Uncovering the Thermodynamics of Monomer Binding for RNA Replication

Enver Cagri Izgu,[†,‡,∇] Albert C. Fahrenbach,[†,‡,§,∇] Na Zhang,[†,∥] Li Li,[†,‡] Wen Zhang,[†,‡] Aaron T. Larsen,[†,‡] J. Craig Blain,[†,⊥] 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

[§]Earth-Life Science Institute, Tokyo Institute of Technology, 2-12-1-IE-1 Ookayama, Meguro-ku, Tokyo 152-8550, Japan

[#]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford St., Cambridge, Massachusetts 02138, United States

ⓢ *Supporting Information*

ABSTRACT: The nonenzymatic replication of primordial RNA is thought to have been a critical step in the origin of life. However, despite decades of effort, the poor rate and fidelity of model template copying reactions have thus far prevented an experimental demonstration of nonenzymatic RNA replication. The overall rate and fidelity of template copying depend, in part, on the affinity of free ribonucleotides to the RNA primer–template complex. We have now used ^{1}H NMR spectroscopy to directly measure the thermodynamic association constants, K_as, of the standard ribonucleotide monophosphates (rNMPs) to native RNA primer–template complexes. The binding affinities of rNMPs to duplexes with a complementary single-nucleotide overhang follow the order C

> G > A > U. Notably, these monomers bind more strongly to RNA primer–template complexes than to the analogous DNA complexes. The relative binding affinities of the rNMPs for complementary RNA primer–template complexes are in good quantitative agreement with the predictions of a nearest-neighbor analysis. With respect to G:U wobble base-pairing, we find that the binding of rGMP to a primer–template complex with a 5′-U overhang is approximately 10-fold weaker than to the complementary 5′-C overhang. We also find that the binding of rGMP is only about 2-fold weaker than the binding of rAMP to 5′-U, consistent with the poor fidelity observed in the nonenzymatic copying of U residues in RNA templates. The accurate K_a measurements for ribonucleotides obtained in this study will be useful for designing higher fidelity, more effective RNA replication systems.
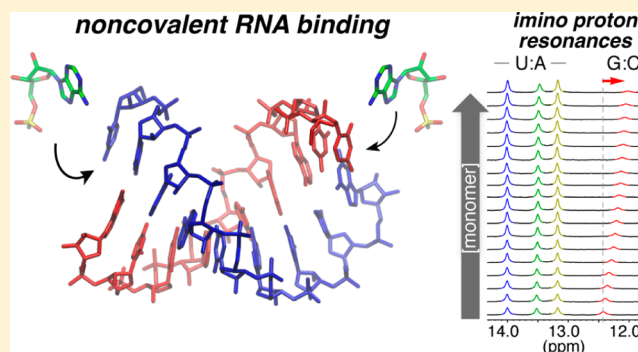
■ INTRODUCTION

RNA has the dual capability of serving as a genetic polymer that mediates the inheritance of information and as a catalyst for chemical transformations and is therefore a promising candidate as an ancestral biopolymer.[1] In early protocells, prior to the emergence of an RNA replicase, nonenzymatic template-directed polymerization has been hypothesized to be responsible for RNA copying.[2−5] In this context, a template is a single-stranded RNA oligomer that, by Watson–Crick base-pairing, directs the sequence-specific polymerization of ribonucleotides to form its complementary strand. This process is frequently modeled as a repeated series of single-nucleotide primer extension events. Each primer extension step is a supramolecular event that takes place at the 3′-end of a primer annealed to a template and is orchestrated by noncovalent binding interactions, particularly hydrogen bonding and nucleobase stacking.

A high rate of primer extension is critical for efficient RNA copying; this rate is affected initially by how tightly nucleotides bind to the complex formed by the template and the growing primer strand. Furthermore, the fidelity of primer extension is influenced by the affinity of nucleotides that form wobble base pairs as well as other mismatches, and the overall rate of polymerization decreases significantly as a result of mismatches that lead to stalling of primer extension.[6] Understanding the thermodynamics of the binding interactions between a ribonucleotide and an RNA primer–template complex is of fundamental importance toward elucidating the mechanism as well as optimizing the rates of prebiotic replication of RNA. However, quantitative physical measurements of mononucleotide binding thermodynamics, as opposed to values derived from reaction kinetics, are limited. Early work by Kanavarioti et
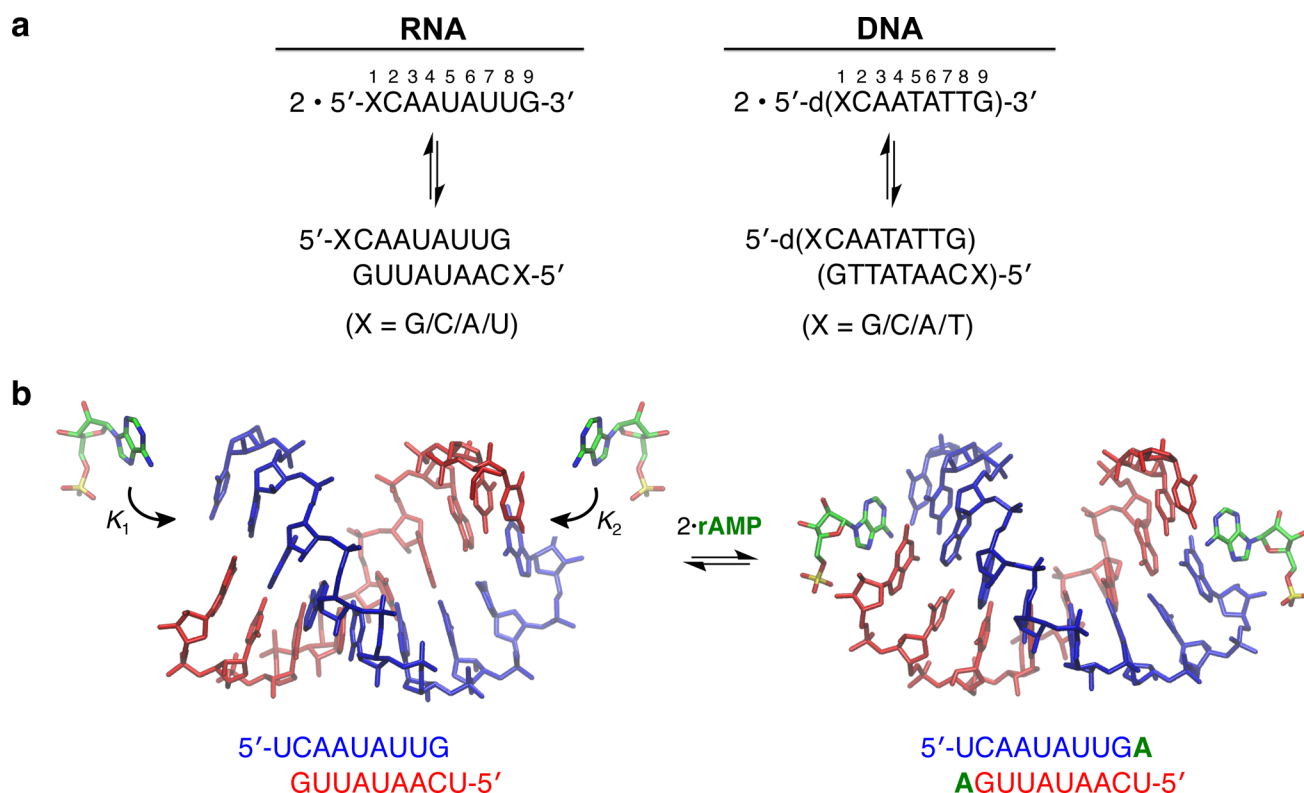
Figure 1. (a) Selected model RNA and DNA sequences used in this study. 5′-XCAAUAUUG-3′ represents 5′-ʳXʳCʳAʳAUʳAUUʳG-3′ RNA sequence and 5′-d(XCAATATTG)-3′ represents 5′-ᵈXᵈCᵈAᵈAATᵈATTᵈG-3′ DNA sequence. (b) Energy-minimized models of the A-form 5′-UCAAUAUUG-3′ RNA duplex in both free and rAMP-bound states. All components were solvated in water boxes, neutralized by Na⁺, and minimized using the program NAMD 2.9 with CHARMM 36 parameter set (see Methods for further details on the modeling and SI for other minimized systems).

al. examined the binding of multiple guanosine monomers on a polycytidylate template using differences in UV−vis absorption.[7] More recently, Richert and co-workers have reported thermodynamic investigations of the binding of nucleotides to modified oligonucleotide duplexes,[8] wherein they employed NMR spectroscopy to study the affinity of rGMP for a model RNA primer−template complex comprised of a hexaethylene glycol (HEG)-linked hairpin, 5′-CCAG(HEG)CUG-3′.

While these pioneering studies provided the physical measurements of rGMP binding to RNA templates, studies of the other ribonucleotides, of context effects, and of non-Watson−Crick binding interactions would be extremely useful in efforts to improve nonenzymatic RNA copying.

Here we present the results of the noncovalent binding of ribonucleotide monophosphates (rNMPs) to both RNA and DNA primer−template complexes in neutral aqueous media using ¹H NMR spectroscopy supplemented by molecular dynamics (MD) structural calculations. We measure the thermodynamic association constants, K_as, of nucleotide monomers for oligonucleotide duplexes by monitoring the resonances of the imino protons of all base-pairs in the duplexes, as the free monomer concentrations are increased incrementally. Our approach provides insight into non-enzymatic template-directed RNA polymerization by allowing the binding event to be separated from the subsequent chemical steps of primer extension.

■ RESULTS

Template Design. All of the RNA and DNA sequences that we studied (Figure 1a) are self-complementary and form stable palindromic duplexes in aqueous media at pH 7 and 12

°C, as verified by circular dichroism (CD) spectroscopy (see Figures S1−S2, SI), UV-melting experiments (Figures S3−S4, SI), and ¹H NMR investigations (Figure S5, SI). The 5′-single-nucleotide overhangs on either side of the duplexes, where "X" represents G, C, A, or U/T, serve as the template to which nucleotides can bind through hydrogen bonding, while the 3′-ends serve as the "primer" providing a stacking surface that enhances nucleotide interactions. The C_2 axis of symmetry in these palindromic duplexes not only allows a straightforward interpretation of the NMR spectra of the imino protons but also enables the use of a simple isotherm model[9] to measure the binding constants of the titrated monomers (see SI for mathematical derivations).

Characterization of the Duplexes. We began by carrying out a detailed structural characterization of all oligonucleotide duplexes that were used for NMR titration experiments. As a representative example, here we focus on the RNA duplex obtained from the 5′-UCAAUAUUG-3′ sequence (Figures 1b and 2). A qualitative analysis to confirm its self-hybridization and to determine the type of helical geometry was performed by variable-temperature CD spectroscopy. The CD spectra (Figure 2a) acquired over a temperature range of 4−80 °C revealed the characteristic features of a global A-form helical geometry, namely a small negative peak at ca. 245 nm and a large positive peak at ca. 265 nm.[10] Likewise, the CD spectra for all other RNA duplexes were consistent with global A-form conformations (Figure S1, SI). As expected, the CD spectra for all DNA duplexes (Figure S2, SI) were consistent with the B-form conformation.[11] We then performed MD simulations to derive energy-minimized models of the primer−template duplexes representing both the free and monomer-bound
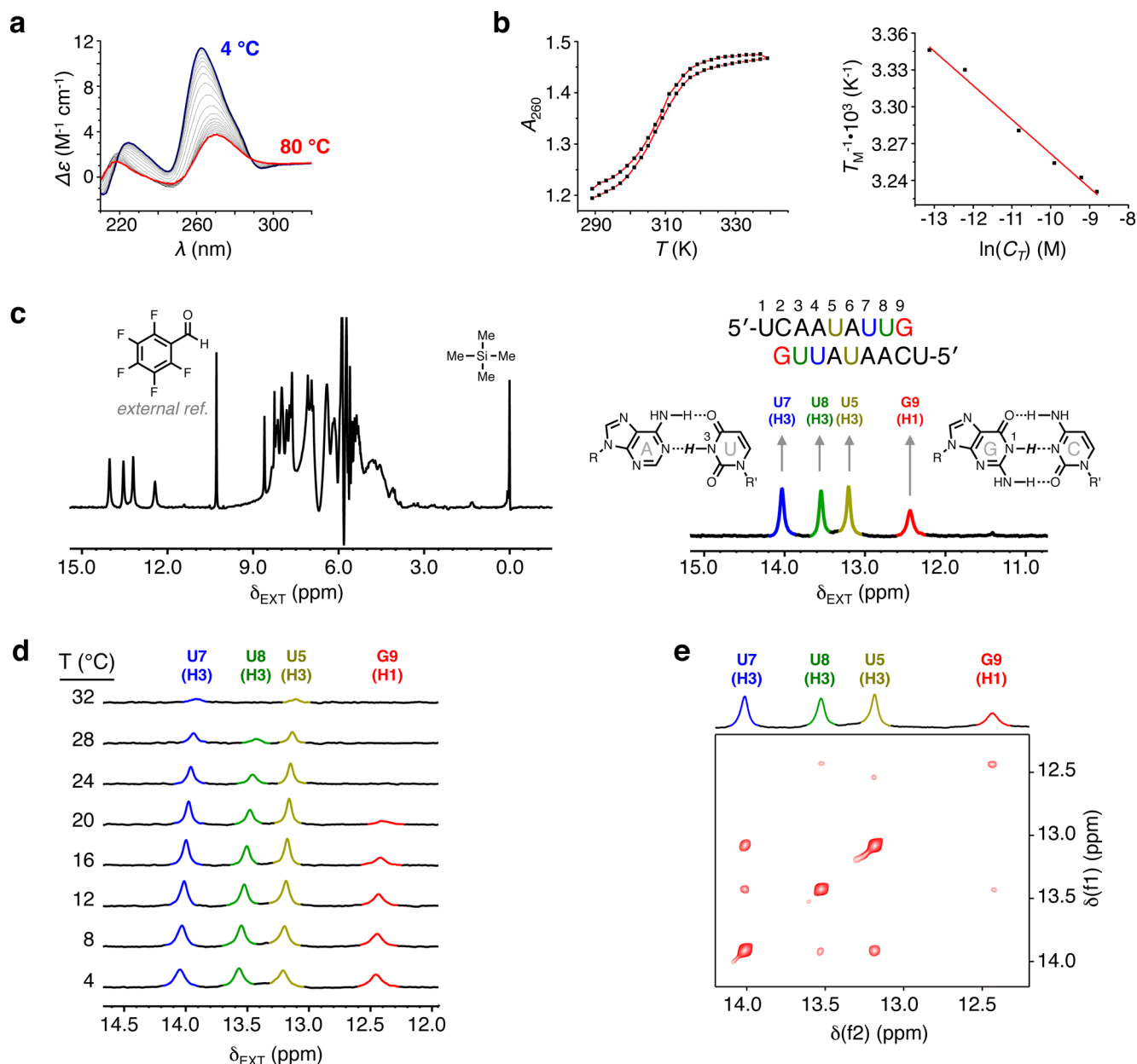
Figure 2. Characterization of the RNA duplex with a 5′-U overhang. (a) CD spectral overlay of the native 5′-UCAAUAUUG-3′ RNA duplex solution (0.2 mM) acquired at variable-temperature [4 °C intervals from 4 (blue) to 80 °C (red)], revealing that the sequence forms an A-form helix at room temperature and below. (b) (Left) UV-melting experiment for this RNA sequence at 0.1 mM concentration, confirming that it forms a stable duplex at pH 7.0. (Right) Van't Hoff analysis of the melting temperature (T_m) measured in the UV-melting experiment as a function of the total concentration of RNA. In this particular analysis, double helix formation is driven by an enthalpic contribution (ΔH) of −71.4 kcal mol^{-1} with an entropic cost (ΔS) of −213 cal mol^{-1} K^{-1}. (c) (Left) ^{1}H NMR (400 MHz) spectrum of the same RNA duplex (1.5 mM) dissolved in H$_2$O/D$_2$O (9:1) in the presence of NaCl (500 mM) at 12 °C and pH 7.0 (±0.1). The spectrum was acquired using the Watergate flip-back sequence to suppress the bulk water peak. Chemical shift values are externally referenced to pentafluorobenzaldehyde (δ_{EXT} = 10.285 ppm), which is dissolved in CDCl$_3$ (0.05% v/v tetramethylsilane) and applied within a thin coaxial NMR tube. (Right) Expanded imino proton region (12−14 ppm), wherein the three U(H3) and one G(H1) protons resonate. (d) Variable-temperature NMR experiment performed for the same solution to assign the imino protons based on water exchange rates. Faster exchange rates, which lead to greater line broadening, are indicative of closer proximity to the duplex termini. (e) Expanded imino proton region of the ^{1}H−^{1}H 2D NOESY NMR spectrum (200 ms mixing time), where the NOE cross-peaks represent the interaction between neighboring imino protons.

states. For all CD spectra and structural MD models, see Figures S1−S2 and S27−S34 in the SI, respectively. In order to evaluate the thermodynamic parameters associated with duplex formation under identical conditions used for the NMR spectroscopic titrations, we obtained a melting profile (Figure 2b, left panel) by carrying out variable-temperature UV absorption experiments at 260 nm, for a range of total strand

concentrations between 1 and 100 μM. A van't Hoff analysis[12] (Figure 2b, right panel) of the melting temperature as a function of the total concentration of RNA revealed enthalpic and entropic contributions to duplex formation of ΔH = −71.4 kcal mol^{-1} and ΔS = −213 cal mol^{-1} K^{-1}, respectively. These parameters allowed us to calculate the concentration of free and duplexed RNA at a total strand concentration of 3 mM at 12

°C, the conditions under which all NMR monomer titrations were performed. Under these conditions, the concentration of single-stranded RNA is on the order of 1 μM (i.e., <0.04 mol %), a fraction small enough to be neglected when considering the mechanism of ribonucleotide binding. For similar calculations carried out on other sequences, see Figures S3−S4 in the SI.

Assignment of the Imino Proton Peaks. The imino protons within Watson−Crick G:C and A:U/T base-pairs have distinct chemical shift values ranging between 12 and 13 ppm for G:C pairs[13] and 13 and 15 ppm for A:U/T pairs.[13,14] In the duplex formed from 5′-UCAAUAUUG-3′, the magnetically nonequivalent imino protons U5(H3), U7(H3), U8(H3), and G9(H1) resonate as expected in the region from 12 to 15 ppm (Figure 2c). In order to assign these resonances, we relied on a combination of variable-temperature (Figure 2d) and ^{1}H−^{1}H 2D NOESY NMR techniques (Figure 2e). We corroborated the assignment of the most upfield signal (red) to G9(H1) based on its greater line width compared to the other resonances, as expected from the fact that the terminal base pair allows faster G9(H1) exchange with H_2O on the time scale of the NMR experiment.[15]

NMR Measurements of rNMP Binding. We analyzed the change in the chemical shift ($\Delta\delta$) and the line width of the duplex imino proton resonances, as we gradually increased the concentration of the free nucleotide monomers. Duplex solutions were titrated with ribonucleotide monomer solutions (up to 250 mM), which contained the same concentration of the duplex (1.5 mM). The total cation (Na^+) concentration (500 mM) was also maintained at a constant value in a similar manner (Methods, SI). All titration experiments were carried out using 90% H_2O/10% D_2O solutions of the rNMP and RNA or DNA duplex. Spectra were recorded using a Watergate flip-back suppression[16,17] pulse sequence to obtain the signals of the imino protons while effectively suppressing the bulk water peak (Figure 2c).

Binding of rAMP to the 5′-UCAAUAUUG-3′ Primer−Template Complex. We monitored the ^{1}H NMR spectra of the 5′-U overhang RNA duplex (Figure 3a) as we titrated rAMP from 0 to 250 mM. No new signals in the 12−15 ppm region emerged throughout the titration, while the G9(H1) resonance (red) moved upfield from ∼12.4 ppm (no monomer present) to 12.0 ppm (near saturation of the duplex). This upfield shift is presumably due to the fact that G9(H1) becomes magnetically shielded by the ring current of adenine upon rAMP binding (see MD predicted structures in Figure S29, SI). This aromatic ring-current effect decreases, as expected, upon moving away from the binding site, such that the second strongest change in the chemical shift is exhibited by U8(H3) (green), and the two most internal protons, U5(H3) (yellow) and U7(H3) (blue), show the smallest changes, which are likely due to subtle alterations in conformation of the duplex upon binding. Significantly, $\delta_{G9(H1)}$ changes hyperbolically as a function of concentration of rAMP. Together, these observations are consistent with the specific binding of the rAMP monomers to both of the 5′-U overhangs of the duplex, a process that is in fast equilibrium on the NMR time scale. In order to confirm the 2:1 monomer-duplex binding stoichiometry, we performed an rAMP titration by continuous variation[18] on the analogous DNA sequence, 5′-d(TCAATATTG)-3′, which shows similar hyperbolic behavior in terms of the $\Delta\delta_{G9(H1)}$. We used a DNA duplex for this experiment because of the large amounts of material required.
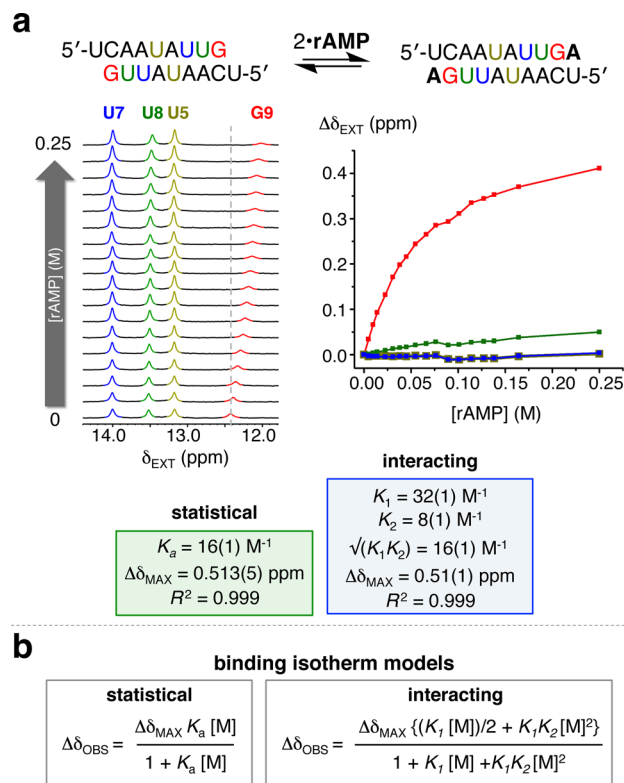

Figure 3. (a) Single titration experiment carried out with rAMP for the 5′-UCAAUAUUG-3′ RNA duplex. Errors obtained from the best fit are shown as the deviation in the least significant digit placed in parentheses (see the SI for the best fit curves). (b) Binding isotherm models used in this study. The statistical model assumes that two binding sites are identical and noninteracting, reducing to a single binding site isotherm. The interacting model assumes that the two binding sites are identical but can also interact (see the SI for details).

We maintained a constant total concentration of the DNA duplex and rAMP and monitored the chemical shift of G9(H1) while varying the ratio of these two species. We constructed a Job plot from these data using $\Delta\delta_{G9(H1)}$ and the mole fraction of the duplex (X_{dupl}). We observed a maximum in the Job plot at a value of $X = 0.35$. This value of X, which is related to the stoichiometric ratio n by the formula $X = 1/(1 + n)$, is consistent with a 2:1 binding stoichiometry of rAMP to the duplex (Figure S23).

We globally fit the chemical shift values of all four imino proton resonances to two binding isotherm models, which we have called "statistical" and "interacting", respectively (Figure 3b). The statistical model assumes that the two binding sites are identical and noninteracting. In this mechanism, the two stepwise macroscopic binding constants K_1 and K_2 (also see Figure 1b) are related to each other statistically through an intrinsic binding constant K_a ($^1/_2 K_1 = 2K_2 = K_a$; $\sqrt{K_1 K_2} = K_a$). This statistical model has the same mathematical form as a single binding site isotherm. The value of K_a using this model was determined to be 16(1) M^{-1} for the single titration experiment shown in Figure 3a. On the other hand, the interacting model places no constraints on the macroscopic binding constants K_1 and K_2 in order to account for the possibility that the binding of the first monomer has an effect on the binding of the second. The fit to the statistical isotherm gave identical results within error to the fit with the interacting model, which yielded values of K_1 and K_2 as 32(1) and 8(1)
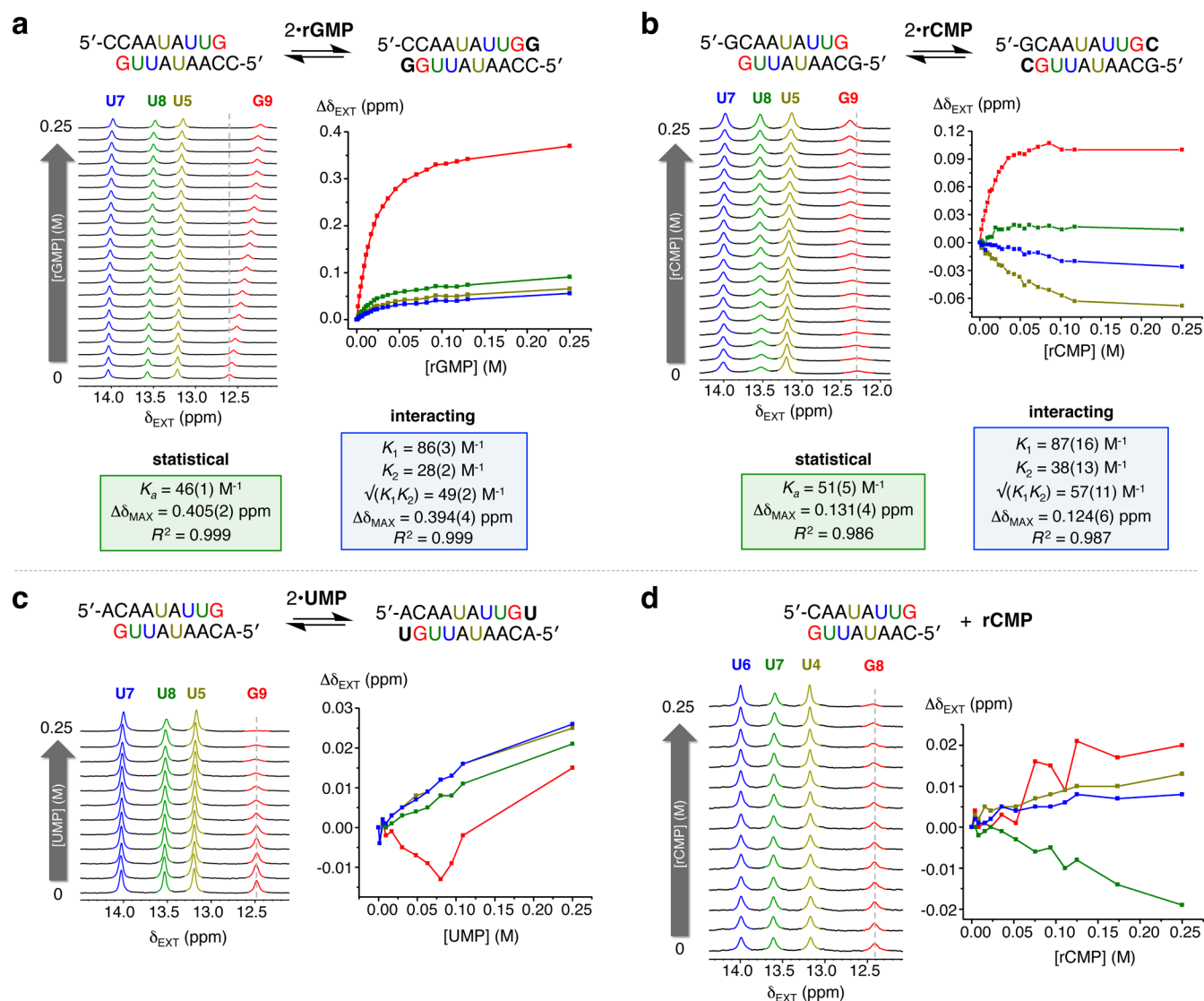
Figure 4. (a–d) Selected single NMR titrations carried out for all-RNA systems (see the SI for best fit curves to the data in panels a and b). Errors obtained from the best fits are shown as the deviations in the least significant digit(s) placed in parentheses.

M^{-1}, respectively. These results are likely explained by the supposition that the binding of the first rAMP does not significantly alter the local conformation of the second binding site on the other end of the duplex. Nearly all of the fits performed on the data obtained from the other RNA and DNA duplexes yielded either indistinguishable results with respect to the values determined from these two binding models (Figure 3b) or only slightly deviated from statistical behavior. The most significant exception to statistical binding was observed in the case of the 3′-C/5′-G RNA primer–template complex (Figure 5a). For a single titration experiment, K_1 and K_2 were measured to be 42(4) and 30(5) M^{-1}, respectively, which suggests that a slight amount of positive cooperativity is at play. This positive cooperativity may be the result of changes in conformation of the duplex that occur after the first rCMP binds, which enhances the binding of the second. Nevertheless, the value of $\sqrt{K_1 K_2}$, 35(3) M^{-1} from the interacting model, is almost the same as value of K_a, 29(1) M^{-1}, lying just outside of the errors from the fits.

Summary of Monomer Binding for RNA Duplexes. Figure 4 shows a compilation of the results of the single

titration experiments carried out on other all-RNA systems, and Table 1 displays a summary of the thermodynamic binding properties measured for monomers including the binding constants K_a (statistical model), the change in Gibbs free energy ΔG, maximum change in the chemical shift of the G-imino proton $\Delta\delta_{maxG9(H1)}$, and the R^2 values obtained from the fits. All values reported in Table 1 represent the averages obtained over three titration experiments, and the associated errors are those calculated from the sample standard deviation, unless otherwise noted. The RNA duplexes with 5′-U, -C and -G overhangs showed hyperbolic $\Delta\delta_{G9(H1)}$ when the complementary monomer was titrated. We obtained a K_a of 52(5) M^{-1} (average of three experiments: Table 1, entry 1) for rGMP (in 500 mM NaCl, pH 7, 12 °C) binding to the 5′-C overhang of the duplex formed by 5′-CCAAUAUUG-3′ (single experiment: Figure 4a). When we measured the binding of 2′-deoxyguanosine monophosphate (dGMP) to the duplex formed by 5′-CCAAUAUUG-3′, we obtained a K_a of 41(5) M^{-1} (Table 1, entry 5), slightly weaker than that of rGMP for the same template. Lastly, we found that rCMP binding to the 5′-G overhang (Figure 4b) exhibited a K_a of 66(15) M^{-1} (Table

Table 1. Summary of the Thermodynamic Binding Properties Measured for the NMPs Titrated into the RNA and DNA duplexes[a]

entry	monomer		sequence	K_a (M^{-1})	ΔG (kcal mol^{-1})	$\Delta\delta_{maxG(H1)}$ (ppm)	R^2
1	rGMP		5′-CCAAUAUUG-3′	52(5)	−2.24(8)	0.41(1)	0.999
2	rCMP		5′-GCAAUAUUG-3′	66(15)	−2.4(1)	0.13(1)	0.986
3	rAMP	RNA	5′-UCAAUAUUG-3′	14(2)	−1.49(8)	0.54(3)	0.999
4	UMP		5′-ACAAUAUUG-3′	ND	ND	ND	ND
5	dGMP		5′-CCAAUAUUG-3′	41(5)	−2.10(7)	0.39(4)	0.999
6	rGMP		5′-d(CCAATATTG)-3′	12(2)	−1.41(9)	0.18(1)	0.988
7	rCMP		5′-d(GCAATATTG)-3′	13(3)	−1.5(1)	0.103(2)	0.970
8	rAMP	DNA	5′-d(TCAATATTG)-3′	7(2)	−1.1(2)	0.46(6)	0.996
9	TMP		5′-d(ACAATATTG)-3′	ND	ND	ND	ND
10	dGMP		5′-d(CCAATATTG)-3′	13(2)	−1.45(9)	0.19(3)	0.991
11	rCMP	RNA	5′-GGAAUAUUC-3′	22(7)	−1.8(2)	0.29(1)	0.992
12	rCMP	DNA	5′-d(GGAATATTC)-3′	ND	ND	ND	ND
13	rGMP	RNA	5′-UCAAUAUUG-3′	6[b]	−1.0	0.22(2)	0.981

[a]H$_2$O/D$_2$O (9:1), NaCl (500 mM) at pH 7.0 (±0.1) and 12 °C. Errors shown are the standard deviations calculated from three independent titrations. [b]Data obtained from a single titration experiment. K_a represents the observed association constant assuming that all ratios of activity coefficients are unity.

1, entry 2). An exception to those studied above was of the titration performed with UMP (Figure 4c), in which the binding constant was too small to be reliably quantified.

In order to examine possible effects of unbound monomer on the chemical environment of a duplex, we carried out a titration of rCMP into the blunt-ended duplex formed from 5′-CAAUAUUG-3′ RNA, which does not possess any monomer binding site (Figure 4d). Only modest changes in the chemical shifts of all four imino proton resonances were observed, ranging from 0.01 to 0.02 ppm, indicating that little to no interaction takes place between rCMP and the fully paired double helix. Notably, the G8(H1) resonance undergoes significant line broadening as the concentration of rCMP increases. This contrasts with the line sharpening of the analogous G9(H1) (cf. Figure 4b) observed when rCMP is titrated into a solution of the 5′-GCAAUAUUG-3′ duplex. This control experiment suggests that excess phosphate monoester likely acts as a proton acceptor and facilitates imino proton exchange,[15] whereas monomer binding to the overhang slows this process down.

Summary of Monomer Binding for DNA Duplexes. The binding constants of rGMP, rCMP and rAMP to the analogous DNA primer−template complexes are all less than their RNA counterparts [K_a of 12(2), 13(3) and 7(2) M^{-1}, respectively, Table 1, entries 6−8]. See the SI for the titration data for the DNA duplexes. In addition to the titration of free ribonucleotides, we also studied the binding affinity of dGMP toward the 5′-C overhang DNA duplex (Table 1, entry 10). Under the same conditions (pH 7, 12 °C), the K_a of dGMP was measured as 13(2) M^{-1}, which is the same within error as that obtained for rGMP, 12(2) M^{-1}.

Context Effects on Monomer Binding. In addition to the titration studies described above, we carried out experiments to ascertain the influence of sequence context on binding affinity. Such context effects would presumably be mediated largely by differences in stacking interactions but could also be affected by sequence dependent changes in helical geometry. As a preliminary effort to measure the significance of such potential effects, we measured the K_a of rCMP for 5′-G overhang RNA and DNA duplexes with a primer terminated by a 3′-C instead

of a 3′-G (Figure 5a,b). The binding constant for rCMP to 5′-GGAAUAUU**C**-3′ RNA was measured to be 22(7) M^{-1} (average of three experiments: Table 1, entry 11), a decrease of about 3-fold compared to that [66(15) M^{-1}] for 5′-GCAAUAUUG-3′. In the case of the 5′-d(GGAATATT**C**)-3′ DNA duplex, the K_a value of rCMP (Table 1, entry 2) was too small to measure, as no hyperbolic change was observed in $\delta_{G2(H1)}$.

G:U Wobble Base Pairing. Finally, we tested the strength of the G:U wobble base pair, the major source of nonenzymatic replication error, by performing a titration of rGMP into a solution of 5′-UCAAUAUUG-3′ RNA duplex (Figure 5c). The binding constant for rGMP was estimated to be 6 M^{-1} (Table 1, entry 13), a value which is approaching the lower limit of our ability to quantify reliably. Nevertheless, this value is only about two times less than the binding constant for rAMP to the same template (Table 1, entry 3). This difference in K_a corresponds to a difference in ΔG of binding of approximately 0.5 kcal mol^{-1}.

■ DISCUSSION

Our titration experiments show that the noncovalent binding affinities between rNMPs and RNA primer−template complexes follow the order of C > G > A > U. Our initial expectation, based upon primer extension experiments, was that rGMP would be the tightest binder instead of rCMP. The rather surprising result that rCMP actually binds more tightly motivated us to carry-out an analysis relying on well-established nearest-neighbor (NN) models[19] in order to gain additional insight into the parameters governing the binding event. In our analysis, we first calculated the duplex energies of the "full-length" double-stranded 10-mers, which serve as models for the noncovalently bound rNMP−RNA complexes. We next calculated the duplex energies of the RNA complexes with rNMP-free overhangs that were used during the titration experiments. The differences in duplex energies between the full-length and dangling-end duplexes are plotted (Figure 6) along with the corresponding experimentally determined rNMP binding energies. The relative order of the experimental binding energies is in good agreement with those predicted from the
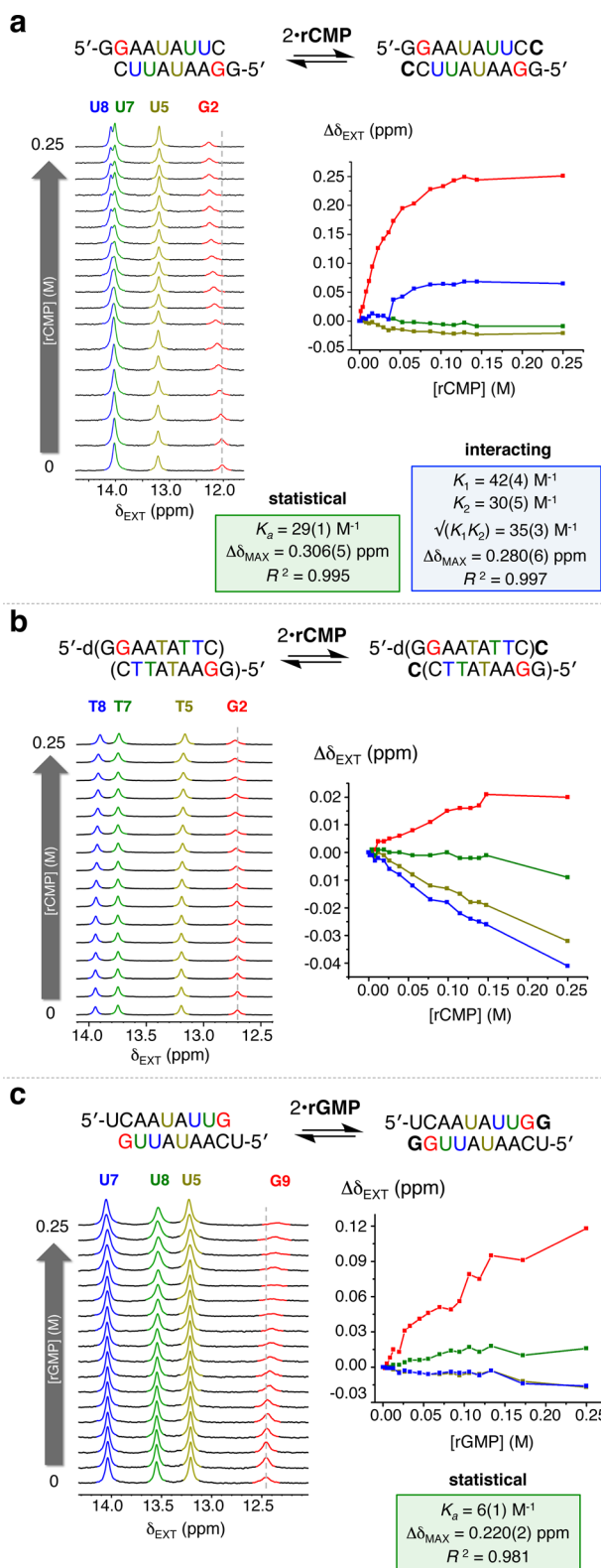
Figure 6. A comparison of experimental to predicted binding energies of the rNMPs to their complementary RNA primer−template complexes. The red bars are for the predicted values obtained from a NN analysis (12 °C), the blue bars are the experimental values obtained from titration data, while the green bars are the differences between experimental and predicted energies. The dashed bars for UMP are assuming a 1.5 kcal mol^{-1} difference between predicted and experimental energies. CMP* corresponds to the 5′-GGAAUAUUC-3′ RNA duplex.

Figure 5. (a,b) Titrations carried out to investigate the effect of primer base stacking surface on the binding strength of free monomer using (a) the 5′-GGAAUAUUC-3′ RNA duplex and (b) the DNA analog of the same sequence. (c) Titration experiment performed to quantify the affinity of rGMP toward a 3′-G/5′-U primer/template in order to determine the strength of the G:U wobble base-pair formation. Errors obtained from the best fits are shown as the deviations in the least significant digit placed in parentheses (see the SI for best fit curves to the data in panels a and c).
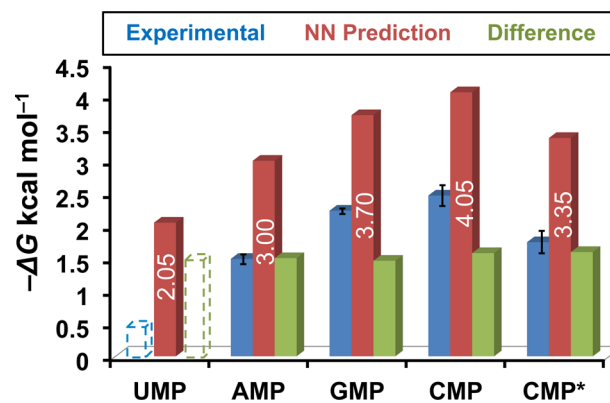
NN analysis, i.e., C > G > A > U. The NN analysis also accurately predicts that the energy of binding for rCMP with the 5′-GGAAUAUUC-3′ duplex (CMP* in Figure 6) lies in between those for rAMP and rGMP. The values predicted by the NN analysis are independent of the core sequence and only depend on the identities of the dangling-end and terminal base pairs (for more details, see S24 in SI). The most significant difference between the experimental and predicted NN binding energies is that the predicted NN energies are all approximately 1.5 kcal mol^{-1} more negative than the experimental values. As a consequence of the fact that the NN analysis does not take into consideration that the rNMP monomers are only noncovalently bound, the predicted energies should be expected to be more negative than what is observed. We hypothesize that this roughly constant difference of 1.5 kcal mol^{-1} is due to the negative change in entropy related to losses in translational and rotational degrees of freedom that occur upon rNMP monomer binding, and which are approximately the same for each monomer. Extrapolating this 1.5 kcal mol^{-1} difference in the case of UMP allows us to estimate the binding energy to be on the order of 0.5 kcal mol^{-1}, corresponding to a K_a of ∼2 M^{-1} at 12 °C, a value which is below our ability to quantify reliably under our experimental setup.

The excellent agreement between the measured binding energies of the rNMPs to the RNA primer−template complexes with the values predicted from the NN analyses suggests that rNMP binding to RNA is subject to similar structural constraints to those that determine duplex stability. Duplex stability is dominated[20] by hydrogen bonding and base-stacking, and these short-range interactions depend mostly on the identities of adjacent nucleotides. The differences in binding energy between rCMP and rGMP, for example, are primarily the result of the greater stability of the GC/CG (ΔG = −4.25 kcal mol^{-1}, 12 °C) propagation sequence in comparison to the GG/CC counterpart (ΔG = −3.74 kcal mol^{-1}, 12 °C). The contribution arising from the differences in energies between the GC/G (ΔG = −0.17 kcal mol^{-1}, 12 °C) and CC/G (ΔG = −0.06 kcal mol^{-1}, 12 °C) dangling ends is much less significant.

The larger binding constant of rCMP contrasts with the observations reported by Wu and Orgel[4] using hairpin templates that the 5′-phosphor-2-methylimidazolide of rGMP (2-MeImpG) has better primer extension reactivity than that of rCMP (2-MeImpC). One possible explanation may be that 2-MeImpG is more tightly bound than 2-MeImpC in the presence of Mg^{2+}, conditions under which the majority of template-directed primer extensions take place. Our data on rGMP and rCMP binding may differ from the binding of the corresponding 2-methylimidazolides. Affinity, however, is not the only factor affecting the rate of template-directed polymerization. For example, it could be that once bound to a primer−template complex, 2-MeImpG is sterically better positioned to react than is 2-MeImpC, giving rise to a larger rate constant. Under conditions in which the primer−template complex is fully occupied with activated monomers, a larger rate constant for 2-MeImpG would lead to a faster primer extension regardless of the value of their respective noncovalent binding constants, K_a. The direct experimental measurement of the binding of 2-methylimidazole-activated monomers is difficult since these compounds hydrolyze on the time scale of the NMR titration experiments. We are therefore currently exploring the possibility of measuring the binding of non-hydrolyzable analogs of the activated nucleotides.

When comparing the rNMP binding energies for the RNA versus the DNA primer−template complexes, the contribution of stacking energy at the primer−template binding site is likely to be affected by the backbone preorganization and helical parameters of the duplex,[21] which are significantly different for A-form RNA and B-form DNA helices. We hypothesize that the local geometry of the 5′-overhangs, which depend primarily on whether the primer−template complexes are in the A- or B-form conformations, is the dominant factor that leads to the differences in binding energies observed between RNA and DNA. The structural models obtained from MD calculations (see Figures S27−S34, SI) lend support to this hypothesis in that these calculations do indeed yield significant differences in the local geometry of the rNMP monomers depending on whether they are bound to A-form RNA or B-form DNA. One of the key differences may be the greater degree of interstrand base-stacking present in A-form RNA in comparison to B-form DNA, a structural feature which is often considered important[20] in providing at least a partial explanation for the greater duplex stability of RNA.

The differences in binding energy between rGMP and dGMP when binding to either the RNA or DNA primer−template complexes are likely accounted for by small differences in the intrinsic conformational preferences of these monomers. The K_a values of rGMP [12(2) M^{-1}] and dGMP [13(2) M^{-1}] for DNA duplexes are most likely the same within error, while the K_a of dGMP [41(5) M^{-1}] is slightly less than that of rGMP [52(5) M^{-1}] for the RNA analogs. We have shown previously[22] that both dGMP and rGMP bind to DNA primer−template complexes in the 2′-*endo* sugar pucker conformation, which likely explains why their binding constants are essentially the same within error. In the case of binding to an RNA primer−template complex, rGMP was observed to make a switch to the 3′-*endo* conformation, while dGMP stayed in the 2′-*endo* conformation. We hypothesize that the preferred 2′-*endo* sugar puckering of the bound dGMP gives rise to the slightly smaller binding energy compared to that of rGMP observed herein.

The result for the G:U mismatch experiment shows that the binding affinity of rGMP to the 5′-U RNA duplex is comparable to that of rAMP to the same duplex. This finding is in agreement with the duplex stability experiments carried out for oligoribonucleotides[23] with G:U mismatches and points to the future challenge of devising a way to improve the selectivity of rAMP over rGMP for a U template in order to avoid an inevitable error catastrophe.[24,25]

The noncovalent binding constant of rGMP that we measured for 5′-CCAAUAUUG-3′ [52(5) M^{-1}] appears to be slightly lower than that (71 M^{-1}) reported by Richert et al. employing the 5′-CCAG(HEG)CUG-3′ RNA hairpin in the presence of Mg^{2+} (2 mM hairpin, 200 mM Na-phosphate, pH 7.0, 400 mM NaCl, 80 mM $MgCl_2$, 20 °C).[8] This difference might not be surprising given that no divalent cations (e.g., Mg^{2+}) are present in our titration conditions. Similarly, the K_a of 6 M^{-1} that we measured for a rG:U wobble base-pair is slightly weaker than that (10 M^{-1}) reported by Richert et al. for a dG:T wobble pair. Finally, our observation that both UMP and TMP monomers bind very poorly to 5′-A RNA and DNA duplexes, respectively, is in agreement with the results of Richert et al. in their DNA system. On the other hand, there are some differences in the relative binding affinities of monomers to the RNA and DNA duplex systems that we studied, compared to those reported by Richert et al. for their HEG-linked DNA hairpins (conditions as above except pH 8.9). For example, their measured K_a values for the deoxy-monomers to the complementary templates are in the order dGMP (100 M^{-1}) > dAMP (26 M^{-1}) ≈ dCMP (25 M^{-1}) > TMP (4 M^{-1}), where the titrations of dGMP, dCMP, and TMP were carried out with DNA hairpins terminated by a 3′-A instead of a 3′-G. In contrast, our measured binding strength of rAMP is always lower than that of rCMP for both the RNA (14 vs 66 M^{-1}) and DNA (7 vs 13 M^{-1}) duplexes we employed, in which the incoming monomer is always stacked against a 3′-G residue. In addition, our K_a values of rGMP for the 5′-d(CCAATATTG)-3′ DNA duplex [12(2) M^{-1}] and of dGMP for the same duplex [13(2) M^{-1}] are both about an order of magnitude lower than that of dGMP reported by Richert et al. for the 5′-d(CTGC(HEG)GCA)-3′ DNA hairpin (100 M^{-1}) and five times lower than that of dGMP for the 5′-d(CTCTGC(HEG)-GCA)-3′ DNA hairpin (63 M^{-1}). These differences could in part be due to the presence of magnesium in the experiments performed by Richert et al., or alternatively could reflect structural differences between our symmetrical native duplexes and their short HEG-linked hairpin duplexes. Further work will be required to fully assess the context dependence of monomer binding to RNA and DNA templates.

■ CONCLUSION

In summary, we have presented results from an NMR-based quantitative approach to the study of the noncovalent interactions between ribonucleotide monophosphates and primer−template complexes, through which we obtained the thermodynamic binding constants, K_as, of the ribonucleotide monomers. Application of solution-phase techniques common for supramolecular and host−guest chemistry, in particular NMR spectroscopy, played an important role in developing the template-directed syntheses that led to a variety of complex macromolecular structures.[26,27] Accordingly, a thorough understanding of monomer−duplex binding interactions will provide a strong foundation[28] on which to optimize the template-directed polymerization of RNA. These same techniques can also be applied to experiments with structurally alternative polymers[29] that may have played a role in the early evolution of

life. The higher affinity of monomers for RNA primer–template complexes than for the corresponding DNA complexes may explain in part the advantage of RNA over DNA templates in nonenzymatic replication. In fact, similar arguments have been made through qualitative studies showing how the ribonucleotide monomers adapt more productive conformations upon binding to RNA templates.[22,30] We are currently extending our ^{1}H NMR experiments to the measurement of binding kinetics, binding affinity of higher order nucleotides (dimers and trimers), and, in the case of the templates with multiple binding sites, the effect of cooperativity on monomer binding.

METHODS

Preparation of the Oligonucleotides and NMR Experiments. RNA duplexes were synthesized by standard solid-phase phosphoramidite chemistry (for detailed procedures see the SI). DNA templates were purchased from Integrated DNA Technologies. Ribonucleotide monomers were purchased from Sigma-Aldrich as disodium salts. Each oligonucleotide duplex was titrated with the selected ribonucleotide monophosphate (up to ca. 250 mM) dissolved in a 9:1 mixture of H_2O/D_2O. Monomer solutions contained the same concentration of the oligonucleotide duplex (1.5 mM) in order to maintain a constant duplex concentration throughout the titration experiments. This was also the case for the total cation (Na^+) concentration (500 mM). The pH of both duplex and monomer solutions was adjusted to 7.0 (±0.1) with either NaOH or HCl, and the NMR spectra were acquired at 12 °C, unless otherwise noted. Monomer titration, pH, and temperature gradient experiments were performed on a Varian INOVA 400 MHz NMR spectrometer. Initial concentrations of the duplex and monomer were determined by UV (NanoDrop) measurements and confirmed by ^{31}P NMR (161 MHz) spectroscopy using a potassium sodium phosphate buffer concentrate (Supelco), which was applied using a coaxial NMR tube. The latter technique was also used to measure monomer concentrations throughout the titrations.

Theoretical Calculations. For each sequence, an ideal A-form 10-mer RNA and B-form 10-mer DNA duplex generated by COOT[31] was used to model both the free and monomer-bound states of the 9-mer duplex with single-nucleotide 5′-overhangs at both ends. Two systems were modeled for the monomer-bound states, each containing two dianionic nucleoside monophosphates. All systems were solvated in ca. $48 \times 64 \times 48$ Å^3 TIP3P[32] water boxes and neutralized by Na^+ using VMD.[33] The final systems contain ca. 2.8×10^4 atoms including nucleic acids, water, and ions.

These systems were minimized using the program NAMD 2.9[34] with a CHARMM 36 parameter set.[35,36] All simulations were performed using periodic boundary conditions. The bonded, non-bonded, and electrostatic interactions were calculated at time steps of 1, 2, and 4 fs, respectively. The switching (cutoff) distance for nonbonded interaction was set at 10 (12) Å. To compute long-range electrostatic interactions, the particle mesh Ewald method[37] with a grid density of at least 1 Å^{-3} was used. A four-step minimization protocol similar to that used by Eargle et al.[38] was applied. In the first 10,000 steps of minimization, all heavy atoms were fixed. In the next 20,000 steps, water molecules were freed. Ions, nucleobases, and part of the sugar were then freed while keeping the nucleic acid backbone fixed for the next 20,000 steps. Finally, everything was set free for the last 30,000 steps of minimization.

ASSOCIATED CONTENT

Ⓢ Supporting Information

Detailed experimental procedures, synthesis, and spectroscopic characterization of oligonucleotide duplexes, mathematical derivation of binding isotherms, ^{1}H NMR spectra and data analysis of monomer titrations, energy-minimized models, error analysis of the binding constants, and a statistical correlation study between K_a and $\Delta\delta_{MAX}$. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b02707.

AUTHOR INFORMATION

Corresponding Author
*szostak@molbio.mgh.harvard.edu

Present Addresses
∥High Magnetic Field Laboratory, Chinese Academy of Sciences, HeFei 230031, China
⊥Ra Pharmaceuticals, Inc., Cambridge, MA 02139.

Author Contributions
∇These authors contributed equally.

Notes
The authors declare no competing financial interest.

REFERENCES

(1) Blain, J. C.; Szostak, J. W. *Annu. Rev. Biochem.* **2014**, *83*, 615–640.
(2) Lohrmann, R.; Orgel, L. E. *Nature* **1976**, *261*, 342–344.
(3) Inoue, T.; Orgel, L. E. *Science* **1983**, *219*, 859–862.
(4) Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, *114*, 5496–5501.
(5) Wu, T.; Orgel, L. E. *J. Am. Chem. Soc.* **1992**, *114*, 7963–7969.
(6) Rajamani, S.; Ichida, J. K.; Antal, T.; Treco, D. A.; Leu, K.; Nowak, M. A.; Szostak, J. W.; Chen, I. A. *J. Am. Chem. Soc.* **2010**, *132*, 5880–5885.
(7) Kanavarioti, A.; Hurely, T. B.; Baird, E. E. *J. Mol. Evol.* **1995**, *41*, 161–168.
(8) Kervio, E.; Claasen, B.; Steiner, U. E.; Richert, C. *Nucleic Acids Res.* **2014**, *42*, 7409–7420.
(9) Connors, K. A. *Binding Constants: The Measure of Molecular Complex Stability*; Wiley-Interscience: New York, 1987.
(10) Gratzer, W. B.; Richards, E. G. *Biopolymers* **1971**, *10*, 2607–2614.
(11) Ivanov, V. I.; Minchenkova, L. E.; Schyolkina, A. K.; Poletayev, A. I. *Biopolymers* **1973**, *12*, 89–110.
(12) Holbrook, J. A.; Capp, M. W.; Saecker, R. M.; Record, M. T., Jr. *Biochemistry* **1999**, *38*, 8409–8422.
(13) Conte, M. R.; Conn, G. L.; Brown, T.; Lane, A. N. *Nucleic Acids Res.* **1996**, *24*, 3693–3699.
(14) Lane, A. N.; Jenkins, T. C.; Frenkiel, T. A. *Biochim. Biophys. Acta* **1997**, *1350*, 205–220.
(15) Guéron, M.; Leroy, J.-L. *Methods Enzymol.* **1995**, *261*, 383–413.
(16) Piotto, M.; Saudek, V.; Sklenár, V. *J. Biomol. NMR* **1992**, *2*, 661–665.
(17) Lippens, G.; Dhalluin, C.; Wieruszeski, J.-M. *J. Biomol. NMR* **1995**, *5*, 327–331.
(18) Renny, J. S.; Tomasevich, L. L.; Tallmadge, E. H.; Collum, D. B. *Angew. Chem., Int. Ed.* **2013**, *52*, 11998–12013.
(19) Turner, D. H.; Sugimoto, N.; Freier, S. M. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 167–192.
(20) Egli, M. On Stacking. In *Structure and Function*; Springer: Dordrecht, Netherlands, 2010, 177–196.
(21) Parker, T. M.; Hohenstein, E. G.; Parrish, R. M.; Hud, N. V.; Sherrill, C. D. *J. Am. Chem. Soc.* **2013**, *135*, 1306–1316.

(22) Zhang, N.; Zhang, S.; Szostak, J. W. *J. Am. Chem. Soc.* **2012**, *134*, 3691−3694.

(23) Chen, J. L.; Dishler, A. L.; Kennedy, S. D.; Yildirim, I.; Liu, B.; Turner, D. H.; Serra, M. J. *Biochemistry* **2012**, *51*, 3508−3522.

(24) Szostak, J. W. *J. Syst. Chem.* **2012**, *3*, 2.

(25) Leu, K.; Obermayer, B.; Rajamani, S.; Gerland, U.; Chen, I. A. *Nucleic Acids Res.* **2011**, *39*, 8135−8147.

(26) Diederich, F.; Stang, P. J. *Templated Organic Synthesis*; Wiley-VCH: Weinheim, 2000.

(27) Forgan, R. S.; Sauvage, J. P.; Stoddart, J. F. *Chem. Rev.* **2011**, *111*, 5434−5464.

(28) Fahrenbach, A. C. *Pure Appl. Chem.* **2015**, *87*, 205−218.

(29) Schöning, K.-U.; Scholz, P.; Guntha, S.; Wu, X.; Krishnamurthy, R.; Eschenmoser, A. *Science* **2000**, *290*, 1347−1351.

(30) Kozlov, I. A.; Orgel, L. E. *Mol. Biol.* **2000**, *34*, 781−789.

(31) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 486−501.

(32) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.

(33) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33−38.

(34) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781−1802.

(35) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(36) Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D., Jr. *J. Comput. Chem.* **2011**, *32*, 1929−1943.

(37) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.

(38) Eargle, J.; Black, A. A.; Sethi, A.; Trabuco, L. G.; Luthey-Schulten, Z. *J. Mol. Biol.* **2008**, *377*, 1382−1405.