Insight into the mechanism of nonenzymatic RNA primer extension from the structure of an RNA-GpppG complex

Wen Zhang[a,b,c,d], Chun Pong Tam[a,b,c,e], Travis Walton[a,b,c,d], Albert C. Fahrenbach[a,b,c,d,f], Gabriel Birrane[g], and Jack W. Szostak[a,b,c,d,e,f,1]

[a]Howard Hughes Medical Institute, Massachusetts General Hospital, Boston, MA 02114; [b]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114; [c]Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, MA 02114; [d]Department of Genetics, Harvard Medical School, Boston, MA 02115; [e]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138; [f]Earth–Life Science Institute, Tokyo Institute of Technology, Tokyo 152-8550, Japan; and [g]Division of Experimental Medicine, Beth Israel Deaconess Medical Center, Boston, MA 02215

Edited by Jerrold Meinwald, Cornell University, Ithaca, NY, and approved June 6, 2017 (received for review March 9, 2017)

The nonenzymatic copying of RNA templates with imidazole-activated nucleotides is a well-studied model for the emergence of RNA self-replication during the origin of life. We have recently discovered that this reaction can proceed through the formation of an imidazolium-bridged dinucleotide intermediate that reacts rapidly with the primer. To gain insight into the relationship between the structure of this intermediate and its reactivity, we cocrystallized an RNA primer–template complex with a close analog of the intermediate, the triphosphate-bridged guanosine dinucleotide GpppG, and solved a high-resolution X-ray structure of the complex. The structure shows that GpppG binds the RNA template through two Watson–Crick base pairs, with the primer 3′-hydroxyl oriented to attack the 5′-phosphate of the adjacent G residue. Thus, the GpppG structure suggests that the bound imidazolium-bridged dinucleotide intermediate would be preorganized to react with the primer by in-line S_N2 substitution. The structures of bound GppG and GppppG suggest that the length and flexibility of the 5′-5′ linkage are important for optimal preorganization of the complex, whereas the position of the 5′-phosphate of bound pGpG explains the slow rate of oligonucleotide ligation reactions. Our studies provide a structural interpretation for the observed reactivity of the imidazolium-bridged dinucleotide intermediate in nonenzymatic RNA primer extension.

RNA self-replication | diguanosine dinucleotide | crystal structure | origin of life

In the RNA world hypothesis, the emergence of RNA-catalyzed RNA replication is thought to have been preceded by a stage in which RNA replication was driven purely through chemical processes (1, 2). The nonenzymatic RNA-templated polymerization of activated nucleotides or oligonucleotides has been extensively studied, with the intent of optimizing the rate, extent, and fidelity of nonenzymatic RNA/DNA polymerization (3, 4). Numerous phosphate-activating groups have been studied in the context of nonenzymatic RNA replication. For example, imidazoles such as 2-methylimidazole (5) and, more recently, 2-aminoimidazole (6), have been found to be useful phosphate activators. The potentially prebiotic synthesis of imidazoles under primitive Earth conditions has been investigated (7). On the other hand, Richert and coworkers reported the use of benzotriazole-activated monomers to improve the rate of primer extension (8). In an alternative approach, the in situ activation of monoribonucleotides, and subsequent template-guided polymerization, has been achieved by Richert and coworkers (9), using a carbodiimide reagent together with *N*-alkyl-imidazole catalysts.

Many of the thermodynamic and kinetic parameters associated with nonenzymatic RNA replication have been quantitatively determined (10–14). Until recently, the general assumption has been that nonenzymatic primer extension with activated mononucleotides involves classical S_N2 nucleophilic substitution, in which the nucleophilic 3′-hydroxyl group of the primer attacks the activated phosphorus center of the adjacent template-bound monomer via an in-line mechanism (15). The reaction is known to be catalyzed by the presence of an activated downstream monomer or oligomer, an effect initially believed to result from a noncovalent interaction between the leaving groups of adjacent nucleotides, such that the reactive site becomes preorganized for in-line displacement (16, 17). However, recent work from the J.W.S. laboratory strongly indicates that an imidazolium-bridged dinucleotide is a covalent intermediate in the reaction (18). In this model of primer extension, two activated monomers in solution react with each other, forming the imidazolium-bridged dinucleotide (Fig. 1*A*), which then binds the RNA template. As a dinucleotide, the intermediate could potentially form two Watson–Crick base pairs and thus bind more tightly than a monomer. An even more important question is whether the structure of the complex could favor nucleophilic attack of the primer on the

Significance

Rudimentary mechanisms of genome replication are essential for the earliest RNA-based cellular life, yet it is unknown how RNA or related polymers could have replicated nonenzymatically. For decades, 2-methylimidazole–activated GMP (2-MeImpG) has been used as a model substrate. We recently showed that two 2-MeImpG monomers react to form an imidazolium-bridged dinucleotide, which then reacts rapidly with the RNA primer. To explore this mechanism, we cocrystallized an RNA primer–template complex with several 5′-5′–linked analogs of the imidazolium-bridged intermediate. The closest analog, GpppG, binds to RNA in a conformation that explains the high reactivity of the imidazolium-bridged intermediate, whereas the structures of other dinucleotide ligands appear less favorable. Our study provides insight into the fundamental mechanism of nonenzymatic RNA self-replication.

Author contributions: W.Z., C.P.T., and J.W.S. designed research; W.Z., C.P.T., and T.W. performed research; A.C.F. and G.B. contributed new reagents/analytic tools; W.Z., C.P.T., T.W., A.C.F., G.B., and J.W.S. analyzed data; and W.Z., C.P.T., T.W., and J.W.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: Crystallography, atomic coordinates, and structure factors have been deposited in the Protein Data Bank, www.pdb.org (PDB ID codes 5UEE, 5UED, 5UEG, and 5UEF).

[1]To whom correspondence should be addressed. Email: szostak@molbio.mgh.harvard.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1704006114/-/DCSupplemental.

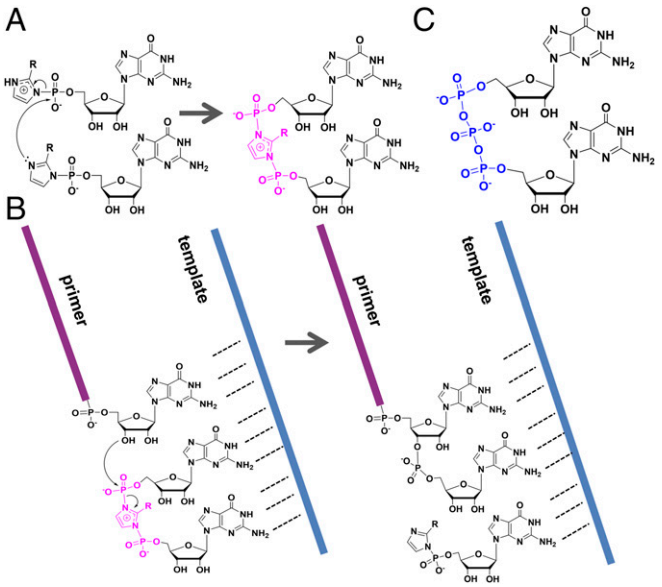

Fig. 1. Mechanism of nonenzymatic primer extension by guanosine phosphoroimidazolide monomers. (*A*) Two monomers react to form the imidazolium-bridged diguanosine intermediate. (*B*) The intermediate binds to the RNA template through two Watson–Crick base pairs. The 3′-hydroxyl of the primer attacks the adjacent phosphorus center, leading to the S_N2 reaction and primer extension. (*C*) Structure of GpppG, a close analog of the imidazolium-bridged intermediate. R = -CH_3 or -NH_2. The bridging moieties of the intermediate and the analog are colored pink and blue, respectively.

intermediate (Fig. 1*B*). As yet, there is no structural evidence as to how the dinucleotide intermediate binds the RNA and whether the conformation would be favorable for nucleophilic attack.

Despite these recent studies, relatively few mechanistic investigations have been based on crystallographic approaches. We have previously used crystallographic approaches to investigate various aspects of nonenzymatic RNA polymerization, including the effect of 2′-5′ phosphodiester linkages and 2-thio-U substitution on RNA structure (19–21). To study either enzymatic or nonenzymatic replication reactions, chemically stable analogs of substrates are typically necessary. For example, the nonhydrolyzable α–β imido ATP analogs have been intensively applied to structural studies of the mechanism of catalysis by DNA polymerase (22). In the J.W.S. laboratory, the nonhydrolyzable guanosine 5′-(3-methyl-1*H*-pyrazol-4-yl)phosphonate (PZG) was designed and synthesized to mimic the activated monomer 2-MeImpG (23). Our crystal structures of RNA-PZG complexes revealed both Watson–Crick and, surprisingly, noncanonical base pairs, suggesting that mismatched template–monomer base pairing may be more common than expected. More significantly, the phosphate and leaving group portions of these structures were always disordered. The absence of any detectable interaction between the leaving group analogs of adjacent monomers suggested that the catalytic effect of a downstream activated nucleotide might not be due to a noncovalent interaction and raised the question of whether the imidazolium-bridged dinucleotide intermediate might exhibit a more ordered structure that could help to explain its high reactivity.

To address the mechanism of nonenzymatic RNA polymerization, we used X-ray crystallography to provide high-resolution structures of RNAs complexed with stable analogs of the reactive imidazolium-bridged intermediate. The readily available compound, P^1,P^3-diguanosine-5′-triphosphate (GpppG), is a reasonable analog of the imidazolium-bridged dinucleotide, Gp-Im-pG. GpppG has a central linker of similar length and flexibility, but the labile N-P bonds are replaced by more stable O-P bonds (Fig. 1*C*). In GpppG, the flanking phosphates are separated by a central phosphate, which is similar in size to the central imidazolium in Gp-Im-pG. Furthermore, the net charge of the imidazolium-bridged intermediate is the same as that of GpppG coordinated to Mg^{2+} because the imidazolium moiety carries one positive charge, whereas the combination of a bound Mg^{2+} ion with the central phosphate of GpppG again results in a net single positive charge. Finally, we have recently shown that 2-aminoimidazole–activated monomers are superior substrates for primer extension. The corresponding 2-aminoimidazolium–bridged intermediate may adopt a conformation in which the exocyclic amino group forms hydrogen bonds with the adjacent nonbridging oxygens of the flanking phosphates, generating a compact p-NH_2Im-p structure that is closely similar to that of the triphosphate in the Mg^{2+}-GpppG complex. GpppG is also an analog of the m7G(5′)ppp(5′)G cap structure, which is present at the 5′-terminus of most eukaryotic and viral mRNAs and promotes translation initiation both in vitro and in vivo (24, 25). However, the ability of 5′-5′ linked dinucleotides to bind RNA has not been investigated through a structural approach. Here we report insights into the mechanism of nonenzymatic RNA primer extension derived from high-resolution crystal structures of GpppG and three other G-G dinucleotides complexed with the same RNA primer–template complex.

Results

Overall Structural Features of the RNA-GpppG Complex. We cocrystallized a series of guanosine dinucleotides with the RNA 5′-**mCmCmC**GACUUAAG-UCG-3′ (23) (*SI Appendix* provides experimental details). The first three nucleotides in bold are 5-methylcytidine LNA residues, designed to favor and rigidify the A-form strand conformation (26) and thereby facilitate RNA crystallization. At each end of the RNA duplex, the **mCmC** overhang serves as the binding site for G-G dinucleotides because the **mC** residues form Watson–Crick base pairs with G (23, 27). The RNA cocrystallized with the GpppG ligand with hexagonal symmetry, as in our previous RNA-GMP and RNA-PZG complexes (23). We determined the structure to 1.9-Å resolution by molecular replacement, with an overall B-factor of 32.42. The space group is $P3_121$, and there is one RNA duplex with two bound GpppG molecules per asymmetric unit. At each end of the RNA duplex, the overhanging **mCmC** binding sites are fully occupied through Watson–Crick base pairs with GpppG (Fig. 2 *A* and *B*). As in our previously determined RNA–monomer complex structures, the RNA double helices are A-form, all sugars are in the C3′-*endo* conformation, and the duplexes slip-stack on each other to form extended columns. Groups of three RNA duplex–ligand complexes form triangular prisms, and the central channel accommodates at least three water molecules that bridge the neighboring duplexes (Fig. 2*C*). Two symmetry-related water molecules form three 2.5-Å H-bond contacts with the three surrounding duplexes via the 2′-hydroxyls of their G4 residues (*SI Appendix*, Fig. S2*A*). The other water (and a possible fourth water molecule at a symmetry-related position, but with very weak density) appears to engage in three 3.1-Å hydrogen bonds with the *pro*-S_P nonbonded oxygen atoms of three G4–A5 phosphodiester linkages. However, this may actually be a time-averaged view of a water molecule that at any given moment is acting as an H-bond donor to two nonbridging phosphate oxygens. In addition, Mg^{2+} ions link the three adjacent complexes by coordinating with both the 2′- and 3′-hydroxyl groups of the guanosine at the primer +1 position of three adjacent GpppG ligands, forming a total of six ∼2.4-Å electrostatic interactions. (Fig. 2*D*).

Structure of GpppG Bound to RNA. A GpppG ligand is Watson–Crick base paired to the 5′-**mCmC** overhang at each end of the RNA duplex; the six hydrogen bond contact distances range from 2.8 Å to 3.0 Å (Fig. 2 *E* and *F*). The entire GpppG

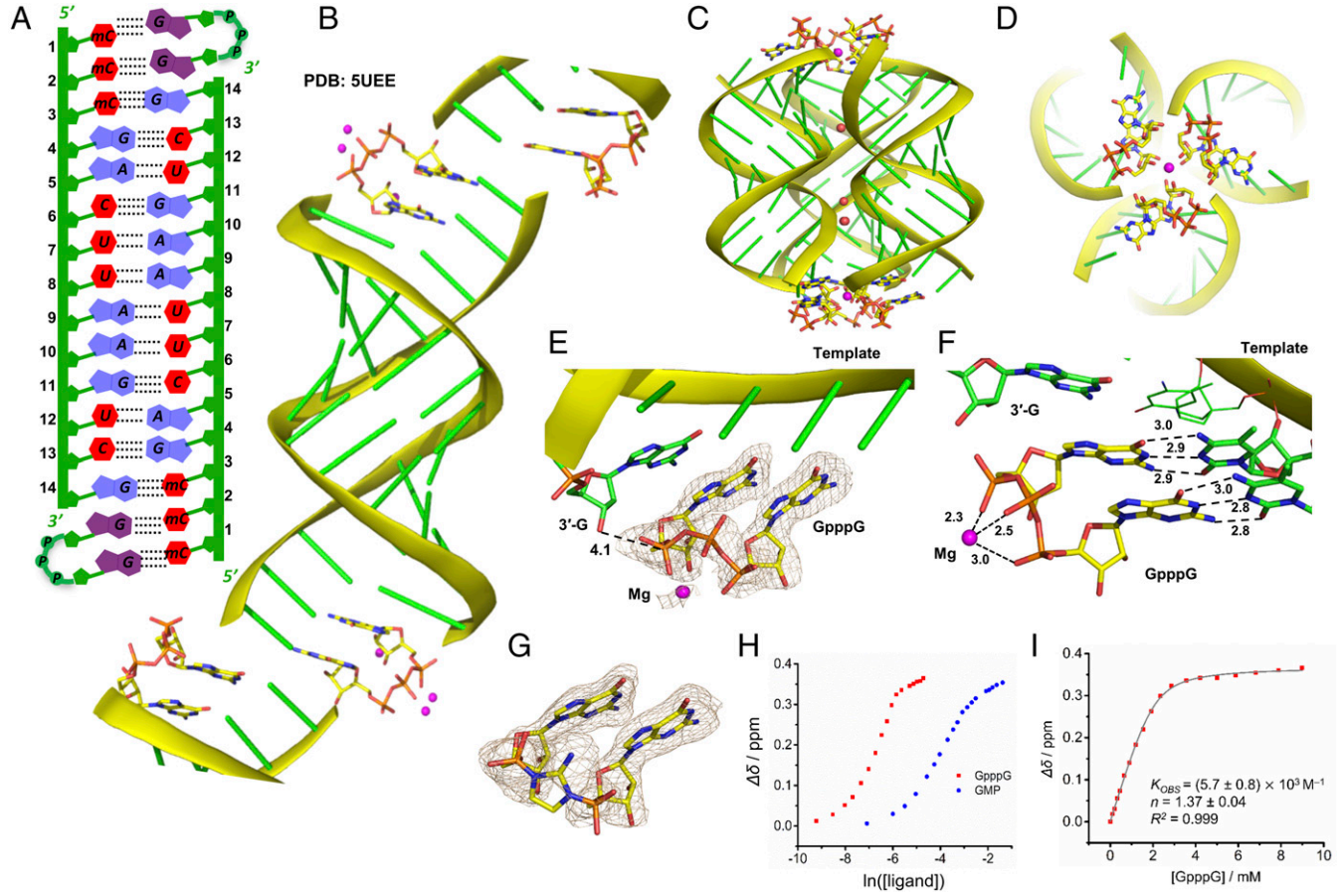

Fig. 2. RNA-GpppG crystal structure and affinity measurement. Yellow stick, GpppG ligand; green stick, RNA molecule; red dot, water molecule; magenta dot, Mg atom. All of the wheat meshes indicate the $F_o - F_c$ omit map contoured at 1.5σ. (*A*) Schematic of the RNA-GpppG complex. (*B*) Overall structure of the RNA-GpppG complex. (*C* and *D*) Ordered water molecules and Mg atoms bridge sets of three adjacent RNA-GpppG complexes. (*E*) The local structure of the GpppG ligand bound to RNA template. The corresponding omit map indicates the ordered GpppG ligand. (*F*) GpppG forms two Watson–Crick base pairs with the template through six hydrogen bonds. (*G*) Gp-NH_2Im-pG ligand is constructed and applied in refinement to fit the density of GpppG. (*H*) The chemical shift of the G imino proton changes significantly as ligand binding takes place. The leftward shift of the sigmoidal plot for GpppG shows that GpppG binding is significantly stronger than GMP. (*I*) The change in chemical shift of the same G imino proton plotted against the concentration of GpppG, fitted to a modified single-site binding isotherm.

molecule is well-ordered. The two guanine nucleobases are coplanar and stacked with the upstream primer and the neighboring duplex with interplanar distances of ∼3.3 Å. Both ribose sugars of the GpppG are in the C3′-*endo* A-form conformation, consistent with our previous observation of a C3′-*endo* sugar pucker when activated guanosine ribonucleotides bind to a template in solution (28). Moreover, the GpppG triphosphate linkage is well-ordered due to a Mg^{2+} ion that is coordinated with three nonbridging oxygen atoms, one from each phosphate of GpppG. The three electrostatic interaction distances are 2.3 Å, 2.5 Å, and 3.0 Å. This coordination with Mg^{2+} results in the triphosphate bridge of GpppG being buckled in a well-defined manner (Fig. 2 *E* and *F*).

Because GpppG was chosen as a close analog of the imidazolium-bridged dinucleotide Gp-Im-pG, we were interested in whether GpppG is bound in a conformation consistent with the observed high reactivity of the Gp-Im-pG intermediate with the primer. The distance between the primer 3′-hydroxyl and the phosphorus atom of the closest phosphate of GpppG is 4.1 Å, and the angle between the 3′-OH and the bridging P-O bond of GpppG is 126° (Table 1). We then asked whether the actual imidazolium-bridged intermediate could potentially adopt a similar conformation to the observed conformation of GpppG when bound to the RNA primer–template. We constructed a 2-aminoimidazolium–bridged diguanosine model (Gp-NH_2Im-pG) and applied it in the restrained refinement in place of GpppG (29). We discovered that the Gp-NH_2Im-pG ligand fit reasonably well to the electron density of GpppG (Fig. 2*G*). The observed electron density fits both the nucleobases and sugars of Gp-NH_2Im-pG very well (B-factors: nucleobase, 27.5; sugar, 36.4), and the Gp-NH_2Im-pG molecule formed two Watson–Crick base pairs with the template in the same manner as GpppG. The B-factors of the two phosphorus atoms in Gp-NH_2Im-pG (37.6 and 45.6) were comparable with the corresponding atoms in GpppG (31.6 and 43.4), and the distance between the P^1 and P^3 atoms in Gp-NH_2Im-pG is only marginally longer than that in GpppG (4.8 Å vs. 4.5 Å). The fact that Gp-NH_2Im-pG can be modeled to fit the density corresponding to GpppG demonstrates the potential structural similarity between the imidazolium-bridged diguanosine intermediate and the stable GpppG analog, when bound to an RNA primer–template complex.

In order for the proposed dinucleotide intermediate in nonenzymatic RNA polymerization to bind and react with the primer, it must compete with the more abundant free monomer for template occupancy. We therefore set out to measure the strength of GpppG binding to the RNA template relative to GMP. NMR methods have been used to measure the affinity of GMP for RNA primer–templates (10, 12–14). Here we used the same methods and RNA primer–template system (10) to determine

Table 1. Crystallographic structure features

RNA–ligand complex	5′-5′ linkage	Binding motifs	3′-O–P distance, Å	3′-O–P-O angle
RNA-GpppG	Triphosphate	Watson–Crick	4.1	126°
RNA-GppG	Pyrophosphate	Watson–Crick/noncanonical	4.8	170°/ND
RNA-GppppG	Tetraphosphate	Watson–Crick	~4.4	ND
RNA-pGpG	Monophosphate (3′-5′)	Watson–Crick	4.9	ND

ND, not detectable.

the affinity of GpppG for an RNA template overhang consisting of two consecutive cytidines. In the experiment, the concentrated GpppG solution was titrated into the RNA primer–template complex solution, and the binding constant was then calculated from the change in the chemical shift of the imino proton of the primer 3′-G vs. the concentration of GpppG (*SI Appendix* provides detailed procedures). Strikingly, the observed K_d of GpppG is ~0.2 mM, which is ~100-fold greater than our previously determined measurement of the affinity of GMP, which has a K_d of ~20 mM (Fig. 2 *H* and *I*). The stronger binding of GpppG most likely reflects its two Watson–Crick base pairs with the template. We then asked whether the aminoimidazolium-bridged diguanosine intermediate Gp-NH_2Im-pG would have a similar affinity to the same RNA substrate. To address this question, we isolated Gp-NH_2Im-pG in ~80% purity, with 2-AmImpG as impurity, and used it as the substrate in a primer extension reaction using the same primer–template complex as for the K_d measurement of GpppG. Interestingly, the K_m is ~0.6 mM, which is comparable to the K_d of the GpppG analog, given the different conditions required for each assay (*SI Appendix* provides details). This affinity measurement suggests that the Gp-NH_2Im-pG intermediate and the structural analog GpppG may compete effectively with monomers for binding to the template during nonenzymatic RNA polymerization reactions.

G(5′)pp(5′)G Binds to RNA Template Through Two Different Motifs. To understand the properties of the 5′-5′ linkage that make the imidazolium-bridged dinucleotide an appropriate intermediate, we cocrystallized the same RNA sequence with other guanosine dinucleotides that have different lengths or types of bridging linkages. Crystals of all the RNA–dinucleotide complexes grew with hexagonal symmetry (*SI Appendix*, Tables S2 and S3). The overall structures were similar to that of the RNA-GpppG complex, with the same molecular packing patterns, including the slip-stacked RNA double helices, the triangular-prism structure formed by groups of three RNA–ligand complexes, and the binding of the dinucleotide ligand with the **mCmC** binding sites at the RNA terminus. We first examined diguanosine-5′-5′-pyrophosphate (GppG), which is known to form in trace amounts in activated monomer solutions as a result of the reaction of GMP, formed by hydrolysis, with an activated monomer. Complementary templates can catalyze the synthesis of pyrophosphate-linked dinucleotides from activated monomers (30). These pyrophosphate-bridged dinucleotides likely inhibit nonenzymatic RNA polymerization, presumably through competitive binding of the RNA template.

The mode of binding and the rigidity of the pyrophosphate linkage of GppG are subtly different from that of GpppG. At one end, GppG forms two Watson–Crick base pairs with the templating **mC**s, but the pyrophosphate linkage and sugars are disordered (Fig. 3*B*), making it difficult to define the geometry of the pyrophosphate and the conformation of the GppG sugars. Remarkably, GppG binds to the template in a distinctly different manner at the other end. The guanosine adjacent to the primer is Watson–Crick base paired with the template **mC** through three hydrogen bonds as expected. However, the second guanosine forms a noncanonical G:C base pair with two hydrogen bonds: a weak 3.5-Å H-bond between the guanine N3 and the exocyclic amine of **mC** and a second 2.9-Å H-bond between the exocyclic amine of the guanine and the N3 of the **mC**. We previously observed the same type of noncanonical G:C base pair in our RNA-GMP and RNA-PZG structures, suggesting that this structure could play an important role in modulating the efficiency and fidelity of nonenzymatic RNA replication (23). This GppG is well-ordered overall, including nucleobases, sugars, and the pyrophosphate linkage. The first ribose is in the C3′-*endo* conformation, whereas the second one is in the C2′-*endo* conformation. The distance between the 3′-hydroxyl group of the primer and the phosphorus center of pyrophosphate linkage is 4.8 Å (Table 1 and Fig. 3*C*), significantly longer than for the GpppG complex. Although the poor solubility of GppG prevented us from measuring the affinity of GppG for the RNA primer–template, the overall structure of the RNA-GppG complex suggests that GppG likely binds the RNA template with a similar affinity as GpppG.

Tetraphosphate of G(5′)pppp(5′)G Bound to an RNA Template is Disordered. We then cocrystallized the same RNA duplex with diguanosine-5′,5′-tetraphosphate (GppppG), in which the linkage between the two nucleosides is longer and possibly more flexible than that of GpppG. In the complex structure, GppppG binds to the RNA through two well-ordered G:C Watson–Crick base pairs at both ends, but the tetraphosphate linkage is disordered. The electron density associated with the tetraphosphate linkage (Fig. 4*B*) is significantly larger and more globular than that of GpppG and can be modeled as two GppppG ligands with different conformations (each with an occupancy of 0.5). The distance between the 3′-hydroxyl group of the primer and the

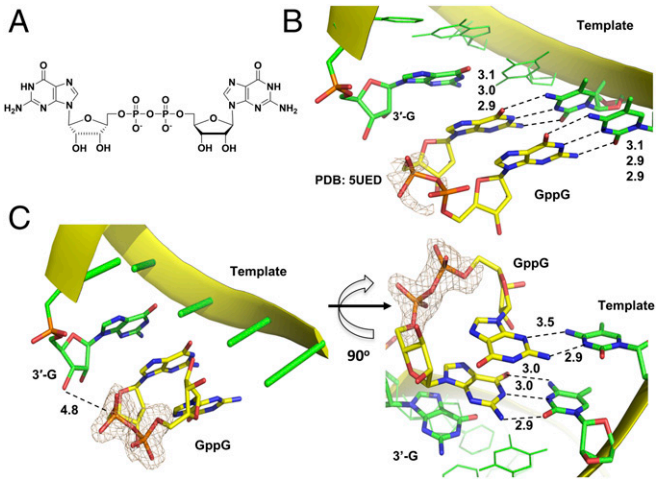

Fig. 3. Structure of the RNA-GppG complex. RNA and ligand structures and $F_o - F_c$ omit map are represented as mentioned above. (*A*) Chemical structure of GppG ligand. (*B*) At one end of RNA duplex, GppG forms two Watson–Crick base pairs with template, but the diphosphate linkage and sugar moieties are disordered. (*C*) At the other end of the duplex, GppG forms two different base pairs with the template. The diphosphate linkage is highly ordered.

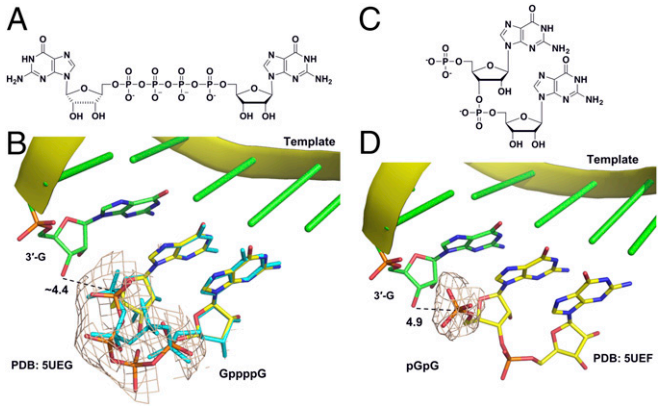

Fig. 4. Chemical and crystal structures of RNA-GppppG and RNA-pGpG complexes. (*A*) Chemical structure of GppppG. (*B*) Local crystal structure of GppppG. The tetraphosphate is disordered. Two potential conformations with 0.5 occupancy each are presented by the yellow and cyan sticks. (*C*) Chemical structure of pGpG. (*D*) Local crystal structure of pGpG bound to the RNA template.

adjacent phosphorus center is about 4.4 Å. The disordered structure of the tetraphosphate bridge suggests that any bridge with similar length and flexibility is unlikely to be optimal for the primer extension reaction due to unfavorable reaction angles for in-line attack by the primer.

pGpG Binds the RNA Template Through Watson–Crick Base Pairs. We cocrystallized the RNA with the 5′-phosphorylated GG dimer pGpG to compare this 3′-5′ linked oligonucleotide with the previously studied 5′-5′ bridged dinucleotides. At both ends of the RNA duplex, the pGpG dimer binds to the **mCmC** binding sites through two Watson–Crick base pairs. However, the two sugars are partially disordered, making it impossible to define the ribose conformation. The 5′-phosphate of the dimer is displaced toward the major groove, and the distance between the 3′-hydroxyl of the primer and the phosphorus atom of the phosphate is 4.9 Å (Fig. 4*D*). This distance is the longest of all the complex structures and is consistent with the previous observation that primer extension by ligation is much slower than by polymerization with activated monomers in the presence of an activated downstream helper monomer or oligonucleotide (16).

Discussion

Understanding the mechanism of template-directed nonenzymatic primer extension may lead to improved and/or more prebiotically realistic ways of driving RNA replication without enzymes. We have recently shown that nonenzymatic primer extension with 2-methylimidazole–activated nucleotides (5) can proceed via a two-step process, in which two activated monomers first react with each other to form an imidazolium-bridged dinucleotide intermediate (18). Once formed and bound to the template, this intermediate reacts rapidly with the primer. This surprising reaction mechanism and the unusual structure of the intermediate raise several important questions. Here we have focused on the interaction of the intermediate with the template: Can the intermediate bind to the template through Watson–Crick base pairing of both of its nucleobases? Once bound, is there some aspect of its conformation that would favor reaction with the primer? To address these questions, we have taken advantage of the structural similarity between the triphosphate-bridged dinucleotide, GpppG, and the imidazolium-bridged diguanosine intermediate, Gp-Im-pG. The 2-methylimidazole/2-aminoimidazole–bridged dinucleotides are not stable enough to directly cocrystallize with RNA, but the triphosphate analog is quite stable. As a result, we were able to obtain a high-resolution crystal structure of a close analog to the true reaction intermediate bound to an RNA primer–template complex.

The crystal structure of the RNA-GpppG complex shows that both G nucleotides interact with the template via Watson–Crick base pairing. We also showed that Gp-NH_2Im-pG could be modeled to fit the electron density of GpppG. Based on the structural similarities of GpppG and the imidazolium-bridged intermediate, it seems highly likely that the true intermediate also binds to the template through two Watson–Crick base pairs. The formation of two Watson–Crick base pairs leads to a much higher affinity of GpppG than GMP for a CC template (K_d of ~0.2 mM vs. ~20 mM, respectively). Additionally, the K_m of purified Gp-NH_2Im-pG in primer extension (~0.6 mM) suggests comparable affinity of the imidazolium-bridged intermediate. The high affinity of the imidazolium-bridged intermediate helps to explain its effectiveness in primer extension reactions because even if only a small fraction of monomer is converted to intermediate, the intermediate would still be able to bind to the primer–template complex in the presence of a large excess of activated monomer. Interestingly the GppG dinucleotide, which has a shorter linker between the two G residues, binds to the template via two Watson–Crick base pairs at one end of the RNA duplex, but at the other end by one Watson–Crick and one noncanonical base pair, suggesting that there is some strain involved in folding GppG into the conformation necessary to allow both Gs to Watson–Crick pair with the template.

Once the dinucleotide, GpppG, is bound to RNA, its overall structure, including the sugars and the triphosphate linker, becomes highly ordered. In the RNA-GpppG complex, the triphosphate linkage is structured due to a Mg^{2+} ion that coordinates with one nonbridging oxygen from each phosphate. As a result, the local structure is preorganized for S_N2 reaction with the primer 3′-hydroxyl, which in the case of the imidazolium-bridged intermediate would result in primer extension by one nucleotide. In the RNA-GpppG structure, the distance between the primer 3′-OH and the adjacent phosphate is 4.1 Å, and the 3′-O–P-O angle is 126°. The distance and angle seen here are comparable to the 4.1 Å and 107° for the S_N2 reaction catalyzed by the eukaryotic RNA polymerase II during transcription initiation, as seen in the corresponding crystal structure (31). Assuming that the imidazolium-bridged intermediate adopts a similar conformation, it is clear that further adjustments in the distance and angle of attack would have to occur before any reaction could take place, suggesting a possible role for divalent metal ion catalysis. Nevertheless, comparison with the structures of dinucleotides with different linkages (either a shorter pyrophosphate linkage or a longer tetraphosphate linkage) revealed greater 3′-O to P distances and/or greater disorder of the phosphate. Therefore, if the imidazolium-bridged intermediate adopts a conformation similar to that of GpppG under primer extension conditions, it will be partially preorganized for S_N2 attack by the primer hydroxyl. One caveat is that the intermediate generated from 2-methylimidazole–activated monomers could not bind Mg^{2+} in the same manner at the GpppG triphosphate, and the p-2MeIm-p bridge might be less ordered and therefore less than optimally reactive. However, 2-aminoimidazole–activated monomers, which lead to 10–100 times faster primer extension than 2-methylimidazole–activated monomers (6), would generate a 2-aminoimidazolium–bridged intermediate. We propose that the 2-amino group of the imidazole could hydrogen bond to the nonbridging oxygens of both flanking phosphates, potentially resulting in a highly ordered structure that is preorganized for nucleophilic attack by the primer hydroxyl. It is also of note that when the dinucleotide pGpG is bound to the same primer–template complex, the primer 3′-O to P distance is 4.9 Å, which is significantly greater than the distance for any of the 5′-5′–linked dinucleotides. This long distance may contribute to the slow rate of oligonucleotide

ligation reactions, compared with primer extension with activated monomers.

Finally, we note that the 5′-5′–linked dinucleotides with pyrophosphate, triphosphate, and tetraphosphate linkages could all be synthesized prebiotically, given the availability of imidazole-activated nucleotides. Pyrophosphate-linked dinucleotides are readily formed by attack of the 5′-phosphate of an unactivated monomer on the activated phosphate of a second monomer (30, 32). Similarly, attack of the β-phosphate of GDP on the phosphate of an activated GMP would generate GpppG. The tight binding of GpppG to a CC template suggests that GpppG might be an ideal primer for the initiation of template copying by primer extension. The resulting RNAs would begin with a 5′ cap-like GpppG moiety, suggesting a potential evolutionary origin for the eukaryotic mRNA 5′-cap structure.

In summary, our structural studies of template-bound GpppG support the model that the structurally similar imidazolium-bridged intermediate binds the template tightly through two Watson–Crick base pairs and that the conformational constraint imposed by the covalent internucleotide bridge helps to preorganize the bound complex for in-line nucleophilic attack by the primer 3′-hydroxyl.

Materials and Methods

The oligonucleotide used for crystallography was custom-synthesized by Exiqon, Inc. Oligonucleotides for affinity measurements were prepared by solid-phase synthesis. Data were collected at the SIBYLS beamlines 8.2.1 and 8.2.2 at Lawrence Berkeley National Laboratory. Datasets were processed using HKL2000 and DENZO/SCALEPACK (33). All structures were solved by molecular replacement. The refinement protocol includes simulated annealing, positional refinement, restrained B-factor refinement, and bulk solvent correction (34). The topologies and parameters for mC(LCC), GpppG(GP3), GppG(GP2), GppppG(GP4), and dinucleotide intermediate (GIM) were constructed and applied. Detailed experimental protocols are provided in *SI Appendix*.

ACKNOWLEDGMENTS. We thank Drs. Li Li, Daniel Duzdevich, and Anna Wang for helpful discussions and insightful commentaries on the manuscript. This research used beamlines 8.2.1 and 8.2.2 of the Advanced Light Source of Lawrence Berkeley National Laboratory, which is a Department of Energy Office of Science User Facility under contract DE-AC02-05CH11231. J.W.S. is an Investigator of the Howard Hughes Medical Institute. This work was supported in part by Grant 290363 from the Simons Foundation (to J.W.S.) and by Grant CHE-1607034 from the NSF (to J.W.S.). A.C.F. was supported by a Research Fellowship from the Earth–Life Science Institute at the Tokyo Institute of Technology.

1. Crick FH (1968) The origin of the genetic code. *J Mol Biol* 38:367–379.
2. Orgel LE (1968) Evolution of the genetic apparatus. *J Mol Biol* 38:381–393.
3. Szostak JW (2012) The eightfold path to non-enzymatic RNA replication. *J Syst Chem* 3:2.
4. Robertson MP, Joyce GF (2012) The origins of the RNA world. *Cold Spring Harb Perspect Biol* 4:a003608.
5. Inoue T, Orgel LE (1981) Substituent control of the poly (C)-directed oligomerization of guanosine 5′-phosphoroimidazolide. *J Am Chem Soc* 103:7666–7667.
6. Li L, et al. (2017) Enhanced nonenzymatic RNA copying with 2-aminoimidazole activated nucleotides. *J Am Chem Soc* 139:1810–1813.
7. Oró J, Basile B, Cortes S, Shen C, Yamrom T (1984) The prebiotic synthesis and catalytic role of imidazoles and other condensing agents. *Orig Life* 14:237–242.
8. Hagenbuch P, Kervio E, Hochgesand A, Plutowski U, Richert C (2005) Chemical primer extension: Efficiently determining single nucleotides in DNA. *Angew Chem Int Ed Engl* 44:6588–6592.
9. Jauker M, Griesser H, Richert C (2015) Copying of RNA sequences without pre-activation. *Angew Chem Int Ed Engl* 54:14559–14563.
10. Tam CP, et al. (2017) Downstream oligonucleotides strongly enhance the affinity of GMP to RNA primer–template complexes. *J Am Chem Soc* 139:571–574.
11. Röthlingshöfer M, et al. (2008) Chemical primer extension in seconds. *Angew Chem Int Ed Engl* 47:6065–6068.
12. Kervio E, Claasen B, Steiner UE, Richert C (2014) The strength of the template effect attracting nucleotides to naked DNA. *Nucleic Acids Res* 42:7409–7420.
13. 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 Res* 44:5504–5514.
14. Izgu EC, et al. (2015) Uncovering the thermodynamics of monomer binding for RNA replication. *J Am Chem Soc* 137:6373–6382.
15. Blain JC, Szostak JW (2014) Progress toward synthetic cells. *Annu Rev Biochem* 83:615–640.
16. Prywes N, Blain JC, Del Frate F, Szostak JW (2016) Nonenzymatic copying of RNA templates containing all four letters is catalyzed by activated oligonucleotides. *eLife* 5:e17756.
17. Wu T, Orgel LE (1992) Nonenzymatic template-directed synthesis on hairpin oligonucleotides. 2. Templates containing cytidine and guanosine residues. *J Am Chem Soc* 114:5496–5501.
18. Walton T, Szostak JW (2016) A highly reactive imidazolium-bridged dinucleotide intermediate in nonenzymatic RNA primer extension. *J Am Chem Soc* 138:11996–12002.
19. Sheng J, Larsen A, Heuberger BD, Blain JC, Szostak JW (2014) Crystal structure studies of RNA duplexes containing s(2)U:A and s(2)U:U base pairs. *J Am Chem Soc* 136:13916–13924.
20. Sheng J, et al. (2014) Structural insights into the effects of 2′-5′ linkages on the RNA duplex. *Proc Natl Acad Sci USA* 111:3050–3055.
21. Shen F, et al. (2017) Structural insights into RNA duplexes with multiple 2′-5′-linkages. *Nucleic Acids Res* 45:3537–3546.
22. Biertümpfel C, et al. (2010) Structure and mechanism of human DNA polymerase η. *Nature* 465:1044–1048.
23. Zhang W, Tam CP, Wang J, Szostak JW (2016) Unusual base-pairing interactions in monomer–template complexes. *ACS Cent Sci* 2:916–926.
24. Banerjee AK (1980) 5′-terminal cap structure in eucaryotic messenger ribonucleic acids. *Microbiol Rev* 44:175–205.
25. Furuichi Y, Shatkin AJ (2000) Viral and cellular mRNA capping: Past and prospects. *Adv Virus Res* 55:135–184.
26. Kaur H, Arora A, Wengel J, Maiti S (2006) Thermodynamic, counterion, and hydration effects for the incorporation of locked nucleic acid nucleotides into DNA duplexes. *Biochemistry* 45:7347–7355.
27. Fox JJ, et al. (1959) Thiation of nucleosides. II. Synthesis of 5-methyl-2′-deoxycytidine and related pyrimidine nucleosides. *J Am Chem Soc* 81:178–187.
28. Zhang N, Zhang S, Szostak JW (2012) Activated ribonucleotides undergo a sugar pucker switch upon binding to a single-stranded RNA template. *J Am Chem Soc* 134:3691–3694.
29. Murshudov GN, et al. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr D Biol Crystallogr* 67:355–367.
30. Majerfeld I, Puthenvedu D, Yarus M (2016) Cross-backbone templating; ribodinucleotides made on poly(C). *RNA* 22:397–407.
31. Cheung AC, Sainsbury S, Cramer P (2011) Structural basis of initial RNA polymerase II transcription. *EMBO J* 30:4755–4763.
32. Kanavarioti A, Rosenbach MT, Hurley TB (1992) Nucleotides as nucleophiles: Reactions of nucleotides with phosphoimidazolide activated guanosine. *Orig Life Evol Biosph* 21:199–217.
33. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol* 276:307–326.
34. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr* 53:240–255.