Unusual Base-Pairing Interactions in Monomer–Template Complexes

Wen Zhang,[†,‡,§] Chun Pong Tam,[†,§,||] Jiawei Wang,[⊥] and Jack W. Szostak*,[†,‡,||]

[†]Howard Hughes Medical Institute, Department of Molecular Biology and Center for Computational and Integrative Biology, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, United States

[‡]Department of Genetics, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, United States

[||]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, United States

[⊥]School of Life Sciences, Tsinghua University, Beijing 100084, China

S Supporting Information

ABSTRACT: Many high-resolution crystal structures have contributed to our understanding of the reaction pathway for catalysis by DNA and RNA polymerases, but the structural basis of nonenzymatic template-directed RNA replication has not been studied in comparable detail. Here we present crystallographic studies of the binding of ribonucleotide monomers to RNA primer–template complexes, with the goal of improving our understanding of the mechanism of nonenzymatic RNA copying, and of catalysis by polymerases. To explore how activated ribonucleotides recognize and bind to RNA templates, we synthesized an unreactive phosphonate-linked pyrazole analogue of guanosine 5′-phosphoro-2-methylimidazolide (2-MeImpG), a highly activated nucleotide that has been used extensively to study nonenzymatic primer extension. We cocrystallized this analogue with structurally rigidified RNA primer–template complexes carrying single or multiple monomer binding sites, and obtained high-resolution X-ray structures of these complexes. In addition to Watson–Crick base pairing, we repeatedly observed noncanonical guanine:cytidine base pairs in our crystal structures. In most structures, the phosphate and leaving group moieties of the monomers were highly disordered, while in others the distance from O3′ of the primer to the phosphorus of the incoming monomer was too great to allow for reaction. We suggest that these effects significantly influence the rate and fidelity of nonenzymatic RNA replication, and that even primitive ribozyme polymerases could enhance RNA replication by enforcing Watson–Crick base pairing between monomers and primer–template complexes, and by bringing the reactive functional groups into closer proximity.

Observed PZG:mC Base Pairs
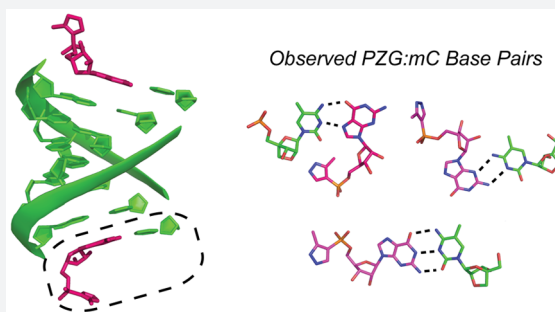
INTRODUCTION

In the RNA world hypothesis as first proposed by Crick,[1] Orgel,[2] and Woese,[3] the genetic continuity of the earliest earthly life forms was ensured by RNA replication. Template-directed nonenzymatic RNA replication[4−6] has unique significance in the RNA world hypothesis in that it facilitates the transition from the prebiotic chemistry of simple organic building blocks to the emergence of the earliest ribozyme catalysts.[4] During nonenzymatic RNA replication, chemically activated mono- or oligoribonucleotides bind to their complementary binding sites on an RNA primer–template complex, and the polymerization or ligation of these monomers and oligomers then follows to furnish the complementary RNA strand.[6] An example of such nonenzymatic RNA template copying is the purely chemically driven but template-directed synthesis of short, 3′–5′-linked guanosine- and cytosine-rich RNA segments from 2-methylimidazole-activated ribonucleotide monomers in the presence of divalent metal ions.[7] Subsequently, the thermodynamic principles that govern the

reversible noncovalent association between RNA templates and activated ribonucleotides,[8,9] the kinetic parameters of the subsequent oligomerization,[10,11] and the reaction conditions that ensure the regioselectivity of the 3′- to 5′-RNA polymerization,[12] have been explored.

The hydrogen-bonding interactions between the nucleobases of the template and the incoming activated mononucleotides are instrumental in ensuring complementarity between template and monomers.[6] Intense experimental efforts have been devoted to the observation of the structural nature of such interactions. Despite the weakness of noncovalent template–monomer interactions in aqueous media,[13,14] studies of complexes composed of a polypyrimidine template and monomeric purine nucleotides have provided evidence, including infrared absorption spectra, that the template–monomer complexes resemble fully complementary RNA

Scheme 1. (A) Structures of 2-MeImpG and PZG Analogue Compared Side-by-Side,[a] (B) Stick Model Representations of 2-MeImpG and PZG, and (C) Synthesis of PZG[b]
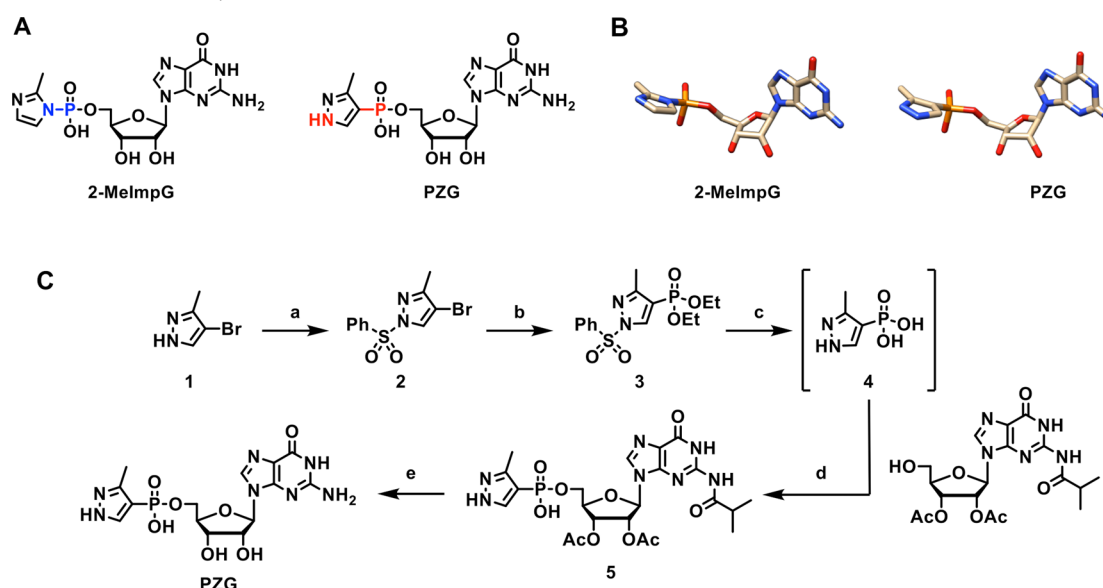
[a]One of the nitrogen atoms of the 3-methylpyrazolyl group (labeled red) is displaced from its corresponding position in 2-MeImpG (labeled blue). [b]Reaction conditions: (a) PhSO$_2$Cl, Et$_3$N, DCM; (b) (EtO)$_2$P(O)H, Pd(PPh$_3$)$_4$, PPh$_3$, Et$_3$N, DMSO, 115 °C, 2 h; (c) TMSBr, Et$_3$N, DCM, 4 h; then MeOH; (d) DIAD, PPh$_3$, DCM, 3 h; (e) 28% NH$_3$ in H$_2$O, MeOH, 65 °C, 4 h.

double helices.[15,16] This observation, together with the high fidelity of G:C template copying, implies that monomers associate with the template with Watson–Crick geometry. Surprisingly, Fujita and colleagues have recently demonstrated that a self-assembling, tripyridyl-triazine capped hydrophobic molecular cage encourages formation of nonconventional anti-Hoogsteen base pairing interactions between adenosine and uridine monophosphate monomers (AMP and UMP), as well as between two self-complementary thymidylyl-adenosine (TA) dinucleotides within the hydrophobic cavity of the cage.[17]

These observations raise the question of whether Watson–Crick interactions indeed dominate the noncovalent association between complementary nucleobases, or whether other types of interaction geometries could be of comparable importance in base pairing. We therefore decided to probe the structural nature of template–monomer recognition under aqueous conditions that are as close as possible to standard reaction conditions, in order to elucidate the structural basis of nonenzymatic RNA replication.

Recent studies have elucidated some of the thermodynamic aspects of the binding interactions between mononucleotides and oligonucleotides. For example, studies from our lab and the Richert lab have used proton NMR spectroscopy to probe the affinity and context dependence of the reversible binding between all four deoxy- and ribonucleotides with complementary oligonucleotide templates, as well as the G:T(U) and C:T mismatches.[8,9] We have also applied transferred nuclear Overhauser effect spectroscopy (TrNOESY) to demonstrate that the sugar pucker of activated guanosine ribonucleotides switches from predominantly C2′-endo in the free state to predominantly C3′-endo when bound to a complementary poly-C RNA template.[18] Interestingly, this phenomenon is observed only in the case of an RNA template and activated ribonucleotide monomers, and not with a DNA template or deoxynucleotide monomers. Despite the fact that these studies provide insight into certain structural aspects of monomer–

template binding, the noncovalent assembly of ribonucleotides onto an RNA template has not yet been probed by X-ray crystallography.

X-ray crystallography remains the dominant technique for revealing the three-dimensional structures of nucleic acids and nucleic acid–ligand complexes at atomic resolution.[19] We have recently used crystallography to study modified RNA duplexes, with the goal of obtaining structural information that could lead to more efficient and accurate nonenzymatic RNA replication. This approach has been successful in demonstrating how the structural perturbations resulting from interspersed 2′–5′-phosphodiester linkages are accommodated by local conformational adjustments in the region flanking the aberrant linkage.[20] We also demonstrated by crystallography that the stabilization of the U:A Watson–Crick base pair by replacing U with 2-thio-U causes minimal perturbation of the overall RNA duplex structure.[21] In order to accurately reflect the structural characteristics of monomer–template binding under conditions resembling standard nonenzymatic RNA replication protocols, we designed and synthesized a stable nonhydrolyzable isosteric 2-MeImpG analogue, and cocrystallized it with LNA-rigidified RNA primer–template complexes. We obtained atomic resolution X-ray crystal structures of this ligand complexed with three variants of self-complementary LNA-modified RNA duplexes that contain either one, two, or three 5-methylcytidine (5-mC) LNA overhangs at their 5′-termini, as binding sites for the guanosine analogue. In addition to canonical Watson–Crick base pairing, we frequently observed base pairing of the guanosine analogue with the templating 5-mC via noncanonical interactions. The activated phosphate moiety of a template-bound monomer was frequently highly disordered, or, in more ordered structures, was too far from the 3′-hydroxyl group of the upstream nucleotide to react without significant conformational change. Our findings suggest that noncanonical base pairing may occur alongside standard Watson–Crick base-pairing interactions in solution, potentially leading to slower

Table 1. Crystal Structure Determination Statistics

	PZG–RNA 15mer	PZG–RNA 7mer	PZG–RNA 14mer	PZG–RNA 13mer	GMP-RNA 15mer	GMP-RNA 7mer	GMP-RNA 14mer
PDB code	5KRG	5HBW	5HBX	5HBY	5L00	5DHB	5DHC
space group	$P3$	$P2_12_12_1$	$P3$	C_12_1	$P3$	$P3$	$P3$
RNA duplexes per asymmetric unit	2	1	2	0.5	2	2	2
resolution range, Å	50−1.60	50−1.90	50−1.70	50−1.18	50−1.25	50−1.80	50−1.55
unique reflns	41492	7254	38632	23597	47622	20608	25146
completeness, %	95.5 (71.3)	99.6 (99)	99.8 (100)	93.6 (54.2)	92.7 (64.5)	98.1 (94.3)	99.9 (100)
R_{work}, %	20.6	21.7	21.3	15.3	20.8	18.7	17.0
R_{free}, %	23.4	25.4	24.9	16.6	23.8	19.2	19.5
av B factors, Å^2	27.54	33.77	29.32	13.29	16.85	31.38	20.59
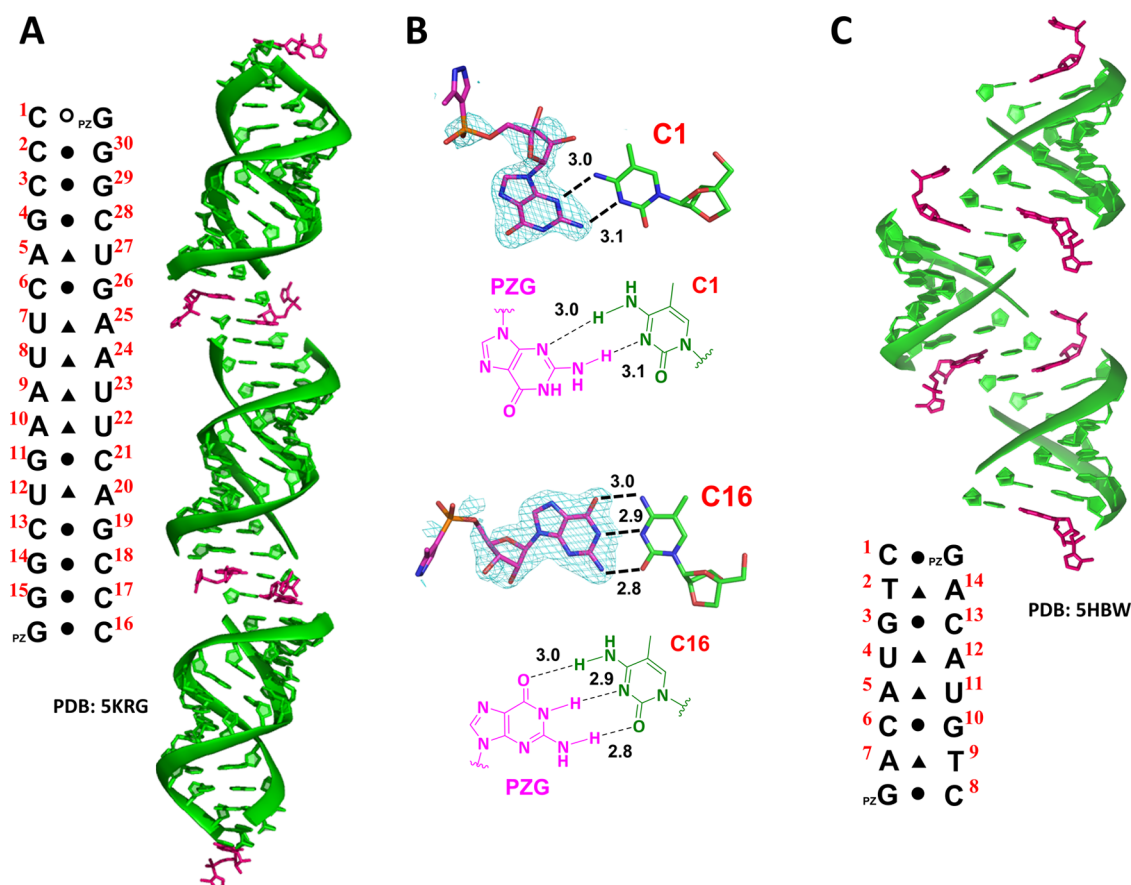
Figure 1. Crystal structures of single binding site RNA−**PZG** complexes. Each duplex has a one nucleotide 5′-overhang at each end that binds one **PZG** residue. Green: RNA duplex. Magenta: **PZG** monomer. Black filled circle: GC Watson−Crick pairs. Black triangle: AU Watson−Crick pairs. Black open circle: noncanonical base pairs with two hydrogen bonds. Hydrogen bonds are labeled as black dashes, and their interatomic distances are indicated. (A) Diagram and structure of duplex formed from a self-complementary RNA 15mer with one 5-mC overhang at each end. The bound **PZG** monomers are stacked between the ends of coaxial RNA duplexes. The sequence of the oligonucleotide is shown and numbered. (B) The **PZG** monomers form two different base pairs with the templating 5-mC residues at opposite ends of the duplex. The corresponding $2F_o − F_c$ omit maps contoured at 1.0σ (cyan mesh) indicate the ordered nucleobases and the disorder of the sugar, phosphate, and pyrazolyl moieties. (C) A self-complementary RNA 7mer with one 5-mC binding site at each end was cocrystallized with **PZG** monomers. The monomers are stacked between coaxial RNA duplexes and Watson−Crick paired to the template.

and less accurate primer extension. Our observations further suggest that primitive ribozyme (and protein) RNA polymerases may enhance the rate and fidelity of template copying by restricting monomer−template interactions to those with Watson−Crick geometry and by imposing structural order and specific conformational changes on the reacting primer terminus and incoming monomer.

■ RESULTS

Synthesis of the 2-MeImpG Pyrazole Analogue PZG. In 2-MeImpG, the 2-methylimidazole leaving group is linked to guanosine monophosphate via a labile nitrogen−phosphorus bond (Scheme 1A, highlighted in blue). Under nonenzymatic RNA copying conditions, divalent metal cations, such as Mg(II) and Mn(II), can behave as Lewis acids and may coordinate with an oxygen atom on the phosphate of 2-MeImpG,[22,23] thus

chemically activating the phosphorus for nucleophilic attack by either the 3′- or 2′-OH group of the RNA primer (leading to primer extension), or by ambient water molecules (leading to hydrolysis), thereby breaking the N−P bond and liberating the 2-methylimidazole leaving group. We designed a novel pseudoactivated monomer by replacing the labile N−P bond with a stable C−P bond. We achieved this by changing the 2-methylimidazoyl group of 2-MeImpG into a 3-methylpyrazoyl group and by connecting the leaving group mimic to the phosphorus via a stable and unreactive phosphonate linkage (Scheme 1A, highlighted in red). The new molecule, guanosine 5′-(3-methyl-1*H*-pyrazol-4-yl)phosphonate (**PZG**), is isosteric to the original 2-MeImpG, as well as having a similar overall structure and aromaticity (Scheme 1B).

The synthetic scheme used to prepare **PZG** is shown in Scheme 1C; the longest linear sequence from commercially available starting materials was five steps. To synthesize **PZG**, two bond-forming events were critical: the C−P linkage of the pyrazolyl-phosphonate, and the O−P bond that bridges the guanosine ribonucleoside and the pyrazolyl-phosphonate. We first protected commercially available bromopyrazole **1** with a benzenesulfonyl protecting group to generate the *N*-phenyl-sulfonyl-protected **2**; the protecting group served to temporarily mask the acidic pyrazole proton, and also to activate the carbon−bromine bond for subsequent C−P bond formation. Next, the protected pyrazolyl-phosphonate **3** was produced by a palladium(0)-catalyzed C−P coupling between **2** and diethyl phosphite, furnishing the critical carbon−phosphorus bond.[24] To connect the pyrazolyl-phosphonate to guanosine, the two ethyl groups on **3** must be removed to unveil the aryl phosphonate in its nucleophilic anionic form. This was achieved by treating **3** with bromotrimethylsilane, followed by decomposing the resultant bis(trimethylsilyl)phosphonate ester with methanol to generate the dianionic aryl phosphonate **4**.[25,26] The guanosine−phosphonate ester linkage was prepared via a Mitsunobu reaction to afford protected guanosine pyrazolyl-phosphonate **5**, followed by global protecting group aminolysis to afford **PZG**.

Structure Determination of RNA Duplexes with Bound PZG. We cocrystallized **PZG** with self-complementary RNA oligonucleotide duplexes which share a common ten-nucleotide duplex core, flanked on both ends by 5′ single-stranded overhangs that provided one, two, or three **PZG** binding sites.[27] The adapted oligonucleotide sequence was originally reported by Beal and co-workers for probing the effects of nucleobase alkynyl and triazolyl modifications on the overall duplex structure.[27] The 5′-overhangs were synthesized with 5-mC LNA nucleotides (denoted as mC in oligomer sequences, *vide infra*). The LNA modification (denoted as bolded nucleotides, *vide infra*) locks the sugar into the 3′-endo conformation and helps to preorganize the single-stranded overhang in the A-form conformation. The pK_a of N3 of 5-mC is close to that of native cytidine (4.45 for C and 4.6 for 5-mC),[28] and 5-mC can form a Watson−Crick base pair with a guanine nucleobase much like canonical C. Together, these modifications facilitated crystallization and allowed us to obtain high-resolution X-ray structures of the resulting oligonucleo-tide−monomer complexes.

We began our exploration of monomer−template binding by examining the crystal structure of the complex formed by **PZG** and the self-dimerizing RNA sequence 5′-**mCmCmC**GA-CUUAAGUCGG-3′, with the single overhanging 5-mC locked-ribonucleotide at the 5′-terminus serving as a binding

site for the **PZG** monomer.[27] The key crystallographic parameters are listed in Table 1 (PDB: 5KRG). There were two RNA duplex−**PZG** molecular complexes per asymmetric unit, thus providing four independent views of **PZG**:template binding. Each **PZG** bound to the 5-mC LNA overhang was sandwiched between the upstream guanine at the 3′-end of the oligonucleotide, as well as the "downstream" 5′-overhanging 5-mC of a neighboring slip-stacked RNA duplex (Figure 1A). The internucleobase spacing between the **PZG** guanine base and both the upstream guanine and the "downstream" 5-mC nucleobase was approximately 3.3 Å based on analysis using the CONTACT program from CCP4.[29] At one end of each duplex, the **PZG** monomer was bound to the templating 5-mC via Watson−Crick base pairing. Remarkably, a noncanonical G:C base pair was observed at the other end of each duplex, mediated by two hydrogen bonds: the guanine N3 of the **PZG** was 3.0 Å from the exocyclic amine of the 5-mC, while the exocyclic amine of the guanine was 3.1 Å from the N3 of the 5-mC[30] (Figure 1B). Unfortunately, the sugar and aryl phosphonate moieties of the **PZG** monomers at both ends of both RNA duplexes were too disordered (nucleobase *B* factor, 30.6; ribose *B* factor, 49; phosphonate *B* factor, 109) to obtain useful information regarding the conformation of the sugar and leaving group mimic moieties of the **PZG** monomer.

In an effort to obtain a structure in which a single bound **PZG** would be more highly ordered, we crystallized a shorter RNA duplex (5′-**mC**TGUACA-3′).[31] The structure was determined at 1.9 Å resolution ($P2_12_12_1$ space group, PDB: 5HBW). An end-to-end helical stacking interaction was also observed as in the previous one-binding-site structure. However, in this case, the **PZG** monomer was bound to the templating 5-mC via Watson−Crick base pairing at both ends of the duplex (H-bond distances: 2.8 to 3.0 Å). Furthermore, the improved electron density fitting of the **PZG** monomers in this structure allowed observation of different ribose con-formations at the two termini of the duplex (Figure 1C). At one end, the **PZG** sugar was in the canonical 3′-endo conformation, as were the sugars in the rest of the A-form RNA duplex. At the other end, however, the ribose was in the 2′-endo conformation. This observation highlights the conformational flexibility of the **PZG** monomer, even when it is bound to the template, and is consistent with our prior TrNOESY spectroscopic studies which showed that 2-methylimidazolide activated guanosine mononucleotide was in the 2′-endo conformation a fraction of the time even when template bound.[18] The distances between the 3′-hydroxyl groups at the two duplex 3′-termini and the phosphorus atoms of **PZG** aryl phosphonates were over 6 Å at both ends (6.3 Å for 3′-endo **PZG** and 6.5 Å for 2′-endo **PZG**); at this distance, the phosphodiester-bond-forming reaction between the RNA primer and the activated monomer would require a major conformational change.

We next cocrystallized the **PZG** monomer with a partially self-complementary RNA 14mer oligonucleotide, 5′-**mCmC**AGACUUAAGUCU-3′,[27] which forms a 12 base-pair duplex with two-nucleotide 5-mCmC 5′-overhangs at both ends. The sequence of the duplex core was similar to that of the 15mer duplex reported above.

The structure was determined to 1.7 Å resolution by molecular replacement using the central core of the 12mer RNA duplex as a search model, with an overall B-factor of 29.32 (PDB: 5HBX). As in the previous structures, the individual double helices were slip-stacked with one another end-to-end (Figure 2A). Overall, two **PZG** monomers were clearly bound
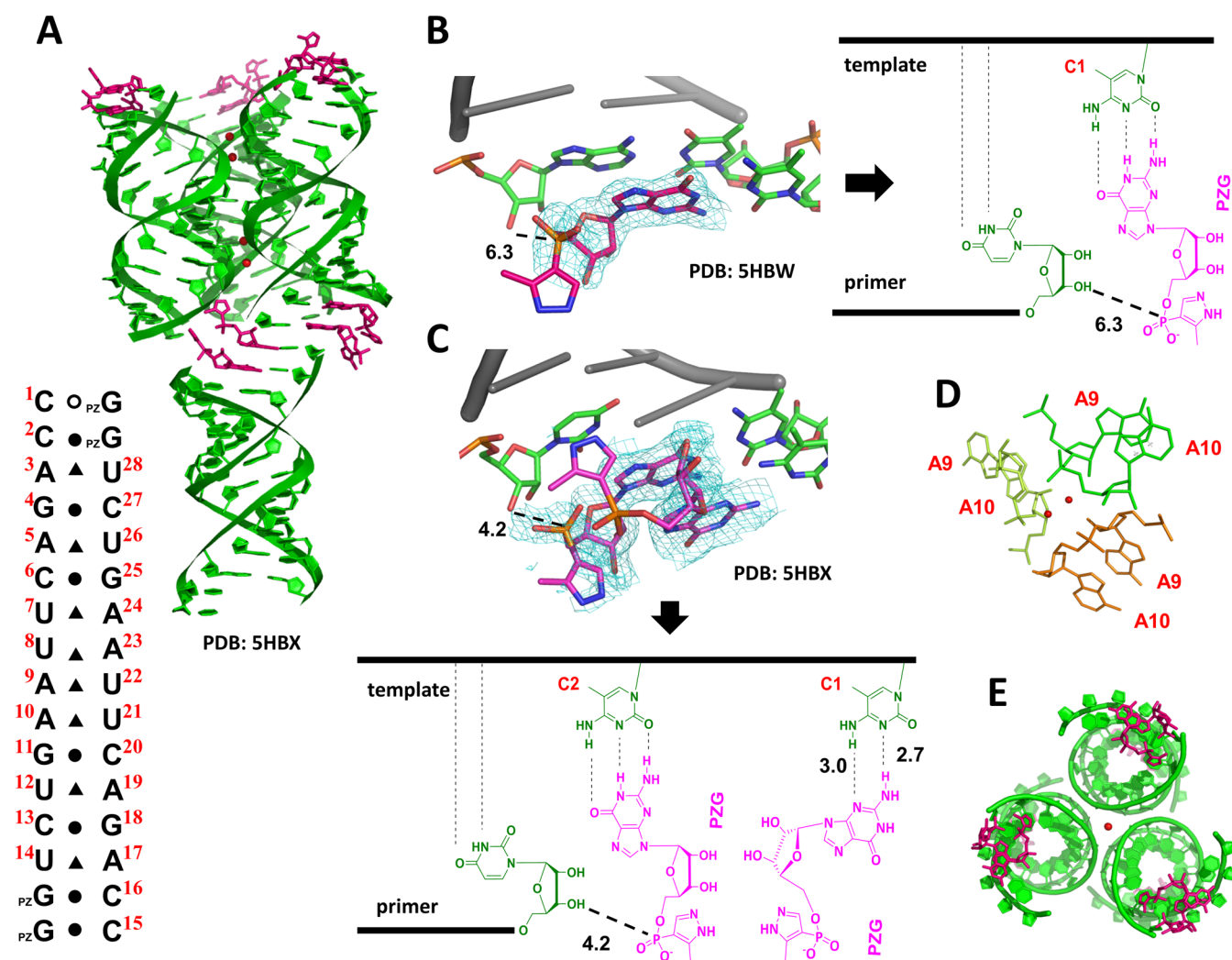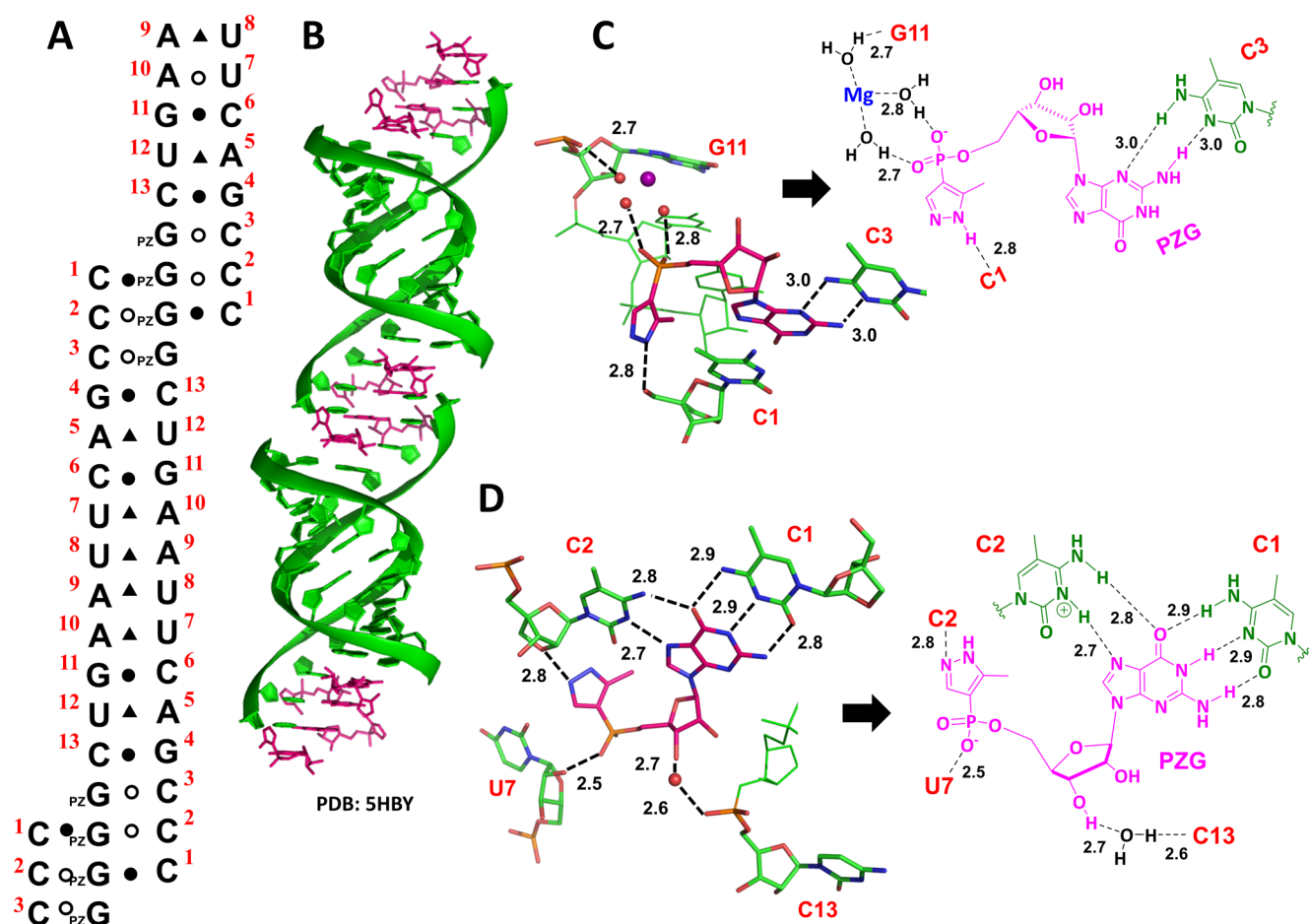
Figure 2. Crystal structures of RNA−**PZG** complexes with 5′-overhangs with two **PZG** binding sites. Green: RNA duplex. Magenta: **PZG** monomer. Red sphere: water molecules. Black filled circle: GC Watson−Crick pairs. Black triangle: AU Watson−Crick pairs. Black open circle: noncanonical base pairs with two hydrogen bonds. (A) A self-complementary RNA 14mer with two 5-mC binding sites was cocrystallized with **PZG** monomer. The monomers are stacked with one another and with the ends of two RNA duplexes. Highly ordered water molecules were observed to bridge and bundle three RNA duplexes into a larger prism-like complex. (B) In the RNA 7mer−**PZG** complex, the distance between the primer 3′-hydroxyl group and the **PZG** phosphorus atom was 6.3 Å (black dashes). The corresponding ChemDraw schematic is shown on the right. (C) In the RNA 14mer−**PZG** complex, the distance between primer 3′-hydroxyl group and the first bound **PZG** phosphorus atom was 4.2 Å (black dashes). The corresponding $2F_o − F_c$ omit maps contoured at 1.0σ (cyan mesh) indicate the ordered nucleobases and partial disorder of the sugar, phosphate, and pyrazolyl moieties. The corresponding ChemDraw schematic is shown below. (D) Water molecules formed hydrogen bonds with three different neighboring RNA duplexes to facilitate molecular assembly. (E) Bird's-eye view of the prism-like complex, with three symmetric RNA duplexes forming a central hollow channel. Water molecules are centrally located within the channel.

at each end of each RNA duplex, fully occupying all available binding sites. The guanine nucleobases of both monomers were bound to the two consecutive 5-mC template sites with clearly defined electron densities, while the sugars and pyrazolyl-phosphonate groups of the monomers were highly disordered (B factors >40).

At one end of each RNA duplex, the guanine nucleobases of the two consecutively bound **PZG**s were well-ordered, coplanar, and Watson−Crick base paired with the consecutive 5-mC LNA nucleotides in the 5′-overhang (hydrogen-bond distances, 2.7 to 2.9 Å; the interplanar distances between the monomer nucleobases and their flanking nucleotides, ∼3.4 Å). The sugars and aryl phosphonates of both monomers were disordered. At the other end of the RNA duplex, however, an alternative base-recognition motif was observed. The nucleobase of the first bound monomer (adjacent to the 3′-end of the

oligonucleotide) was base paired with the template in Watson−Crick fashion (hydrogen-bond lengths, 3.1, 2.8, 2.6 Å; nucleobase B factor, 26.8), and the ribose sugar was in the 3′-endo A-form conformation (B factor: 38.9). The second bound monomer was paired with 5-mC via the previously described noncanonical G:C N3-amino, amino-N3 interactions (Figure 2C).[30] Unfortunately, the sugar and aryl phosphonate moieties of the second bound **PZG** monomer were disordered and not clearly visible. B factor calculation indicated that the first monomer was more structurally ordered than the second one at both ends of the RNA duplex. This phenomenon is consistent with our NMR spectroscopic data[32] suggesting that, when nucleotide monomers bind sequentially to two consecutive binding sites, the affinity of the second bound monomer is lower than that of the first bound monomer. It is noteworthy that the P−O interatomic distance between the aryl

Figure 3. Crystal structures of RNA−**PZG** complexes with 5′-overhangs with three **PZG** binding sites. (A) Sequence of the RNA duplex with three 5-mC binding sites at each end, to which **PZG** monomers could bind. Black filled circle: GC Watson−Crick pairs. Black triangle: AU Watson−Crick pairs. Black open circle: noncanonical base pairs with two hydrogen bonds. (B) The self-complementary RNA duplex was cocrystallized with **PZG** monomers. The monomers bound to the second and third 5-mC binding site bridge two duplexes, constituting a local triplex region. (C) At the first binding site, the **PZG** monomer pairs with a templating 5-mC via a noncanonical base pair, and was connected to the 5′-terminus of another downstream duplex via a hydrogen-bond contact with the pyrazolyl group of **PZG**. A magnesium(II) cation and its chelating water molecules were observed in the major groove, and coordinated the **PZG** monomer with the RNA backbone. The base pairing motif and the additional stabilizing hydrogen-bonding interactions are highlighted in the corresponding ChemDraw schematic. (D) At both the second and the third 5-mC binding sites, a base triple was observed, with the **PZG** monomer forming both Watson−Crick and Hoogsteen base pairs with 5-mC on two different template strands. The conformation of the **PZG** is stabilized by hydrogen-bond contacts with neighboring RNAs. The chemical structure of the base triple is illustrated.

phosphonate of the first bound monomer and the 3′-hydroxyl group of the primer 3′-terminus was only 4.2 Å, compared to the 6.3 Å distance as observed in the one-binding-site 7mer duplex structure (Figure 2B and 2C). While there are significant differences in sequence composition and in the crystallization conditions for the two duplexes, we suggest that the stacking interactions afforded by the second bound monomer may have favored a structural reorganization such that the activated phosphate moiety of the first monomer moved closer to the upstream primer terminus.

In order to explore the possible influences of the pyrazolyl-phosphonate groups on the diversity of monomer−template base-pairing motifs, we decided to revisit the crystal structures of the above RNA duplexes, using guanosine monophosphate (GMP) in lieu of **PZG**. The structures of the 15-, 7-, and 14mer RNA:GMP cocrystals were determined to 1.25 Å, 1.8 Å, and 1.55 Å resolution, respectively (PDB: 5L00, 5DHB, and 5DHC respectively). These structures were all very similar to their respective RNA−**PZG** complex counterparts, including the

GMP-assisted interduplex helical slip-stacking, as well as the presence of both Watson−Crick and noncanonical GC base-pairing motifs (Figure S1). For the 15mer:GMP complex, the ribose and phosphate moieties of GMPs at the terminal binding sites remained disordered, as was the case of the corresponding RNA−**PZG** complex. For the 7mer:GMP complex, however, the conformation of the ribose and phosphate moieties of both template-bound GMPs were much more clearly defined. Unlike the bound **PZG**s, the two monomers were both in the 3′-*endo* conformation, as were the ribose sugars in the RNA duplex region. Finally, in the 14mer RNA−GMP structure, with two GMP molecules binding at each end of the duplex, the ribose and phosphate of the first bound monomer was better ordered than the second bound GMP on both ends of the duplex. As in the corresponding RNA−**PZG** complexes, the phosphate group of the first bound GMP was found to be much closer to the primer 3′-terminus (roughly 4.2 Å) than in the one-binding-site structure (6.3 Å) (Figure S2). Our results indicate that our observations of both (i) diverse internucleobase binding motifs
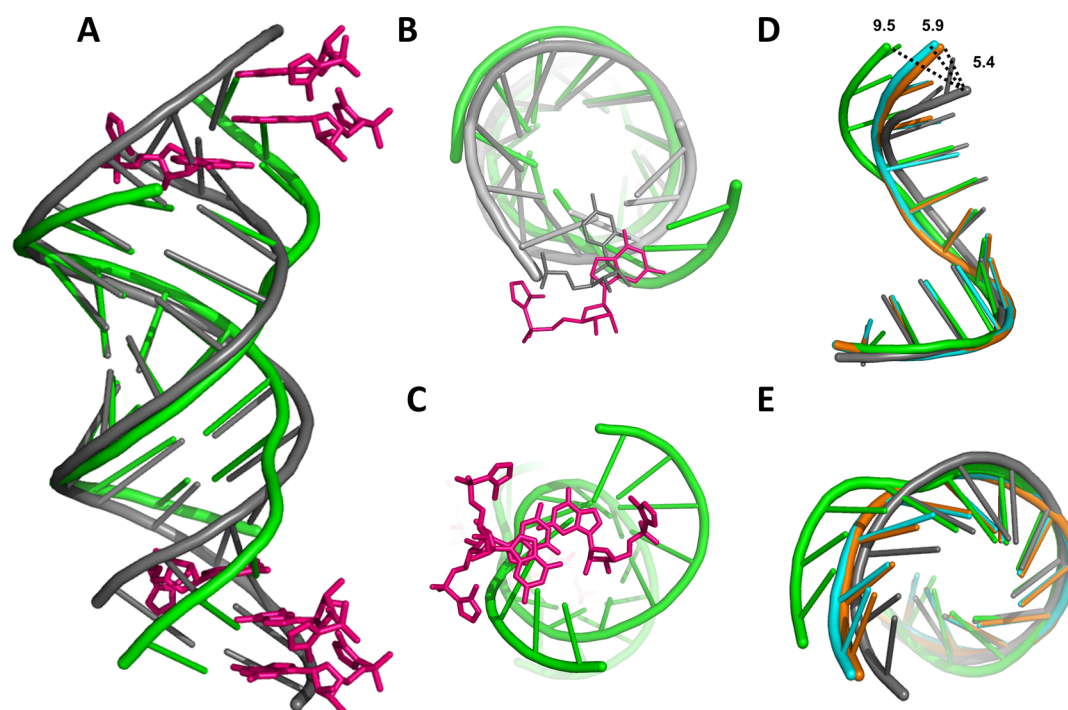
Figure 4. Superimposed structures of the native full-length RNA and the RNA−**PZG** complex. Magenta: **PZG** monomer. Gray: native full-length RNA. Green: RNA duplex with three 5-mC binding sites. Cyan: RNA duplex with two 5-mC binding sites. Orange: RNA duplex with one 5-mC binding site. (A) Superimposed structures of the native full-length RNA and the RNA−**PZG** complex with 3 binding sites. (B) Bird's-eye view of the superimposed structures showing location of the second bound **PZG** monomer within the minor groove. (C) The three bound **PZG**s orient their leaving group mimics in different trajectories. (D, E) Side view (D) and bird's-eye view (E) of the superimposed structures comprising of full-length RNA, complex with one binding site, complex with two binding sites, and complex with three binding sites. The 5′-hydroxyl termini of the **PZG**− RNA complexes are unwound from the central helical axis, and are shifted by 5.4, 5.9, and 9.5 Å from the corresponding position in the native full-length structure (one, two, and three **PZG**s bound, respectively). This generated a wider helix diameter with more open space for monomers.

between the template and bound monomer and (ii) the decreased primer−monomer distance afforded by a downstream bound monomer can probably occur irrespective of the chemical constituents present on the phosphate moiety of the bound monomers.

Three Distinct PZG Binding Modes on a Three-Nucleotide Template. We cocrystallized the **PZG** monomer with a partially self-complementary RNA 13mer oligonucleotide with a triple 5-mC 5′-overhang (5′-**mCmCmC**GACUUA-AGUC-3′, Figure 3A). To our surprise, the crystal diffracted to 1.2 Å resolution despite having a relatively flexible three-nucleotide 5′-overhang. The high-resolution diffraction data allowed us to solve the structure by *ab initio* methods instead of by molecular replacement as performed previously.[33] An initial model, comprising a single chimeric RNA strand and two GMP molecules, was derived solely from our high-resolution diffraction data, and was subsequently refined. The overall RMSD value between the initial model and the final structure was 1.6 Å. In contrast to the one- and two-binding-site structures elucidated above, the RNA 13mer−**PZG** complex crystallized with space group C_121; one asymmetric unit contained one single-stranded RNA and two **PZG** monomer molecules (PDB: 5HBY). Each self-complementary RNA duplex bound to a total of six **PZG** molecules, three at each end, with all six monomers fitting very well to their corresponding electron densities (*B* factor <20 for all six monomers) and with all ribose moieties in the 3′-*endo* conformation. Two of the three **PZG** molecules bound to each end of the duplex were shared with an adjacent duplex,

thus explaining the ratio of one oligonucleotide to two **PZG** monomers.

We observed multiple noncanonical base-pairing interactions between the bound **PZG** monomers and the template 5-mC residues. At the first binding site, adjacent to the 3′-end of the oligonucleotide (mC3 in Figure 3A), the guanine nucleobase of **PZG** paired with the 5-mC template via the same N3-amino, amino-N3 interaction[34] as described above. However, in this higher resolution structure, we were able to observe additional interactions that may also contribute to the stabilization of this unusual structure. The N2 of the pyrazolyl group made a contact with the 5′-OH of the downstream 5-mC nucleobase from another strand (distance 2.8 Å). In addition, a highly ordered hexaaqua-magnesium(II) cation was observed in the major groove of the duplex. Two of the magnesium-coordinated water molecules hydrogen bonded with the two oxygen atoms of the phosphonate moiety of **PZG**, while another coordinating water molecule hydrogen bonded with the *pro*-R_P oxygen atom of the A10−G11 phosphodiester linkage (Figure 3C). These observations demonstrate that divalent magnesium ions, together with inner shell coordinated water molecules, could serve to stabilize the binding of the monomer via interactions with the backbone of the duplex RNA to which the monomer is bound.

At the second 5-mC binding site (mC2 in Figure 3A), the guanine nucleobase of the **PZG** was base paired with the template 5-mC in a Hoogsteen fashion (the two hydrogen bonds are 2.7 and 2.8 Å, Figure 3). In addition, N1 of the **PZG** pyrazolyl phosphonate was within hydrogen-bonding distance of O2′ of the 4′-C-methylene-linked bridge of the mC-LNA

ribose. Additional hydrogen-bond contacts from nucleotides of adjacent RNA strands also appeared to stabilize this unusual mode of monomer:template binding: the *pro*-R$_P$ nonbonded oxygen from the **PZG** aryl phosphonate interacted with the 2′-hydroxyl oxygen of U7 from an adjacent duplex, and a water molecule bridged the 3′-OH of **PZG** and the *pro*-S$_P$ nonbonded oxygen of the U12−C13 phosphodiester linkage from another duplex. Only at the third and terminal binding site (mC1 in Figure 3A) was the nucleobase of the **PZG** monomer observed to base pair with the template mC in Watson−Crick fashion (with distances of ∼2.8 Å, Figure 3D). This **PZG** monomer was shared by two RNA strands in the same manner as the **PZG** at the second binding site. It was striking that, at the primer +2 and +3 positions, the **PZG** monomers bridged two RNA strands by utilizing all available hydrogen-bond donors and acceptors on the Watson−Crick and Hoogsteen faces of the guanine nucleobase (N1, N7, the exocyclic amine, and the carbonyl group), so as to constitute two stacked planar C·G·C$^+$ triplex motifs.[13,35,36] These two base triples are related by symmetry, and therefore possess identical H-bond interactions and geometry. This local triple helix motif, consisting of both Hoogsteen and Watson−Crick base pairing, a local hydrogen-bonding network, and expanded internucleotide stacking interactions, constitute a highly ordered platform for monomer binding at the +1 position, and for nucleation of the crystallization process (Figure 3D).

Overall Topology of the One-, Two-, and Three-Binding-Site Complexes. Following elucidation of various monomer−template base-pairing motifs, we have compared the overall topology and crystallization pattern of the duplexes containing one, two, and three 5-mC binding sites with the corresponding native full-length RNA duplex. The RNA 15mer and 14mer cocrystallized with either **PZG** or GMP in an almost identical manner, with the resulting RNA−monomer complexes sharing the same space group and asymmetric unit presumably because of the overall sequence and length identity. For example, in the RNA 14mer−**PZG** complex structure, groups of three adjacent duplexes coordinated four highly ordered water molecules (Figure 2A, red dots) within the central channel formed by the triangular prism (Figure 2E). These structured water molecules bridged adjacent duplexes by hydrogen bonding, thus ensuring facile assembly of these duplexes into larger structures. Two of the four water molecules formed three 2.6 Å hydrogen-bond contacts with the three surrounding duplexes via the 2′-hydroxyl of A9 of the RNA sequence, while the remaining two water molecules each formed three 2.8 Å hydrogen-bond contacts with the *pro*-S$_P$ nonbonded oxygen atoms of the A10−G11 phosphodiester linkages of the three prism-forming duplexes. The local structure and coordination environment for these water molecules is depicted in Figure 2D. The RNA 15mer−**PZG** complex crystallized in a similar manner, sharing comparable molecular contacts and overall structural assembly.

In contrast to the 15mer and 14mer structures, the 13mer RNA duplex, with three binding sites on its two ends, crystallized with different symmetry and a distinctive packing pattern. Instead of forming water-bridged RNA prisms, four highly ordered water molecules were observed to bridge and stabilize **PZG** monomers, as well as connecting them to both a neighboring RNA and the RNA duplex they were bound to (Figure 3C and 3D).

A detailed comparison of the overall topology of these RNA−monomer structures against the original full length RNA[27] was then generated by superimposing their RNA duplex backbones. Figures 4A and 4B show the superimposed structures of the previously determined full-length RNA 16mer (PDB ID: 4NFO) and our three-binding-site RNA:**PZG** complex. Figures 4A (side view), 4B, and 4C (bird's-eye view) show that the first incoming **PZG** monomer bound to the template by having the guanine base slide directly underneath the primer in a nucleobase-first, "head-on" fashion to engage in non-Watson−Crick base−base recognition. The second and third monomers clearly reside on the wide and shallow minor groove of the RNA duplex, with the nucleobases of **PZG** twisted and tilted so to open up the minor groove and unwind the duplex (Figures 4A and 4B). The ribose and pyrazolyl phosphonate groups of the first bound monomer remained buried inside the major groove of the duplex, while those of the second and third monomers were solvent-exposed and pointed toward the periphery of the duplex with different trajectories (Figure 4C). In this conformation, the **PZG** monomers can interact with neighboring RNAs or water molecules. Compared with the native and fully complementary RNA duplex, the 5′-terminus of the one-binding-site duplex overhang shifted by 5.4 Å, and that of the two-binding-site complex shifted by 5.9 Å, while the 5′-terminus of the three-binding-site complex shifted by 9.5 Å to create additional space in the minor groove for monomer binding. The superimposed structures, together with the measured distances between terminal 5′-hydroxyl groups, are shown in Figures 4D and 4E.

■ DISCUSSION

Given the dominance of the Watson−Crick base-pairing motif in nucleic acid replication chemistry, it has long been thought that the binding of activated monomers to RNA templates would occur predominantly via Watson−Crick base pairing.[6] Indeed, in the context of protein polymerase-assisted RNA or DNA copying, a multitude of strategically located hydrophobic, electrostatic, and hydrogen-bonding interactions constrain the noncovalent interaction of the RNA template and incoming mononucleotide substrates,[37] and only Watson−Crick base pairing geometry is seen, because this recognition motif is the basis of information transfer during nucleic acid replication. However, under the prebiotic conditions of nonenzymatic RNA replication, where these enzyme assisted monomer−template interactions are absent, the many hydrogen-bond donor and acceptor groups on the nucleobases allow pairing to occur in a multitude of ways in addition to the usual Watson−Crick constraints. This is especially true when multiple mononucleotides are binding to a single-stranded template, and the lack of spatial and/or conformational constraints allows nucleotide binding in unusual structural modes. In fact, hydrogen bonding between complementary base pairs can occur via nonconventional motifs even when spatial constraints and/or assistive hydrophobic interactions are present: for pairs of complementary monoribonucleotides and deoxyribodinucleotides that were stabilized within the hydrophobic cavity formed by the triazine caps and bipyridyl pillars of a molecular cage, Fujita and co-workers exclusively observed non-Watson−Crick base pairing between the nucleobases of the interacting molecules.[17] Arguably, in the context of monomer:template interactions, the perceived energetic and/or conformational preference of Watson−Crick base pairing over other noncanonical mechanisms may not be as pronounced as previously thought. Our crystallographic studies suggest that the binding of activated monomers to RNA templates might involve, at least transiently,

non-Watson—Crick base pairing during nonenzymatic RNA copying in solution. If they do occur in solution, some of the various binding modes of incoming monomers could, in the worst cases, lead to errors in nonenzymatic RNA polymerization and subsequent stalling of primer extension.[38] Alternatively, some non-Watson—Crick binding modes could also potentially lead to unproductive conformations which displace the activated phosphate of the incoming monomer away from the primer terminus, thus retarding the rate of nonenzymatic RNA polymerization. Strategies to reduce the high error rates seen in nonenzymatic RNA replication are necessary to enable the expansion of the genome, so that larger numbers of ribozymes could be encoded. Ribozyme polymerases, by enforcing Watson—Crick pairing of incoming monomers to the template via geometrical constraints, could significantly increase the fidelity of early RNA replication, thus increasing the coding capacity of primitive RNA genomes.

The experiments described herein represent our initial attempt to derive a structural basis for understanding some of the fundamental mechanisms involved in nonenzymatic RNA copying. In the partial duplexes with one or two nucleotide overhangs, the bound **PZG** monomers are in some (but not all) cases bound in Watson—Crick fashion to the template C residue. In these cases, the distance from the O3′ of the last nucleotide of the RNA oligonucleotide to the phosphorus of the incoming monomer ranges from 4.5 to 6.5 Å. This distance is far too long to allow for attack of the 3′-hydroxyl on the activated phosphate, as required for primer extension. This suggests that our structures represent a snapshot of the ground state conformations of the bound monomer, and that a significant conformational rearrangement that brings O3′ closer to the reactive P atom must precede initiation of the primer extension reaction. This conformational change could be driven, for example, by binding of the catalytic divalent metal ion, which could potentially bridge the 3′-hydroxyl of the primer and the phosphate of the incoming monomer. Another contributing factor could be a difference in conformation when one vs two monomers are bound to the template. It is well-known that the rate of the primer extension reaction is much faster when two or more monomers are bound to the template than when only a single monomer is present.[39−41] In our structure of the 14mer duplex with single nucleotide overhangs at either end, the monomer pyrazolyl-phosphate is too disordered to measure the O3′−P distance, however in the shorter 6mer duplex with single nucleotide overhangs this distance is 6.5 Å. In our structure of the 12mer duplex flanked by two nucleotide overhangs, the end in which both **PZG**s are Watson—Crick paired with the template is again too disordered to determine the O3′−P distance. However, at the other end, the first **PZG** is Watson—Crick paired, with an O3′−P distance of 4.5 Å, still far too long for any reaction to occur, but significantly shorter than in the case with a single bound **PZG**. The presence of the second, non-Watson—Crick-paired **PZG** appears to have pushed the first **PZG** closer to the 3′-OH of the "primer", possibly representing an initial step along the reaction coordinate. Another possibility is that the reactive conformation is only achieved upon binding, or synthesis on the template, of the imidazolium-bridged dinucleotide reaction intermediate that we have recently described.[42] Additional structures may clarify the role of downstream monomers or oligonucleotides and intermediates, in preorganizing the primer—template—monomer complex for reaction.

We have recently shown that primer extension with an activated monomer can be catalyzed by a downstream activated monomer or oligonucleotide.[41] These observations, together with related earlier observations by Orgel,[39,40] strongly suggested the possibility of a physical interaction between the leaving groups of adjacent template-bound activated monomers. In the primer—template—**PZG** complexes described above, the pyrazolyl-phosphonate moieties of most template-bound **PZG** monomers are disordered, making it impossible to observe any potential noncovalent interactions between the pyrazolyl moieties of consecutively bound monomers. Such interactions, if they occur, may be too transient to see in a crystal structure. Alternatively, the **PZG** analogue, in spite of mimicking the overall structure of 2-MeImpG (Figures 1A and 1B), may still be too different from 2-MeImpG to exhibit the catalytically relevant interactions. For example, the pK_a values for pyrazoles and imidazoles are quite different.[43] In addition, pyrazole has a charge distribution skewed toward the NH—N single bond, while the charge distribution in the imidazole ring is symmetrically shared between the two nitrogen atoms. The electronic and chemical differences between **PZG** and 2-MeImpG could prevent adjacent **PZG**s from forming stable leaving group—leaving group interactions *in crystallo*. We are currently examining other inert 2-MeImpG structural analogues that better reflect the chemical and electronic characteristics of 2-MeImpG, and may serve to illustrate the precise role of leaving group—leaving group interactions in primer extension reactions. Finally, as noted above, we have recently obtained evidence suggesting that primer extension proceeds through a covalent intermediate generated by the reaction of two monomers to form an imidazolium-bridged dinucleotide, which is subsequently attacked by the primer 3′-hydroxyl, resulting in primer extension by one nucleotide with concomitant release of an activated nucleotide as the leaving group.[42] We are currently examining the structure of stable analogues of the proposed intermediate when bound to an RNA template.

In this report, we used locked nucleic acid and 5-methylcytidine-modified RNA as the monomer binding template to enhance crystallization efficiency and complex stability. Although these modifications provide useful models for the study of monomer—template binding during nonenzymatic RNA replication, it is also possible that these structural modifications may introduce artifacts that do not necessarily reflect solution-phase activated monomer-driven RNA replication. For example, the higher conformational rigidity of the LNA overhang could facilitate the formation of weak non-Watson—Crick base pairs between **PZG** and the template. Finally, the high effective molarity of RNA duplexes and guanosine monomers in the RNA crystals implies that some of the *in crystallo* interactions we observed—for example, the stabilization of **PZG** monomers in the 15mer RNA complex by a neighboring slip-stacked RNA duplex, or the simultaneous noncovalent interactions observed between the **PZG** monomer and two RNA strands in the three-binding-site complex—result from the close intermolecular contacts during crystallization, and may not be relevant to monomer—template recognition in dilute aqueous solution.

One of the noncanonical base pairs that we observed several times is the C·G N3-amino, amino-N3 interaction.[30,34] This base-pairing motif belongs to the family of *trans* Watson—Crick/sugar-edge interactions,[34] and has been observed in larger RNA complexes such as the group I self-splicing intron of

Tetrahymena thermophila,[44,45] and the peptidyl transferase center[46] of the 23*S* rRNA embedded within the 50S large ribosomal subunit from *Haloarcula marismortui*.[47] On the other hand, Hoogsteen base pairing,[48] as observed in the three-binding-site RNA complex, is arguably more prevalent in nature, as evidenced by its integral roles in the formation of various triplex and quadruplex motifs. Recently, NMR studies have demonstrated that, at any given time, ∼0.1−1% of Watson−Crick base pairs in a duplex DNA can transiently open up and switch to a Hoogsteen-like conformation, with the Hoogsteen-like base pairs having lifetimes on the order of milliseconds.[48] Considered together, these experimental observations demonstrate the possible omnipresence of non-canonical base-pairing motifs in the prebiotic world and contemporary biology.

In summary, we have designed and synthesized the pyrazolyl-phosphonate nucleotide **PZG**, which is an inert structural mimic of 2-MeImpG. High-resolution X-ray crystal structures of this ligand bound to RNA duplexes with various overhang lengths revealed an unexpected variety of monomer−template binding motifs. Our work reveals that, in addition to canonical Watson−Crick base pairing, the guanine base of **PZG** can engage with RNA templates via noncanonical binding motifs by taking advantage of all available hydrogen-bond donor and acceptor groups of the nucleobase. Our work demonstrates the utility of rational analogue design and X-ray crystallography in teasing out mechanistic aspects of nonenzymatic RNA replication. We are currently expanding our use of crystallography to explore the mechanistic roles of divalent metal ions and monomer−monomer interactions during phosphodiester bond formation.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acscentsci.6b00278.

> Synthetic procedures, descriptions, and characterizations of compounds (PDF)

Crystallographic data of the duplexes described herein are deposited on the RCSB PDB with the following accession codes: 5KRG, 5HBW, 5HBX, 5HBY, 5L00, 5DHB, 5DHC.

AUTHOR INFORMATION

Corresponding Author

*E-mail: szostak@molbio.mgh.harvard.edu.

Author Contributions

§W.Z. and C.P.T. contributed equally.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Crick, F. H. The origin of the genetic code. *J. Mol. Biol.* **1968**, *38*, 367−379.

(2) Orgel, L. E. Evolution of the genetic apparatus. *J. Mol. Biol.* **1968**, *38*, 381−393.

(3) Woese, C. R. *The Genetic Code: The Molecular Basis for Genetic Expression*; Harper & Row: New York, 1967; pp 179−195.

(4) Szostak, J. W. The eightfold path to non-enzymatic RNA replication. *J. Syst. Chem.* **2012**, *3*, 2.

(5) Robertson, M. P.; Joyce, G. F. The origins of the RNA world. *Cold Spring Harbor Perspect. Biol.* **2012**, *4*, a003608.

(6) Orgel, L. E. Prebiotic chemistry and the origin of the RNA world. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39*, 99−123.

(7) Inoue, T.; Orgel, L. E. Oligomerization of (guanosine 5′-phosphor)-2-methylimidazolide on poly (C): an RNA polymerase model. *J. Mol. Biol.* **1982**, *162*, 201−207.

(8) Izgu, E. C.; Fahrenbach, A. C.; Zhang, N.; Li, L.; Zhang, W.; Larsen, A. T.; Blain, J. C.; Szostak, J. W. Uncovering the Thermodynamics of Monomer Binding for RNA Replication. *J. Am. Chem. Soc.* **2015**, *137*, 6373−6382.

(9) Kervio, E.; Claasen, B.; Steiner, U. E.; Richert, C. The strength of the template effect attracting nucleotides to naked DNA. *Nucleic Acids Res.* **2014**, *42*, 7409−7420.

(10) Kanavarioti, A.; Bernasconi, C. F.; Alberas, D. J.; Baird, E. E. Kinetic dissection of individual steps in the poly (C)-directed oligoguanylate synthesis from guanosine 5′-monophosphate 2-methylimidazolide. *J. Am. Chem. Soc.* **1993**, *115*, 8537−8546.

(11) Kanavarioti, A.; Bernasconi, C. F.; Baird, E. E. Effects of monomer and template concentration on the kinetics of nonenzymatic template-directed oligoguanylate synthesis. *J. Am. Chem. Soc.* **1998**, *120*, 8575−8581.

(12) Inoue, T.; Orgel, L. E. Substituent control of the poly(C)-directed oligomerization of guanosine 5′-phosphoroimidazolide. *J. Am. Chem. Soc.* **1981**, *103*, 7666−7667.

(13) Saenger, W. *Principles of Nucleic Acid Structure*; Springer-Verlag: New York, 1984; pp 116−158.

(14) Philp, D.; Stoddart, J. F. Self-assembly in natural and unnatural systems. *Angew. Chem., Int. Ed. Engl.* **1996**, *35*, 1154−1196.

(15) Howard, F. B.; Frazier, J.; Lipsett, M. N.; Miles, H. Y. Infrared demonstration of two- and three-strand helix formation between poly C and guanosine mononucleotides and oligonucleotides. *Biochem. Biophys. Res. Commun.* **1964**, *17*, 93−102.

(16) Howard, F. B.; Frazier, J.; Singer, M. F.; Miles, H. T. Helix formation between polyribonucleotides and purines, purine nucleosides and nucleotides. II. *J. Mol. Biol.* **1966**, *16*, 415−439.

(17) Sawada, T.; Yoshizawa, M.; Sato, S.; Fujita, M. Minimal nucleotide duplex formation in water through enclathration in self-assembled hosts. *Nat. Chem.* **2009**, *1*, 53−56.

(18) Zhang, N.; Zhang, S.; Szostak, J. W. Activated ribonucleotides undergo a sugar pucker switch upon binding to a single-stranded RNA template. *J. Am. Chem. Soc.* **2012**, *134*, 3691−3694.

(19) Egli, M. Nucleic acid crystallography: current progress. *Curr. Opin. Chem. Biol.* **2004**, *8*, 580−591.

(20) Sheng, J.; Li, L.; Engelhart, A. E.; Gan, J.; Wang, J.; Szostak, J. W. Structural insights into the effects of 2′-5′ linkages on the RNA duplex. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 3050−3055.

(21) Sheng, J.; Larsen, A.; Heuberger, B. D.; Blain, J. C.; Szostak, J. W. Crystal structure studies of RNA duplexes containing s2U:A and s2U:U base pairs. *J. Am. Chem. Soc.* **2014**, *136*, 13916−13924.

(22) Lohrmann, R.; Bridson, P. K.; Orgel, L. E. Efficient metal-ion catalyzed template-directed oligonucleotide synthesis. *Science* **1980**, *208*, 1464−1465.

(23) Dunn, M. F. Mechanisms of zinc ion catalysis in small molecules and enzymes. *Struct. Bonding (Berlin)* **1975**, *23*, 61−122.

(24) Kalek, M.; Ziadi, A.; Stawinski, J. Microwave-assisted palladium-catalyzed cross-coupling of aryl and vinyl halides with H-phosphonate diesters. *Org. Lett.* **2008**, *10*, 4637−4640.

(25) McKenna, C. E.; Higa, M. T.; Cheung, N. H.; McKenna, M.-C. The facile dealkylation of phosphonic acid dialkyl esters by bromotrimethylsilane. *Tetrahedron Lett.* **1977**, *18*, 155−158.

(26) Błażewska, K. M. McKenna reaction − which oxygen attacks bromotrimethylsilane? *J. Org. Chem.* **2014**, *79*, 408−412.

(27) Phelps, K. J.; Ibarra-Soza, J. M.; Tran, K.; Fisher, A. J.; Beal, P. A. Click modification of RNA at adenosine: structure and reactivity of 7-ethynyl-and 7-triazolyl-8-aza-7-deazaadenosine in RNA. *ACS Chem. Biol.* **2014**, *9*, 1780−1787.

(28) Fox, J. J.; Van Praag, D.; Wempen, I.; Doerr, I. L.; Cheong, L.; Knoll, J. E.; Eidinoff, M. L.; Bendich, A.; Brown, G. B. Thiation of Nucleosides. II. Synthesis of 5-methyl-2′-deoxycytidine and Related Pyrimidine Nucleosides. *J. Am. Chem. Soc.* **1959**, *81*, 178−187.

(29) Krissinel, E.; Winn, M.; Ballard, C.; Ashton, A.; Patel, P.; Potterton, E.; McNicholas, S.; Cowtan, K.; Emsley, P. The new CCP4 Coordinate Library as a toolkit for the design of coordinate-related applications in protein crystallography. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2250−2255.

(30) Lemieux, S.; Major, F. RNA canonical and non-canonical base pairing types: a recognition method and complete repertoire. *Nucleic Acids Res.* **2002**, *30*, 4250−4263.

(31) Thota, N.; Li, X. H.; Bingman, C.; Sundaralingam, M. High-resolution refinement of the hexagonal A-DNA octamer d-(GTGTACAC) at 1.4 Å. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1993**, *49*, 282−291.

(32) Tam, C. P.; Fahrenbach, A. C.; Björkbom, A.; Prywes, N.; Izgu, E. C.; Szostak, J. W. Downstream oligonucleotides strongly enhance the affinity of GMP to RNA primer-template complexes. Manuscript submitted.

(33) Foadi, J.; Woolfson, M. M.; Dodson, E. J.; Wilson, K. S.; Jia-xing, Y.; Chao-de, Z. A flexible and efficient procedure for the solution and phase refinement of protein structures. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2000**, *56*, 1137−1147.

(34) Leontis, N. B.; Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **2001**, *7*, 499−512.

(35) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I. *Nucleic Acids: Structures, Properties, and Functions*; University Science Books: Sausalito, CA, 2000; pp 259−334.

(36) Neidle, S. *Principles of Nucleic Acid Structure*, 1st ed.; Academic Press: London, 2008; pp 81−131.

(37) Westover, K. D.; Bushnell, D. A.; Kornberg, R. D. Structural basis of transcription: nucleotide selection by rotation in the RNA polymerase II active center. *Cell* **2004**, *119*, 481−489.

(38) Rajamani, S.; Ichida, J. K.; Antal, T.; Treco, D. A.; Leu, K.; Nowak, M. A.; Szostak, J. W.; Chen, I. A. Effect of stalling after mismatches on the error catastrophe in nonenzymatic nucleic acid replication. *J. Am. Chem. Soc.* **2010**, *132*, 5880−5885.

(39) Wu, T.; Orgel, L. E. Nonenzymatic template-directed synthesis on hairpin oligonucleotides. 2. Templates containing cytidine and guanosine residues. *J. Am. Chem. Soc.* **1992**, *114*, 5496−5501.

(40) Wu, T.; Orgel, L. E. Nonenzymatic template-directed synthesis on hairpin oligonucleotides. 3. Incorporation of adenosine and uridine residues. *J. Am. Chem. Soc.* **1992**, *114*, 7963−7969.

(41) Prywes, N.; Blain, J. C.; Del Frate, F.; Szostak, J. W. Nonenzymatic copying of RNA templates containing all four letters is catalyzed by activated oligonucleotides. *eLife* **2016**, DOI: 10.7554/eLife.17756.

(42) Walton, T.; Szostak, J. W. A highly reactive imidazolium-bridged dinucleotide intermediate in nonenzymatic RNA primer extension. *J. Am. Chem. Soc.* **2016**, *138*, 11996−12002.

(43) Joule, J. A.; Mills, K. *Heterocyclic Chemistry*; 5th ed.; Wiley-Blackwell: Hoboken, NJ, 2013; pp 56−57.

(44) Juneau, K.; Podell, E.; Harrington, D. J.; Cech, T. R. Structural basis of the enhanced stability of a mutant ribozyme domain and a detailed view of RNA-solvent interactions. *Structure* **2001**, *9*, 221−231.

(45) Battle, D. J.; Doudna, J. A. Specificity of RNA-RNA helix recognition. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 11676−11681.

(46) Puglisi, E. V.; Green, R.; Noller, H. F.; Puglisi, J. D. Structure of a conserved RNA component of the peptidyl transferase centre. *Nat. Struct. Biol.* **1997**, *4*, 775−778.

(47) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* **2000**, *289*, 905−920.

(48) Nikolova, E. N.; Zhou, H.; Gottardo, F. L.; Alvey, H. S.; Kimsey, I. J.; Al-Hashimi, H. M. A historical account of Hoogsteen base-pairs in duplex DNA. *Biopolymers* **2013**, *99*, 955−968.