotides, to probe essential nucleoside functional groups in ribozymes and other functional RNAs (Cochrane & Strobel, 2004; Jansen *et al.* 2006).

Deep sequencing analysis of a Twister ribozyme variant delivered a mutational landscape, by probing all single and double mutants, and provided a quantitative insight into the structure–function relationship of this ribozyme (Kobori & Yokobayashi, 2016). An interesting outcome of this mutational study was the discovery of its robustness to mutation, with mutations outside the catalytic cleft widely tolerated. These findings are entirely consistent with previous results for other small nucleolytic ribozymes (Kun *et al.* 2005), where again mutations in the stem regions were widely tolerated, as long as the helix context and hence the overall fold of the ribozyme were not strongly perturbed, demonstrating the relaxed sequence requirements (and low error threshold for replication) of the small ribozymes.

In the context of the origin of life, both simplicity of sequence requirements and robustness to mutations emerge as clear advantages for RNA. Indeed, the seemingly disadvantageous compositional simplicity of nucleic acids compared with proteins (with only four structurally and chemical similar nucleobase building blocks compared with 20 structurally and chemically diverse amino acid side-chains) might in fact be critical for early evolution, enabling both high mutational tolerance as well as rapid adaptive trajectories across a lower complexity sequence space facilitating evolution.

3.2 In vitro *selected ribozymes*

Why is RNA cleavage by transesterification the only reaction catalysed by natural small ribozymes? A putative RNA world would have required a more diverse range of reactions, but given the narrow range of chemical transformations performed by today's natural ribozymes, it was not obvious that ribozymes would be able to support a putative RNA world metabolism. Following the advent of *in vitro* selection technologies, the principal capability of RNA catalysing diverse chemical reactions likely necessary in an RNA world could be explored (Chen *et al.* 2007; Martin *et al.* 2015; Muller, 2015).

Apart from RNA cleavage and ligation, one likely fundamental reaction in an RNA world (as in organic chemistry) would have been the formation of carbon–carbon (C–C) bonds. Accordingly, inspired by current organic chemistry, ribozymes catalysing C–C bond formation by either Diels-Alder cyclo-addition (Seelig & Jaschke, 1999; Tarasow *et al.* 1997), Michael

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

addition (Sengle *et al.* 2001) or aldol condensation (Fusz *et al.* 2005) were identified. Other reactions catalysed by *in vitro* selected ribozymes and likely necessary at the onset of the RNA world include pyrimidine nucleotide synthesis (Unrau & Bartel, 1998), polynucleotide phosphorylation (kinase activity) (Lorsch & Szostak, 1994) and carbon–nitrogen bond formation (N-alkylation) (Wilson & Szostak, 1995) (for a more complete overview see Chen *et al.* 2007; Silverman, 2008; Wilson & Szostak, 1999).

The transition from an RNA world to the more protein-based biology of today would have required RNA-catalysed amide bond (Wiegand *et al.* 1997) or more specifically peptide bond (Zhang & Cech, 1997) formation and at a later stage the coordinated execution of all the processes comprising today's translation cycle. While modern day proteinaceous aminoacyl-tRNA synthetases (aaRS) combine activation and amino acid transfer, *in vitro* selected ribozymes are capable of catalysing amino acid activation in two separate steps. Amino acids can be activated as aminoacyl-guanylates (Kumar & Yarus, 2001) chemically similar to natural activation as aminoacyl-adenylates, and the transfer of the activated amino acid to the 2′ or 3′ hydroxyl terminus of an acceptor RNA (aminoacylation) can be rapidly catalysed by *in vitro* selected ribozymes (Illangasekare *et al.* 1995; Lee *et al.* 2000), even reduced to the smallest ribozyme ever described (Turk *et al.* 2010) comprising only five nucleotides (nt) reacting with a tetranucleotide substrate (Turk *et al.* 2011). Ribozymes were also selected catalysing the transfer of an amino acid (Met) on their own 5′-hydroxyl or -amino terminus forming either ester or amide bonds using 3′-acylated RNA as amino acid donor (Lohse & Szostak, 1996), similar to catalysis in the P site of the ribosome. Finally, a range of ribozymes was developed (Flexizymes) (Morimoto *et al.* 2011) that are able to couple activated amino-acids to given tRNAs *in vitro* with applications in e.g. peptide selections by DNA display (Roberts & Szostak, 1997). What is, however, lacking, so far, are ribozymes able to charge RNAs with specific amino acids, or otherwise link the identity of the amino acid to a coding triplet (or other) sequence unit to manifest a genetic code. Demonstrating control in implementation of catalytic phenotypes is as important as the catalytic phenotypes themselves when understanding RNA's capacity to form a functional translation system.

In the present-day biochemistry, nucleosides are activated as high-energy triphosphates (NTPs) to be used as substrates for nucleic acid synthesis and replication. Therefore, the *in vitro* selected RNA polymerase ribozyme (RPR) (see below), a molecular analogue of a postulated RNA replicase, was selected using nucleoside triphosphates as substrates (Ekland & Bartel, 1996). Nucleoside triphosphates have some key advantages over more highly activated nucleotides such as phosphorimidazolides. While the latter are highly reactive, they also hydrolyse readily in aqueous solution and therefore need to be continuously replenished. Nucleotide triphosphates (NTPs) on the other hand, while thermodynamically unstable, show a remarkable kinetic stability at neutral pH and therefore, once synthesized would accumulate. However, currently no prebiotic synthesis of NTPs has been described. This has motivated the search for a triphosphorylating ribozyme, which was recently discovered, using the prebiotically plausible trimetaphosphate as phosphate source (Dolan *et al.* 2015; Moretti & Muller, 2014). The identified TPR1/TPR1e ribozyme catalyses the formation of triphosphorylated RNA from trimetaphosphate, and a 5′-hydroxyl RNA oligonucleotide with a catalytic rate of 6·8 min^{-1} under optimal conditions. Originally 96 nt long, a recently derived fragmented variant can be constructed from oligonucleotides no longer than 34 nt (Akoopie & Muller, 2016), approaching the range of RNA oligomers accessible by non-enzymatic RNA polymerization (Ferris *et al.* 1996). However, none of the current variants is capable of directly triphosphorylating nucleoside monomers and relies on attachment as part of a polynucleotide for 5′ positioning; general nucleoside substrate binding may be a challenging trait to evolve due to the tendency for RNA molecules to harness base-pairing for molecular recognition.

Even though proteinogenic amino acids exhibit a broader chemical diversity, more than half of modern day protein enzymes use cofactors with a large variety of functional groups, often based around a nucleoside 'handle', in particular adenosine (Chen *et al.* 2007), potentially representing remnants from RNA world metabolism (White, 1976). As nucleic acids exhibit high affinity and specificity for binding metal cations and small ligands, there is, in principle, no obstacle to ribozymes recruiting cofactors to broaden their chemical functionality and catalytic potential. Nevertheless, except for the *glmS* ribozyme, none of the natural ribozymes performs cofactor-assisted catalysis (e.g. by applying one of the typical protein cofactors such as coenzyme A (CoA), nicotinamide adenine dinucleotide (NAD) or flavin adenine dinucleotide (FAD)). However, *in vitro* evolution experiments have established that there are no functional obstacles to ribozymes utilizing cofactors. Examples include, e.g. an alcohol dehydrogenase ribozyme using NAD^+ (Tsukiji *et al.* 2003) or a ribozyme that decarboxylates a pyruvate-like substrate using thiamin as cofactor (Cernak & Sen, 2013). *In vitro* selected ribozymes are also capable of catalysing the synthesis of the common cofactors CoA, NAD and FAD from their precursors 4-phosphopantetheine, nicotinamide mononucleotide (NMN) and flavin mononucleotide (FMN) respectively (Huang *et al.* 2000).

The RNA 4-base 'code' is both informationally and chemically much simpler than the 20 amino acid protein composition. Nevertheless, one may ask if an even simpler ternary or even a binary code could support RNA catalysis. Joyce and coworkers explored this question using ribozyme catalysed RNA ligation as a model system. To perform selection experiments in the

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

absence of C (comprising sequences with only A, G and U), all C residues in the original random RNA library were deaminated to U by sodium bisulphite treatment (Rogers & Joyce, 1999). From this ternary code RNA library, functional RNA ligases could be isolated, but reselection with the inclusion of C resulted in an increase of the catalytic rate by a factor of 20 (Rogers & Joyce, 2001). Ribozyme selections with only two nucleotides [2,6-diaminopurine (replacing the natural adenine for higher base-pairing stability) and uridine] led to a functional ligase variant, however showing only low catalytic rates and yields (8% ligation yield in 80 h, $k_{obs} = 0{\cdot}05\ h^{-1}$) (Reader & Joyce, 2002). Thus, it seems (at least judging from these three examples) that although catalysts can be isolated from simple binary repertoires, catalytic power seems to scale with informational complexity. Nevertheless, in an early environment without competition by efficient ribozymes or protein enzymes even a small rate enhancement over the uncatalysed reaction might have resulted in a substantial selective advantage.

In vitro selected ribozymes not only show a broad spectrum of different reaction parameters depending on the chemical transformation they catalyse, but also on the applied selection conditions, including strong variations in catalytic rates and yields, catalysis in a *cis*- and/or *trans* format and the ability for multi-turnover catalysis. Reaction conditions are also often prebiotically implausible including high concentrations of reactants and/or high Mg^{2+} concentrations detrimental to the half-life of ribozymes. However, the selected ribozymes represent at best a fraction of the potential prebiotic sequence and phenotype space, and therefore should simply be considered as proof-of-principle for the potential of ribozyme-catalysed reactions. Nevertheless, lack of efficient reaction rates and yields with ideally multi-turnover catalysis remain one of the main shortcomings of many *in vitro* selected ribozymes.

Although substrate selectivity is an essential requirement for catalysts, a degree of substrate promiscuity would provide a mechanism to evolve new ribozyme functions rapidly as has been observed for protein enzymes (Khersonsky & Tawfik, 2010). A related question is whether ribozymes have to adopt different structural folds to catalyse different chemical transformations. Bartel and co-workers (Schultes & Bartel, 2000) explored this question using a RNA sequence derived from the HDV self-cleaving ribozyme and the class III self-ligating ribozyme (catalysing 2′–5′ linked bond formation from 5′-triphosphorylated and 2′,3′-diol substrate RNAs) (Ekland & Bartel, 1996) by a number of iterative mutational steps reaching a 'hybrid sequence', which is 42 and 44 mutational steps away from the parent ligase or HDV sequence, respectively. This hybrid sequence was able to fold into two distinct folds, catalysing either RNA cleavage or ligation, but with reduced catalytic rates compared with the original variants that fold into only one catalytic active fold. On the other hand, the conversion of a self-aminoacylating ribozyme, that aminoacylates its 3′ terminus using adenylated phenylalanine (Illangasekare *et al.* 1995) into a self-kinase ribozyme that phosphorylates its own 5′-end using GTPγS (Lorsch & Szostak, 1994) by *in-vitro* evolution required on average only 14 mutations, with an increased likelihood to find catalytic activity for the new substrate the more distant the RNA moved from the original fold, indicating the necessity to escape the parent fold (Curtis & Bartel, 2005).

The application of deep sequencing technology has allowed a more in-depth analysis of the adaptive fitness landscapes of functional RNAs and therefore also the distribution of a specific catalytic function in RNA sequence space (Pitt & Ferre-D'Amare, 2010). A recent *in vitro* selection experiment starting from two different ligase ribozymes, the class I ligase (Ekland & Bartel, 1996) and the DSL ligase (Ikawa *et al.* 2004), both catalysing 3′–5′ bond formation between 5′-triphosphorylated RNA and 2′,3′-hydroxyl RNA substrates, resulted in variants clustered around each parent sequence, indicating a RNA fitness landscape with isolated fitness peaks (Petrie & Joyce, 2014). At least for these ribozymes this study deemphasizes the function of neutral drift as primary source of genetic change, but rather as a provider of a reservoir of sequences on which selective adaptation can be based.

While high-resolution structures for all currently known natural ribozymes are available (see above) only few crystal structures of *in-vitro* selected ribozymes, such as the leadzyme (Wedekind & McKay, 1999) and the Diels-Alder ribozyme (Serganov *et al.* 2005) have been determined. The latter adopts a fold that forms a binding pocket for enantioselective catalysis with a combination of different factors such as shape complementarity, electronic effects, stacking interactions (in particular to the anthracene substrate) and hydrogen bonding (mainly to the maleimide substrate) all contributing to the catalysed C–C bond formation.

To expand the chemical functionality beyond the four standard ribonucleotides, modified nucleotides, in particular with modifications to the C5 position of uracil, have been introduced. Substituents attached to the C5 position project into the major groove and cause minimal steric clashes with the polymerase and are therefore well tolerated by most DNA/RNA polymerases and reverse transcriptases. Furthermore, there is a reasonably facile chemical synthesis of C5-modified U-triphosphates. The selection of a Diels-Alder ribozyme (Tarasow *et al.* 1997) was one of the first ribozyme selections including a base-modified nucleoside triphosphates (5-pyridylmethyl-carboxamide-UTP), with the pyridine contributing to increased stacking interactions. A later selection without modified triphosphates resulted in another Diels-Alder ribozyme variant (Seelig & Jaschke, 1999), with a likely different catalytic fold (Serganov *et al.* 2005). Other selections performed with nucleobase-modified

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

triphosphates include, e.g. an amide synthase ribozyme (Wiegand *et al.* 1997) with 5-imadozolyl-UTP and an RNA ligase ribozyme with N6-aminohexyl modified adenine residues (Teramoto *et al.* 2000). Nevertheless none of the modifications are *per se* essential for the catalysed chemical transformation and other ribozymes without modifications are not inferior in their catalytic activity.

3.3 DNA catalysis

For a long-time RNA was only seen as an information carrier from genes to proteins, while the role for DNA was manifested in its function for long-term storage of genetic information. The capacity of DNA for information storage and the possibility of catalytic activity were considered mutually exclusive. Indeed, DNA is generally depicted in the famous double helical form (Watson & Crick, 1953) which, with its rigid linear structure, seems unlikely to support catalysis. It therefore came as a surprise, when, in 1994, the first deoxyribozyme/DNAzyme was identified by *in vitro* selection by Breaker and Joyce (Breaker & Joyce, 1994). This first deoxyribozyme catalysed the Pb^{2+} assisted cleavage of a single ribonucleotide linkage inside an all DNA substrate strand with a rate enhancement of $\sim 10^5$-fold over the uncatalysed reaction.

So far no *bona fide* deoxyribozymes have been found in nature and therefore the question of whether catalytic DNA has functions *in vivo* remains unanswered. Recently, a short Zn^{2+} dependent DNA cleaving deoxyribozyme was identified (Gu *et al.* 2013), and sequence comparison with natural genomes yielded a number of hits with consensus sequences showing DNA cleavage activity under the selection conditions. Further studies will be needed to establish, if this is merely a fortuitous sequence similarity or, if it reflects true *in vivo* functionality.

Since this first example, the catalytic potential of DNA has been explored by *in vitro* selection and many DNAzymes identified that catalyse a diverse range of chemical reactions similar to their RNA counterparts (Hollenstein, 2015; Silverman, 2009, 2016). Indeed, it seems that in a number of ways DNA is not catalytically inferior compared to RNA, despite the absence of the 2′-hydroxyl functionality that can assist in acid/base catalysis or act as a nucleophile in RNA (Silverman, 2008). Rather, deoxyribozymes come with a number of (technical) advantages including easier (and less costly) synthesis and greater resistance to chemical and enzymatic degradation. Nevertheless the deoxyribose in DNA leads to a preferential C3′ endo sugar pucker *versus* a C2′ endo pucker for ribose in RNA, which also results in a preferential B- *versus* A-form helical conformation for double-stranded DNA compared with RNA. This, together with altered base-pairing energetics prevents the direct conversion of ribozymes into deoxyribozymes (or vice versa) leading instead to inactive variants. However, using *in vitro* evolution, one ribozyme could be transformed into the corresponding deoxyribozyme (Paul *et al.* 2006) requiring only seven mutations suggesting that active ribo- and deoxyribozymes may be proximal in sequence space at least in some cases. Indeed, even HHR variants with a mixed ribo-/deoxyribonucleotide backbone can be catalytically active (Perreault *et al.* 1990).

Similar to ribozymes, DNA catalysts show a strong preference for phosphodiester transfer reactions and for nucleic acids substrates in general. Rather than the true catalytic potential, this may again reflect the biases introduced by selection strategies, which are facilitated by the easy positioning of substrates through Watson–Crick base-pairing.

Mechanistic analysis of ribo- and deoxyribozymes suggests that the catalytic potential of RNA and DNA is realized by comparable catalytic strategies. However, while the 3D arrangement of catalytic residues and aspects of the catalytic mechanism of many naturally occurring ribozymes are known in some detail due to high-resolution structures (see above), deoxyribozymes so far lag behind in structural understanding. Nevertheless, there is hope that this might change in the near future. The recent landmark publication of the first atomic resolution structure of a deoxyribozyme (Ponce-Salvatierra *et al.* 2016) paves the way for a more detailed understanding of deoxyribozyme catalysis. The crystal structure was obtained of the 44 nt (of which 31 nt form the catalytic core) comprising minimal RNA-ligating 9DB1 deoxyribozyme (Purtha *et al.* 2005; Wachowius *et al.* 2010) bound to its 15 nt RNA substrate in the post-catalytic state. The structure resembles the Greek letter λ with the two DNA–RNA duplexes of the binding arms forming an angle of 120° to each other and both lying above and tightly attached to the catalytic core. The catalytic domain consists of a 4 and a 2 nt base-pair stem and two nucleotides in the catalytic core (dT29 and dT30), which directly base pair with the RNA nucleotides A1 and G1 at the ligation junction leading to a double pseudoknot structure of the deoxyribozyme RNA substrate complex (Fig. 5).

The original 9DB1 sequence shows a strong preference for purines (A, G) at the 5′ end of the triphosphorylated RNA substrates. Interestingly, as a result of the observed base-pairing between the two DNA nucleobases in the catalytic core with the RNA nucleotides at the ligation junction, a single mutation in the catalytic loop of dT29 to either dG29 or dA29 allows an exchange of the nucleobase at the 5′ position of the triphosphorylated RNA substrate to C or U respectively. This enables ligation of substrates with all 4 RNA nucleobases and demonstrates how structural data may allow the reengineering of deoxyribozymes.

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

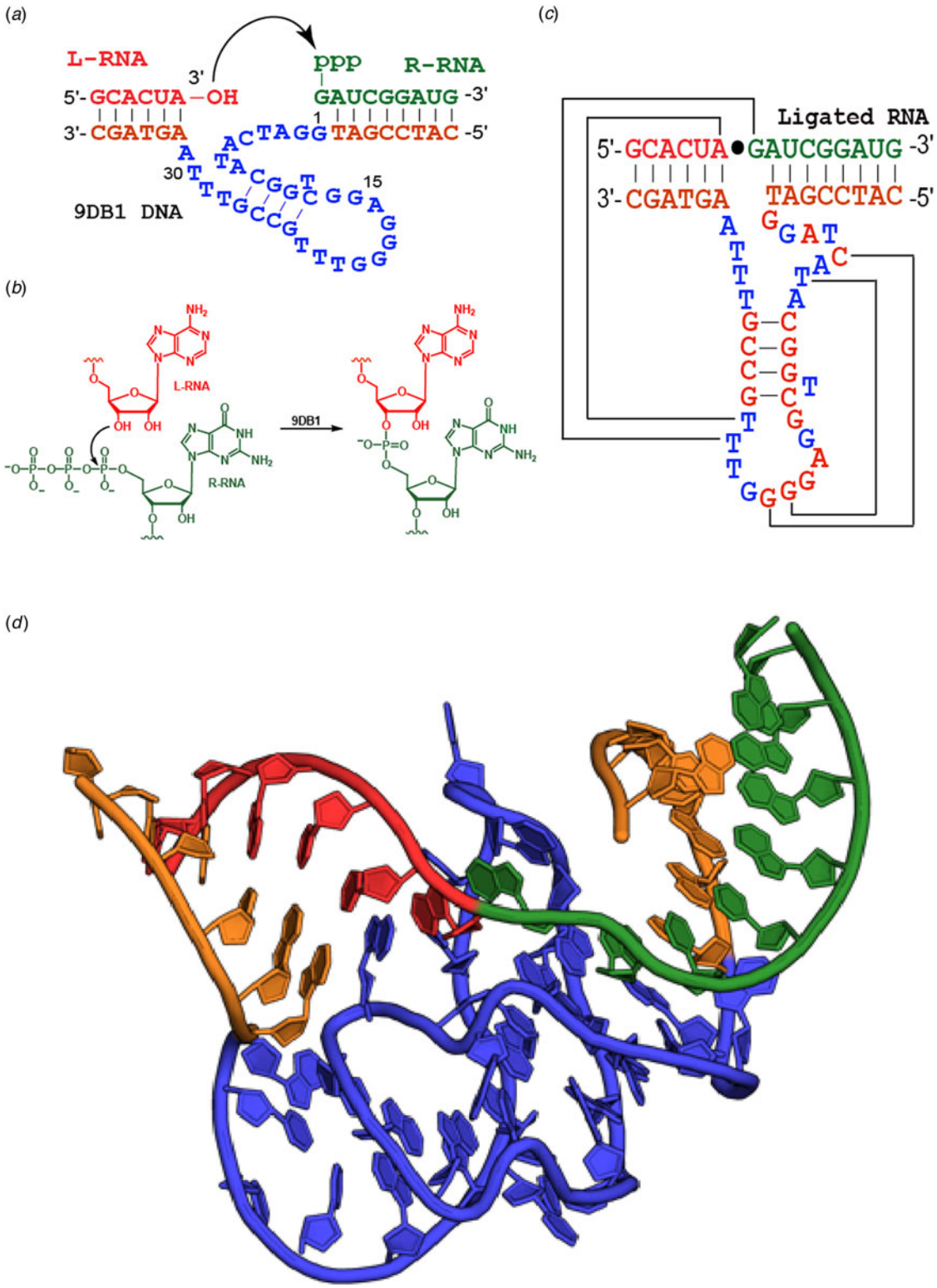

Fig. 5. Mechanism and structure of the RNA-ligating deoxyribozyme 9DB1. (*a*) Secondary structure of the minimized 9DB1 variant, displaying the catalytic core in blue, the RNA binding regions in orange and the RNA substrates 5′ and 3′ of the ligation junction in red and green, respectively. (*b*) Chemical mechanism of 9DB1 catalysed 3′–5′-RNA ligation. The nucleophilic attack of a 3′-hydroxyl of a 2′,3′-diol terminated RNA on a 5′-triphosphorylated RNA substrate generates regioselective 3′–5′-RNA phosphodiester linkages. (*c*) Secondary structure of the 9DB1 crystal structure illustrating the double pseudoknot interactions, red marked nucleotides in the catalytic core are sensitive to mutations. (*d*) Ribbon representation (including the nucleobases) of the crystal structure of the 9DB1 deoxyribozyme bound to its ligated RNA substrate (PDB: 5cck). The colour code corresponds to (*a*).

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

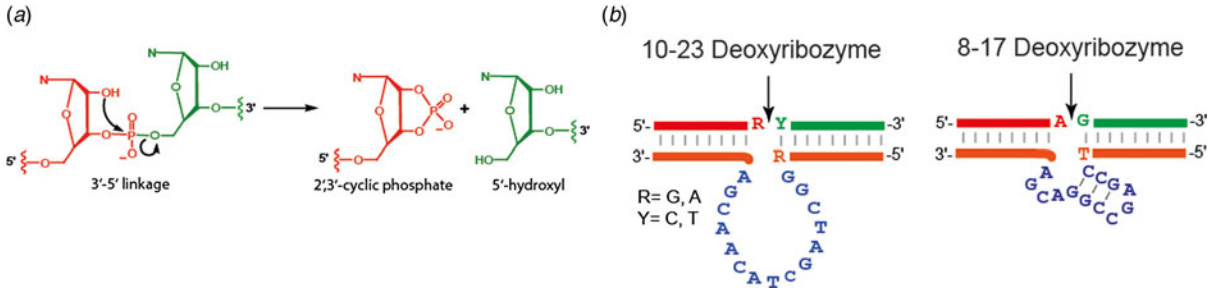

Fig. 6. Deoxyribozyme catalysed RNA cleavage. (*a*) The nucleophilic attack of the 2′-hydroxyl on the adjacent phosphorus of the phosphodiester bond generates 2′,3′-cyclic phosphate and 5′-hydroxyl termini. (*b*) Secondary structure of the most prominent RNA cleaving deoxyribozymes 10-23 and 8-17. The catalytic core is shown in blue, the substrate-binding arms in orange and the RNA strand 5′ and 3′ of the cleavage junction (arrow) are displayed in red and green, respectively.

The structure also provides a first glimpse of how DNA compensates for its 'missing' 2′-OH to perform with comparable catalytic efficiency as RNA. This appears to be achieved by the broad range of the pseudorotation phase angles of nucleotides in the DNAzyme. In particular, the DNA nucleotides in the catalytic loop of 9DB1 show a much broader flexibility of the sugar phosphate backbone compared with ribozymes. There are 20 (out of 31) forming south (S)-type and eight north (N)-type sugar puckers, with the remaining three nucleotides adopting sugar conformations outside typical N-/S-conformations enabling positioning of active residues for catalysis.

The most prominent and widely used deoxyribozymes are RNA cleaving deoxyribozymes (Silverman, 2005). Almost all RNA cleaving deoxyribozymes catalyse RNA cleavage by a transesterification mechanism similar to the small nucleolytic ribozymes, involving an intramolecular attack of the 2′-hydroxyl on the adjacent phosphodiester linkage forming a 2′, 3′-cyclic phosphate and a 5′-hydroxyl terminus. Interestingly, other catalytic mechanisms are possible. Recently, a deoxyribozyme was selected that catalyses RNA cleavage by the normally disfavoured hydrolysis mechanism, e.g. attack of a water molecule on a phosphodiester linkage forming either a 5′-phosphate and 3′-hydroxyl or a 5′-hydroxyl and 3′-phosphate (Parker *et al.* 2013).

The most prominent and best-studied representatives of RNA cleaving deoxyribozymes are the 10–23 and 8–17 deoxyribozymes (Santoro & Joyce, 1997) that catalyse RNA cleavage by transesterification, with multiple-turnover capability (Fig. 6).

Variants of the 8–17 motif have been selected independently a number of times (Schlosser & Li, 2010), making the 8–17 sequence motif the most likely solution for RNA cleavage in DNA sequence space, similar to the HHR in RNA space (Salehi-Ashtiani & Szostak, 2001). The catalytic mechanism of deoxyribozyme-catalysed RNA cleavage is likely similar to that of ribozymes involving one or a combination of the following four catalytic strategies: (a) in-line nucleophilic attack, (b) deprotonation of the 2′-hydroxyl group, (c) neutralization of the negative charge at a non-bridging phosphate or (d) at the 5′ oxygen (Emilsson *et al.* 2003). The preference for divalent metal ions may also reflect their availability during the *in vitro* selection process with their identity having a strong impact on the catalytic rate of deoxyribozymes. Indeed, not only are some deoxyribozymes very selective concerning identity and concentration of the metal ion (whereas others are more relaxed), but also different metal ions can lead to different DNA folding arrangements and reaction rates as demonstrated for the 8–17 deoxyribozyme using FRET (Kim *et al.* 2007). 8–17-catalysed RNA cleavage in the presence of Zn^{2+} and Mg^{2+} proceeds via DNA folding followed by catalysis (i.e. the cleavage reaction), but in the presence of Pb^{2+} the cleavage reaction occurred without a folding step, rationalizing the fast rate of the Pb^{2+} assisted cleavage. This points towards a pre-arranged structural DNA scaffold of 8–17 in the presence of Pb^{2+} ions, but not for Zn^{2+} and Mg^{2+} ions (Kim *et al.* 2007; Liu & Sen, 2010). An interesting recent finding is the influence of trivalent lanthanide ions on deoxyribozyme catalysis (Dokukin & Silverman, 2012; Huang *et al.* 2014; Javadi-Zarnaghi & Hobartner, 2013). A number of lanthanide-dependent RNA-cleaving deoxyribozymes were recently reported (Liu, 2015), including variants depending on two metal ions (Torabi & Lu, 2015; Zhou *et al.* 2016b).

The recent finding of a deoxyribozyme independent of divalent metal ions with a fast catalytic rate ($k_{obs} = 0{\cdot}1\ min^{-1}$ in 400 mM Na^+, 20 °C) and additionally with an astonishing selectivity for Na^+ over competing monovalent cations (Torabi *et al.* 2015) underlines the similarity between ribozyme and deoxyribozyme catalysis and points towards the possibility of nucleobase assisted general acid–base catalysis also for deoxyribozymes. This is similar to earlier findings of RNA-cleaving deoxyribozymes that perform catalysis independent of divalent metal ions (Carrigan *et al.* 2004; Faulhammer & Famulok, 1997; Geyer & Sen, 1997). The Na8 deoxyribozyme has a $k_{obs} = 0{\cdot}007\ min^{-1}$ (0·5 M M^+, pH 7 and 25 °C), where the identity of the monovalent cation (M) is largely irrelevant (Geyer & Sen, 1997). Another deoxyribozyme shows divalent metal

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

independent RNA cleavage at pH3 (Liu *et al.* 2003). As the N1 of adenine, N3 of cytosine and N7 of guanine are expected to be protonated at pH3 (Blackburn *et al.* 2006), the positive charge from the protonated bases likely fulfills the function of the divalent metal ions.

DNA catalysis is also possible with a reduced set of nucleotides, albeit with a substantial decrease in activity. A RNA cleaving deoxyribozyme consisting of only C and G showed a $\sim 10^4$ times reduced cleavage activity compared with the parent one with all four nucleotides, but still with an increase by a factor of ~5000 over the uncatalysed background reaction (Schlosser & Li, 2009). This parallels findings for ribozymes with a reduced nucleobase composition (Reader & Joyce, 2002; Rogers & Joyce, 1999).

Apart from RNA cleavage, DNA-catalysed RNA ligation represents another important reaction type, mainly pursued by Silverman and co-workers. Initial selection efforts identified deoxyriboyzmes catalysing non-native 2′–5′ ligation using Mg^{2+} as cofactor (Flynn-Charlebois *et al.* 2003). Interestingly, using Zn^{2+} instead of Mg^{2+} during the selection process yielded deoxyribozymes catalysing the formation of native 3′–5′ linkages (Hoadley *et al.* 2005), illustrating the important contribution of the metal ion cofactor, not only to catalytic rates but to regioselectivity. Another selection strategy led to Mg^{2+} dependent 3′–5′ RNA-ligating deoxyribozymes with a broader sequence generality and good catalytic efficiencies (Purtha *et al.* 2005). In addition to linear RNA ligation, the 5′-end of one RNA substrate could be ligated to an internal 2′-hydroxyl forming a 2′,5′ branched RNA or as a special case of branch formation a lariat RNA, where the RNA reacts on itself in an intramolecular fashion, forming a closed loop. This reaction type is naturally catalysed by group II introns and the spliceosome. The first RNA 2′,5′ branch-forming deoxyribozymes were identified using the 5′-triphosphate/2′,3′-diol RNA substrate combination, albeit with a rather strong sequence requirement at the ligation junction (Wang & Silverman, 2003). Further selection efforts identified the 7S11 deoxyribozyme, that catalyses 2′,5′-branch formation by ligating a 5′-triphosphorylated G to an internal A residue, which is flanked by Watson–Crick duplex regions, in a similar fashion as the first step of natural RNA splicing (Coppins & Silverman, 2004). 7S11 and later identified 2′,5′ branch-forming deoxyribozymes (Lee *et al.* 2011) all form a three-helix-junction (3HJ) with their RNA and DNA substrates. This structural arrangement is similar to ribozymes that also frequently include multiple helix junction structures.

Deoxyribozymes are also capable of using DNA as substrates and catalysing DNA cleavage and ligation reactions. However, as DNA is much less reactive compared with RNA due to the absence of the 2′-hydroxyl group, DNA substrates have to be activated for ligation to achieve similar catalytic rates as their RNA counterparts. The first deoxyribozyme that catalysed DNA ligation was reported soon after the initial description of the first RNA-cleaving deoxyribozyme (Cuenoud & Szostak, 1995). This deoxyribozyme catalyses the ligation of a 5′-hydroxyl DNA substrate with a 3′-phosphoimidazole activated DNA substrate and is an obligate metalloenzyme, requiring Zn^{2+} (or Cu^{2+}) and Mg^{2+} for activity. Similarly, a deoxyribozyme was identified that uses a 5′-adenylate/3′-hydroxyl substrate combination for DNA ligation, mimicking the final step of protein T4 DNA ligase catalysed DNA ligation (Sreedhara *et al.* 2004). The 5′-adenylate substrate was itself synthesized by a capping deoxyribozyme (Li *et al.* 2000) that forms a 5′,5′-pyrophosphate linkage from ATP and a DNA substrate, which is remarkably different to a phosphorylating deoxyribozyme that uses NTPs to catalyse the 5′ phosporylation of DNA (Li & Breaker, 1999).

Due to the absence of an internal nucleophile (as the 2′-OH in RNA) DNA cleavage is much more difficult to achieve. The first DNA cleaving deoxyribozyme described cleaves DNA in a non-specific manner by a Cu^{2+}-dependent oxidative mechanism (Carmi *et al.* 1996). A completely different mechanism for DNA strand cleavage was achieved by the deoxyribozyme catalysed N-glycosylation of a particular G residue, leading to strand scission at the apurinic site (Sheppard *et al.* 2000). Later, the 10MD5 bimetallic deoxyribozyme was identified, requiring both Zn^{2+} and Mn^{2+} for activity, that cleaves single-stranded DNA by a hydrolysis mechanism with multi-turnover kinetics and an astonishing rate enhancement of 10^{12}, albeit with a rather strong sequence dependence (ATG^T) at the cleavage site (Chandra *et al.* 2009). Only two mutations in the original 10MD5 sequence changed the metal ion requirements from bimetallic Mn^{2+}/Zn^{2+} to Zn^{2+} only, suggesting a simple structural role for Mn^{2+} and a catalytic function for Zn^{2+} (Xiao *et al.* 2011). Further selection efforts identified different DNA cleaving deoxyribozymes with different dinucleotide sequence requirements at the cleavage junction (Xiao *et al.* 2012).

Apart from cleavage/ligation reactions of nucleic acid substrates, deoxyribozymes – just like their ribozyme counterparts – are capable of catalysing a diverse array of other reaction types. Nevertheless, due to design of the selection strategies and the selectivity and convenient ease of programming interactions by Watson–Crick base-pairing, almost all reactions occur on substrates tethered to nucleic acids. Exceptions include the Diels-Alder cycloaddition (Chandra & Silverman, 2008) and porphyrin metallation, e.g. the deoxyribozyme catalysed insertion of Cu^{2+} and Zn^{2+} into mesoporphyrin (Li & Sen, 1996). The Silverman group in particular has been expanding the scope of deoxyribozyme catalysis and their current focus lies on

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

peptide/protein modifying deoxyribozymes (Silverman, 2015). Initially, the first deoxyribozyme that catalysed a RNA nucleopeptide linkage was formed between a 5′-triphosphate RNA and the hydroxyl of a tyrosine residue that was replacing the branch site A in the 7S11 3HJ structural context (Pradeepkumar *et al.* 2008). The less reactive aliphatic hydroxyl of serine required a slightly more flexible arrangement by introduction of a tripeptide sequence (Sachdeva & Silverman, 2010) and for the lysine amino acid side-chain, the more reactive 5′-imidazolide RNA substrate was required (Brandsen *et al.* 2014).

The initial selection trial for amide bond hydrolysis led instead to DNA-hydrolysing deoxyribozymes (Chandra *et al.* 2009). The intended deoxyribozyme catalysed cleavage of amide bonds was finally discovered by a clever selection scheme including a 5′-amino oligonucleotide capture tag, capturing the free carboxyl group that is formed by amide or ester cleavage, but not by DNA phosphodiester bond hydrolysis (Brandsen *et al.* 2013). The chemically more favourable cleavage of aromatic amide bonds was achieved with a standard DNA pool, but for the cleavage of an aliphatic amide bond, a selection scheme including modified deoxyuridines with amino acid type side chains at their 5 position (5-aminoallyl, 5-hydroxymethyl and 5-carboxyvinyl) were used, leading to deoxyribozyme variants with amide bond hydrolase activity for all three modifications and demonstrating the principal ability of DNAzymes to cleave peptidic amide bonds (Zhou *et al.* 2016a).

Apart from the cleavage chemistry, the sequence-specific recognition of amino acids and therefore peptides and proteins has been another challenge. Deoxyribozymes are capable of phosphomonoester hydrolysis; hence, phosphatase activity was established by applying an additional selection step, including a RNA capture oligo and a previously identified deoxyribozyme capable of forming a covalent bond between the free hydroxyl of a tyrosine and the 5′ triphosphorylated RNA capture oligo (Chandrasekar & Silverman, 2013). This Zn^{2+}-dependent phosphatase deoxyribozyme is capable of sequence-specific dephosphorylation of phosphotyrosine and phosphoserine inside a hexapeptide and most importantly also within a protein context. Deoxyribozymes are also capable of catalysing the reverse (phosphorylation) reaction. Deoxyribozymes with tyrosine-specific kinase activity were identified by again using a capture deoxyribozyme catalysing the ligation of only phosphor-Tyr (and not Tyr) with a 5′-triphosporylated RNA or GTP (Walsh *et al.* 2013). Another recently described kinase deoxyribozyme is able to catalyse the 3′-phosphorylation of DNA by using 5′-triphosphorylated RNA (Camden *et al.* 2016) a reaction not catalysed by natural occurring protein enzymes.

3.3.1 Modified deoxyribozymes

Another strategy for M^{2+}- independent deoxyribozymes relies on expanded chemical functionality. In particular, the imidazole function of histidine (His), the amino function of lysine (Lys) and the guanidinium function of arginine (Arg) are often involved in the catalytic centre of protein enzymes, with imidazole assisting in acid/base catalysis, while the cationic functionalities of Lys and Arg provide charge stabilization or a nucleophile in the case of Lys. Amino acids can be either added as external cofactors (as was shown for L-His, which likely acts as a general base in the DNA catalysed cleavage of RNA) (Roth & Breaker, 1998) or covalently linked to the nucleobases (Hollenstein *et al.* 2009; Perrin *et al.* 2001; Santoro *et al.* 2000; Sidorov *et al.* 2004). The main rationale behind M^{2+}-independent deoxyribozymes lies in their *in vivo* application for RNA cleavage or sensor applications, aiming at fast catalytic rates under physiological low M^{2+} conditions as in the blood plasma or intercellular fluid (0.5-1mM free M^{2+}, ~150 mM M^{+}, mainly Na^{+}).

A highly functionalized deoxyribozyme bearing three different nucleobases (dA, dC, dU) with three different amino acid-like functional groups (His, Lys, Arg) by incorporating the deoxynucleoside triphosphates 8-(4-imidazolyl)ethylamino-2′-dATP, 5-aminoallyl-2′-deoxycytidine and 5-guanidiniumallyl-2′-deoxyuridine, led to deoxyribozyme 9–86 with an *in cis* k_{obs} of ~0·13 min^{-1} for cleavage of a rC residue under physiological conditions (200 mM M^{+}, 0·2 mM Mg^{2+}, 37 °C) (Hollenstein *et al.* 2009). The observed catalytic rate is very similar to 10–23 (k_{cat} = 0·15 min^{-1}) under simulated physiological conditions (2 mM Mg^{2+}, 150 mM NaCl, pH 7·5, 37 °C) (Santoro & Joyce, 1997), which shows that RNA cleavage under low M^{2+} concentrations can be achieved with and without extended chemical functionality, but likely relying on different catalytic mechanisms. It will be interesting to see, if the introduction of additional functional groups (or improved positioning of the catalytic side-chains within the (deoxy)ribozyme catalytic centres) can be harnessed to not only improve the catalytic efficiency of already reported reactions, but also expand the catalytic repertoire of (deoxy)ribozyme catalysis. A recent report from the Silverman group (Zhou *et al.* 2016a) describing amide bond hydrolysis by introducing amino acid-like modifications (hydroxy, carboxy and amino) at the 5 position of dU led to deoxyribozymes relying on these modifications, although surprisingly a variant without any modification also showed catalytic activity.

A particularly interesting reaction is the deoxyribozyme catalysed cyclobutane pyrimidine dimer (CPD) photolyase chemistry, identified by Sen and colleagues (Chinnapen & Sen, 2004). The selected UV1C deoxyriboyzme is cofactor independent, but forms a G-quadruplex structure that is capable of harnessing UV-light (~305 nm) and acts as an electron shuttle to the CPD in the DNA substrate, which is subsequently cleaved. In a recent study, the authors showed that replacement of certain G

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

residues inside the UV1C structure by the G analogue 6-methylisoxanthopterin (6MI) (Barlev & Sen, 2013) can induce photolyase activity of UV1C at longer wavelengths (~345 nm). In particular, one G to 6MI mutation (G23) leads to efficient pyrimidine dimer repair in the wavelength range 305–400 nm. In addition, mutation of G23 to the long wavelength nucleoside chromophore DSS (7-(2,2-bithien-5-yl)-imidazo-[4,5-b]pyridine) enabled deoxyribozyme photolyase activity at 420 nm (Barlev & Sen, 2013). The same authors also reported a pyrimidine photolyase deoxyribozyme (Sero1C), using the tryptophan analogue serotonin as catalytic cofactor (Thorne *et al.* 2009). Therefore, the evolutionarily important pyrimidine photodimer repair reaction can be catalysed by a rather simple DNA motif either with a cofactor or without. Given the preponderance of G-quadruplex motifs within genomic DNA, it might be of interest to investigate if parts of the genome itself have an inherent capability of repairing photodamage.

In summary, DNA and RNA can act both as catalysts and information coding molecules, and both use Watson–Crick base-pairing for selective recognition making DNA to RNA and RNA to DNA information transfer possible. RNA and DNA show broadly similar catalytic scopes with DNA not (clearly) inferior to RNA in either catalytic range or efficiency. In the context of the origins of nucleic acid catalysis and the RNA world, one may therefore ask why the hydrolytically less stable RNA would have been preferable. A number of (not mutually exclusive) explanations seem possible, including a potentially more efficient prebiotic synthesis of RNA compared with DNA nucleotides or potentially a greater robustness of RNA-catalysed RNA cleavage and ligation under a wider range of conditions. Furthermore, the propensity of even very simple RNA motifs for self-cleavage and ligation reactions, making RNA more flexible regarding multiple transesterification reactions may have been important to support exploration of sequence space through recombination. Finally, the very instability of RNA to hydrolysis may have been crucial, providing (together with recombination reactions) an evolutionary driving force for folding and stability in the nascent pools of RNA oligomers.

4. RNA self-replication

4.1 Prebiotic synthesis of RNA monomers

Self-replication may be considered a specialized form of catalysis coupled to information transfer. The emergence of RNA self-replication has often been considered as a key transition in the origin of life (Gilbert 1986). However, self-replication in a prebiotic setting requires a template molecule to initiate a replication cycle. Thus, nucleic acid polymers need to be first generated by *de novo* assembly from activated precursors and such activated precursors need in turn be generated from simple prebiotic feedstock molecules. However, a convincing prebiotic synthesis of RNA nucleosides or preferably suitably chemically activated nucleotides had proven elusive for a long time. While individual nucleobases could plausibly be assembled from prebiotic building blocks such as HCN, urea or cyanoacetylene, their linkage to ribose or phosphoribose sugars or indeed the synthesis of such sugars in reasonable yield and purity proved challenging with the most plausible reaction, the so-called formose reaction from formaldehyde, yielding mostly indescribably complex mixtures. Nevertheless, the simple presence of borate salts can selectively stabilize 1,2-*cis*-diol compounds (Ricardo *et al.* 2004) demonstrating a possible path to enrich ribose-containing compounds from such mixtures.

The difficulties in describing credible prebiotic syntheses of ribonucleotides and specifically the apparently intractable problem of N-glycosidic bond formation between ribose and nucleobase in an aqueous environment led to investigation of plausible chemical and genetic precursors of RNA. This 'pre-RNA world' or 'proto-RNA' chemistry is based on alternative genetic polymers with a different backbone chemistry such as TNA (Schoning *et al.* 2000) or PNA (Ura *et al.* 2009) or the exploration of completely different sugar nucleobase combinations (Benner *et al.* 2016; Cafferty *et al.* 2016; Winnacker & Kool, 2013) as possible RNA precursors. Both approaches consider the emergence of RNA not as singular abiotic event from simple organic precursors, but instead as the endpoint of a chemical and evolutionary trajectory from more facile, or seemingly prebiotically easier accessible information systems that were gradually transforming into RNA (Hud *et al.* 2013).

However, the need for direct N-glycosidic bond formation between ribose and pyrimidine nucleobase was elegantly circumvented by the landmark discovery of a prebiotic synthesis of activated RNA pyrimidine nucleotides (C, U) in high yields from simple prebiotically-accessible precursor molecules and inorganic phosphate via amino-oxazolines (Powner *et al.* 2009). In a different pathway, a recently described synthesis of the RNA purine nucleosides (A, G) from formamido-pyrimidines and ribose yielded the correct N9 regioisomer and ribose β-anomer, also avoiding the direct coupling of the full nucleobase and ribose (Becker *et al.* 2016) and its associated problems in yield and stereoselectivity (Fuller *et al.* 1972).

These syntheses provide proof of principle that a prebiotic synthesis of the four RNA building blocks from simple organic precursors is possible and lessens the need for pre-RNA and/or proto-RNA world scenarios. Indeed, one potentially fatal pitfall of pre- or proto-RNA world scenarios concerns the problem of genetic 'handover'. While genotypes (i.e. base sequence)

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

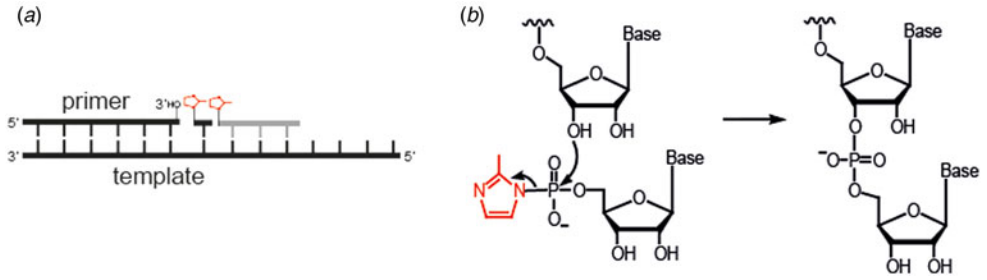

Fig. 7. Non-enzymatic templated polymerization of RNA. (*a*) A templated primer is extended at its 3′ end by 5′-methylimidazolide activated (or other activation chemistries, see text) RNA nucleotides. Polymerization is facilitated by transient binding of 5′-activated short oligonucleotides ('helper' oligomers), coloured in grey, upstream of the template strand. (*b*) The polymerization reaction is based on the nucleophilic attack of the primer 3′-hydroxyl on the 2-methylimidazolide activated 5′-phosphorus of the incoming RNA nucleotide, resulting mainly in canonical 3′–5′-RNA linkages.

are readily transferred between different genetic polymer systems as long as base-pairing properties are not massively distorted (as shown for DNA/RNA and some DNA/XNAs), phenotypes (3D structure/folding/function, e.g. catalytic activity) are generally either substantially impacted or non-transferable. The latter is illustrated by the polymer-specific sequence motifs emerging from *in vitro* evolution experiments and the failure in interconverting active catalysts even between closely related genetic polymer systems, such as DNA and RNA or DNA and ANA (Paul *et al.* 2006; Taylor *et al.* 2015). Finally, while TNA, PNA and other proposed pre-RNA systems show in principle similar information storage capabilities compared with RNA, they nevertheless likely exhibit a different catalytic potential compared with RNA, in particular with regards to transesterification and recombination reactions [as with DNA (see above)], which may have been important for early evolution.

Remarkably, the above described pyrimidine RNA nucleobase synthesis yields 2′,3′-cyclic phosphate activated cytidine and uridine (N > ps) as their final products with similar yields (Powner *et al.* 2009). Assuming that such 2′,3′-cyclic phosphate ribonucleotides are readily accessible from prebiotic chemistry, they could polymerize into short oligonucleotides under favourable conditions (Verlander & Orgel, 1974) (although with preferential formation of the non-canonical 2′–5′ linkages). While a certain amount of sporadic 2′–5′ linkages (within a predominantly 3′–5′ context) are not incompatible with RNA function (Engelhart *et al.* 2013) (see below) it is currently unknown if (and how) a predominantly 2′–5′ RNA polymer could evolve and eventually transition to a 3′–5′ RNA polymer while retaining function. Chemoselective acetylation of the 2′ hydroxyl of ribose may provide a solution: such protection mechanisms can lead to the selective formation of canonical 3′–5′ linkages (Bowler *et al.* 2013).

4.2 Non-enzymatic polymerization of RNA

Non-templated polymerization mediated by substrate alignment and concentration in montmorillonite clays or eutectic ice phases, using the more reactive 5′-phosphorimidazole activated ribonucleotides, can yield RNA oligonucleotides between ~17 nts [with mixed base composition] (Monnard *et al.* 2003) up to 50-mers (homopolymers) (Ferris *et al.* 1996). The prebiotic plausibility of this form of activation is yet to be demonstrated; nucleotide condensation requires phosphate activation arising from either synthesis (e.g. N > ps) or an external electrophile. Oligonucleotide 5′-polyphosphates (including triphosphates) can be formed from polynucleotide mono-phosphates and sodium trimetaphosphate, although given its reactivity the availability and persistence of this agent needs justification. The ideal activating agent or conditions remain to be characterized, but alternative approaches that promote condensation using dehydrating conditions can be imagined. Nucleoside-5′-phosphates can be assembled into polymers by heating and wet/dry cycles in lamellar lipid phases or at acidic pH (Deamer, 2012; DeGuzman *et al.* 2014) though the products of apparent 100 nucleotide length that are observed in gel electrophoresis appear to contain a substantial number of abasic sites (presumably caused by depurination during temperature cycling or at low pH) (Mungi & Rajamani, 2015). Furthermore, due to the inherent chemical fragility of RNA, harsh temperature or chemical/pH gradients are unlikely to be compatible with an early RNA genetic system. Milder conditions for polymerization are likely required to build polymers that retain an intrinsic capability of both information storage and propagation as described below.

RNA templates can pre-organize activated mononucleotides for non-enzymatic polymerization as first explored by Orgel and colleagues for nucleotide phosphorimidazolides and 2-methyimidazolides (Fig. 7).

In particular, the polymerization of guanosine 5′-phosphor-2-methylimidazolides on a polyC template is efficient, resulting in extensions up to 50 nt (Inoue & Orgel, 1982). Nevertheless, guanosine presents the best-case scenario, by combining the two traits of three Watson-Crick hydrogen bonds and a purine ring system, leading to favourable stacking interactions.

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

The analogous polymerization reactions with the three other nucleobases are much less efficient and particularly poor for uridine. Activated ribonucleotides can react with higher efficiency when aided by montmorillonite clay catalysts (Ferris *et al.* 1996), more reactive leaving groups such as 1-methyladenine (Huang & Ferris, 2006) or oxyazabenzotriazolide (Deck *et al.* 2011). More substantial boosts come from tuning the substrate milieu, for example by removing inhibitory hydrolysed monomers by repeated substrate exchange (Deck *et al.* 2011) or through promoting monomer binding by stacking with short downstream 'helper' oligomers (Fig. 7), which recently resulted in the synthesis of an active strand of the HHR (Prywes *et al.* 2016a).

Interactions between leaving groups can substantially alter template-binding affinity (Kervio *et al.* 2016) and polymerization efficiency of nucleotides for example though the local creation of highly reactive intermediates (Walton & Szostak, 2016). The latter strategy relies upon imidazolium-bridged dinucleotide intermediates between adjacent imidazole-activated nucleotide monomer substrates in non-enzymatic templated primer extension and thus may be specific to this activation chemistry. Replication efficiency can also be increased by altering the chemistry of the monomer building blocks, e.g. by replacing the 2′- (or 3′) -hydroxyl with the more potent NH_2-nucleophile, or UTP with the stronger stacking analogue 5-propargyl-UTP. However, this generates nucleic acids with unnatural chemistries, and with the drawback of a reduced replication fidelity (Zhang *et al.* 2013).

Altered template chemistries that pre-organize conformation to RNA-like C3′-endo conformation such as HNA and Alitrol-nucleic acids (AtNA) render non-enzymatic RNA polymerization more efficient than on RNA templates, but their replication would be problematic as HNA- and AtNA-phosphorimidazolides are inefficient substrates for polymerization on RNA templates despite highly stable duplex formation (Kozlov *et al.* 1999a, b, 2000). Fidelity of non-enzymatic replication remains one of the main hurdles, though misincorporations may be depleted in the final products as they lead to stalling of extension and non-templated addition (Leu *et al.* 2013). Some altered nucleotides can improve fidelity, as is the case for 2-thioU (or 2-thio-T), which due to the steric bulk of the C2 sulphur atom have a much reduced tendency to form G·U wobble pairs both in non-enzymatic RNA synthesis (Heuberger *et al.* 2015) as well as in single nucleotide incorporations by the b1–233t polymerase ribozyme (Prywes *et al.* 2016b). Unfortunately, the resulting minor groove modification by the C2 sulphur atom can impact upon downstream synthesis activity by polymerase ribozymes (Attwater *et al.* 2013a). The above described advances in non-enzymatic polymerization starting from the highly activated phosphorimidazolide nucleotides in some cases begin to reach an efficiency (and fidelity) compatible with the templated synthesis and replication of simple ribozymes, therefore closing the conceptual gap between pools of short oligomers created by prebiotic chemistry and the more complex ribozymes thought to have established the RNA world.

Non-templated polymerization of nucleotides activated by the prebiotically more plausible 2′,3′-cyclic phosphate chemistry (>p) tends to generate RNA polymers comprising a substantial fraction of non-canonical 2′–5-linkages (Verlander *et al.* 1973). These linkages also predominate when using 5′-activated nucleotides due to the higher reactivity of the 2′- *versus* the 3′-hydroxyl group. Non-canonical 2′−5′-linkages are highly destabilizing to canonical 3′–5′ linked RNA helical structure (Sheng *et al.* 2014) due to a reduction in both Watson–Crick base-pairing and base-stacking due to a lateral displacement of the base from the helical base-stack and a preference for non-canonical C-2′-endo puckering (Li & Szostak, 2014; Premraj & Yathindra, 1998; Sheng *et al.* 2014). Nevertheless, even fully 2′–5′ linked RNA is able to form specific duplexes with complementary 3′–5′ RNA and (although weaker) with complementary 2′–5′ RNA (Wasner *et al.* 1998). A modest percentage (<25%) of such 2′–5′ linkages are even compatible with ribozyme function (Engelhart *et al.* 2013) and, due to their lower stability to hydrolysis, might over time become depleted in RNA duplex structures; thus, sporadic 2′–5′ linkages have been suggested to reduce product inhibition and aid primordial RNA replication and evolution by transient duplex destabilization (Engelhart *et al.* 2013) at least at low substitution levels. However, due to the ability of 2′–5′ linked RNA strands to self-hybridize and form stable helices (although less stable than 3′–5′ RNA), as well as the altered structural and conformational parameters of 2′–5′ RNA, the possibility that a 2′–5′ RNA sequence space might also contain ligands and catalysts cannot be discounted. Engineering of RNA polymerases capable of synthesizing 2′–5′ linked RNA (or DNA) might allow the exploration of such a sequence space and a testing of this hypothesis (Cozens *et al.* 2015). Nevertheless, it seems unlikely that canonical 3′–5′ RNA catalysts or ligands could emerge from pools of wholly non-canonical 2′–5′ RNA. Nevertheless, a step-wise transition from a mixed population of 3′–5′/2′–5′ RNA to predominantly and wholly 3′–5′ RNA seems more plausible than a wholesale polymer take-over as postulated for a pre-RNA (or protoRNA) world scenario (see above).

4.3 Ribozyme ligases

While non-enzymatic polymerization provides potential avenues for the generation of pools of short RNA oligomers from prebiotic precursor molecules, it is currently unclear, how the longer RNA oligomers likely needed to encode informational functions such as catalysis of ligation or recombination reactions could have emerged from such pools. It is also unknown how frequent such functional sequences are within the RNA sequence space. Indeed, *in vitro* selection experiments suggest

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

that functional sequences are extremely rare (Szostak, 2003) although some very small RNAs can display catalytic function such as the aminoacylating 5 nt ribozyme (Turk *et al.* 2011). Furthermore, larger ribozymes such as the hairpin ribozyme (Vlassov *et al.* 2004) and a triphosphorylation ribozyme (Akoopie & Muller, 2016) can retain function and near wild-type catalytic rates when fragmented into 20–30 nt pieces, which are within the size range accessible from prebiotic chemistry and non-enzymatic replication. Thus, simple ribozymes, may be able to emerge from pools of short oligomers either directly or by non-covalent assembly into functional units and this might allow the bootstrapping of oligomer pools towards the higher compositional and functional complexity needed for self-replication.

So far, enzymatic templated RNA synthesis from mononucleotides appears likely to require quite large catalytic RNAs. This is supported both by theoretical considerations, which suggest a sharp drop off of stable secondary structures (most likely required to form stable active sites) below 30 nts (Briones *et al.* 2009) and *in vitro* evolution experiments aimed at generating ribozymes capable of self-replication. RNA catalysts capable of iterative and template assembly reactions with ligase, recombinase and/or polymerase activity isolated from nature or by *in vitro* evolution are all substantially larger than 20–30 nts. One of the most striking systems is based on two variants of the R3C RNA ligase ribozyme (Lincoln & Joyce, 2009). These are capable of cross-catalytic self-ligation (see below).

Split variants of the *Azoarcus* SSI can also self-assemble into both covalent and non-covalent active complexes and can form cross-catalytic assembly networks (Hayden & Lehman, 2006). Furthermore, both the *sun*Y SSI and a cross-chiral RNA ligase generated by *in vitro* evolution can assemble their complement/mirror chirality sequences from activated oligonucleotides, but require a preformed template strand (Doudna *et al.* 1991; Sczepanski & Joyce, 2014). Finally, RPRs based on the R18 polymerase ribozyme (Johnston *et al.* 2001) (itself derived from the class I ligase ribozyme) (Bartel & Szostak, 1993) are capable of templated synthesis using NTPs as substrates, and some improved variants are able to synthesize other ribozymes, aptamers, tRNAs (Horning & Joyce, 2016; Wochner *et al.* 2011) or RNA oligomers exceeding their own size on favourable template sequences (Attwater *et al.* 2013b). Therefore, there remains a compositional gap between the short RNA oligomer pools and the larger, phenotypically complex ribozymes likely to be required for self-replication, although recent experiments suggest that catalytic cooperation between small ligase and fragmented polymerase ribozymes might be able to close this gap (Mutschler *et al.* 2015).

However, even these complex ribozymes are (currently) not capable of self-replication. One might therefore ask, if self-replication can be implemented by using RNA components alone as postulated in the original (strong) RNA world hypothesis (Neveu *et al.* 2013) and if not, what further functions might be required to realize RNA self-replication. The dramatic demonstration of cross-catalytic RNA self-assembly by Lincoln and Joyce provides an efficient RNA replication system (Lincoln & Joyce, 2009). Starting from two variants of the evolved R3C ligase ribozyme that were engineered to operate in a cross-catalytic format, each ribozyme variant catalysed the formation of the other by ligating two oligonucleotide substrates together. Thus, given a supply of the four component RNAs, an initial catalytic spike of ligase initiated exponential self-assembly.

This quasibiological growth behaviour in a simple and elegant molecular system might be leveraged to assemble other synthetic system components – but can it evolve? Ligase assembly requires pre-defined oligomer substrates with substantial homology to the ribozyme core that can only be supplied externally and this constrains the ability of this system to explore sequence space. Indeed, when substrates with variation in pairing sites are supplied, new ligase variants with better pairing dynamics for exponential amplification can emerge (Lincoln & Joyce, 2009), but the information transmission and hence adaptation can only occur through direct substrate hybridization at these specific loci, and is thus constrained to these small parts of the ribozyme. Other parts of the substrate – including the future catalytic site – are not interrogated during assembly, and if random sequences were supplied, only a negligible fraction of ligatable substrates would yield ligase activity. An elegant split-and-pool substrate synthesis scheme forcing catalytic and recognition regions to co-vary can restore some selection for activity (Sczepanski & Joyce, 2012), but the evolutionary scope of the system remains constrained. Fundamentally, emergence of new functions when assembling long sequences is confounded by the nature of such activities: ligases use less information to choose substrates than is required to define the ligase activity itself, so cannot copy themselves (or other components) from sequences lacking that information, i.e. random sequence. Unconstrained evolution is likely to require more complete information transfer between generations, i.e. encoded RNA from smaller oligonucleotide or mononucleotide building blocks using informationally-complete complementary RNA templates.

4.4 RNA polymerase ribozymes

The emergence of replicases in the RNA world cannot be addressed without understanding mechanisms of non-enzymatic replication. Prior to the emergence of a replicase, non-enzymatic replication would have amplified not just individual

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

sequences but diverse nucleic acid pools. Initially such pools of sequences would evolve to maximize their own abilities as templates (Chen & Nowak, 2012), priming sequence space with sequences (together with their complements) that would likely be amenable to enzymatic replication. Any RNA sequence then able to fold up and catalyse the pre-existing replication process would access new dimensions of selective advantage, without necessarily having to invent a new replication mechanism.

RNA polymerization need not be limited to monomer-building blocks; natural recombinase ribozymes have been harnessed to link together short oligomers in a templated manner extending down to trimers, although with rather low accuracy (Doudna *et al.* 1993). Similar approaches have also been explored for unnatural nucleic acids like glycerol nucleic acids (GNA) (non-enzymatic template-dependent polymerization of apGNA-dinucleotides (Chen *et al.* 2009) and PNA tetra- and penta-oligomers (Brudno *et al.* 2010), where monomer hybridization is weak. However, all oligomer assembly strategies face a challenge in that, while oligomers are easier to assemble than monomers and require fewer catalytic steps (for a given sequence), energetic differences in template binding between cognate and non-cognate substrates rapidly diminish in significance with increasing oligomer lengths thus limiting fidelity.

For this reason and due to the analogies with extant polymerases, achieving RNA-catalysed templated RNA synthesis from mononucleotide building blocks has been a goal ever since the discovery of the first catalytic RNAs. The recombinase activity of group I introns can be leveraged to assemble functional RNAs on RNA templates (Doudna & Szostak, 1989; Green & Szostak, 1992), but the active sites of these natural ribozymes were poorly suited to controlling the identity of the synthesized sequences (Bartel *et al.* 1991; Doudna *et al.* 1993).

New active sites were needed, and a pioneering *in vitro* selection experiment (Bartel & Szostak, 1993) unearthed these *de novo* from pools of random RNA sequences by selecting for the ability to seal a nick in an RNA duplex from 5′-triphosphate and 2′,3′-diol. Among an array of novel ribozyme ligases recovered was the class I ligase, which achieved ligation forming the canonical 3′–5′ linkage. An optimized version of the class I ligase exhibited a remarkable k_{cat} of 100 min^{-1}, still the fastest all-RNA catalyst described. An engineered version of the class I ligase could polymerize a limited number of nucleoside triphosphates (NTPs) on a constrained template (Ekland & Bartel, 1996). Further development of this activity through a combination of *in vitro* evolution and RNA engineering opened up a path towards general ribozyme-catalysed templated RNA replication (Johnston *et al.* 2001), and resulted in the first true polymerase ribozyme (R18) able to add up to 14 nucleotides on a separate primer/template duplex.

R18 polymerase activity was improved by different evolutionary strategies by selecting for the synthesis of longer sequences (Wochner *et al.* 2011; Zaher & Unrau, 2007) (Fig. 8). In the course of these selections, Holliger and colleagues discovered a mode of template hybridization by the polymerase ribozyme via a cognate hexanucleotide motif, akin to the binding and recognition of mRNAs by the prokaryotic ribosome through interactions with the Shine-Dalgarno sequence. Such a mode of cognate RNA recognition may also suggest the potential for RNA kin recognition and selection in early RNA replication, which may have been able to promote phenotype–genotype linkage and keep replication parasites in check prior to effective forms of compartmentalization (see below).

Further evolutionary refinement (based on an in-ice evolution strategy) yielded the tC9Y polymerase ribozyme, which, on a favourable template sequence is able to synthesize RNAs >200 nts long, creating RNA polymers longer than itself (Attwater *et al.* 2013b). tC9Y demonstrates the potential synthetic power of ribozymes, but is currently restricted to favourable RNA template sequences; long extensions remain inefficient upon templates comprising challenging or structured sequences, including those encoding the ribozyme itself. Recently Horning & Joyce described a new polymerase ribozyme variant with improved sequence generality and efficiency, particularly on purine-rich templates, culminating in its ability to perform simple ‘Ribo-PCR’ reactions (Horning & Joyce, 2016). This shows the capability of RNA to catalyse exponential amplification at least of short sequences. The new polymerase ribozyme 24-3 (evolved in 24 rounds of *in vitro* selection from the R18-derived Z RPR as a starting point) also displays an increased ability to read through short template hairpin structures, although at the cost of reduced fidelity of 92%. The increased ability of 24-3 to cope with template secondary structures may be both due to increased speed and efficiency on a wider range of templates.

For RNA templates exhibiting more stable secondary structures alternative strategies may be needed or be helpful. These may include auxiliary factors such as helper strands or helicase ribozymes. However, although the evolution of auxiliary ribozymes like a RNA helicase ribozyme may be possible, it is likely to be challenging and such ribozymes would also need to be replicated, increasing the synthetic burden on the replicase. A more parsimonious approach may be to engineer/evolve a strand-displacement activity in the polymerase ribozyme akin to some proteinaceous polymerases by coupling the energy released from NTP incorporation to strand invasion. Alternatively, one may seek to define conditions or media that would promote a (partial) unfolding of template secondary structures while maintaining ribozyme structure.

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

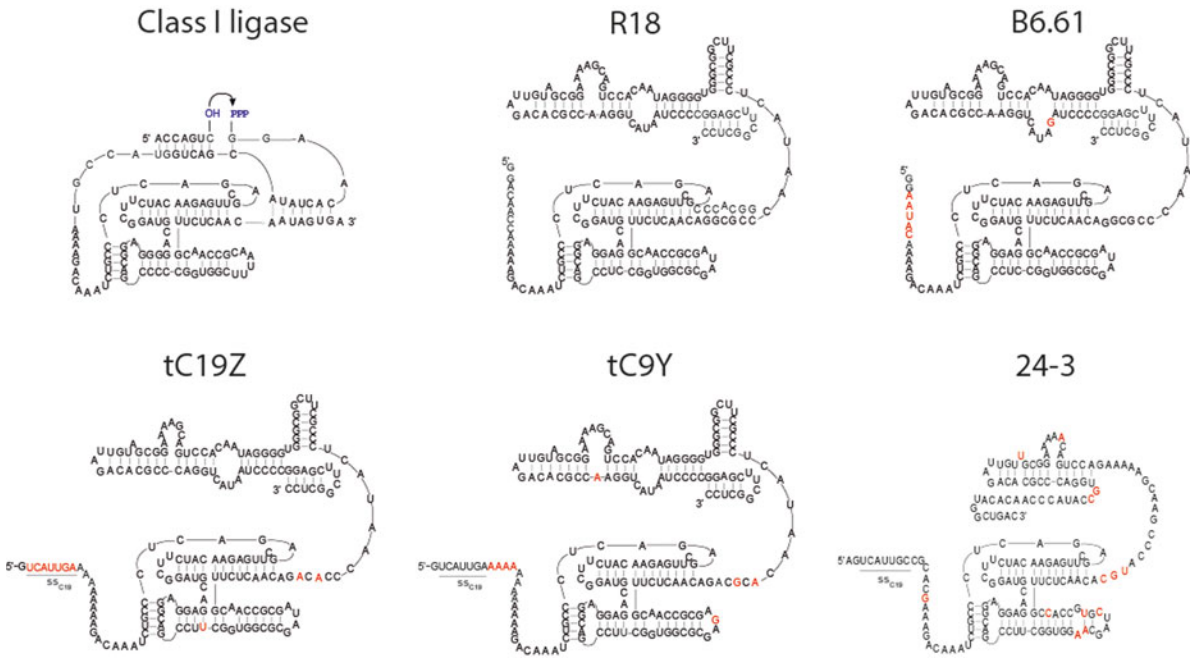

Fig. 8. Ribozyme RNA polymerase (RPR) development. The *in vitro* selected class I ligase catalyses the regioselective formation of canonical 3′−5′-RNA linkages. The addition of an accessory domain at the 3′ end of the class I ligase generated the R18 RNA polymerase. Further *in vitro* selection experiments resulted in the B6·61, tC19Z, tC9Y and 24-3 ribozyme RNA polymerases; the latter three variants include a short tag sequence (ss19) at their 5′ end complementary to the 3′ end of the template sequence. Residues in red are indicating mutations in comparison with R18 for B6·61 and tC19Z or in comparison to tC19Z for tC9Y and 24-3.

Physicochemical cycles (Budin & Szostak, 2010) including thermal, pH, ionic strength as well as wet–dry and freeze–thaw cycles (Mutschler *et al.* 2015) or episodic exposure to high concentrations of denaturants might be able to effect such unfolding – although both thermal and pH cycles harsh enough to disrupt RNA structures would also be likely to accelerate RNA degradation especially in the presence of divalent metal cations. It may be possible to lessen the destructive impact of necessary thermal and pH cycling by reducing the Mg^{2+} requirements of the polymerase ribozymes. Different denaturing cycles such as denaturants and heat or pH and freezing could also be combined in order to lessen the harshness of each individual treatment. Yet, another interesting approach involves the addition of molecular factors that selectively destabilize the duplex form of RNA (or stabilize ssRNA). Indeed, RiboPCR combines high concentrations (0.9 M) of tetrapropyl-ammonium chloride (TPA) to reduce RNA duplex stability with thermocycling (Horning & Joyce, 2016). In another approach, an arginine decapeptide (R10) (Jia *et al.* 2016) selectively binds to ssRNA upon denaturation of a RNA duplex and may aid RNA replication cycles by facilitating repriming. Finally, while it is not clear how severe a problem + and – strand cross-inhibition presents, a possible solution involved a cross-chiral ligase system, wherein a D-RNA ligase assembled its L-RNA equivalent on an L-template (and vice versa) (Sczepanski & Joyce, 2014). As enzyme and substrate (i.e. replicase and replicase template) are of opposing chirality and thus cannot form complementary RNA duplexes, + strands of opposing chirality can be assembled from supplied oligonucleotides (although in any full replication scheme each chiral enzyme would still be exposed to its homochiral template).

A critical strategy towards self-replication by an RNA replicase involves fragmentation of the replicase template at the replication stage. Shorter template strands are not only more accessible to ribozyme-catalysed synthesis (or non-enzymatic replication) due to a lower tendency to contain secondary structure, but, if sufficiently short (i.e. <30 nt long), can be more easily separated into product and template strands after replication. While some simple ribozymes are able to self-assemble from RNA fragments in this size range (Akoopie & Muller, 2016; Vlassov *et al.* 2004), this does not appear to be generally the case, in particular for more complex ribozymes. Indeed, fragmentation and non-covalent assembly of the R18-derived RPR into multiple fragments dramatically reduces activity, and therefore the covalent assembly through a ligase (or recombinase) ribozyme would be required. Recently the assembly of the full-length polymerase ribozyme from seven fragments by an itself fragmented hairpin ligase ribozyme could be demonstrated. The assembly process was performed in the eutectic phase of water-ice in the absence of divalent metal ions and was driven by freeze–thaw cycles, which were found to increase assembly yields by an order of magnitude (Mutschler *et al.* 2015).

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

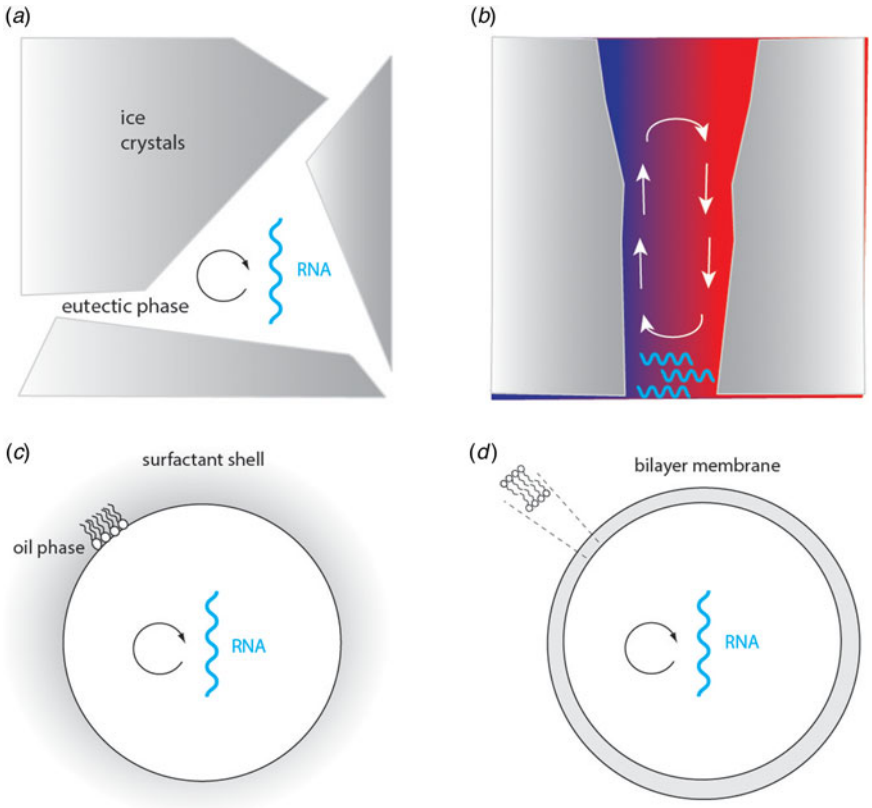

Fig. 9. Possible modes of compartmentalization for RNA at the origin of life. (*a*) Compartmentalization could occur (*a*) in the eutectic phase of water-ice, (*b*) at the bottom of temperature convective pores, (*c*) inside micelles generated by water/oil emulsions or (*d*) inside protocells generated from lipid bilayers.

5. Compartmentalization

Another ancient trait shared throughout extant biology is compartmentalization. Diffusion limitation through confinement inside a molecular compartment or, at the very least, spatial co-localization on a surface (Szabo *et al.* 2002) is a prerequisite for Darwinian evolution and the control of replication parasites (fast replicating sequences that do not contribute to the phenotype). Even preceding such membranous protocells, a wide range of 'membrane-less' forms of compartmentalization could have aided and shaped early evolution.

For a replicase system to evolve requires a form of genetic linkage, whereby a replicase and its offspring remain physically or dynamically linked to ensure kin selection and genotype–phenotype linkage. Such linkages may be spatial, either in the form of compartmentalization or co-localization, or through covalent or non-covalent dynamic interactions. Without such spatial, physical or dynamic linkage self-replication will dissipate as the replicase will replicate unrelated (and most likely inactive) sequences, rather than its own kin. Free-living replicases relying upon covalent template linkage and co-synthetic folding are conceivable (Pace & Marsh, 1985), but physical colocalization through compartmentalization seems a more parsimonious solution with clear parallels to extant biology. Compartmentalization has multiple other potential advantages beyond kin selection and parasite restriction, including diffusion limitation, solute concentration and protection from chemical agents and shearing forces, as well as passive noise filtering thereby protecting self-replication from environmental fluctuations (Stoeger *et al.* 2016).

5.1 Compartmentalization without membranes

Several forms of 'membrane-less' compartmentalization are conceivable and some may have played a role in the context of early evolution. Of particular interest are porous or layered minerals (e.g. clays such as montmorillonite), eutectic ice phases or porous rocks (Fig. 9). Montmorillonite clays and eutectic ice have furthermore been shown to promote both the formation of RNA oligomers from activated nucleotide-building blocks as well as vesicle assembly. It is conceivable that some of these were important in supporting pre-cellular RNA replication. Alternatively, porous rocks in combination with temperature

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

gradients (such as might occur close to hydrothermal systems) have been shown to be able to promote extreme solute concentration (Baaske *et al.* 2007) as well as drive DNA ligation and replication through thermophoresis (Kreysing *et al.* 2015). Thermophoretic systems are of particular interest as they promote the selective concentration of large molecules, i.e. longer RNA oligomers over shorter ones thus providing an unique way of overcoming the 'tyranny of the shortest' in replication. Such a size sorting mechanism could also provide some protection against the (generally) small replication parasites, even in the absence of complete compartmentalization.

Formation of liquid–liquid demixing phases and/or coacervates with highly crowded and charged interiors, which occurs spontaneously at critical concentrations of small biologically relevant cations and anions has been shown to promote RNA catalysis (Jia *et al.* 2016; Strulson *et al.* 2012). Of particular interest are the interactions and the resulting membrane-free microdroplets formed between RNA and simple peptides due to molecular simplicity of the components and the prebiotic context. Indeed, the importance of these phase separation mechanisms is echoed in modern biology, where liquid–liquid demixing gives rise to membrane-free fluidic intracellular compartments rich in DNA, RNA and proteins that are molecularly distinct from the surrounding cytoplasm or nucleus. However, the effects of liquid–liquid demixing and compartment formation on preserving, activating or enhancing RNA activity are still poorly understood.

Another potentially attractive system for both reagent concentration and compartmentalization is the eutectic phase of water-ice. An eutectic phase is formed when aqueous solutions comprising ions, RNA or other solutes are cooled below their freezing point. As freezing proceeds, solutes are excluded from the growing ice crystals and concentrated in an interstitial brine: the eutectic phase. Eutectic phase formation also goes hand in hand with reduced water activity (i.e. dehydration), solute concentration (up to 200-fold) and temperature reduction all of which promote synthetic (over degradative) processes. Indeed, ice phases have been shown to promote some chemical reactions and the formation of RNA oligomers by non-enzymatic polymerization of activated nucleotides (Monnard & Szostak, 2008; Monnard *et al.* 2003). Eutectic ice phases have also been found to stabilize RPR structure and activity (Attwater *et al.* 2010) and enable RPR evolution and adaptation (Attwater *et al.* 2013b). In addition, freeze–thaw cycles have been shown to act akin to modern-day RNA chaperones in promoting refolding of kinetically trapped misfolded RNAs to allow assembly of a complex polymerase ribozyme from small fragments (Mutschler *et al.* 2015).

Although not widely considered as likely forms of prebiotic compartmentalization, emulsions provide an efficient model system to explore the linkage of genotype and phenotype (Fig. 9). Emulsions are formed from mixtures of immiscible liquid phases (e.g. an aqueous and a hydrocarbon oil phase), leading to the dispersion of one of the phases in the other as droplets of microscopic size. Although thermodynamically unstable, emulsion phases can be kinetically stable and persist for long periods of time (even at high termperatures) if stabilized by surfactants.

Of particular interest are water-in-oil (W/O) emulsions, in which the disperse phase forms a suspension of, aqueous cell-like droplets within an inert oil phase. W/O emulsions are experimentally easily tractable model compartments, and have been used for exploring the evolutionary behaviour of model systems of self-replication such in polymerase evolution approaches (Ghadessy *et al.* 2001) and to explore the evolutionary impact of compartmentalization in the Qβ replication system; indeed, the Qβ replicase phenotype can only outlast fast-replicating parasites when replication is compartmentalized within the compartments of a W/O emulsion (Ichihashi *et al.* 2013).

5.2 Compartmentalization with membranes: protocells

Protocellular compartments formed from amphiphilic lipids assemble spontaneously under the right conditions (Fig. 9). These are of paramount importance because of their clear connection to extant biology. As with other forms of compartmentalization the confinement of macromolecules inside membrane-bound vesicles guarantees coupling between genotype and phenotype, while containing the spread of replication parasites. In addition, the physico-chemical properties of the fluid membranes may influence localization and organization of encapsulated polynucleotides and could alter both folding and higher order RNA functions such as RNA catalysis and replication. Membrane properties such as curvature and permeability to solutes as well as vesicle volume, growth and stability may itself be modified in turn by such interactions.

The past decade has seen detailed study of potential host vesicles formed from simple fatty acids (FA), which are moderately permeable, can grow and divide independently, support template non-enzymatic nucleic acid synthesis and maintain stability at high temperatures (Mansy & Szostak, 2008; Mansy *et al.* 2008).

Yet, incompatibilities remain. FA vesicles have a low tolerance for the divalent cations needed by many ribozymes and required for non-enzymatic replication. Such ions, specifically Mg^{2+}, cause FA membrane destabilization, leakage and ultimately FA precipitation. Potential solutions include adaptation of ribozymes to operate without such cations, the inclusion

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

of chelators such as citrate to buffer free Mg^{2+} (Adamala & Szostak, 2013) and the modification of membrane compositions to cope with divalent cations (Namani & Deamer, 2008). Furthermore, membranes are poorly permeable to some replicase substrates and highly charged species such as NTPs are unable to passively diffuse across such membranes. Potential solutions may be found by studying physicochemical cycling of protocells between a permeable and impermeable state (e.g. thermal or freeze–thaw cycles), inclusion of membrane permeability modifiers or the use simpler permeable building blocks that are activated inside the protocell (e.g. by a separate ribozyme such as a triphosphorylating ribozyme) (Moretti & Muller, 2014).

Finally, an enclosed dynamic system must contend with a build-up of potentially inhibitory replication products (pyrophosphate, misextended primers or degraded ribozyme fragments). Nuclease processing would enable clearing of monomers from the protocell by diffusion, but it may be more profitable to recycle such products. Mg^{2+}-catalysed RNA degradation yields 2′,3′-cyclic phosphate termini, and these are potentially directly amenable to religation by the right catalyst, or through regioselective activation chemistry. As a result, degraded ribozymes as well as incomplete extension products could be fed back into synthesis. This would circumvent the need to synthesize full-length ribozymes faster than any backbone breaks occur, and therefore would only require individual ligation synthesis rates to outperform occurrence of backbone breaks, a far more favourable proposition. It might therefore be beneficial to endow protocells with a simple metabolism of substrate activation (Martin *et al.* 2015) or RNA repair and ligation. Indeed, metabolism need not be constrained to mimicking extant biology (Adamala & Szostak, 2013; Rasmussen *et al.* 2016).

6. RNA and peptides: the RNP world

The evidence for an ancient origin of the functional cooperation between RNA and peptides is compelling. A key example is provided by the structure of the inner cores of the large and small ribosomal subunits conserved in all biology (Schmeing & Ramakrishnan, 2009), where ribosomal RNAs are interspersed with unstructured polypeptides (Smith *et al.* 2008) with a highly biased amino acid content. In the context of hierarchical 'accretion' models of ribosome evolution (Bokov & Steinberg, 2009) these peptide 'fingers' appear to have replaced Mg^{2+} as counterions early in ribosome evolution (Hsiao *et al.* 2009).

How could a nascent synthetic system move beyond RNA and harness the enormous potential of peptides and proteins? Short peptides, likely of biased composition, could have catalysed simple metabolic reactions, modify protocell membrane permeability or prove useful cofactors for ribozymes. These peptides could be generated by prebiotic chemistry, by simple ribozymes or the ribozyme ancestor of the peptidyl transfer centre (PTC) of the ribosome. Such simple peptides would likely be limited in their heredity and evolution as encoded protein synthesis requires the vastly more complex multicomponent molecular machinery of the ribosome. Biological components from *Escherichia coli* can be marshaled to generate *in vitro* translation systems (Shimizu *et al.* 2001), and more ambitious proposals seek to integrate translation with DNA and RNA synthesis components to engineer self-sustaining synthetic cells (Forster & Church, 2006). Nevertheless such systems require more than 100 molecular components (most of which are proteins themselves) and are therefore unlikely to illuminate the very origins of translation. Ribozymes have been generated by *in vitro* evolution (see above) that can accelerate some of the chemistries involved in critical aspects of translation (Lohse & Szostak, 1996; Turk *et al.* 2010; Zhang & Cech, 1997), but the key process with regards to evolution, i.e. the decoding of RNA base sequence into a amino acid sequence has not been reproduced by an all RNA system and indeed looks quite complex.

In the absence of encoded protein synthesis and evolution, these simpler peptides likely functioned primarily in stabilizing complex RNA structures. In modern biology, RNA complexion with (poly)peptides to form RNPs is central to both RNA structure, folding and function and to RNA's key roles in genetic information transfer, processing and translation. Indeed, the activity of RNaseP, the spliceosome and the ribosome are critically dependent on association with cognate protein factors despite an all RNA catalytic site. Small cationic peptides can accelerate catalysis in ribozymes that do not depend on protein cofactors, e.g. RNA cleavage by the HHR (Atkins *et al.* 2011; Herschlag *et al.* 1994) or in specifically designed or evolved peptide-dependent ribozymes (Atsumi *et al.* 2001; Robertson *et al.* 2004). In all of these cases the (poly)peptides seems to be function mainly as a counterion, i.e. to overcome electrostatic repulsion during RNA folding and as RNA chaperones to sculpt RNA structure and promote attainment of active conformations. Other potential functions include RNA replication as described recently (Jia *et al.* 2016) in the case of a homo-arginine decapeptide (R10), which selectively binds to ssRNA potentially facilitating non-enzymatic RNA replication cycles. Homopolymeric lysine decapeptides (K10) as well as homo-decapeptides of the non-proteinogenic lysine analogues ornithine (Orn10) and (to a lesser extent) diaminobutyric acid (Dba10), can enhance RPR function irrespective of chirality or chiral purity (Tagami *et al.* 2017). The K10 peptides appear to boost RPR activity by promoting RNA primer-template docking and assembly of the active RPR holoenzyme.

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

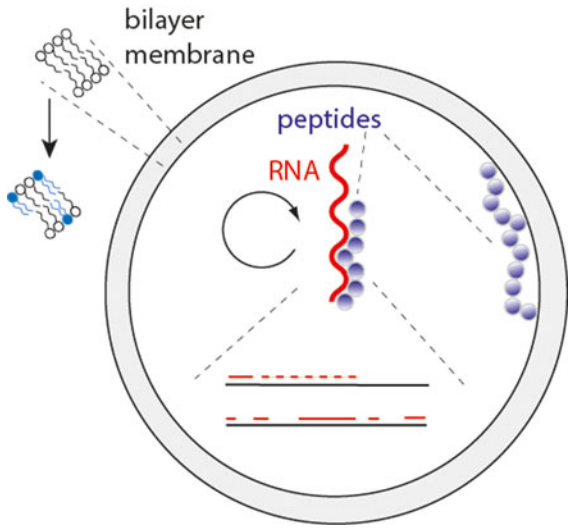

Fig. 10. Possible interactions of biomolecules (RNA, peptides and lipids) at the origin of life. RNA sequences are synthesized non-enzymatically or enzymatically (ribozymes) in a templated manner inside protocells that are generated from lipid bilayers. Ribozyme catalysis would include self-replication by a possible RNA replicase. Peptides are assisting in ribozyme stability and catalysis and membrane stability and integrity.

They also appear to accelerate RPR evolution towards lower Mg^{2+} requirements and enable RPR activity at near physiological ($\geqslant$1 mM) Mg^{2+} concentrations. This allowed the encapsulation of templated RNA synthesis by a RPR within membranous protocells (Tagami *et al.* 2017). Thus, simple cationic peptides may have aided RNA folding, evolution and the formation of the first protocellular entities early on in the RNA world, even preceding the emergence of encoded protein synthesis by the ribosome.

A key question in this context is how such peptides could have provided a beneficial heritable phenotype in the absence of encoded synthesis. Compositionally simple peptides such as the homo-arginine (R10) (Jia *et al.* 2016) and homo-lysine (K10) (Tagami *et al.* 2017) or mixed arginine–tryptophan peptides promoting RNA membrane localization (Kamat *et al.* 2015) might have been generated without complex decoding, but derived from non-templated peptide synthesis by simple peptidyl-transferase ribozymes with a narrow substrate specificity (akin to the modern-day D-Ala-D-Ala ligase enzymes) providing the missing link to heredity (in the form of the peptidyl-transferase ribozymes themselves) as proposed by Cech (Cech, 2009).

7. Synthesizing life

While there are undeniable functional and conceptual arguments for placing nucleic acids at life's origin, the choice between different forms of nucleic acids, be it RNA, DNA or XNAs, is less clear. While historical arguments clearly favour RNA, due to its centrality in the central dogma and its role in catalysing both translation and splicing, functional arguments are less compelling as both RNA and DNA (and XNAs, at least at the basic level so far explored) are able to encode and propagate information and form ligands and catalysts with comparable efficiency. Nevertheless, there are unique aspects of RNA that may be critical such as the vicinal diol arrangement on the ribofuranose ring, with important implications for RNA stability, folding, recombination, polymerization and membrane uptake (Sacerdote & Szostak, 2005).

While the relative importance of this and other divergent traits for 'booting up' life's first genetic system remains unclear, they are increasingly within reach of experimental exploration. Efforts towards the *de novo* assembly of chemical systems displaying life-like properties are closely bound up with the quest to demonstrate a plausible mechanism for the origin of life from prebiotic chemistry (Sutherland, 2016). Such a true synthetic biology aims to demonstrate evolution towards complexity – the capacity to gain ever more complex phenotypes – in a simple system far closer to chemical processes than modern biology (for a more detailed discussion see Attwater & Holliger, 2014; Pinheiro & Holliger, 2014; Szostak *et al.* 2001).

Of particular interest in this regard will be the nascent informational and catalytic capabilities of simple RNA oligomer pools emerging from prebiotic processes as well as ribozymes arising from and building upon early self-replication processes. Construction of synthetic life through engineering and *in vitro* selection represents a stepping-stone towards evolving systems that could have emerged and operated under plausible prebiotic environments on the early Earth.

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

RNA-based replication likely did not function in isolation but occurred in the context of a complex molecular environment involving not just RNA but simple peptides and lipids as provided by prebiotic chemistry (Fig. 10). Only within this unique combination of RNA acting as information carrier and catalyst within a network of interactions among prebiotic chemical compounds may the full potential of each molecular system be realized. Indeed, an emerging molecular symbiosis among different prebiotic molecular entities may be at the heart of the transition from prebiotic chemistry to early biology.

The investigation of such RNA-based quasibiological systems, with chemistries allowed to develop under varying conditions, may begin to reveal the reasons for the primacy of RNA at the onset of life and thereby establish a unique evidentiary connection between synthetic life in modern laboratory conditions and the primordial biosphere.

Acknowledgements

This work was supported by the Medical Research Council (program no. U105178804) (to P.H., F.W. and J.A.) and a grant (no. 293387) from the Simons Foundation (to F.W.).

References

Adamala, K. & Szostak, J. W. (2013). Nonenzymatic template-directed RNA synthesis inside model protocells. *Science* **342**, 1098–1100.

Adami, C. & Labar, T. (2015). From entropy to information: biased typewriters and the origin of life. arXiv:1506.06988.

Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J. & Strobel, S. A. (2004). Crystal structure of a self-splicing group I intron with both exons. *Nature* **430**, 45–50.

Akoopie, A. & Muller, U. F. (2016). Lower temperature optimum of a smaller, fragmented triphosphorylation ribozyme. *PhysChemChemPhys* **18**, 20118–20125.

Ameta, S., Winz, M. L., Previti, C. & Jaschke, A. (2014). Next-generation sequencing reveals how RNA catalysts evolve from random space. *Nucleic Acids Research* **42**, 1303–1310.

Anosova, I., Kowal, E. A., Dunn, M. R., Chaput, J. C., Van Horn, W. D. & Egli, M. (2016). The structural diversity of artificial genetic polymers. *Nucleic Acids Research* **44**, 1007–1021.

Atkins, J. F., Gesteland, R. F. & Cech, T. R. (eds.) (2011). *RNA Worlds. From Life's Origins to Diversity in Gene Regulation*. CSH Laboratory Press, New York.

Atsumi, S., Ikawa, Y., Shiraishi, H. & Inoue, T. (2001). Design and development of a catalytic ribonucleoprotein. *EMBO Journal* **20**, 5453–5460.

Attwater, J. & Holliger, P. (2014). A synthetic approach to abiogenesis. *Nature Methods* **11**, 495–498.

Attwater, J., Tagami, S., Kimoto, M., Butler, K., Kool, E. T., Wengel, J., Herdewijn, P., Hirao, I. & Holliger, P. (2013a). Chemical fidelity of an RNA polymerase ribozyme. *Chemical Sciences* **4**, 2804–2814.

Attwater, J., Wochner, A. & Holliger, P. (2013b). In-ice evolution of RNA polymerase ribozyme activity. *Nature Chemistry* **5**, 1011–1018.

Attwater, J., Wochner, A., Pinheiro, V. B., Coulson, A. & Holliger, P. (2010). Ice as a protocellular medium for RNA replication. *Nature Communications* **1**, 76.

Baaske, P., Weinert, F. M., Duhr, S., Lemke, K. H., Russell, M. J. & Braun, D. (2007). Extreme accumulation of nucleotides in simulated hydrothermal pore systems. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 9346–9351.

Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2·4 A resolution. *Science* **289**, 905–920.

Barlev, A. & Sen, D. (2013). Catalytic DNAs that harness violet light to repair thymine dimers in a DNA substrate. *Journal of the American Chemical Society* **135**, 2596–2603.

Bartel, D. P., Doudna, J. A., Usman, N. & Szostak, J. W. (1991). Template-directed primer extension catalyzed by the Tetrahymena ribozyme. *Molecular and Cellular Biology* **11**, 3390–3394.

Bartel, D. P. & Szostak, J. W. (1993). Isolation of new ribozymes from a large pool of random sequences [see comment]. *Science* **261**, 1411–1418.

Becker, S., Thoma, I., Deutsch, A., Gehrke, T., Mayer, P., Zipse, H. & Carell, T. (2016). A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* **352**, 833–836.

Benner, S. A. (2004). Understanding nucleic acids using synthetic chemistry. *Accounts of Chemical Research* **37**, 784–797.

Benner, S. A., Karalkar, N. B., Hoshika, S., Laos, R., Shaw, R. W., Matsuura, M., Fajardo, D. & Moussatche, P. (2016). Alternative Watson-Crick synthetic genetic systems. *Cold Spring Harbor Perspectives in Biology* **8**, 1–26.

Betz, K., Malyshev, D. A., Lavergne, T., Welte, W., Diederichs, K., Romesberg, F. E. & Marx, A. (2013). Structural insights into DNA replication without hydrogen bonds. *Journal of the American Chemical Society* **135**, 18637–18643.

Bissette, A. J. & Fletcher, S. P. (2013). Mechanisms of autocatalysis. *Angewandte Chemie (International ed. in English)* **52**, 12800–12826.

Blackburn, G. M., Gait, M. J., Loakes, D. & Williams, D. M. (2006). *Nucleic Acids in Chemistry and Biology*. Cambridge: The Royal Society of Chemistry.

Bokov, K. & Steinberg, S. V. (2009). A hierarchical model for evolution of 23S ribosomal RNA. *Nature* **457**, 977–980.

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

Bowler, F. R., Chan, C. K., Duffy, C. D., Gerland, B., Islam, S., Powner, M. W., Sutherland, J. D. & Xu, J. (2013). Prebiotically plausible oligoribonucleotide ligation facilitated by chemoselective acetylation. *Nature Chemistry* **5**, 383–389.

Brandsen, B. M., Hesser, A. R., Castner, M. A., Chandra, M. & Silverman, S. K. (2013). DNA-catalyzed hydrolysis of esters and aromatic amides. *Journal of the American Chemical Society* **135**, 16014–16017.

Brandsen, B. M., Velez, T. E., Sachdeva, A., Ibrahim, N. A. & Silverman, S. K. (2014). DNA-catalyzed lysine side chain modification. *Angewandte Chemie (International ed. in English)* **53**, 9045–9050.

Breaker, R. R. (2012). Riboswitches and the RNA world. *Cold Spring Harbour Perspectives in Biology* **4**, 1–17.

Breaker, R. R. & Joyce, G. F. (1994). A DNA enzyme that cleaves RNA. *Chemistry & Biology* **1**, 223–229.

Briones, C., Stich, M. & Manrubia, S. C. (2009). The dawn of the RNA World: toward functional complexity through ligation of random RNA oligomers. *RNA* **15**, 743–749.

Brudno, Y., Birnbaum, M. E., Kleiner, R. E. & Liu, D. R. (2010). An *in vitro* translation, selection and amplification system for peptide nucleic acids. *Nature Chemical Biology* **6**, 148–155.

Budin, I. & Szostak, J. W. (2010). Expanding roles for diverse physical phenomena during the origin of life. *Annual Review in Biophysics* **39**, 245–263.

Cafferty, B. J., Fialho, D. M., Khanam, J., Krishnamurthy, R. & Hud, N. V. (2016). Spontaneous formation and base pairing of plausible prebiotic nucleotides in water. *Nature Communication* **7**, 11328.

Cairns-Smith, A. G. (1966). The origin of life and the nature of the primitive gene. *Journal of Theoretical Biology* **10**, 53–88.

Camden, A. J., Walsh, S. M., Suk, S. H. & Silverman, S. K. (2016). DNA oligonucleotide 3′-phosphorylation by a DNA enzyme. *Biochemistry* **55**, 2671–2676.

Carmi, N., Shultz, L. A. & Breaker, R. R. (1996). *In vitro* selection of self-cleaving DNAs. *Chemistry & Biology* **3**, 1039–1046.

Carrigan, M. A., Ricardo, A., Ang, D. N. & Benner, S. A. (2004). Quantitative analysis of a RNA-cleaving DNA catalyst obtained via *in vitro* selection. *Biochemistry* **43**, 11446–11459.

Cech, T. R. (2009). Evolution of biological catalysis: ribozyme to RNP enzyme. *Cold Spring Harbor Symposia on Quantitative Biology* **74**, 11–16.

Cech, T. R. (2012). The RNA worlds in context. *Cold Spring Harbour Perspectives in Biology* **4**, a006742.

Cech, T. R. & Steitz, J. A. (2014). The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* **157**, 77–94.

Cernak, P. & Sen, D. (2013). A thiamin-utilizing ribozyme decarboxylates a pyruvate-like substrate. *Nature Chemistry* **5**, 971–977.

Chandra, M., Sachdeva, A. & Silverman, S. K. (2009). DNA-catalyzed sequence-specific hydrolysis of DNA. *Nature Chemical Biology* **5**, 718–720.

Chandra, M. & Silverman, S. K. (2008). DNA and RNA can be equally efficient catalysts for carbon-carbon bond formation. *Journal of the American Chemical Society* **130**, 2936–2937.

Chandrasekar, J. & Silverman, S. K. (2013). Catalytic DNA with phosphatase activity. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 5315–5320.

Chen, I. A. & Nowak, M. A. (2012). From prelife to life: how chemical kinetics become evolutionary dynamics. *Accounts of Chemical Research* **45**, 2088–2096.

Chen, J. J., Cai, X. & Szostak, J. W. (2009). N2′→P3′ Phosphoramidate Glycerol Nucleic Acid as a Potential Alternative Genetic System. *Journal of the American Chemical Society* **131**, 2119–2121.

Chen, X., Li, N. & Ellington, A. D. (2007). Ribozyme catalysis of metabolism in the RNA world. *Chemistry & Biodiversity* **4**, 633–655.

Chen, Y. J., Groves, B., Muscat, R. A. & Seelig, G. (2015). DNA nanotechnology from the test tube to the cell. *Nature Nanotechnology* **10**, 748–760.

Chinnapen, D. J. & Sen, D. (2004). A deoxyribozyme that harnesses light to repair thymine dimers in DNA. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 65–69.

Church, G. M., Gao, Y. & Kosuri, S. (2012). Next-generation digital information storage in DNA. *Science* **337**, 1628.

Cochrane, J. C. & Strobel, S. A. (2004). Probing RNA structure and function by nucleotide analog interference mapping. Current Protocols in Nucleic Acid Chemistry Chapter 6, Unit 6 9, Wiley, Hoboken, New Jersey.

Conn, M. M., Wintner, E. A. & Rebek, J. (1994). Template effects in new self-replicating molecules. *Angewandte Chemie (International ed. in English)* **33**, 1577–1579.

Coppins, R. L. & Silverman, S. K. (2004). A DNA enzyme that mimics the first step of RNA splicing. *Nature Structural & Molecular Biology* **11**, 270–274.

Cozens, C., Mutschler, H., Nelson, G. M., Houlihan, G., Taylor, A. I. & Holliger, P. (2015). Enzymatic synthesis of nucleic acids with defined regioisomeric 2′−5′ linkages. *Angewandte Chemie (International ed. in English)* **54**, 15570–15573.

Cozens, C., Pinheiro, V. B., Vaisman, A., Woodgate, R. & Holliger, P. (2012). A short adaptive path from DNA to RNA polymerases. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 8067–8072.

Cronin, L. & Walker, S. I. (2016). ORIGIN OF LIFE. Beyond prebiotic chemistry. *Science* **352**, 1174–1175.

Cuenoud, B. & Szostak, J. W. (1995). A DNA metalloenzyme with DNA ligase activity. *Nature* **375**, 611–614.

Curtis, E. A. & Bartel, D. P. (2005). New catalytic structures from an existing ribozyme. *Nature Structural & Molecular Biology* **12**, 994–1000.

Das, S. R. & Piccirilli, J. A. (2005). General acid catalysis by the hepatitis delta virus ribozyme. *Nature Chemical Biology* **1**, 45–52.

Deamer, D. (2012). Liquid crystalline nanostructures: organizing matrices for non-enzymatic nucleic acid polymerization. *Chemical Society Reviews* **41**, 5375–5379.

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

Deck, C., Jauker, M. & Richert, C. (2011). Efficient enzyme-free copying of all four nucleobases templated by immobilized RNA. *Nature Chemistry* **3**, 603–608.

Deguzman, V., Vercoutere, W., Shenasa, H. & Deamer, D. (2014). Generation of oligonucleotides under hydrothermal conditions by non-enzymatic polymerization. *Journal of Molecular Evolution* **78**, 251–262.

Denesyuk, N. A. & Thirumalai, D. (2015). How do metal ions direct ribozyme folding? *Nature Chemistry* **7**, 793–801.

Dokukin, V. & Silverman, S. K. (2012). Lanthanide ions as required cofactors for DNA catalysts. *Chemical Science* **3**, 1707–1714.

Dolan, G. F., Akoopie, A. & Muller, U. F. (2015). A faster triphosphorylation ribozyme. *PLoS ONE* **10**, e0142559.

Doudna, J. A., Couture, S. & Szostak, J. W. (1991). A multisubunit ribozyme that is a catalyst of and template for complementary strand RNA synthesis. *Science* **251**, 1605–1608.

Doudna, J. A. & Szostak, J. W. (1989). RNA-catalysed synthesis of complementary-strand RNA. *Nature* **339**, 519–522.

Doudna, J. A., Usman, N. & Szostak, J. W. (1993). Ribozyme-catalyzed primer extension by trinucleotides: a model for the RNA-catalyzed replication of RNA. *Biochemistry* **32**, 2111–2115.

Eckstein, F. (2014). Phosphorothioates, essential components of therapeutic oligonucleotides. *Nucleic Acid Therapeutics* **24**, 374–387.

Eiler, D., Wang, J. & Steitz, T. A. (2014). Structural basis for the fast self-cleavage reaction catalyzed by the twister ribozyme. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 13028–13033.

Ekland, E. H. & Bartel, D. P. (1996). RNA-catalysed RNA polymerization using nucleoside triphosphates. *Nature* **382**, 373–376.

Ellington, A. D. & Szostak, J. W. (1990). *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822.

Emilsson, G. M., Nakamura, S., Roth, A. & Breaker, R. R. (2003). Ribozyme speed limits. *RNA* **9**, 907–918.

Engelhart, A. E., Powner, M. W. & Szostak, J. W. (2013). Functional RNAs exhibit tolerance for non-heritable 2′–5′ *versus* 3′–5′ backbone heterogeneity. *Nature Chemistry* **5**, 390–394.

Eschenmoser, A. (1999). Chemical etiology of nucleic acid structure. *Science* **284**, 2118–2124.

Famulok, M. & Mayer, G. (2014). Aptamers and SELEX in Chemistry & Biology. *Chemistry & Biology* **21**, 1055–1058.

Faulhammer, D. & Famulok, M. (1997). Characterization and divalent metal-ion dependence of *in vitro* selected deoxyribozymes which cleave DNA/RNA chimeric oligonucleotides. *Journal of Molecular Biology* **269**, 188–202.

Fekry, M. I., Tipton, P. A. & Gates, K. S. (2011). Kinetic consequences of replacing the internucleotide phosphorus atoms in DNA with arsenic. *ACS Chemical Biology* **6**, 127–130.

Ferre-D'amare, A. R. & Scott, W. G. (2010). Small self-cleaving ribozymes. *Cold Spring Harbor Perspectives in Biology* **2**, a003574.

Ferre-D'amare, A. R., Zhou, K. & Doudna, J. A. (1998). Crystal structure of a hepatitis delta virus ribozyme. *Nature* **395**, 567–574.

Ferris, J. P., Hill, Jr., A. R., Liu, R. & Orgel, L. E. (1996). Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* **381**, 59–61.

Fica, S. M., Tuttle, N., Novak, T., Li, N. S., Lu, J., Koodathingal, P., Dai, Q., Staley, J. P. & Piccirilli, J. A. (2013). RNA catalyses nuclear pre-mRNA splicing. *Nature* **503**, 229–234.

Flynn-Charlebois, A., Wang, Y., Prior, T. K., Rashid, I., Hoadley, K. A., Coppins, R. L., Wolf, A. C. & Silverman, S. K. (2003). Deoxyribozymes with 2′–5′ RNA ligase activity. *Journal of the American Chemical Society* **125**, 2444–2454.

Forster, A. C. & Church, G. M. (2006). Towards synthesis of a minimal cell. *Molecular Systems Biology* **2**, 45.

Frank, J. (2016). Whither ribosome structure and dynamics research? (A Perspective). *Journal of Molecular Biology* **428**, 3565–3569.

Fuller, W. D., Sanchez, R. A. & Orgel, L. E. (1972). Studies in prebiotic synthesis. VI. Synthesis of purine nucleosides. *Journal of Molecular Biology* **67**, 25–33.

Fusz, S., Eisenfuhr, A., Srivatsan, S. G., Heckel, A. & Famulok, M. (2005). A ribozyme for the aldol reaction. *Chemistry & Biology* **12**, 941–950.

Gaines, C. S. & York, D. M. (2016). Ribozyme catalysis with a twist: active state of the twister ribozyme in solution predicted from molecular simulation. *Journal of the American Chemical Society* **138**, 3058–3065.

Georgiadis, M. M., Singh, I., Kellett, W. F., Hoshika, S., Benner, S. A. & Richards, N. G. (2015). Structural basis for a six nucleotide genetic alphabet. *Journal of the American Chemical Society* **137**, 6947–6955.

Gesteland, R. F., Cech, T. R. & Atkins, J. F. (2005). *The RNA World*, Cold Spring Harbor Laboratory Press.

Geyer, C. R. & Sen, D. (1997). Evidence for the metal-cofactor independence of an RNA phosphodiester-cleaving DNA enzyme. *Chemistry & Biology* **4**, 579–593.

Ghadessy, F. J., Ong, J. L. & Holliger, P. (2001). Directed evolution of polymerase function by compartmentalized self-replication. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 4552–4557.

Gilbert, W. (1986). Origin of life: The RNA world. *Nature* **319**, 618.

Grabow, W. W. & Jaeger, L. (2014). RNA self-assembly and RNA nanotechnology. *Accounts of Chemical Research* **47**, 1871–1880.

Green, R. & Szostak, J. W. (1992). Selection of a ribozyme that functions as a superior template in a self-copying reaction. *Science* **258**, 1910–1915.

Gu, H., Furukawa, K., Weinberg, Z., Berenson, D. F. & Breaker, R. R. (2013). Small, highly active DNAs that hydrolyze DNA. *Journal of the American Chemical Society* **135**, 9121–9129.

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**, 849–857.

Guo, P. (2010). The emerging field of RNA nanotechnology. *Nature Nanotechnology* **5**, 833–842.

Gutierrez, J. M., Hinkley, T., Taylor, J. W., Yanev, K. & Cronin, L. (2014). Evolution of oil droplets in a chemorobotic platform. *Nature Communications* **5**, 5571.

Hammann, C., Luptak, A., Perreault, J. & De La Pena, M. (2012). The ubiquitous hammerhead ribozyme. *RNA* **18**, 871–885.

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

Hanczyc, M. M. & Szostak, J. W. (2004). Replicating vesicles as models of primitive cell growth and division. *Current Opinion in Chemical Biology* **8**, 660–664.

Harris, K. A., Lunse, C. E., Li, S., Brewer, K. I. & Breaker, R. R. (2015). Biochemical analysis of pistol self-cleaving ribozymes. *RNA* **21**, 1852–1858.

Harris, T. K. & Turner, G. J. (2002). Structural basis of perturbed pKa values of catalytic groups in enzyme active sites. *IUBMB Life* **53**, 85–98.

Hayden, E. J. & Lehman, N. (2006). Self-assembly of a group I intron from inactive oligonucleotide fragments. *Chemistry & Biology* **13**, 909–918.

Hayden, E. J., von Kiedrowski, G. & Lehman, N. (2008). Systems chemistry on ribozyme self-construction: evidence for anabolic autocatalysis in a recombination network. *Angewandte Chemie (International ed. in English)* **47**, 8424–8428.

Herschlag, D., Khosla, M., Tsuchihashi, Z. & Karpel, R. L. (1994). An RNA chaperone activity of non-specific RNA binding proteins in hammerhead ribozyme catalysis. *EMBO Journal* **13**, 2913–2924.

Heuberger, B. D., Pal, A., Del Frate, F., Topkar, V. V. & Szostak, J. W. (2015). Replacing uridine with 2-thiouridine enhances the rate and fidelity of nonenzymatic RNA primer extension. *Journal of the American Chemical Society* **137**, 2769–2775.

Hirao, I., Kimoto, M. & Yamashige, R. (2012). Natural *versus* artificial creation of base pairs in DNA: origin of nucleobases from the perspectives of unnatural base pair studies. *Accounts of Chemical Research* **45**, 2055–2065.

Hoadley, K. A., Purtha, W. E., Wolf, A. C., Flynn-Charlebois, A. & Silverman, S. K. (2005). Zn2+-dependent deoxyribozymes that form natural and unnatural RNA linkages. *Biochemistry* **44**, 9217–9231.

Hollenstein, M. (2015). DNA catalysis: the chemical repertoire of DNAzymes. *Molecules* **20**, 20777–20804.

Hollenstein, M., Hipolito, C. J., Lam, C. H. & Perrin, D. M. (2009). A self-cleaving DNA enzyme modified with amines, guanidines and imidazoles operates independently of divalent metal cations (M^{2+}). *Nucleic Acids Research* **37**, 1638–1649.

Horning, D. P. & Joyce, G. F. (2016). Amplification of RNA by an RNA polymerase ribozyme. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 9786–9791.

Hsiao, C., Chou, I. C., Okafor, C. D., Bowman, J. C., O'Neill, E. B., Athavale, S. S., Petrov, A. S., Hud, N. V., Wartell, R. M., Harvey, S. C. & Williams, L. D. (2013). RNA with iron(II) as a cofactor catalyses electron transfer. *Nature Chemistry* **5**, 525–528.

Hsiao, C., Mohan, S., Kalahar, B. K. & Williams, L. D. (2009). Peeling the onion: ribosomes are ancient molecular fossils. *Molecular Biology and Evolution* **26**, 2415–2425.

Huang, F., Bugg, C. W. & Yarus, M. (2000). RNA-Catalyzed CoA, NAD, and FAD synthesis from phosphopantetheine, NMN, and FMN. *Biochemistry* **39**, 15548–15555.

Huang, P. J., Vazin, M. & Liu, J. (2014). *In vitro* selection of a new lanthanide-dependent DNAzyme for ratiometric sensing lanthanides. *Analytical Chemistry* **86**, 9993–9999.

Huang, W. & Ferris, J. P. (2006). One-step, regioselective synthesis of up to 50-mers of RNA oligomers by montmorillonite catalysis. *Journal of the American Chemical Society* **128**, 8914–8919.

Hud, N. V., Cafferty, B. J., Krishnamurthy, R. & Williams, L. D. (2013). The origin of RNA and "my grandfather's axe". *Chemistry & Biology* **20**, 466–474.

Ichihashi, N., Usui, K., Kazuta, Y., Sunami, T., Matsuura, T. & Yomo, T. (2013). Darwinian evolution in a translation-coupled RNA replication system within a cell-like compartment. *Nature Communication* **4**, 2494.

Ikawa, Y., Tsuda, K., Matsumura, S. & Inoue, T. (2004). *De novo* synthesis and development of an RNA enzyme. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 13750–13755.

Illangasekare, M., Sanchez, G., Nickles, T. & Yarus, M. (1995). Aminoacyl-RNA synthesis catalyzed by an RNA. *Science* **267**, 643–647.

Inoue, T. & Orgel, L. E. (1982). Oligomerization of (guanosine 5′-phosphor)-2-methylimidazolide on poly(C). An RNA polymerase model. *Journal of Molecular Biology* **162**, 201–217.

Jansen, J. A., Mccarthy, T. J., Soukup, G. A. & Soukup, J. K. (2006). Backbone and nucleobase contacts to glucosamine-6-phosphate in the glmS ribozyme. *Nature Structural & Molecular Biology* **13**, 517–523.

Jauker, M., Griesser, H. & Richert, C. (2015). Spontaneous Formation of RNA Strands, Peptidyl RNA, and Cofactors. *Angewandte Chemie (International ed. in English)* **54**, 14564–14569.

Javadi-Zarnaghi, F. & Hobartner, C. (2013). Lanthanide cofactors accelerate DNA-catalyzed synthesis of branched RNA. *Journal of the American Chemical Society* **135**, 12839–12848.

Jia, T. Z., Fahrenbach, A. C., Kamat, N. P., Adamala, K. P. & Szostak, J. W. (2016). Oligoarginine peptides slow strand annealing and assist non-enzymatic RNA replication. *Nature Chemistry* **8**, 915–921.

Jimenez, 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. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 14984–14989.

Jimenez, R. M., Polanco, J. A. & Luptak, A. (2015). Chemistry and biology of self-cleaving ribozymes. *Trends in Biochemical Sciences* **40**, 648–661.

Johnston, W. K., Unrau, P. J., Lawrence, M. S., Glasner, M. E. & Bartel, D. P. (2001). RNA-catalyzed RNA polymerization: accurate and general RNA-templated primer extension. *Science* **292**, 1319–1325.

Kamat, N. P., Tobe, S., Hill, I. T. & Szostak, J. W. (2015). Electrostatic localization of RNA to protocell membranes by cationic hydrophobic peptides. *Angewandte Chemie (International ed. in English)* **54**, 11735–11739.

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

Kath-Schorr, S., Wilson, T. J., Li, N. S., Lu, J., Piccirilli, J. A. & Lilley, D. M. (2012). General acid–base catalysis mediated by nucleobases in the hairpin ribozyme. *Journal of the American Chemical Society* **134**, 16717–16724.

Kauffman, S. A. (1996). *At Home in the Universe: The Search for the Laws of Self-Organization and Complexity*. Oxford University Press, Oxford.

Kazantsev, A. V., Krivenko, A. A., Harrington, D. J., Holbrook, S. R., Adams, P. D. & Pace, N. R. (2005). Crystal structure of a bacterial ribonuclease P RNA. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13392–13397.

Ke, A., Zhou, K., Ding, F., Cate, J. H. & Doudna, J. A. (2004). A conformational switch controls hepatitis delta virus ribozyme catalysis. *Nature* **429**, 201–205.

Kervio, E., Sosson, M. & Richert, C. (2016). The effect of leaving groups on binding and reactivity in enzyme-free copying of DNA and RNA. *Nucleic Acids Research* **44**, 5504–5514.

Khersonsky, O. & Tawfik, D. S. (2010). Enzyme promiscuity: a mechanistic and evolutionary perspective. *Annual Review of Biochemistry* **79**, 471–505.

Kim, H. K., Rasnik, I., Liu, J., Ha, T. & Lu, Y. (2007). Dissecting metal ion-dependent folding and catalysis of a single DNAzyme. *Nature Chemical Biology* **3**, 763–768.

Klein, D. J., Been, M. D. & Ferre-D'amare, A. R. (2007). Essential role of an active-site guanine in glmS ribozyme catalysis. *Journal of the American Chemical Society* **129**, 14858–14859.

Klein, D. J. & Ferre-D'amare, A. R. (2006). Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science* **313**, 1752–1756.

Koboldt, D. C., Steinberg, K. M., Larson, D. E., Wilson, R. K. & Mardis, E. R. (2013). The next-generation sequencing revolution and its impact on genomics. *Cell* **155**, 27–38.

Kobori, S., Nomura, Y., Miu, A. & Yokobayashi, Y. (2015). High-throughput assay and engineering of self-cleaving ribozymes by sequencing. *Nucleic Acids Research* **43**, e85.

Kobori, S. & Yokobayashi, Y. (2016). High-throughput mutational analysis of a twister ribozyme. *Angewandte Chemie (International ed. in English)* **55**, 10354–10357.

Kosutic, M., Neuner, S., Ren, A., Flur, S., Wunderlich, C., Mairhofer, E., Vusurovic, N., Seikowski, J., Breuker, K., Hobartner, C., Patel, D. J., Kreutz, C. & Micura, R. (2015). A mini-twister variant and impact of residues/cations on the phosphodiester cleavage of this ribozyme class. *Angewandte Chemie (International ed. in English)* **54**, 15128–15133.

Kozlov, I. A., De Bouvere, B., Van Aerschot, A., Herdewijn, P. & Orgel, L. E. (1999a). Efficient transfer of information from hexitol nucleic acids to RNA during nonenzymatic oligomerization. *Journal of the American Chemical Society* **121**, 5856–5859.

Kozlov, I. A., Politis, P. K., Van Aerschot, A., Busson, R., Herdewijn, P. & Orgel, L. E. (1999b). Nonenzymatic synthesis of RNA and DNA oligomers on hexitol nucleic acid templates: the importance of the A structure. *Journal of the American Chemical Society* **121**, 2653–2656.

Kozlov, I. A., Zielinski, M., Allart, B., Kerremans, L., Van Aerschot, A., Busson, R., Herdewijn, P. & Orgel, L. E. (2000). Nonenzymatic template-directed reactions on altritol oligomers, preorganized analogues of oligonucleotides. *Chemistry–A European Journal* **6**, 151–155.

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* **7**, 203–208.

Kruger, K., Grabowski, P. J., 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**, 147–157.

Kumar, R. K. & Yarus, M. (2001). RNA-catalyzed amino acid activation. *Biochemistry* **40**, 6998–7004.

Kun, A., Santos, M. & Szathmary, E. (2005). Real ribozymes suggest a relaxed error threshold. *Nature Genetics* **37**, 1008–1011.

Lam, B. J. & Joyce, G. F. (2009). Autocatalytic aptazymes enable ligand-dependent exponential amplification of RNA. *Nature Biotechnology* **27**, 288–292.

Leamy, K. A., Assmann, S. M., Mathews, D. H. & Bevilacqua, P. C. (2016). Bridging the gap between *in vitro* and *in vivo* RNA folding. *Quarterly Reviews of Biophysics* **49**, e10.

Lee, C. S., Mui, T. P. & Silverman, S. K. (2011). Improved deoxyribozymes for synthesis of covalently branched DNA and RNA. *Nucleic Acids Research* **39**, 269–279.

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

Lee, N., Bessho, Y., Wei, K., Szostak, J. W. & Suga, H. (2000). Ribozyme-catalyzed tRNA aminoacylation. *Nature Structural Biology* **7**, 28–33.

Legault, P. & Pardi, A. (1997). Unusual dynamics and pKa shift at the active site of a lead-dependent ribozyme. *Journal of the American Chemical Society* **119**, 6621–6628.

Lescrinier, E., Esnouf, R., Schraml, J., Busson, R., Heus, H., Hilbers, C. & Herdewijn, P. (2000). Solution structure of a HNA–RNA hybrid. *Chemistry & Biology* **7**, 719–731.

Leu, K., Kervio, E., Obermayer, B., Turk-Macleod, R. M., Yuan, C., Luevano, Jr., J. M., Chen, E., Gerland, U., Richert, C. & Chen, I. A. (2013). Cascade of reduced speed and accuracy after errors in enzyme-free copying of nucleic acid sequences. *Journal of the American Chemical Society* **135**, 354–366.

Li, L. & Szostak, J. W. (2014). The free energy landscape of pseudorotation in 3′–5′ and 2′–5′ linked nucleic acids. *Journal of the American Chemical Society* **136**, 2858–2865.

Li, P., Sergueeva, Z. A., Dobrikov, M. & Shaw, B. R. (2007). Nucleoside and oligonucleoside boranophosphates: chemistry and properties. *Chemical Reviews* **107**, 4746–4796.

Li, S., Lunse, C. E., Harris, K. A. & Breaker, R. R. (2015). Biochemical analysis of hatchet self-cleaving ribozymes. *RNA* **21**, 1845–1851.

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

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

Li, X., Zhan, Z. Y., Knipe, R. & Lynn, D. G. (2002). DNA-catalyzed polymerization. *Journal of the American Chemical Society* **124**, 746–747.
Li, Y. & Breaker, R. R. (1999). Phosphorylating DNA with DNA. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 2746–2751.
Li, Y., Liu, Y. & Breaker, R. R. (2000). Capping DNA with DNA. *Biochemistry* **39**, 3106–3114.
Li, Y. & Sen, D. (1996). A catalytic DNA for porphyrin metallation. *Nature Structural Biology* **3**, 743–747.
Lilley, D. M. (2005). Structure, folding and mechanisms of ribozymes. *Current Opinion in Structural Biology* **15**, 313–323.
Lilley, D. M. (2011). Catalysis by the nucleolytic ribozymes. *Biochemical Society Transactions* **39**, 641–646.
Lilley, D. M. J. & Eckstein, F. (2008). *Ribozymes and RNA Catalysis*. Cambridge: The Royal Society of Chemistry.
Lincoln, T. A. & Joyce, G. F. (2009). Self-sustained replication of an RNA enzyme. *Science* **323**, 1229–1232.
Lipfert, J., Doniach, S., Das, R. & Herschlag, D. (2014). Understanding nucleic acid-ion interactions. *Annual Review of Biochemistry* **83**, 813–841.
Liu, J. (2015). Lanthanide-dependent RNA-cleaving DNAzymes as metal biosensors. *Canadian Journal of Chemistry* **93**, 273–278.
Liu, Y. & Sen, D. (2010). Local rather than global folding enables the lead-dependent activity of the 8–17 deoxyribozyme: evidence from contact photo-crosslinking. *Journal of Molecular Biology* **395**, 234–241.
Liu, Y., Wilson, T. J., Mcphee, S. A. & Lilley, D. M. (2014). Crystal structure and mechanistic investigation of the twister ribozyme. *Nature Chemical Biology* **10**, 739–744.
Liu, Z., Mei, S. H., Brennan, J. D. & Li, Y. (2003). Assemblage of signaling DNA enzymes with intriguing metal-ion specificities and pH dependences. *Journal of the American Chemical Society* **125**, 7539–7545.
Lohse, P. A. & Szostak, J. W. (1996). Ribozyme-catalysed amino-acid transfer reactions. *Nature* **381**, 442–444.
Lorsch, J. R. & Szostak, J. W. (1994). *In vitro* evolution of new ribozymes with polynucleotide kinase activity. *Nature* **371**, 31–36.
Luther, A., Brandsch, R. & von Kiedrowski, G. (1998). Surface-promoted replication and exponential amplification of DNA analogues. *Nature* **396**, 245–248.
Malyshev, D. A., Dhami, K., Lavergne, T., Chen, T., Dai, N., Foster, J. M., Correa, Jr., I. R. & Romesberg, F. E. (2014). A semi-synthetic organism with an expanded genetic alphabet. *Nature* **509**, 385–388.
Mansy, S. S., Schrum, J. P., Krishnamurthy, M., Tobe, S., Treco, D. A. & Szostak, J. W. (2008). Template-directed synthesis of a genetic polymer in a model protocell. *Nature* **454**, 122–125.
Mansy, S. S. & Szostak, J. W. (2008). Thermostability of model protocell membranes. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 13351–13355.
Martick, M. & Scott, W. G. (2006). Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* **126**, 309–320.
Martin, L. L., Unrau, P. J. & Muller, U. F. (2015). RNA synthesis by *in vitro* selected ribozymes for recreating an RNA world. *Life (Basel)* **5**, 247–268.
Micklefield, J. (2001). Backbone modification of nucleic acids: synthesis, structure and therapeutic applications. *Current Medicinal Chemistry* **8**, 1157–1179.
Mir, A. & Golden, B. L. (2016). Two active site divalent ions in the crystal structure of the hammerhead ribozyme bound to a transition state analogue. *Biochemistry* **55**, 633–636.
Mondragon, A. (2013). Structural studies of RNase P. *Annual Review of Biophysics* **42**, 537–557.
Monnard, P. A., Kanavarioti, A. & Deamer, D. W. (2003). Eutectic phase polymerization of activated ribonucleotide mixtures yields quasi-equimolar incorporation of purine and pyrimidine nucleobases. *Journal of the American Chemical Society* **125**, 13734–13740.
Monnard, P. A. & Szostak, J. W. (2008). Metal-ion catalyzed polymerization in the eutectic phase in water-ice: a possible approach to template-directed RNA polymerization. *Journal of Inorganic Biochemistry* **102**, 1104–1111.
Moretti, J. E. & Muller, U. F. (2014). A ribozyme that triphosphorylates RNA 5′-hydroxyl groups. *Nucleic Acids Research* **42**, 4767–4778.
Morimoto, J., Hayashi, Y., Iwasaki, K. & Suga, H. (2011). Flexizymes: their evolutionary history and the origin of catalytic function. *Accounts of Chemical Research* **44**, 1359–1368.
Muller, S. (2015). Engineering of ribozymes with useful activities in the ancient RNA world. *Annals of the New York Academy of Sciences* **1341**, 54–60.
Mungi, C. V. & Rajamani, S. (2015). Characterization of RNA-like oligomers from lipid-assisted nonenzymatic synthesis: implications for origin of informational molecules on early earth. *Life (Basel)* **5**, 65–84.
Murray, J. B., Seyhan, A. A., Walter, N. G., Burke, J. M. & Scott, W. G. (1998). The hammerhead, hairpin and VS ribozymes are catalytically proficient in monovalent cations alone. *Chemistry & Biology* **5**, 587–595.
Mustoe, A. M., Brooks, C. L. & Al-Hashimi, H. M. (2014). Hierarchy of RNA functional dynamics. *Annual Review of Biochemistry* **83**, 441–466.
Mutschler, H., Wochner, A. & Holliger, P. (2015). Freeze-thaw cycles as drivers of complex ribozyme assembly. *Nature Chemistry* **7**, 502–508.
Nakano, S., Chadalavada, D. M. & Bevilacqua, P. C. (2000). General acid–base catalysis in the mechanism of a hepatitis delta virus ribozyme. *Science* **287**, 1493–1497.
Namani, T. & Deamer, D. W. (2008). Stability of model membranes in extreme environments. *Origins of Life and Evolution of the Biosphere* **38**, 329–341.

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

Nauwelaerts, K., Fisher, M., Froeyen, M., Lescrinier, E., Aerschot, A. V., Xu, D., Delong, R., Kang, H., Juliano, R. L. & Herdewijn, P. (2007). Structural characterization and biological evaluation of small interfering RNAs containing cyclohexenyl nucleosides. *Journal of the American Chemical Society* **129**, 9340–9348.

Neveu, M., Kim, H. J. & Benner, S. A. (2013). The "strong" RNA world hypothesis: fifty years old. *Astrobiology* **13**, 391–403.

Nguyen, T. H., Galej, W. P., Bai, X. C., Savva, C. G., Newman, A. J., Scheres, S. H. & Nagai, K. (2015). The architecture of the spliceosomal U4/U6.U5 tri-snRNP. *Nature* **523**, 47–52.

Nguyen, T. H., Galej, W. P., Fica, S. M., Lin, P. C., Newman, A. J. & Nagai, K. (2016). CryoEM structures of two spliceosomal complexes: starter and dessert at the spliceosome feast. *Current Opinion in Structural Biology* **36**, 48–57.

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**, 920–930.

Nielsen, P. E. (1995). DNA analogues with nonphosphodiester backbones. *Annual Review of Biophysics and Biomolecular Structure* **24**, 167–183.

Nielsen, P. E. (2007). Peptide nucleic acids and the origin of life. *Chemistry & Biodiversity* **4**, 1996–2002.

Nogales, E. & Scheres, S. H. (2015). Cryo-EM: a unique tool for the visualization of macromolecular complexity. *Molecular Cell* **58**, 677–689.

Noller, H. F., Hoffarth, V. & Zimniak, L. (1992). Unusual resistance of peptidyl transferase to protein extraction procedures. *Science* **256**, 1416–1419.

Oberholzer, T., Albrizio, M. & Luisi, P. L. (1995a). Polymerase chain reaction in liposomes. *Chemistry & Biology* **2**, 677–682.

Oberholzer, T., Wick, R., Luisi, P. L. & Biebricher, C. K. (1995b). Enzymatic RNA replication in self-reproducing vesicles: an approach to a minimal cell. *Biochemical and Biophysical Research Communications* **207**, 250–257.

Orgel, L. E. (2008). The implausibility of metabolic cycles on the prebiotic Earth. *PLoS Biology* **6**, e18.

O'Rourke, S. M., Estell, W. & Scott, W. G. (2015). Minimal hammerhead ribozymes with uncompromised catalytic activity. *Journal of Molecular Biology* **427**, 2340–2347.

Pace, N. R. & Marsh, T. L. (1985). RNA catalysis and the origin of life. *Origins of Life and Evolution of the Biosphere* **16**, 97–116.

Parker, D. J., Xiao, Y., Aguilar, J. M. & Silverman, S. K. (2013). DNA catalysis of a normally disfavored RNA hydrolysis reaction. *Journal of the American Chemical Society* **135**, 8472–8475.

Patel, B. H., Percivalle, C., Ritson, D. J., Duffy, C. D. & Sutherland, J. D. (2015). Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nature Chemistry* **7**, 301–307.

Paul, N. & Joyce, G. F. (2002). A self-replicating ligase ribozyme. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 12733–12740.

Paul, N., Springsteen, G. & Joyce, G. F. (2006). Conversion of a ribozyme to a deoxyribozyme through *in vitro* evolution. *Chemistry & Biology* **13**, 329–338.

Perreault, J. P., Wu, T. F., Cousineau, B., Ogilvie, K. K. & Cedergren, R. (1990). Mixed deoxyribo- and ribo-oligonucleotides with catalytic activity. *Nature* **344**, 565–567.

Perrin, D. M., Garestier, T. & Helene, C. (2001). Bridging the gap between proteins and nucleic acids: a metal-independent RNAseA mimic with two protein-like functionalities. *Journal of the American Chemical Society* **123**, 1556–1563.

Peselis, A. & Serganov, A. (2014). Themes and variations in riboswitch structure and function. *Biochimica et Biophysica Acta* **1839**, 908–918.

Petrie, K. L. & Joyce, G. F. (2014). Limits of neutral drift: lessons from the *in vitro* evolution of two ribozymes. *Journal of Molecular Evolution* **79**, 75–90.

Pfeiffer, F. & Mayer, G. (2016). Selection and biosensor application of aptamers for small molecules. *Frontiers in Chemistry* **4**, 25.

Pinheiro, V. B. & Holliger, P. (2012). The XNA world: progress towards replication and evolution of synthetic genetic polymers. *Current Opinion in Chemical Biology* **16**, 245–252.

Pinheiro, V. B. & Holliger, P. (2014). Towards XNA nanotechnology: new materials from synthetic genetic polymers. *Trends in Biotechnology* **32**, 321–328.

Pinheiro, V. B., Loakes, D. & Holliger, P. (2013). Synthetic polymers and their potential as genetic materials. *BioEssays* **35**, 113–122.

Pinheiro, V. B., Taylor, A. I., Cozens, C., Abramov, M., Renders, M., Zhang, S., Chaput, J. C., Wengel, J., Peak-Chew, S. Y., Mclaughlin, S. H., Herdewijn, P. & Holliger, P. (2012). Synthetic genetic polymers capable of heredity and evolution. *Science* **336**, 341–344.

Pitt, J. N. & Ferre-D'amare, A. R. (2010). Rapid construction of empirical RNA fitness landscapes. *Science* **330**, 376–379.

Ponce-Salvatierra, A., Wawrzyniak-Turek, K., Steuerwald, U., Hobartner, C. & Pena, V. (2016). Crystal structure of a DNA catalyst. *Nature* **529**, 231–234.

Powner, M. W., Gerland, B. & Sutherland, J. D. (2009). Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242.

Pradeepkumar, P. I., Höbartner, C., Baum, D. A. & Silverman, S. K. (2008). DNA-catalyzed formation of nucleopeptide linkages. *Angewandte Chemie(International ed. in English)* **47**, 1753–1757.

Premraj, B. J. & Yathindra, N. (1998). Stereochemistry of 2′,5′ nucleic acids and their constituents. *Journal of Biomolecular Structure and Dynamics* **16**, 313–328.

Pressman, A., Blanco, C. & Chen, I. A. (2015). The RNA world as a model system to study the origin of life. *Current Biology* **25**, 953–963.

Prywes, N., Blain, J. C., Del Frate, F. & Szostak, J. W. (2016a). Nonenzymatic copying of RNA templates containing all four letters is catalyzed by activated oligonucleotides. *Elife* **5**, 1–14.

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

Prywes, N., Michaels, Y. S., Pal, A., Oh, S. S. & Szostak, J. W. (2016b). Thiolated uridine substrates and templates improve the rate and fidelity of ribozyme-catalyzed RNA copying. *Chemical Communications (Cambridge)* **52**, 6529–6532.

Purtha, W. E., Coppins, R. L., Smalley, M. K. & Silverman, S. K. (2005). General deoxyribozyme-catalyzed synthesis of native 3′–5′ RNA linkages. *Journal of the American Chemical Society* **127**, 13124–13125.

Pyle, A. M. (2016). Group II intron self-splicing. *Annual Review of Biophysics* **45**, 183–205.

Rasmussen, S., Constantinescu, A. & Svaneborg, C. (2016). Generating minimal living systems from non-living materials and increasing their evolutionary abilities. *Philosophical Transactions of the Royal Society of London B Biological Sciences* **371**, 1–10.

Reader, J. S. & Joyce, G. F. (2002). A ribozyme composed of only two different nucleotides. *Nature* **420**, 841–844.

Ren, A., Kosutic, M., Rajashankar, K. R., Frener, M., Santner, T., Westhof, E., Micura, R. & Patel, D. J. (2014). In-line alignment and Mg^{2+} coordination at the cleavage site of the env22 twister ribozyme. *Nature Communication* **5**, 5534.

Ricardo, A., Carrigan, M. A., Olcott, A. N. & Benner, S. A. (2004). Borate minerals stabilize ribose. *Science* **303**, 196.

Roberts, R. W. & Szostak, J. W. (1997). RNA-peptide fusions for the *in vitro* selection of peptides and proteins. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 12297–12302.

Robertson, D. L. & Joyce, G. F. (1990). Selection *in vitro* of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature* **344**, 467–468.

Robertson, M. P. & Joyce, G. F. (2014). Highly efficient self-replicating RNA enzymes. *Chemistry and Biology* **21**, 238–245.

Robertson, M. P., Knudsen, S. M. & Ellington, A. D. (2004). *In vitro* selection of ribozymes dependent on peptides for activity. *RNA* **10**, 114–127.

Rogers, J. & Joyce, G. F. (1999). A ribozyme that lacks cytidine. *Nature* **402**, 323–325.

Rogers, J. & Joyce, G. F. (2001). The effect of cytidine on the structure and function of an RNA ligase ribozyme. *RNA* **7**, 395–404.

Rosenbaum, D. M. & Liu, D. R. (2003). Efficient and sequence-specific DNA-templated polymerization of peptide nucleic acid aldehydes. *Journal of the American Chemical Society* **125**, 13924–13925.

Roth, A. & Breaker, R. R. (1998). An amino acid as a cofactor for a catalytic polynucleotide. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 6027–6031.

Roth, A., Weinberg, Z., Chen, A. G., Kim, P. B., Ames, T. D. & Breaker, R. R. (2014). A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nature Chemical Biology* **10**, 56–60.

Rupert, P. B. & Ferre-D'amare, A. R. (2001). Crystal structure of a hairpin ribozyme–inhibitor complex with implications for catalysis. *Nature* **410**, 780–786.

Sacerdote, M. G. & Szostak, J. W. (2005). Semipermeable lipid bilayers exhibit diastereoselectivity favoring ribose. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 6004–6008.

Sachdeva, A. & Silverman, S. K. (2010). DNA-catalyzed serine side chain reactivity and selectivity. *Chemical Communications (Cambridge)* **46**, 2215–2217.

Saenger, W. & Egli, M. (1984). *Principles of Nucleic Acid Structure*. Springer, New York.

Salehi-Ashtiani, K., Luptak, A., Litovchick, A. & Szostak, J. W. (2006). A genomewide search for ribozymes reveals an HDV-like sequence in the human CPEB3 gene. *Science* **313**, 1788–1792.

Salehi-Ashtiani, K. & Szostak, J. W. (2001). *In vitro* evolution suggests multiple origins for the hammerhead ribozyme. *Nature* **414**, 82–84.

Santoro, S. W. & Joyce, G. F. (1997). A general purpose RNA-cleaving DNA enzyme. *Proceedings of the National Academy of Sciences of the United States of America* **94**, 4262–4266.

Santoro, S. W., Joyce, G. F., Sakthivel, K., Gramatikova, S. & Barbas, III, C. F. (2000). RNA cleavage by a DNA enzyme with extended chemical functionality. *Journal of the American Chemical Society* **122**, 2433–2439.

Schlosser, K. & Li, Y. (2009). DNAzyme-mediated catalysis with only guanosine and cytidine nucleotides. *Nucleic Acids Research* **37**, 413–420.

Schlosser, K. & Li, Y. (2010). A versatile endoribonuclease mimic made of DNA: characteristics and applications of the 8–17 RNA-cleaving DNAzyme. *Chembiochem* **11**, 866–879.

Schmeing, T. M. & Ramakrishnan, V. (2009). What recent ribosome structures have revealed about the mechanism of translation. *Nature* **461**, 1234–1242.

Schoning, K., Scholz, P., Guntha, S., Wu, X., Krishnamurthy, R. & Eschenmoser, A. (2000). Chemical etiology of nucleic acid structure: the alpha-threofuranosyl-(3′→2′) oligonucleotide system. *Science* **290**, 1347–1351.

Schultes, E. A. & Bartel, D. P. (2000). One sequence, two ribozymes: implications for the emergence of new ribozyme folds. *Science* **289**, 448–452.

Scott, W. G., Finch, J. T. & Klug, A. (1995). The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. *Cell* **81**, 991–1002.

Sczepanski, J. T. & Joyce, G. F. (2012). Synthetic evolving systems that implement a user-specified genetic code of arbitrary design. *Chemistry & Biology* **19**, 1324–1332.

Sczepanski, J. T. & Joyce, G. F. (2014). A cross-chiral RNA polymerase ribozyme. *Nature* **515**, 440–442.

Seelig, B. & Jaschke, A. (1999). A small catalytic RNA motif with Diels-Alderase activity. *Chemistry & Biology* **6**, 167–176.

Segre, D. & Lancet, D. (2000). Composing life. *EMBO Reports* **1**, 217–222.

Sengle, G., Eisenfuhr, A., Arora, P. S., Nowick, J. S. & Famulok, M. (2001). Novel RNA catalysts for the Michael reaction. *Chemistry & Biology* **8**, 459–473.

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

Serganov, A., Keiper, S., Malinina, L., Tereshko, V., Skripkin, E., Hobartner, C., Polonskaia, A., Phan, A. T., Wombacher, R., Micura, R., Dauter, Z., Jaschke, A. & Patel, D. J. (2005). Structural basis for Diels-Alder ribozyme-catalyzed carbon-carbon bond formation. *Nature Structural & Molecular Biology* **12**, 218–224.

Serganov, A. & Nudler, E. (2013). A decade of riboswitches. *Cell* **152**, 17–24.

Sharma, C. & Awasthi, S. K. (2016). Versatility of peptide nucleic acids (PNAs): role in chemical biology, drug discovery, and origins of life. *Chemical Biology & Drug Design* **89**, 16–37.

Sheng, J., Li, L., Engelhart, A. E., Gan, J., Wang, J. & Szostak, J. W. (2014). Structural insights into the effects of 2′−5′ linkages on the RNA duplex. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 3050–3055.

Sheppard, T. L., Ordoukhanian, P. & Joyce, G. F. (2000). A DNA enzyme with N-glycosylase activity. *Proceedings of the National Academy of Sciences of the United States of America* **97**, 7802–7807.

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K. & Ueda, T. (2001). Cell-free translation reconstituted with purified components. *Nature Biotechnology* **19**, 751–755.

Sidorov, A. V., Grasby, J. A. & Williams, D. M. (2004). Sequence-specific cleavage of RNA in the absence of divalent metal ions by a DNAzyme incorporating imidazolyl and amino functionalities. *Nucleic Acids Research* **32**, 1591–1601.

Sievers, D. & von Kiedrowski, G. (1994). Self-replication of complementary nucleotide-based oligomers. *Nature* **369**, 221–224.

Sigel, A., Sigel, H. & Sigel, K. O. (2012). *Interplay between Metal Ions and Nucleic Acids*, Springer, Dordrecht, Heidelberg, London, New York.

Silverman, S. K. (2005). *In vitro* selection, characterization, and application of deoxyribozymes that cleave RNA. *Nucleic Acids Research* **33**, 6151–6163.

Silverman, S. K. (2008). *Nucleic Acid Enzymes (Ribozymes and Deoxyribozymes): In Vitro Selection and Application*. Wiley Encyclopedia of Chemical Biology, Hoboken, New Jersey.

Silverman, S. K. (2009). Deoxyribozymes: selection design and serendipity in the development of DNA catalysts. *Accounts of Chemical Research* **42**, 1521–1531.

Silverman, S. K. (2015). Pursuing DNA catalysts for protein modification. *Accounts of Chemical Research* **48**, 1369–1379.

Silverman, S. K. (2016). Catalytic DNA: scope, applications, and biochemistry of deoxyribozymes. *Trends in Biochemical Sciences* **41**, 595–609.

Smith, T. F., Lee, J. C., Gutell, R. R. & Hartman, H. (2008). The origin and evolution of the ribosome. *Biology Direct* **3**, 16.

Sreedhara, A., Li, Y. & Breaker, R. R. (2004). Ligating DNA with DNA. *Journal of the American Chemical Society* **126**, 3454–3460.

Stoeger, T., Battich, N. & Pelkmans, L. (2016). Passive noise filtering by cellular compartmentalization. *Cell* **164**, 1151–1161.

Strulson, C. A., Molden, R. C., Keating, C. D. & Bevilacqua, P. C. (2012). RNA catalysis through compartmentalization. *Nature Chemistry* **4**, 941–946.

Sullenger, B. A. & Nair, S. (2016). From the RNA world to the clinic. *Science* **352**, 1417–1420.

Suslov, N. B., Dasgupta, S., Huang, H., Fuller, J. R., Lilley, D. M., Rice, P. A. & Piccirilli, J. A. (2015). Crystal structure of the Varkud satellite ribozyme. *Nature Chemical Biology* **11**, 840–846.

Sutherland, J. D. (2016). The origin of life – out of the blue. *Angewandte Chemie (International ed. in English)* **55**, 104–121.

Szabo, P., Scheuring, I., Czaran, T. & Szathmary, E. (2002). *In silico* simulations reveal that replicators with limited dispersal evolve towards higher efficiency and fidelity. *Nature* **420**, 340–343.

Szostak, J. W. (2003). Functional information: molecular messages. *Nature* **423**, 689.

Szostak, J. W., Bartel, D. P. & Luisi, P. L. (2001). Synthesizing life. *Nature* **409**, 387–390.

Tagami, S., Attwater, J. & Holliger, P. (2017). Simple peptides derived from the ribosomal core potentiate RNA polymerase ribozyme function. *Nature Chemistry*, doi:10.1038/nchem.2739.

Tarasow, T. M., Tarasow, S. L. & Eaton, B. E. (1997). RNA-catalysed carbon-carbon bond formation. *Nature* **389**, 54–57.

Taylor, A. I., Beuron, F., Peak-Chew, S. Y., Morris, E. P., Herdewijn, P. & Holliger, P. (2016). Nanostructures from synthetic genetic polymers. *Chembiochem* **17**, 1107–1110.

Taylor, A. I., Pinheiro, V. B., Smola, M. J., Morgunov, A. S., Peak-Chew, S., Cozens, C., Weeks, K. M., Herdewijn, P. & Holliger, P. (2015). Catalysts from synthetic genetic polymers. *Nature* **518**, 427–430.

Teramoto, N., Imanishi, Y. & Ito, Y. (2000). *In vitro* selection of a ligase ribozyme carrying alkylamino groups in the side chains. *Bioconjugate Chemistry* **11**, 744–748.

Thorne, R. E., Chinnapen, D. J., Sekhon, G. S. & Sen, D. (2009). A deoxyribozyme, Sero1C, uses light and serotonin to repair diverse pyrimidine dimers in DNA. *Journal of Molecular Biology* **388**, 21–29.

Toor, N., Keating, K. S., Taylor, S. D. & Pyle, A. M. (2008). Crystal structure of a self-spliced group II intron. *Science* **320**, 77–82.

Torabi, S. F. & Lu, Y. (2015). Identification of the same $Na^{(+)}$-specific DNAzyme motif from two *in vitro* selections under different conditions. *Journal of Molecular Evolution* **81**, 225–234.

Torabi, S. F., Wu, P., Mcghee, C. E., Chen, L., Hwang, K., Zheng, N., Cheng, J. & Lu, Y. (2015). *In vitro* selection of a sodium-specific DNAzyme and its application in intracellular sensing. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 5903–5908.

Tsukiji, S., Pattnaik, S. B. & Suga, H. (2003). An alcohol dehydrogenase ribozyme. *Nature Structural Biology* **10**, 713–717.

Tuerk, C. & Gold, L. (1990). Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510.

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

Turk, R. M., Chumachenko, N. V. & Yarus, M. (2010). Multiple translational products from a five-nucleotide ribozyme. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 4585–4589.

Turk, R. M., Illangasekare, M. & Yarus, M. (2011). Catalyzed and spontaneous reactions on ribozyme ribose. *Journal of the American Chemical Society* **133**, 6044–6050.

Unrau, P. J. & Bartel, D. P. (1998). RNA-catalysed nucleotide synthesis. *Nature* **395**, 260–263.

Ura, Y., Beierle, J. M., Leman, L. J., Orgel, L. E. & Ghadiri, M. R. (2009). Self-assembling sequence-adaptive peptide nucleic acids. *Science* **325**, 73–77.

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**, 72–77.

Vasas, V., Fernando, C., Santos, M., Kauffman, S. & Szathmáry, E. (2012). Evolution before genes. *Biology Direct* **7**, 1–14.

Verlander, M. S., Lohrmann, R. & Orgel, L. E. (1973). Catalysts for the self-polymerization of adenosine cyclic 2′, 3′-phosphate. *Journal of Molecular Evolution* **2**, 303–316.

Verlander, M. S. & Orgel, L. E. (1974). Analysis of high molecular weight material from the polymerization of adenosine cyclic 2′, 3′-phosphate. *Journal of Molecular Evolution* **3**, 115–120.

Vinothkumar, K. R. & Henderson, R. (2016). Single particle electron cryomicroscopy: trends, issues and future perspective. *Quarterly Reviews of Biophysics* **49**, e13.

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 Research* **32**, 2966–2974.

Voorhees, R. M. & Ramakrishnan, V. (2013). Structural basis of the translational elongation cycle. *Annual Review of Biochemistry* **82**, 203–236.

Wachowius, F., Javadi-Zarnaghi, F. & Höbartner, C. (2010). Combinatorial mutation interference analysis reveals functional nucleotides required for DNA catalysis. *Angewandte Chemie (International ed. in English)* **49**, 8504–8508.

Wahl, M. C., Will, C. L. & Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. *Cell* **136**, 701–718.

Walsh, S. M., Sachdeva, A. & Silverman, S. K. (2013). DNA catalysts with tyrosine kinase activity. *Journal of the American Chemical Society* **135**, 14928–14931.

Walton, T. & Szostak, J. W. (2016). A highly reactive imidazolium-bridged dinucleotide intermediate in nonenzymatic RNA primer extension. *Journal of the American Chemical Society* **138**, 11996–12002.

Wan, R., Yan, C., Bai, R., Wang, L., Huang, M., Wong, C. C. & Shi, Y. (2016). The 3·8 A structure of the U4/U6.U5 tri-snRNP: insights into spliceosome assembly and catalysis. *Science* **351**, 466–475.

Wang, Y. & Silverman, S. K. (2003). Deoxyribozymes that synthesize branched and lariat RNA. *Journal of the American Chemical Society* **125**, 6880–6881.

Wasner, M., Arion, D., Borkow, G., Noronha, A., Uddin, A. H., Parniak, M. A. & Damha, M. J. (1998). Physicochemical and biochemical properties of 2′,5′-linked RNA and 2′,5′-RNA:3′,5′-RNA "hybrid" duplexes. *Biochemistry* **37**, 7478–7486.

Watson, J. D. & Crick, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737–738.

Webb, C. H., Riccitelli, N. J., Ruminski, D. J. & Luptak, A. (2009). Widespread occurrence of self-cleaving ribozymes. *Science* **326**, 953.

Wedekind, J. E. & Mckay, D. B. (1999). Crystal structure of a lead-dependent ribozyme revealing metal binding sites relevant to catalysis. *Nature Structural Biology* **6**, 261–268.

Weinberg, Z., Kim, P. B., Chen, T. H., Li, S., Harris, K. A., Lunse, C. E. & Breaker, R. R. (2015). New classes of self-cleaving ribozymes revealed by comparative genomics analysis. *Nature Chemical Biology* **11**, 606–610.

Westheimer, F. H. (1987). Why nature chose phosphates. *Science* **235**, 1173–1178.

White III, H. B. (1976). Coenzymes as fossils of an earlier metabolic state. *Journal of Molecular Evolution* **7**, 101–104.

Wiegand, T. W., Janssen, R. C. & Eaton, B. E. (1997). Selection of RNA amide synthases. *Chemistry & Biology* **4**, 675–683.

Wilcox, J. L. & Bevilacqua, P. C. (2013). A simple fluorescence method for pK(a) determination in RNA and DNA reveals highly shifted pK(a)'s. *Journal of the American Chemical Society* **135**, 7390–7393.

Wilson, C. & Szostak, J. W. (1995). *In vitro* evolution of a self-alkylating ribozyme. *Nature* **374**, 777–782.

Wilson, D. S. & Szostak, J. W. (1999). *In vitro* selection of functional nucleic acids. *Annual Review of Biochemistry* **68**, 611–647.

Wilson, T. J., Liu, Y., Domnick, C., Kath-Schorr, S. & Lilley, D. M. (2016a). The novel chemical mechanism of the twister ribozyme. *Journal of the American Chemical Society* **138**, 6151–6162.

Wilson, T. J., Liu, Y. & Lilley, D. M. (2016b). Ribozymes and the mechanisms that underlie RNA catalysis. *Frontiers of Chemical Science and Engineering* **10**, 178–185.

Winnacker, M. & Kool, E. T. (2013). Artificial genetic sets composed of size-expanded base pairs. *Angewandte Chemie (International ed. in English)* **52**, 12498–12508.

Wochner, A., Attwater, J., Coulson, A. & Holliger, P. (2011). Ribozyme-catalyzed transcription of an active ribozyme. *Science* **332**, 209–212.

Xiao, Y., Allen, E. C. & Silverman, S. K. (2011). Merely two mutations switch a DNA-hydrolyzing deoxyribozyme from heterobimetallic (Zn^{2+}/Mn^{2+}) to monometallic (Zn^{2+}-only) behavior. *Chemical Communication (Camb)* **47**, 1749–1751.

Xiao, Y., Wehrmann, R. J., Ibrahim, N. A. & Silverman, S. K. (2012). Establishing broad generality of DNA catalysts for site-specific hydrolysis of single-stranded DNA. *Nucleic Acids Research* **40**, 1778–1786.

Yan, C., Hang, J., Wan, R., Huang, M., Wong, C. C. & Shi, Y. (2015). Structure of a yeast spliceosome at 3·6-angstrom resolution. *Science* **349**, 1182–1191.

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press

Zaher, H. S. & Unrau, P. J. (2007). Selection of an improved RNA polymerase ribozyme with superior extension and fidelity. *RNA* **13**, 1017–1026.

Zhang, B. & Cech, T. R. (1997). Peptide bond formation by *in vitro* selected ribozymes. *Nature* **390**, 96–100.

Zhang, F., Nangreave, J., Liu, Y. & Yan, H. (2014). Structural DNA nanotechnology: state of the art and future perspective. *Journal of the American Chemical Society* **136**, 11198–11211.

Zhang, S., Zhang, N., Blain, J. C. & Szostak, J. W. (2013). Synthesis of N3′-P5′-linked phosphoramidate DNA by nonenzymatic template-directed primer extension. *Journal of the American Chemical Society* **135**, 924–932.

Zhou, C., Avins, J. L., Klauser, P. C., Brandsen, B. M., Lee, Y. & Silverman, S. K. (2016a). DNA-catalyzed amide hydrolysis. *Journal of the American Chemical Society* **138**, 2106–2109.

Zhou, W., Zhang, Y., Huang, P. J., Ding, J. & Liu, J. (2016b). A DNAzyme requiring two different metal ions at two distinct sites. *Nucleic Acids Research* **44**, 354–363.

https://doi.org/10.1017/S0033583517000038 Published online by Cambridge University Press