Chemical fidelity of an RNA polymerase ribozyme

Article *in* Chemical Science · July 2013

Impact Factor: 9.21 · DOI: 10.1039/c3sc50574j

9 authors, including:

Eric T Kool
Stanford University
288 PUBLICATIONS **14,114** CITATIONS
SEE PROFILE

Jesper Wengel
University of Southern Denmark
452 PUBLICATIONS **12,009** CITATIONS
SEE PROFILE

Piet Herdewijn
University of Leuven
780 PUBLICATIONS **14,921** CITATIONS
SEE PROFILE

Ichiro Hirao
Institute of Bioengineering and Nanotechn…
163 PUBLICATIONS **3,226** CITATIONS
SEE PROFILE

Chemical Science

RSC Publishing

EDGE ARTICLE

Chemical fidelity of an RNA polymerase ribozyme†

Cite this: *Chem. Sci.*, 2013, **4**, 2804

James Attwater,[a] Shunsuke Tagami,[a] Michiko Kimoto,[b] Kyle Butler,[c] Eric T. Kool,[c] Jesper Wengel,[d] Piet Herdewijn,[e] Ichiro Hirao[b] and Philipp Holliger*[a]

The emergence of catalytically active RNA enzymes (ribozymes) is widely believed to have been an important transition in the origin of life. In the context of a likely heterogeneous chemical environment, substrate specificity and selectivity of these primordial enzymes would have been critical for function. Here we have explored the chemical fidelity, *i.e.* substrate selectivity and specificity for both single and multiple catalytic steps of the Z RNA polymerase ribozyme – a modern day analogue of the primordial RNA replicase. Using a wide range of nucleotide analogues and ionic conditions, we observe strong energetic but weak geometric discrimination at the incorporation step, indicative of an open active site. In contrast, stringent selectivity is exerted at the extension steps through specific down and upstream interactions with the 3′-terminal nucleoside as well as nascent product and template strands. Our results indicate specificity mechanisms that are found in functionally analogous forms in natural polymerases. They also reveal a level of chemical fidelity over multiple catalytic steps that is remarkable for a comparatively unoptimized enzyme developed *de novo* from a random sequence pool. The convergent evolution of specificity mechanisms in phylogenetically unrelated proteinaceous polymerases and polymerase ribozymes suggests that chemical as well as informational fidelity are emergent properties of polymerase enzymes.

Received 28th February 2013
Accepted 3rd May 2013

DOI: 10.1039/c3sc50574j

www.rsc.org/chemicalscience

Introduction

The emergence of catalytic RNA molecules capable of self-replication[1] on the early Earth is thought to have been a key event in the transition from chemistry to biology. Those simple primordial replicators would most likely have been heterotrophic, selecting substrates from the pool of nucleotide feedstock molecules present on the early Earth, instead of synthesising their own. Both prebiotic chemistry as well as the organic chemical inventories of meteoritic impactors suggest that the chemosphere of the early Earth could have presented significant chemical heterogeneity in the composition, purity and abundance of 'feedstock' molecules.[2–4] In such an environment, a putative replicase ribozyme would require not only informational fidelity, *i.e.* an ability to replicate its RNA nucleobase sequences with an accuracy above the error threshold,[1] but also chemical fidelity, an ability to select the correct substrate and stereochemistry for incorporation to preserve the chemical and functional integrity of its 'offspring'. In a similar vein, a nascent replicase might be exposed to and be required to function in a range of heterogeneous reaction media.[5]

Although descendants of the ancestral replicase ribozymes appear to be absent from modern biology, we can study the activity of modern-day equivalents generated by *in vitro* evolution,[6] such as the Z polymerase ribozyme (Fig. 1a).[7] Z is a member of the class of ribozymes derived from the paradigmatic R18 RNA polymerase ribozyme,[8] and is capable of accurate and template-dependent RNA synthesis. Z exhibits more sequence-general activity than some of its relatives,[7] but processivity and pausing still vary between different template sequences and within each template sequence due to its low affinity for the target primer/template duplex.

Here we have sought to assess the challenges facing primordial ribozymes by exploring the chemical fidelity of RNA replication in a processive RNA synthesis context, specifically by investigating the ability of the Z polymerase ribozyme to polymerise modified nucleotides in primer extension reactions and to function in environments of varying salinity, including alternative reaction media such as the eutectic phase of water ice.[9] This screen of chemical space of ribozyme-catalyzed RNA synthesis begins to define the functional tolerance and chemical fidelity range of polymerase ribozymes, uncovering strategies the Z ribozyme employs to achieve synthesis of diverse RNA sequences, with implications for RNA replication in a primordial context.

[a]*MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge, CB2 0QH, UK. E-mail: ph1@mrc-lmb.cam.ac.uk*

[b]*RIKEN Systems and Structural Biology Center (SSBC), 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan*

[c]*Stanford University, Department of Chemistry, Stanford, CA 94305, USA*

[d]*Nucleic Acid Center, Department of Physics, Chemistry and Pharmacy, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark*

[e]*REGA Institute, Katholieke Universiteit Leuven, Minderbroederstraat 10, B 3000, Leuven, Belgium*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c3sc50574j

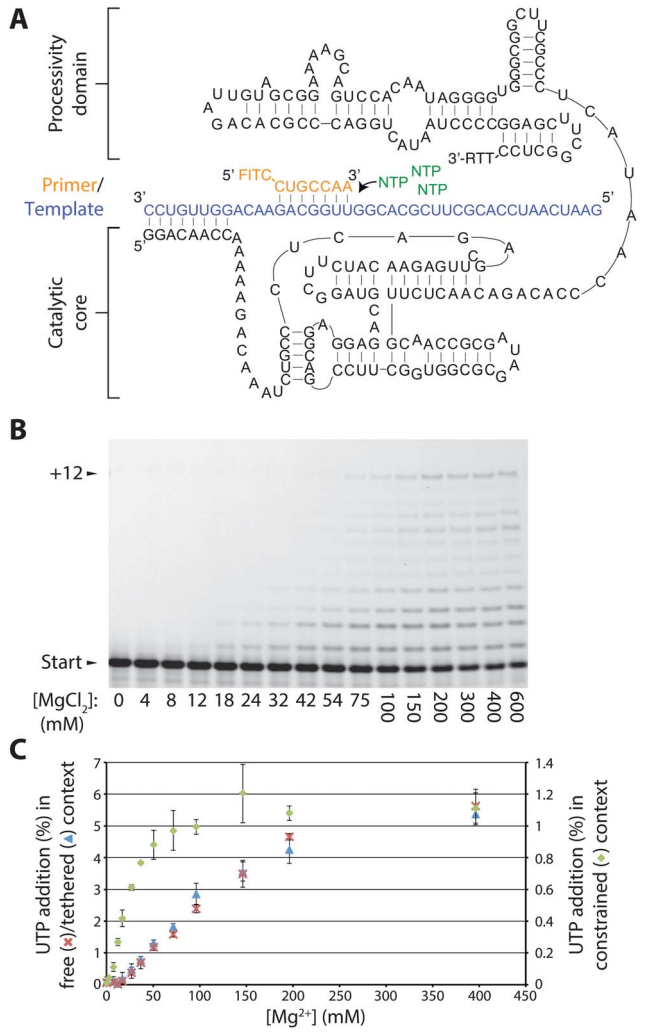

Fig. 1 (A) Secondary structure of the Z polymerase ribozyme. The catalytic core domain is linked to the processivity domain via a flexible single-stranded linker; the ribozyme is depicted surrounding a primer/template duplex (A7/I), to which it is tethered via hybridisation. Nucleoside triphosphate substrates (NTPs, green) are added to the 3′-end of the growing primer. 3′-RTT = 3′-run-though transcript (see ESI†). (B) Denaturing PAGE of primer A10 extended upon template I by Z, at a range of Mg^{2+} concentrations achieved by varying the concentration of $MgCl_2$ in the buffer (17 °C, 1 day, 0.5 mM each NTP). (C) Observed incorporation of a single UTP (at 2 mM; error bars represent s.d., $n = 3$) under varying Mg^{2+} concentrations by the Z ribozyme arranged with the primer/template duplex in three different formats (Fig. S1c in ESI†). Incubation time (at 17 °C) was tailored for each format to approximate initial extension rates; tethered = 25 min, free = 2 h, constrained = 15 s.

Results

The burden of polymerase ribozyme magnesium dependence

One of the key features of the activity of R18-derived polymerase ribozymes is their dependence on high concentrations of magnesium (Mg^{2+}) ions. This 'addiction' to Mg^{2+} has implications not only for activity, but also for ribozyme stability (Mg^{2+} ions catalyse the hydrolysis of the RNA phosphodiester backbone[10]) and replicase protocell compatibility (even low Mg^{2+} concentrations (~4 mM) cause precipitation of membrane-forming fatty acids[11,12]).

We determined the magnesium dependence of the Z polymerase ribozyme by measuring primer extension as a function of Mg^{2+} concentration ($[Mg^{2+}]$). As previously described, our extension assay setup partially compensates for the low affinity of the ribozyme for the primer/template duplex by hybridisation of Z to the template upstream of the primer[9] (Fig. 1a). Such tethering provides a flexible linkage to the duplex, increasing local ribozyme concentration and boosting extension. Nevertheless, the ribozyme must still complete an independent duplex docking event to extend the primer.

We find that RNA primer extension by Z in this context proceeds to ≥12 nucleotides only at high $[Mg^{2+}]$ (>0.2 M) (Fig. 1b). Quantification of the incorporation of a single UTP nucleotide as a function of $[Mg^{2+}]$ (Fig. 1c) and determination of the average number of nucleotides added per primer (Fig. S1a and b in ESI†) both indicate a sigmoidal Mg^{2+} dependence of Z polymerase activity. A similar dependence is observed in a fully *in trans*, *i.e.* 'free' untethered context (Fig. 1c and S1c in ESI†), and persists even after correction for the 1–2 Mg^{2+} ions presumed chelated by each NTP substrate. Significantly, this magnesium dependence of the Z ribozyme differs from that of the class I RNA ligase ribozyme, the precursor and catalytic core of the polymerase ribozyme.[13,14] The catalytic rate of the class I ligase had previously been investigated and showed a Mg^{2+} dependence shaped by a single low-affinity (K_D ~ 70–100 mM) binding site.[15] In contrast, estimates of Hill coefficients for these polymerase ribozymes (~1.7 to 1.9, Fig. S1d in ESI†) at high Mg^{2+} concentrations indicate cooperative contributions to activity (through at least two low-affinity binding sites for functionally important Mg^{2+} ions). This sigmoidal Mg^{2+} dependence implies the existence of additional roles for Mg^{2+} ions in the Z polymerase ribozyme.

The recent crystal structures of the class I ligase, coupled to biochemical interrogation, have provided important insights into ligase function and a basis for the inference of the polymerase ribozyme's likely structure, duplex binding and catalysis.[16,17] For example, in the class I ligase, the sequence region analogous to the template strand in the Z polymerase ribozyme is positioned and constrained by tertiary interactions. In Z, by contrast, the ribozyme must achieve docking with the polyanionic template–primer RNA duplex. To test if it was this requirement for intermolecular docking that impacted Mg^{2+} dependence, we designed a Z variant where the primer/template duplex formed part of ribozyme, to promote constraints through tertiary interactions analogous to the substrate complex in the class I RNA ligase. This constrained complex allows fast nucleotide addition, but is limited to addition of only three nucleotides. When we investigated its RNA polymerase activity $[Mg^{2+}]$ response, this constrained Z ribozyme exhibited a $[Mg^{2+}]$ dependence (Fig. 1c) reminiscent of that of the ligase core, with a Hill coefficient estimation of ≤1 (Fig. S1d in ESI†).[15] This suggests no cooperative contributions of Mg^{2+} to constrained polymerase activity at high Mg^{2+} concentrations: here, activity must be governed by a single low-affinity Mg^{2+} binding site. This differed substantially from the standard Z ribozyme constructs described above – which must dock to the duplex (with or without help from upstream tethering to the

duplex through hybridisation, Fig. S1c in ESI†). Thus we deduce that the Z polymerase dependence on multiple Mg^{2+} ions at high [Mg^{2+}] arises from the requirement for sequence-general docking interactions between the polymerase and the primer/template duplex, and is likely not due to mechanistic differences with the ligase, *e.g.* the need to bind single nucleoside triphosphate molecules in the active site of the polymerase ribozyme. This also suggests that mechanistic details of the phosphotransfer reaction as determined in the structure of the class I ligase core are likely to apply to the derived polymerase ribozymes.[17]

Might a dependence on high [Mg^{2+}] be a prerequisite of polymerase ribozyme function? Charge screening is likely to represent a fundamental mechanistic requirement for two polyanionic polymers (ribozyme and primer–template duplex) to interact specifically without resorting to Watson–Crick base-pairing. In addition to its electrostatic contacts, magnesium's non-electrostatic interactions with phosphate oxygens in the phosphodiester backbone make it well suited amongst metal ions to fulfil such a role.[18] Indeed, when thoroughly investigating the cation dependence of the ligase core, Bartel and colleagues did not find any cation capable of satisfactorily substituting for Mg^{2+} ions.[15]

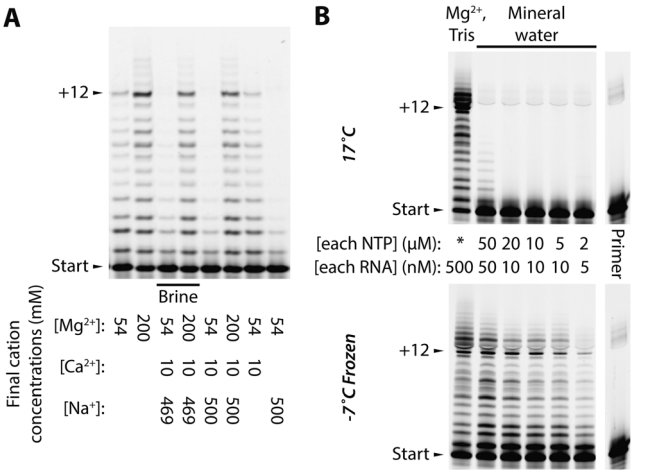

Fig. 2 (A) Extension by Z (17 °C, 2 days) upon the A10/I duplex under different salt regimes. For the brine reactions, the final concentrations of only selected cations are shown. In non-brine reactions, the chloride salts of cations were used. Reactions also contained 50 mM Tris pH 8.3, 0.5 mM each NTP. (B) Extension to completion (36 days) by Z upon the BioA10/I duplex under low salt/nucleotide regimes, both at 17 °C and at −7 °C in frozen reactions. [Each RNA] includes ribozyme, primer and template; * = standard extension buffer (200 mM $MgCl_2$, 50 mM Tris pH 8.3, 4 mM each NTP (17 °C)/1 mM each NTP (ice)). The principal components of the mineral water (pH 8.65) are 4.5 mM Mg^{2+}, 0.2 mM Ca^{2+}, 0.4 mM Na^+, 0.24 mM SO_4^{2-}, 0.13 mM Cl^-, and 7.2 mM bicarbonate.

Challenges of environmental cationic diversity

However, plausible primordial environments could fail to provide the high concentrations of Mg^{2+} ions (and other components) critical for this polymerase ribozyme function. Instead of matching or approaching a polymerase ribozyme's preferred extension conditions (0.2 M $MgCl_2$, 4 mM each NTP, pH 8.3), such environments would likely by dominated by heterogeneous mixtures of other solutes. To mimic this challenge and investigate the tolerance of ribozyme polymerase activity to the presence of non-native solutes, we investigated ribozyme activity in a sea-salt brine (containing, amongst other components, 54 mM Mg^{2+}, 469 mM Na^+, 10 mM Ca^{2+}, 11 mM K^+, 543 mM Cl^-, and 28 mM SO_4^{2-}). We observed a substantial decrease in primer extension activity (Fig. 2a), to below the level observed with 54 mM Mg^{2+} alone. However, supplementing the brine with $MgCl_2$ to 0.2 M was sufficient to restore full ribozyme activity, indicating that this heterogeneous mixture of ions is compatible with RNA replication provided the ribozyme's Mg^{2+} requirements are met. Indeed, a simpler 'mock brine' comprising just the chloride salts of Na^+, Ca^{2+} and Mg^{2+} at concentrations similar to in the brine yielded the same behaviour. Surprisingly, even though Bartel and colleagues did find substantial competitive inhibition of ligase activity by Ca^{2+} ions,[15] the inhibition of polymerase activity observed in unsupplemented brine did not originate from Ca^{2+} (which may be at too low a concentration for competitive inhibition of the ribozyme (Fig. S2a in ESI†)) but instead from the high concentrations of NaCl (Fig. 2a). To better understand this effect, we investigated polymerase ribozyme activity as a function of NaCl concentration (at a fixed [$MgCl_2$] of 0.1 M). We observed a reduction of ribozyme activity between 0.2 and 1 M NaCl, although interestingly activity recovered as [NaCl] rose to 3 M (Fig. S2b in ESI†). Inhibition by NaCl is more severe at lower magnesium concentrations such as those found in the sea-salt brine (Fig. S2c in ESI†), suggesting that Na^+ ions may compete for some of the polymerase's low-affinity Mg^{2+} binding sites.

The primordial oceans of the early Earth are thought to have been salty[19] (though their precise composition, like many aspects of primordial geochemical scenarios, is a matter of debate). However, the inhibition by high salt conditions suggests that RNA self-replication might have been more likely to arise and persist in environments of lower salinity. Intriguingly, high salt concentrations have also been observed to negatively affect other potentially prebiotic processes, such as nucleotide monomer condensation and vesicle formation.[11,12] For these reasons, a freshwater environment on early landmasses might represent a potentially attractive alternative locale for the emergence of RNA self-replication. However, while relaxing high salt inhibition, freshwater environments present other potential challenges such as low availability of essential Mg^{2+} ions, and the presumed scarcity of nucleotide building blocks required by a ribozyme.

A plausible scenario for the emergence of RNA self-replication in a freshwater environment therefore needs to include effective mechanisms for substrate and counterion concentration. Many such mechanisms ranging from desiccation to mineral adsorption are conceivable.[20] Here, we have concentrated on water ice formation as a potent form of solute and substrate concentration within a supercooled eutectic phase, which we had previously shown to be compatible with RNA polymerase ribozyme activity.[9] The composition of primordial fresh water bodies is unknown but may reasonably be assumed to be not too dissimilar to present-day fresh water sources from

areas low in biochemical sedimentary rocks such as limestone. To test whether freezing could enable RNA replicase activity in such a fresh water environment, we challenged the Z polymerase to synthesise RNA in a reaction buffer comprising a commercially available mineral water (with pH 8.65 and low Ca^{2+} content) supplemented with micromolar (μM) concentrations of nucleotides (Fig. 2b). Remarkably, while no extension was observed in the solution phase at ambient temperatures, freezing allowed RNA synthesis comparable to that in optimal extension buffer to be achieved, using mineral water supplemented with just 10 μM of each NTP, with synthesis detectable using as little as 2 μM of each NTP – facilitated by the concentration effect of ice crystal growth. Thus freshwater environments in conjunction with diurnal or seasonal ice formation could provide a plausible environment for RNA self-replication away from inhibitory high salinity conditions of the primordial seas or evaporite lakes.

Influence of ribonucleotide 2′-moieties

Having investigated solute dependence, we next sought to establish the chemical fidelity of the Z polymerase ribozyme by examining substrate selection and specificity using a range of defined non-cognate nucleoside triphosphate analogues. Of particular interest in this context are modifications to the 2′-position of the ribofuranose ring. The 2′-OH group of RNA is responsible for many of its distinctive properties, and incorporation of substrate variants with altered 2′-groups yields RNAs with altered conformational and catalytic properties.[21–23] We first examined incorporation of UTP variants bearing 2′-modifications using an RNA template in which the first twelve positions (Fig. 1a) encode the addition of only one uridine, allowing an assessment of the modification's effect upon multiple stages of RNA synthesis in a processive context. We examined five 2′-moieties of increasing steric bulk compared to the cognate 2′-OH of RNA: –H, –F, $–NH_2$, $–OCH_3$, and $–N_3$ (Fig. 3a). We found that the Z polymerase ribozyme readily incorporated uridines with any of these five groups (in the presence of riboC, riboG and riboA) – though only 2′-F with full efficiency (Fig. 3b). These results indicate that only limited discrimination is exerted by the Z polymerase ribozyme upon the 2′-group at the incorporation step. The crystal structure of the class I ligase ribozyme that forms the catalytic core of Z[17] suggests a potential hydrogen bond between the 2′-OH of U47 and the 2′-OH of the incoming nucleotide. This interaction might be partially preserved with a 2′-F, though F rarely acts as a good hydrogen bond acceptor.[24] However, the tolerance of bulky 2′-groups like -N_3 that are unable to hydrogen bond to U47 renders it unlikely that this interaction exerts a strong steric or energetic discriminatory effect upon the incoming nucleotide via its 2′-group.

In contrast to the incorporation step, we observed in these same reactions significant variation in the efficiency of subsequent extension of the 2′-modification-containing extended primers by ribonucleotides. The presence of different 2′-modifications at the 3′-terminus of the synthesised RNA induced widely different degrees of pausing (Fig. 3b and c). For example, while a 2′-deoxyribonucleotide (2′-H) was very poorly extended, a 2′-NH_2 effected more complete extension than the cognate 2′-OH. 2′-F yielded a modest decrease in extension, with more pronounced reductions resulting from 2′-N_3 and $–OCH_3$ groups. No obvious structural contacts or interactions to the primer terminal 2′-OH are observed in the ligase core structure.[16] Instead, the effects of different 2′-groups appear to correlate to their ability to lower the pK_a of the proximal 3′-OH and enhance its reactivity. In analogy to non-enzymatic polymerisation[25] a vicinal 2′-OH (or -NH_2) group may stabilise the deprotonated 3′-O^- necessary for nucleophilic attack upon the α-phosphate of the incoming NTP.

As the ribozyme continues extension beyond the incorporated 2′-modified nucleotide, further significant pausing is observed when the 2′-modification reaches the -3 position of the nascent RNA duplex (after two further extension steps beyond the initial incorporation) (Fig. 3b and c). In the class I ligase, the 2′-OH group of the equivalent -3 nucleotide (in the substrate strand) interacts with A26 and A27 as part of an A-minor triad motif.[16] Modifications of 2′-hydroxyls on the primer strand are likely to disrupt these sequence-general contacts and impact upon duplex affinity and translocation. Indeed, the pausing data recapitulate results obtained previously using primers containing 2′-deoxy modifications at defined positions, which indicated a dependence of extension upon the presence of 2′-OH groups in the template strand.[26] In our experiments, all the different modifications triggered a similar level of pausing, suggestive of disruption of a specific interaction. We observed negligible pausing relative to 2′-hydroxyls at almost all other distances from the primer 3′-end (with the exception of the bulky 2′-OCH_3 group, which also caused pausing at the adjacent -2 position). When performed in ice, the incorporation and extension experiments again yielded similar patterns of pausing, although the impact of modifications at the -2 and -3 positions disrupting ribozyme contacts was reduced. This indicates a similar chemical fidelity of ribozyme catalyzed RNA synthesis in ice with regards to 2′-modifications (Fig. S3a in ESI†).

Despite clear interference with ribozyme function both at the level of incorporation and extension, the cumulative effects were moderate enough to allow incorporation of multiple modified nucleotides for some 2′-modifications of GTP (2′-F GTP, 2′-NH_2 GTP; Fig. S3b in ESI†) though with greatly reduced yields *vs.* 2′-OH. These experiments illustrate that the Z RNA polymerase ribozyme is capable of synthesising a narrow selection of RNA polymers of mixed chemical composition.

Response to nucleotide sugar heterogeneity

Next we investigated the incorporation of nucleotides in which the canonical ribofuranose (ribo) of RNA is replaced by alternative sugar ring structures or congeners (2′-deoxyxylose (d-xylo),[27] arabinose (ara),[28] 2′-deoxy-2′-fluoro-arabinose (F-ara),[29] β-D-glucopyranose (gluco),[30] 2′,3′-dideoxy-β-D-glucopyranose (homo),[30] 1,5-anhydrohexitol (hex),[31] cyclohexenyl (Ce),[31] 2′-O,4′-C-methylene-β-D-ribose (locked),[32] and an acyclic ribofuranose mimic – 2′,3′-seco-ribose (unlocked)[32] (Fig. 3a)). We found that the Z polymerase ribozyme incorporated a range of

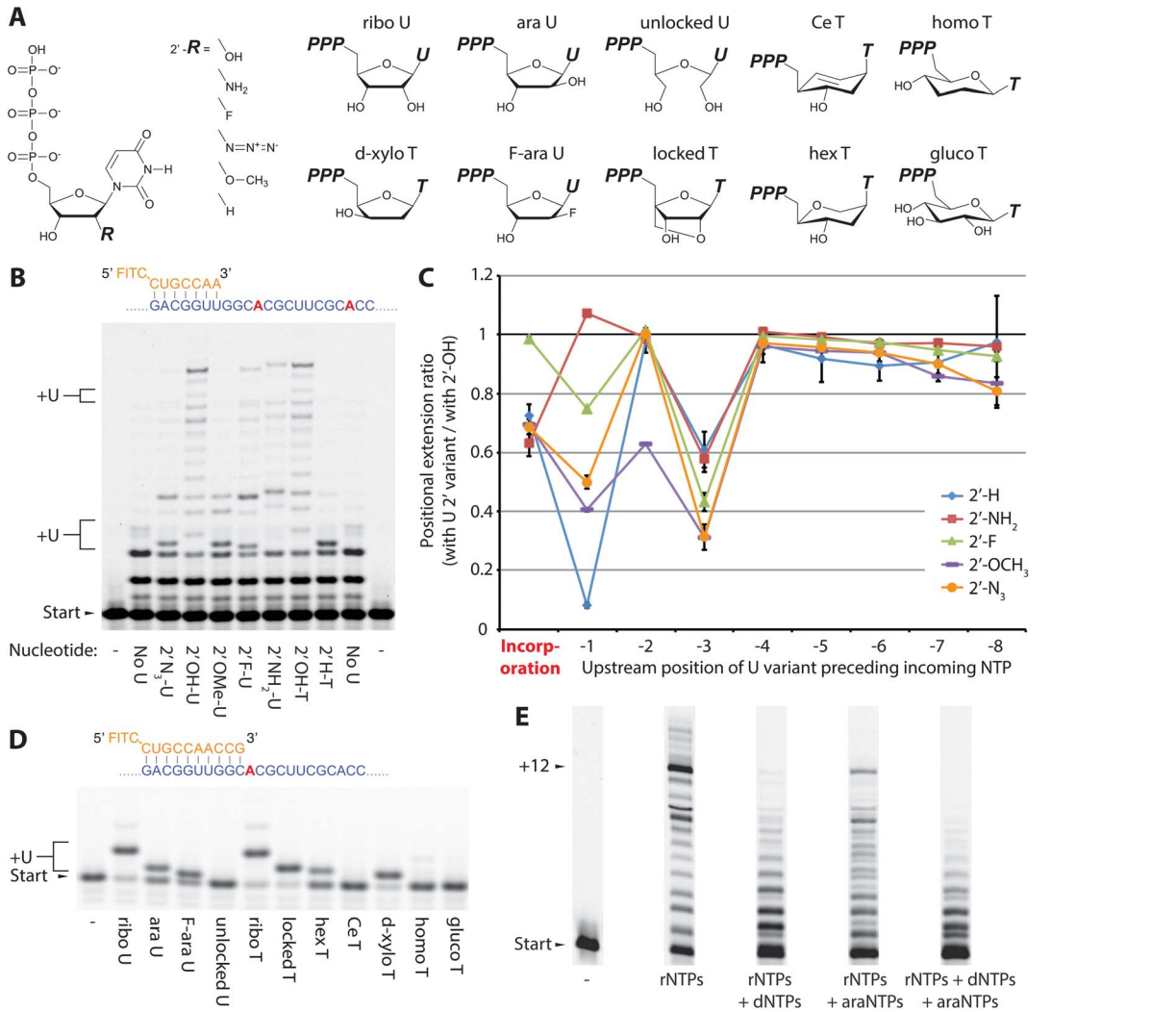

Fig. 3 Influence of nucleotide sugar variants or congeners upon extension by Z. (A) The structures of ribouridine triphosphate 2′-modifications (2′-R) (left panel) and a set of sugar variants or congener nucleotides. (B) Extension upon the A7/I duplex (above panel, 17 °C, 2 days) using GTP, ATP, CTP and a 2′-variant of UTP or TTP (- = no NTPs). The positions in the product ladder where the template encodes U incorporation are indicated, and are highlighted in the template sequence in red. (C) The fractions of extension products in (B) that were successfully extended beyond each successive position (numbered relative to the first 2′-modified U incorporation) were calculated, and compared to the fractions obtained using rUTP (or rTTP for 2′-H) to obtain positional extension ratios (error bars represent s.d.; $N = 3$). (D) Single-nucleotide incorporation of sugar variants/congeners of U/T alone opposite A (highlighted in red in the template sequence) in the A10/I duplex (above panel, 17 °C, 10 days). (E) Extension (17 °C, 2 days) using rNTPs upon the BioA10/I duplex in the presence of equimolar mixtures of dNTPs, araNTPs or both (- = no NTPs).

sugar-modified nucleotides to varying extents (with locked > d-xylo ≈ ribo > F-ara > ara > hex, Fig. 3d). Strikingly, a locked-NTP is actually incorporated with greater efficiency (Fig. S4d in ESI†) while the d-xyloNTP is incorporated to a similar extent to the canonical riboNTP. However, other NTPs with a more substantial degree of structural and/or conformational deviation from RNA (Ce, unlocked, gluco, homo) were incorporated very poorly or not at all. For example, CeTTP and unlocked-UTP showed no evidence of incorporation opposite A at all, although very limited CeGTP and unlocked-GTP incorporation opposite C was detectable (Fig. S4a in ESI†), presumably due to the improved stacking and hydrogen bonding of guanosine compared to uridine.

Thus the polymerase ribozyme appears not to display stringent discrimination between different ring structures at the incorporation step. However, all variants effectively terminated further extension (Fig. S4b and c in ESI†). Thus, while some variation in ring pucker, helix conformational preference and hydroxyl reactivity/positioning are tolerated at the incorporation step, subsequent extension appears to be exquisitely sensitive to deviations from the cognate ribofuranose parameters. Therefore, in analogy to results obtained for the 2′-modified nucleotides, the polymerase ribozyme displays stringent RNA synthesis specificity at the extension step.

Such chemical discrimination at the extension step is an effective strategy to prevent the synthesis of mosaic oligomers comprising a non-heritable mix of different nucleotide chemistries. However, in the absence of an editing mechanism, this same behaviour also leads to premature termination of synthesis for substrate mixtures containing substantial amounts of terminating sugar variants alongside ribonucleotides. In a prebiotic context, deoxyribonucleotides[33] and

arabinonucleotides[34] are of particular interest. While a riboNTP does have a modest incorporation advantage when present together with a dNTP and is strongly preferred over an araNTP (Fig. S4d in ESI†), over the course of multiple extensions, the presence of competing araNTPs and/or dNTPs significantly reduces synthesis of full-length RNA products through iterative termination (Fig. 3e), potentially providing a strong selection pressure for improved chemical selectivity in a heterogeneous chemical environment.

Tolerance of variation in the nucleobase

To compound these challenges, chemical diversity may not have been limited to the sugar portion of the nucleotide. To carry information, modern RNA is built from only four nucleobases U, C, G and A. Where added chemical diversity is needed (such as in ribosomal and transfer RNAs), modifications to the canonical nucleotides are introduced post-synthetically by dedicated protein enzymes. In a prebiotic setting, both atmospheric and surface chemistry[35] as well as other processes (such as meteoritic or cometary impacts[36]) could conceivably have generated a much wider range of nucleobases that as nucleotide substrates would hold the potential to hinder (or benefit) RNA replication. We therefore investigated the chemical fidelity of the polymerase ribozyme with respect to a range of model nucleobase modifications, beginning with incorporation of a range of ribonucleotide base variants opposite template adenine (Fig. 4a and b).

In analogy to proteinaceous polymerases, we found that small substituents such as methyl at the 5 position of uracil are well tolerated; the thymine (T) ribonucleotide was added and extended with comparable efficiency to the cognate uracil. 5-Iodouracil (5-iodo-U), however, was incorporated with markedly reduced efficiency, despite a similar steric bulk of the iodine substituent. Halogen atoms at the 5 position are known to affect the keto/enol equilibrium of uracil,[37] though terminal 5-iodo-U residues have been shown to modestly stabilise RNA duplexes.[38]

5,6-Dihydrouracil (5,6-DHU) incorporation was inefficient and subsequent extension was reduced (Fig. 4c), presumably due to the deviations from planarity in the non-aromatic 5,6-DHU ring disrupting base stacking interactions.[39] In contrast, pseudouridine (ψ) was incorporated by the ribozyme even more efficiently than UTP, though causing a slight pausing effect before subsequent extension. ψ is able to coordinate a water molecule between its free N1–H and the phosphates of itself and the preceding base in an RNA strand promoting local RNA stacking,[40] which may explain its enhanced incorporation. Pausing induced by incorporation of modified bases relative to UTP during extension in ice was broadly similar, though less pronounced in several instances (Fig. S5a in ESI†).

Next we investigated modifications to the H-bonding interface. When we replaced the uracil base of UTP with a series of halogenated isosteres (benzene, 2,4-difluorobenzene (DFB), 2,4-dichlorobenzene (DCB), 2,4-dibromobenzene (DBB), and 2,4-diiodobenzene (DIB)) which are largely unable to H-bond with template adenine,[41,42] we observed little or no incorporation (Fig. S5b in ESI†). Only after very long incubations could incorporation of the isosteres with larger halogen substitutions (DCB, DBT, DIT) be observed. In this the polymerase ribozyme resembles proteinaceous DNA polymerases of the polY family,

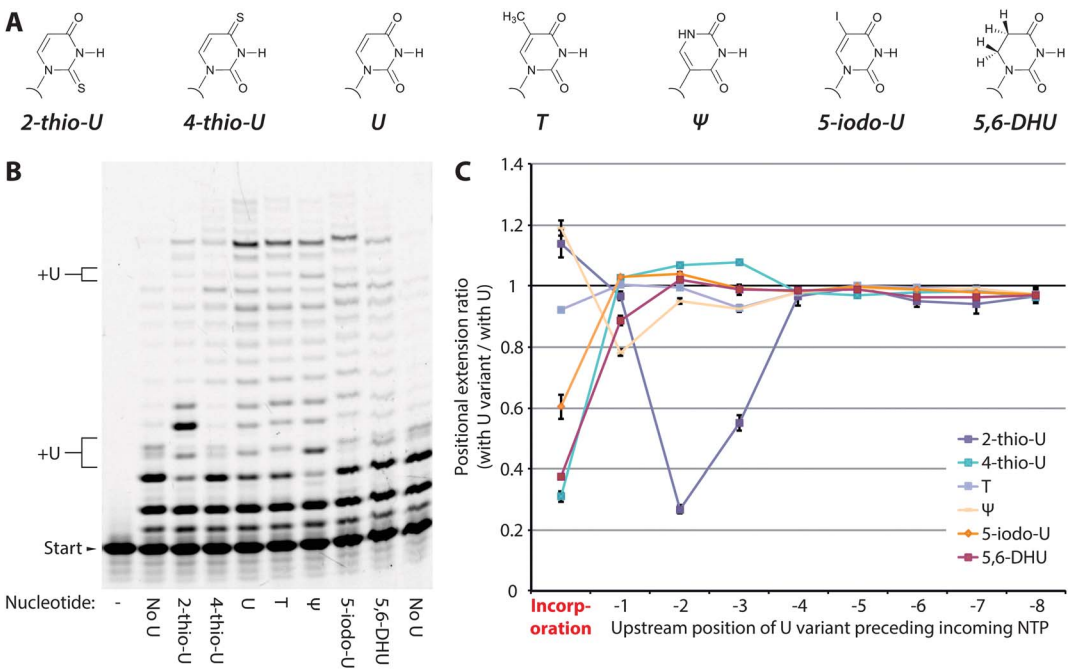

Fig. 4 Influence of nucleotide base variants upon extension by Z. (A) Structures of a range of uracil-variant bases, investigated as ribonucleoside triphosphates. (B) Extension upon the A7/I duplex (see Fig. 3B; 17 °C, 2 days) using GTP, ATP, CTP and a ribonucleotide variant of UTP (- = no NTPs). The positions in the product ladder where the template encodes U incorporation are indicated. (C) The positional extension ratios relative to UTP incorporation when using each U variant (error bars represent s.d.; $N = 3$).

involved in translesion synthesis and DNA repair. Polymerases from this family possess geometrically unconstrained active sites to enable trans-lesion synthesis of conformationally distorted or bulky template lesions and adducts. Consequently, these polymerases rely on hydrogen-bonding rather than geometric control for substrate selection and are thus also unable to use isosteres efficiently as substrates.[43]

These data indicate that the polymerase ribozyme relies primarily on nucleotide binding energy (from H-bonding and base stacking) for substrate selection at the incorporation step. Indeed, even subtle modifications to the hydrogen bonding groups exert significant effects. The thio substitution in 4-thio-U weakens the U4O-A6NH_2 hydrogen bond during Watson–Crick pairing,[44] reducing incorporation of the nucleotide opposite A by the Z ribozyme. In 2-thio-U, however, the thio substitution does not disrupt H-bonding; in fact, the A:2-thio-U pair has been observed to be slightly stronger than the canonical A:U pair.[44] Indeed, we found that Z incorporates 2-thio-U more efficiently than U, perhaps also due to the nucleotide's propensity to adopt a C3′-endo sugar pucker before template binding.[45] Furthermore, the 2-thio modification weakens hydrogen bonding to guanine during formation of a U·G wobble pair;[44,46] indeed, we observed less misincorporation of 2-thio-U than uracil opposite G (although, unexpectedly, misincorporation of 4-thio-U was also reduced; Fig. S6 in ESI†).

A reduced formation of U·G wobble pairs when U is replaced with 2-thio-U has been proposed as a means to increase the fidelity of non-enzymatic RNA replication[47] and might similarly reduce the formation of mutagenic U·G wobble pairs during enzymatic replication. However, we observed an unexpected block in ribozyme extension after 2-thio-U incorporation (Fig. 4b) with substantial pausing with 2-thio-U at the -2 and -3 position in the nascent strand, suggestive of disruption of important contacts between ribozyme and primer/template duplex reminiscent of the extension block observed with 2′-substitutions. Substitution of oxygen for sulphur has the potential to weaken interactions with both hydrogen bond donors as well as hard metal ions like Mg^{2+}; inspection of the class I ligase structure suggested that the 6-NH_2 of A26 of the ribozyme could form a hydrogen bond to the minor groove carbonyl of pyrimidine bases at the -3 position in the primer.[16] However, the ligase structure provides no obvious explanation for the stronger block at the -2 position. As the pausing observed is strong and specific, it seems likely that there is a key interaction between the ribozyme and the duplex at this position. This might involve a water molecule or weakly bound magnesium ion not visible in the ligase structure or new interactions arising from conformational changes in the polymerase compared to the ligase ribozyme structure.

The minor groove carbonyl (2O) of the two natural pyrimidines can act a hydrogen bond acceptor, as can the equivalent N3 position on purines, providing a sequence-independent means of monitoring base pairs[48] – a feature exploited by many DNA polymerases.[49] It seems plausible that disruption of a similar minor groove interaction in the Z polymerase ribozyme by the use of 2-thio-U is at the root of the observed extension block.

Expansion of the ribozyme genetic alphabet

We also screened the potential for misincorporation of purines opposite pyrimidines (and *vice versa*). Results from this screen confirm the dominant role of hydrogen bonding in substrate selection at the incorporation step (Fig. S6a in ESI†). Many bases capable of wobble pairing or reverse wobble pairing with the template base using two hydrogen bonds[50] are incorporated in the absence of competing correct nucleotide. Other bases only show incorporation if tautomeric forms allow it. In agreement with the results obtained with non-hydrogen bonding isosteres, these data build up a picture where the affinity of the nucleotide for the correct position in the primer/template duplex rather than base-pair geometry is the main selectivity determinant at the incorporation step, with hydrogen bonding likely the dominant factor.

To further probe the energetic and geometric requirements of the polymerase ribozyme we assessed whether Z would be capable of replicating unnatural base pairs designed to expand the genetic alphabet. We first analysed the IsoC:IsoG pair,[51] which displays Watson–Crick-type pairing through hydrogen bonding, but with the order of H-bond donor–acceptor groups inverted compared to the canonical C:G pair (Fig. 5a). Performing best in ice, the polymerase ribozyme Z efficiently added IsoGTP opposite template IsoC. However, misincorporation of IsoG opposite template U (presumably due to the IsoG equilibrium with an alternative U-pairing enol tautomer[51]) is comparable to that of G opposite U through wobble pairing (Fig. S6a in ESI†). We also observed a significant amount of misincorporation of GTP opposite template IsoC (Fig. S6b in ESI†), perhaps through reverse wobble pairing.[50] Indeed, in the presence of all four natural NTPs (and absence of IsoGTP), significant extension occurred across and beyond an IsoC residue in the template (Fig. 5b), although addition of IsoGTP boosts extension. The presence of this third base pair in a primordial genetic system could expand the informational and catalytic capacity of RNA, but these data demonstrate the risks to encoded information of an expanded genetic alphabet: a reduction in the fidelity of information transfer, further exacerbated by the propensity for IsoC to deaminate to yield U.[51]

Although extension continued after IsoG incorporation, we again observed significant pausing when the IsoG:IsoC base-pair reached the -3 position in the nascent duplex (Fig. 5b). As observed for 2-thio-U, modifications to the cognate base-pairs in this region appear to interfere with some critical yet-to-be-identified minor groove interaction of the ribozyme with the primer–template duplex. Indeed, while the major groove face of the unnatural IsoG:IsoC base pair resembles that of the A:U pair, the minor groove face is different from both natural pairs. In particular, a 2-amino group is present in place of the universal hydrogen bond acceptors of natural bases (described above). Incorporation of 2-thio-U immediately after IsoG:IsoC triggers more pausing than with either modification alone (Fig. 5c). We suggest that modification of these minor groove positions – hydrogen bond acceptors in natural Watson–Crick pairs – disrupts the sequence-independent interactions with the nascent minor groove of RNA duplex necessary for ribozyme

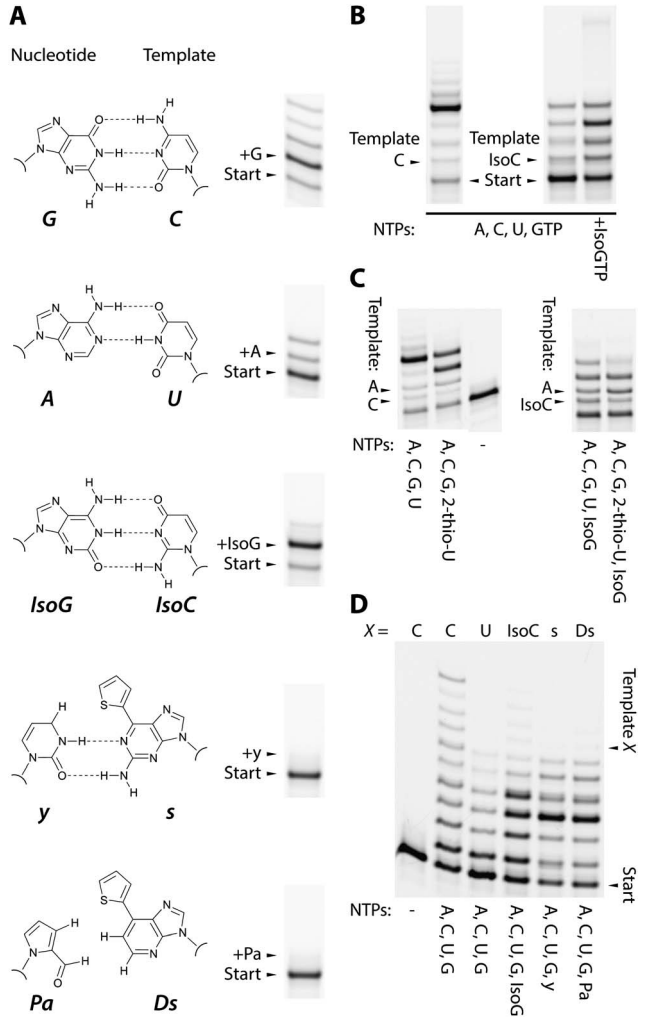

Fig. 5 Effect of artificial base pairs upon ribozyme polymerisation by Z. (A) Extension of primer A18 (−7 °C (frozen), 7 days) using a range of single nucleoside triphosphates opposite their complementary bases (pair structures shown) upon template I–X (X = C, U, IsoC, s, Ds). (B) Extension of primer A18 (−7 °C frozen 7 days) opposite I–C or I–IsoC by ribonucleotides with or without IsoGTP. The indicated +1 extension products were generated by extension opposite the named template base. (C) Extensions as in (B), demonstrating the combined influence of IsoG and 2-thio-U incorporation upon pausing after +3 additions. (D) Extension of primer A12 (−7 °C frozen 7 days) opposite templates I–X containing a range of bases at position 'X'.

duplex binding and processivity. Other synthetic base pairs (such as the Z–P pair, established in a DNA system) can maintain the minor groove pattern of hydrogen bond acceptors and might be more conducive to ribozyme-catalysed replication.[52]

In contrast to the isoG–isoC pair, the unnatural s–y[53] and Pa–Ds[54] base pairs maintain the cognate minor groove hydrogen bond acceptor pattern. Unfortunately, the ribozyme Z was completely unable to incorporate yTP opposite a template s residue (in the same template context as IsoC above (Fig. 5a)) despite being capable of misincorporating yTP opposite G through standard wobble-pairing (Fig. S6a in ESI†). This suggests that the ribozyme's active site does not tolerate the presence of the bulky major groove 6-(2-thienyl) group of s. The same effect may also explain Z's inability to incorporate the analogous PaTP opposite a template Ds, although, in this case, the absence of interstrand H-bonding may be sufficient to abolish incorporation (as suggested by the non-hydrogen-bonding isosteres of U, Fig. S5b in ESI†).

In these experiments, we had directly assayed formation of the synthetic pair as the first nucleotide to be incorporated ('standing start'). Next we also performed 'running start' experiments, whereby the polymerase ribozyme first incorporates several standard NTPs before incorporation of the synthetic nucleotide under investigation. Surprisingly, we observed that extension was attenuated several bases before the ribozyme reached the modified base in the template (Fig. 5d). When a cytosine in the standard RNA template was replaced by either s, Ds, or IsoC, extension slowed between 2 and 4 bases upstream of the modified template base. Intriguingly, this behaviour was not limited to the unnatural bases, as replacement by U caused similar pausing. Effects on template secondary structure could be ruled out by the persistence of the effect when using the tC19Z ribozyme (which sequesters downstream template and thus prevents secondary structure formation).[7] We propose that to bind the primer/template substrate, the ribozyme makes key interactions not just with the upstream primer/template duplex, but also with C positions in the unpaired downstream template sequence. Indeed, a significant reliance upon such interactions by this ribozyme may contribute to its pronounced preference for some template sequences over others and the variable template-specific pausing patterns seen during primer extension with standard ribonucleotides.

Nevertheless, ribozyme behaviour at the most significant extension pause observed on the standard RNA template sequence can be affected by adjustment of extension conditions and substrate chemistry. This pause occurs upstream of a purine-rich template stretch, at G_{12} (Fig. 6), prior to addition of ATP. Under standard conditions, very little extension progresses beyond incorporation of the preceding G despite successful addition of A after G earlier in the sequence. A partial bypass of the block can be achieved by extension in ice using ATP, but the

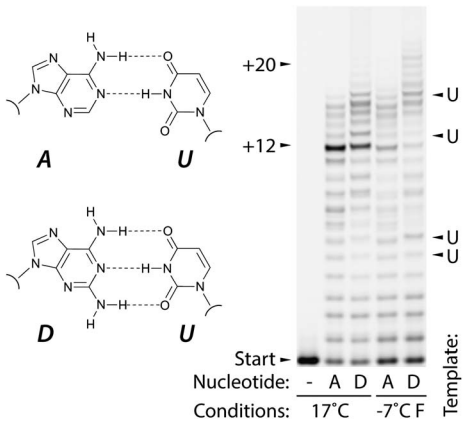

Fig. 6 Extension by Z on the A10/I duplex for 15 days at 17 °C or −7 °C (frozen), using UTP, GTP, CTP, and ATP (A) or diaminopurineriboside 5′-triphosphate (D); base-pairs to U are shown.

block can be almost eliminated (in ice) by replacing ATP with diaminopurine ribonucleoside triphosphate (DTP) yielding extension by over 20 nucleotides to the end of the template (Fig. 6). After incorporation, diaminopurine induced no pausing at the -2 and -3 positions, as expected as its minor groove features are shared amongst natural nucleobases. We speculate that stronger substrate–template interactions in the DTP–U pair may be able to compensate for suboptimal ribozyme–primer–template interactions in this context and thus allow the ribozyme to overcome this extension block.

Discussion

A comparison of the substrate discrimination and chemical fidelity of the polymerase ribozyme with the evolutionarily unrelated proteinaceous polymerases in biology is illustrative. Many natural polymerases display a strong discrimination against deviation from the cognate geometry in the nascent base-pair but are surprisingly tolerant of lack of Watson–Crick hydrogen-bonding, enabling not only the efficient incorporation of non-hydrogen bonding isosteres of the natural nucleobases but the inception of novel non-hydrogen bonding base pairs based on shape alone.[55] However, some polymerases, notably those of the polY family of translesion polymerases, diverge from this pattern and display a strong dependence on hydrogen-bonding at the incorporation step while at the same time displaying a relaxed discrimination towards geometrically aberrant substrates indicative of an open, relatively unconstrained active site.[43] The behaviour of the polymerase ribozyme at the incorporation step parallels this pattern with incorporation strongly dependent on hydrogen-bonding and stacking energy with relatively little geometric discrimination except against nucleotides that deviate strongly from the canonical 3′-endo conformation or display large substituents. It is worth noting that even the highly evolved replicative polymerases of present-day biology retain 'leakiness' with respect to chemical fidelity, as shown by the sporadic incorporation of ribonucleotides into the genome[56] and the susceptibility of dividing cells to nucleotide analogues such as the arabino nucleotide-based cancer drug Cytarabine.

In many polymerases extension of the incorporated base also depends on a correct geometric fit as well as the formation of a stable 3′-pair and the presence of cognate minor groove interactions, with aberrations causing stalling up to several bases after incorporation.[57] Again, we observe striking analogies with the behaviour of the polymerase ribozyme, which does not extend conformationally aberrant pairs like those formed by nucleotides in which the ribofuranose ring is replaced by congeners. Their conformational preferences differ significantly from the ribofuranose C3′-endo conformation (though stalling likely arises through compromised nucleotide binding rather than steric monitoring). Extension also stalls in the absence of cognate minor groove interactions such as in the case of 2-thio-U or the unnatural isoG–isoC pair.

Finally, the polymerase ribozyme displays a simple form of 'read-ahead' function through interactions with the upstream template, which stalls extension in the absence of cognate bases. Again, this is reminiscent of the case of archaeal polymerases of the polB family that stall extension upon detection of upstream template uracil.[58]

A generally accepted view of enzyme evolution proposes that primitive enzymes start out with substantial substrate promiscuity and then evolve towards greater specificity.[59] Although promiscuous at the incorporation step, the polymerase ribozyme displays a surprisingly stringent level of specificity at the extension step that is unexpected in an enzyme that has only recently emerged from a random sequence pool[7,8,13] and has never experienced adaptive pressure for RNA synthesis specificity.

We propose that this level of specificity in the polymerase ribozyme is an emergent property that arises from the complex and interdependent nature of the polymerase catalytic mechanism as well as the iterative nature of the RNA synthesis. This requires the correct positioning and interaction of four biomolecular entities (polymerase, primer, template and NTP substrate) as well as the iterative execution of multiple interlocking molecular steps including substrate recognition, incorporation, extension and translocation, where even minor deviations or suboptimal interactions add up to substantial differences in RNA synthesis yields, to the extent that substrates that may be incorporated are inefficiently extended resulting in pausing and ultimately termination of synthesis. As discussed above, similar strategies are found to varying degrees in present-day proteinaceous polymerases from different families, where they contribute to both informational as well as chemical fidelity. We therefore argue that such strategies are universal and that a basic level of informational fidelity and chemical specificity are an inevitable consequence of the polymerase mechanism and of the additive effect of iterative steps of recognition, positioning, catalysis and translocation that must be carried out with molecular precision. A basic level of both informational and chemical fidelity are therefore likely to be emergent properties of all polymerase enzymes.

Our work also demonstrates how features of natural ribonucleotides – such as the ribose 2′-OH and purine N3/pyrimidine 2-carbonyl groups – can be harnessed as sequence-general duplex contacts by polymerase ribozymes such as Z to enable processive RNA synthesis. Nevertheless, the chemical fidelity of the polymerase ribozyme is at present insufficient to preclude the incorporation of non-cognate nucleotides into the nascent strand. Such promiscuity might have compromised both function and information storage in life's early genetic systems. Although recent work has shown that the presence of some non-heritable chemical heterogeneity (RNA/DNA) is not incompatible with ligand binding function,[60] most non-cognate incorporations cause premature termination of RNA synthesis and therefore reduce RNA yields rather than 'corrupt' the chemical makeup of the 'offspring'. In a primordial replication environment of a heterogeneous chemical nature, some degree of chemical fidelity would likely have been essential to prevent excessive incorporation of non-cognate and non-heritable nucleotide chemistries into nascent ribozymes. However, some proposed abiotic chemistries and physicochemical processes could generate enantiopure ribonucleotide precursors.[61,62] In

locales where such substrates accumulate, the need for chemical fidelity is ameliorated, and simpler polymerase ribozymes with comparable specificity to Z could successfully synthesise RNA: replicators would be most likely to emerge in such environments. Subsequent evolution of improved chemical fidelity could allow colonisation of environments with a more heterogeneous chemical inventory.

Early simple replicases may also have been constrained by their requirement for sufficient concentrations of nucleotide substrates and metal ions. One environment, water ice, can plausibly provide these concentration of reactants and counterions even from levels found in freshwater sources, and is very much compatible with RNA polymerase ribozyme activity.[9] Significantly, scenarios for substrate concentration and enantiomeric enrichment would benefit both ribozyme-catalysed and nonenzymatic replication. Thus the search for self-replicating ribozymes must be complemented by a search for physicochemical mechanisms and plausible geophysical settings[5] that can promote cognate substrate enrichment, concentration, ribozyme stabilization and ultimately compartmentalization as a prerequisite for self-replication and evolution.

Acknowledgements

This work was supported by MRC U105178804 (PH, JA, ST) and Homerton College, Cambridge (JA). JW acknowledges funding from EU (ERC-2010-AdG_20100317).

Notes and references

1 M. P. Robertson and G. F. Joyce, *Cold Spring Harbor Perspect. Biol.*, 2012, **4**, a003608.
2 L. E. Orgel, *Crit. Rev. Biochem. Mol. Biol.*, 2004, **39**, 99.
3 S. A. Benner, H. J. Kim, M. J. Kim and A. Ricardo, *Cold Spring Harbor Perspect. Biol.*, 2010, **2**, a003467.
4 S. Pizzarello and E. Shock, *Cold Spring Harbor Perspect. Biol.*, 2010, **2**, a002105.
5 I. Budin and J. W. Szostak, *Annu. Rev. Biophys.*, 2010, **39**, 245.
6 A. D. Ellington, X. Chen, M. Robertson and A. Syrett, *Int. J. Biochem. Cell Biol.*, 2009, **41**, 254.
7 A. Wochner, J. Attwater, A. Coulson and P. Holliger, *Science*, 2011, **332**, 209.
8 W. K. Johnston, P. J. Unrau, M. S. Lawrence, M. E. Glasner and D. P. Bartel, *Science*, 2001, **292**, 1319.
9 J. Attwater, A. Wochner, V. B. Pinheiro, A. Coulson and P. Holliger, *Nat. Commun.*, 2010, **1**, 76.
10 Y. Li and R. R. Breaker, *J. Am. Chem. Soc.*, 1999, **121**, 5364.
11 I. A. Chen, K. Salehi-Ashtiani and J. W. Szostak, *J. Am. Chem. Soc.*, 2005, **127**, 13213.
12 P. A. Monnard, C. L. Apel, A. Kanavarioti and D. W. Deamer, *Astrobiology*, 2002, **2**, 139.
13 D. P. Bartel and J. W. Szostak, *Science*, 1993, **261**, 1411.
14 E. H. Ekland and D. P. Bartel, *Nucleic Acids Res.*, 1995, **23**, 3231.
15 M. E. Glasner, N. H. Bergman and D. P. Bartel, *Biochemistry*, 2002, **41**, 8103.
16 D. M. Shechner, R. A. Grant, S. C. Bagby, Y. Koldobskaya, J. A. Piccirilli and D. P. Bartel, *Science*, 2009, **326**, 1271.
17 D. M. Shechner and D. P. Bartel, *Nat. Struct. Mol. Biol.*, 2011, **18**, 1036.
18 J. C. Bowman, T. K. Lenz, N. V. Hud and L. D. Williams, *Curr. Opin. Struct. Biol.*, 2012, **22**, 262.
19 C. E. J. De Ronde, D. M. d. Channer, K. Faure, C. J. Bray and E. T. C. Spooner, *Geochim. Cosmochim. Acta*, 1997, **61**, 4025.
20 P. A. Monnard, *Cell. Mol. Life Sci.*, 2005, **62**, 520.
21 B. E. Eaton and W. A. Pieken, *Annu. Rev. Biochem.*, 1995, **64**, 837.
22 D. M. Williams, W. A. Pieken and F. Eckstein, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 918.
23 L. H. Lauridsen, J. A. Rothnagel and R. N. Veedu, *ChemBioChem*, 2012, **13**, 19.
24 J. D. Dunitz and R. Taylor, *Chem.–Eur. J.*, 1997, **3**, 89.
25 R. Lohrmann and L. E. Orgel, *J. Mol. Biol.*, 1977, **113**, 193.
26 U. F. Muller and D. P. Bartel, *Chem. Biol.*, 2003, **10**, 799.
27 B. R. Babu, Raunak, N. E. Poopeiko, M. Juhl, A. D. Bond, V. S. Parmar and J. Wengel, *Eur. J. Org. Chem.*, 2005, 2297.
28 A. M. Noronha, C. J. Wilds, C. N. Lok, K. Viazovkina, D. Arion, M. A. Parniak and M. J. Damha, *Biochemistry*, 2000, **39**, 7050.
29 C. J. Wilds and M. J. Damha, *Nucleic Acids Res.*, 2000, **28**, 3625.
30 M. Renders, M. Abramov, M. Froeyen and P. Herdewijn, *Chem.–Eur. J.*, 2009, **15**, 5463.
31 P. Herdewijn, *Chem. Biodiversity*, 2010, **7**, 1.
32 M. A. Campbell and J. Wengel, *Chem. Soc. Rev.*, 2011, **40**, 5680.
33 M. W. Powner, S. L. Zheng and J. W. Szostak, *J. Am. Chem. Soc.*, 2012, **134**, 13889.
34 M. W. Powner and J. D. Sutherland, *Angew. Chem., Int. Ed.*, 2010, **49**, 4641.
35 S. Miyakawa, H. J. Cleaves and S. L. Miller, *Origins Life Evol. Biospheres*, 2002, **32**, 209.
36 M. P. Callahan, K. E. Smith, H. J. Cleaves, 2nd, J. Ruzicka, J. C. Stern, D. P. Glavin, C. H. House and J. P. Dworkin, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 13995.
37 S. Asano, Y. Kurashina, Y. Anraku and D. Mizuno, *J. Biochem.*, 1971, **70**, 9.
38 K. Ziomek, E. Kierzek, E. Biala and R. Kierzek, *Biophys. Chem.*, 2002, **97**, 233.
39 J. J. Dalluge, T. Hashizume, A. E. Sopchik, J. A. McCloskey and D. R. Davis, *Nucleic Acids Res.*, 1996, **24**, 1073.
40 M. Charette and M. W. Gray, *IUBMB Life*, 2000, **49**, 341.
41 T. W. Kim, J. C. Delaney, J. M. Essigmann and E. T. Kool, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 15803.
42 A. Somoza, A. P. Silverman, R. M. Miller, J. Chelliserrykattil and E. T. Kool, *Chem.–Eur. J.*, 2008, **14**, 7978.
43 M. T. Washington, S. A. Helquist, E. T. Kool, L. Prakash and S. Prakash, *Mol. Cell. Biol.*, 2003, **23**, 5107.
44 R. Vormbrock, R. Morawietz and H. G. Gassen, *Biochim. Biophys. Acta, Nucleic Acids Protein Synth.*, 1974, **340**, 348.
45 N. Zhang, S. Zhang and J. W. Szostak, *J. Am. Chem. Soc.*, 2012, **134**, 3691.

46 S. M. Testa, M. D. Disney, D. H. Turner and R. Kierzek, *Biochemistry*, 1999, **38**, 16655.
47 J. W. Szostak, *J. Syst. Chem.*, 2012, **3**, 2.
48 N. C. Seeman, J. M. Rosenberg and A. Rich, *Proc. Natl. Acad. Sci. U. S. A.*, 1976, **73**, 804.
49 T. E. Spratt, *Biochemistry*, 2001, **40**, 2647.
50 C. R. Geyer, T. R. Battersby and S. A. Benner, *Structure*, 2003, **11**, 1485.
51 C. Y. Switzer, S. E. Moroney and S. A. Benner, *Biochemistry*, 1993, **32**, 10489.
52 Z. Yang, F. Chen, J. B. Alvarado and S. A. Benner, *J. Am. Chem. Soc.*, 2011, **133**, 15105.
53 I. Hirao, *et al.*, *Nat. Biotechnol.*, 2002, **20**, 177.
54 I. Hirao, M. Kimoto, T. Mitsui, T. Fujiwara, R. Kawai, A. Sato, Y. Harada and S. Yokoyama, *Nat. Methods*, 2006, **3**, 729.
55 E. T. Kool, *Annu. Rev. Biochem.*, 2002, **71**, 191.
56 S. A. Nick McElhinny, D. Kumar, A. B. Clark, D. L. Watt, B. E. Watts, E. B. Lundstrom, E. Johansson, A. Chabes and T. A. Kunkel, *Nat. Chem. Biol.*, 2010, **6**, 774.
57 H. Miller and A. P. Grollman, *Biochemistry*, 1997, **36**, 15336.
58 M. A. Greagg, M. J. Fogg, G. Panayotou, S. J. Evans, B. A. Connolly and L. H. Pearl, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 9045.
59 R. A. Jensen, *Annu. Rev. Microbiol.*, 1976, **30**, 409.
60 S. G. Trevino, N. Zhang, M. P. Elenko, A. Luptak and J. W. Szostak, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 13492.
61 M. W. Powner, B. Gerland and J. D. Sutherland, *Nature*, 2009, **459**, 239.
62 J. E. Hein, E. Tse and D. G. Blackmond, *Nat. Chem.*, 2011, **3**, 704.