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*

Supporting Information

Table of Contents

1. General Methods

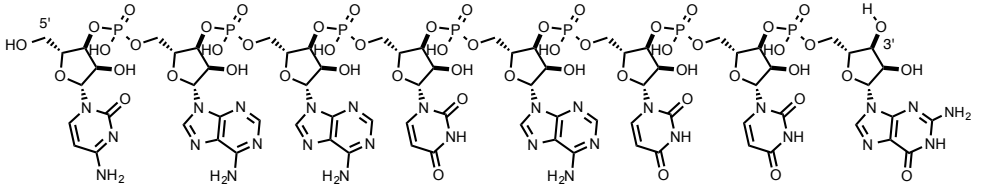
RNA sequences were prepared by solid-phase phosphoramidite chemistry on an Expedite RNA/DNA oligonucleotide synthesizer. 2′-TBDMS-protected RNA phosphoramidites were purchased from Chemgenes. Reagents used for the synthesis protocol, which include 5-(benzylthio)-1*H*-tetrazole (0.25 M in acetonitrile), iodine oxidizing solution (0.02 M in THF/H_2O/pyridine), cap mix A (THF/acetic anhydride/pyridine 8:1:1) and cap mix B (methyl imidazole/THF/pyridine 8:1:1), were obtained from Glen Research. Cleavage of the 5′-DMTr-protected oligomerization products from the solid support and the deprotection of the nucleobases were carried out by treatment with a 1:1 mixture of ammonium hydroxide and 40% aqueous methylamine for 1 h at 65 °C. The reaction mixtures were concentrated to dryness under reduced pressure and without external heating, and the residues were re-dissolved in 100-200 μL of DMSO and 60 μL of triethylamine. The resulting solutions were gently mixed in the presence of 75 μL of TEA•3HF for 2.5 h at 65 °C facilitating the removal of TBDMS group, and then allowed to cool to room temperature and treated with 1.8 μL of a pH 7.5 Tris-quenching buffer solution (Glen Research). Purification of the desired products including removal of 5′-DMTr-deprotecting groups was achieved by means of a series of buffer elutions on a Glen-Pak^{TM} RNA purification cartridge following the protocol recommended by the manufacturer. The excess salt used in the eluting media was removed by size-exclusion chromatography (G-15 Sephadex gel matrix). Finally, the oligonucleotides were converted to their sodium salts using a Dowex-Na^+ ion exchange resin. Chemical identity and degree of purity of the isolated RNA oligomers were confirmed using a high resolution Agilent HPLC-MS-MS instrument. DNA templates were purchased from Integrated DNA Technologies and used as received. Ribonucleotide monomers were purchased from Sigma-Aldrich as disodium salts, except for the Me_4N^+ salt of rGMP, which was prepared from rGMP-free acid purchased from Santa Cruz Biotechnology Inc. This particular sample was intended to be used to prevent monomer aggregation resulting from the formation of G-quadruplex species at high concentrations of GMP. A titration experiment using the Me_4N^+ salt of rGMP for the 5′-CCAAUAUUG-3′ RNA duplex showed very similar spectral behaviors to those carried out using the Na^+ salt analog. 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 duplex (ca. 1.5 mM) in order to maintain a constant duplex concentration throughout the titration experiments. This was

also the case for the total concentration of Na^+, the most abundant cation present. In particular, a 250 mM solution of the monomer (disodium salt form), which was titrated into the duplex solution, was assumed to contain 500 mM of Na^+. In order to keep the Na^+ concentration constant during the titrations, we prepared the monomer-free duplex solutions with 500 mM added NaCl. The pH of both duplex and monomer solutions was adjusted to 7.0 (±0.1) using trace amounts of either NaOH or HCl. NMR spectra were acquired on a Varian INOVA 400 MHz NMR spectrometer equipped with a broadband PFG (z-gradient) probe. Suppression of the bulk water resonance was achieved by the Watergate technique.[1] Each spectrum was recorded at 12 °C after 64–128 transients (nt), with an optimized delay period on pulse sequence (d1) of 1.5s and a pulse width (pw) of 15, unless otherwise noted. Initial concentrations of the duplex and monomer solutions were determined by both UV (NanoDrop™) and ^{31}P NMR (161 MHz) spectroscopy measurements. After each titration, the total concentration of the monomer present in the duplex solution was measured using ^{31}P NMR spectroscopy (nt = 64–128, d1 = 3s, and pw = 15). The ratio of duplex to monomer in the concentrated monomer samples was confirmed using a potassium sodium phosphate buffer concentrate (Supelco), which was applied within a physically separated coaxial NMR tube.

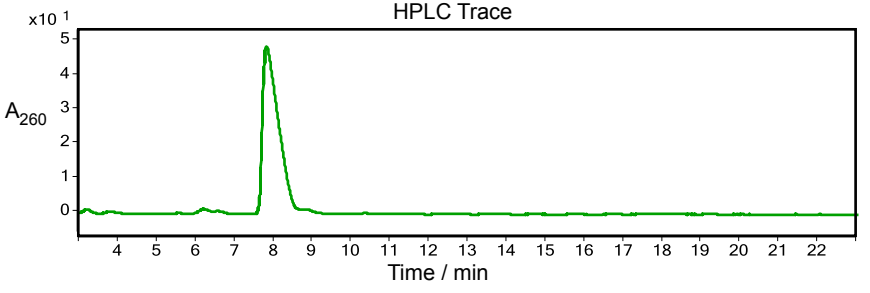
2. RNA Characterization by LCMS

The chemical integrity of the synthesized RNA sequences was evaluated by an Agilent HPLC-MS-MS instrument equipped with a C_{18} analytical column. The results are shown below.

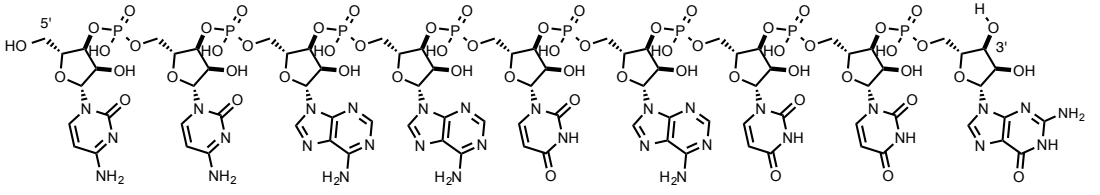
5´-CAAUAUUG-3´



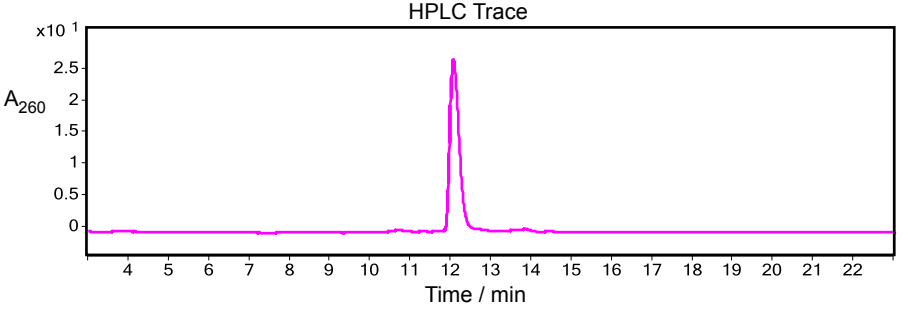
chemical formula: $C_{76}H_{94}N_{29}O_{54}P_7$; calculated exact mass: 2493.3664; measured accurate mass: 2493.3594, error: -2.8 ppm



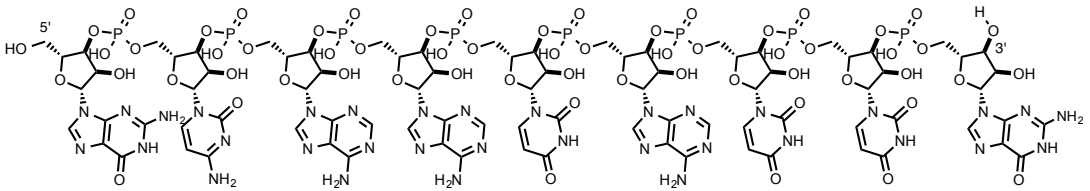
5´-CCAAUAUUG-3´



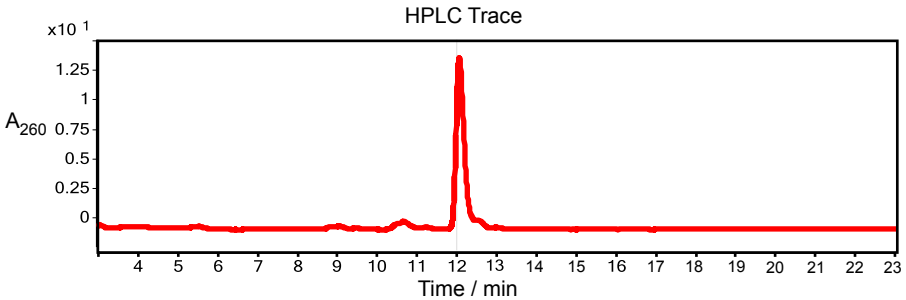
chemical formula: $C_{85}H_{106}N_{32}O_{61}P_8$; calculated exact mass: 2798.4077; measured accurate mass: 2798.3999, error: -2.8 ppm



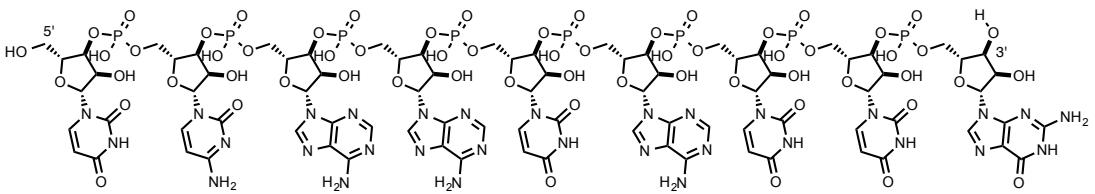
5´-GCAAUAUUG-3´



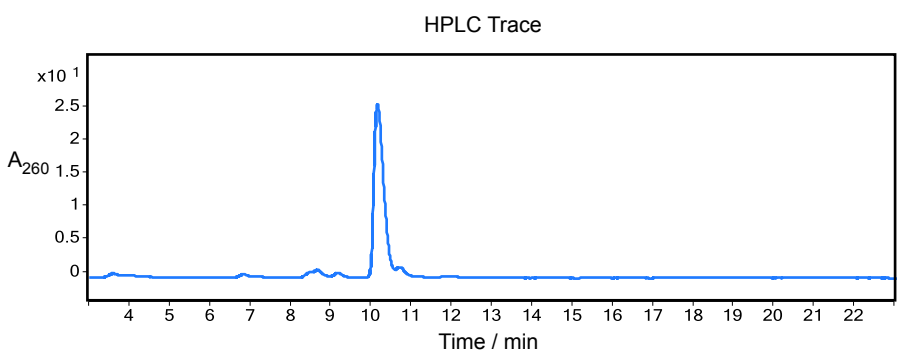
chemical formula: $C_{86}H_{106}N_{34}O_{61}P_8$; calculated exact mass: 2838.4139; measured accurate mass: 2838.4044, error: -3.3 ppm



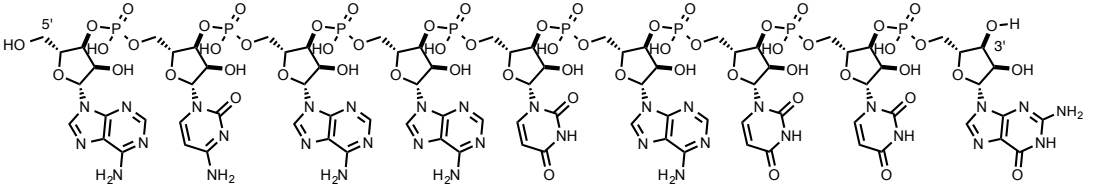
5´-UCAAUAUUG-3´



chemical formula: $C_{85}H_{105}N_{31}O_{62}P_8$; calculated exact mass: 2799.3917; measured accurate mass: 2799.3851, error: -2.4 ppm

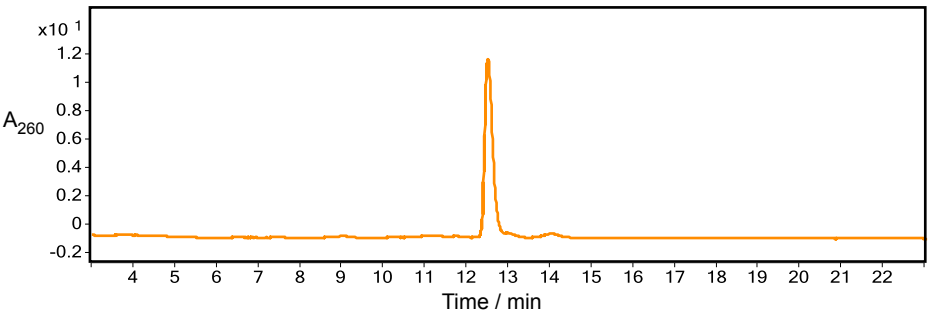


5´-ACAAUAUUG-3´

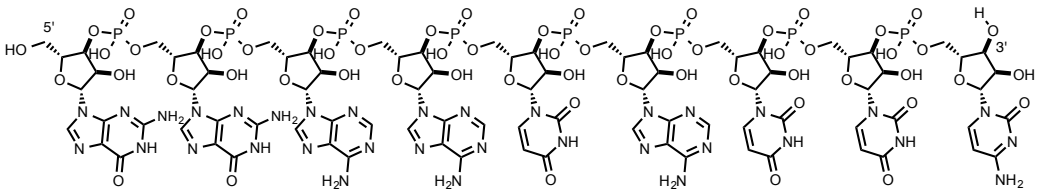


chemical formula: $C_{86}H_{106}N_{34}O_{60}P_8$; calculated exact mass: 2822.4189; measured accurate mass: 2822.4104, error: -3.0 ppm

HPLC Trace

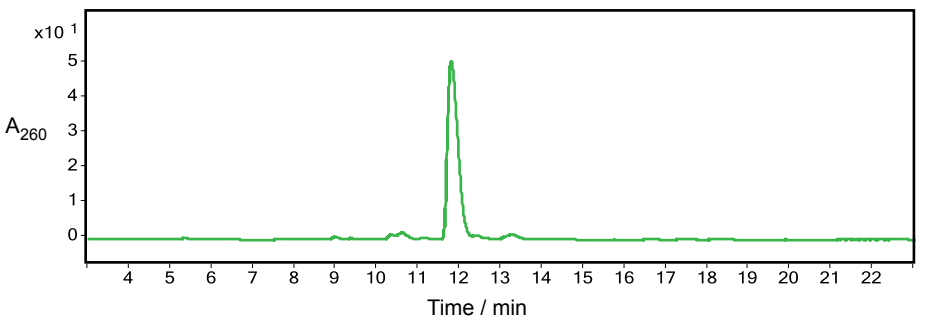


5´-GGAAUAUUC-3´



chemical formula: $C_{86}H_{106}N_{34}O_{61}P_8$; calculated exact mass: 2838.4139; measured accurate mass: 2838.4063, error: -2.7 ppm

HPLC Trace



3. Duplex Characterization by CD Spectroscopy

All RNA and DNA duplexes were characterized by CD spectroscopy in order to determine their conformations under conditions relevant to the ^{1}H NMR titrations. All CD spectra were obtained at 12 °C, pH 7 and 500 mM NaCl at a spectroscopically optimal total strand concentration of ~100–200 μM in quartz cuvettes of 1 mm pathlength. All RNA duplexes (Fig. S1) provided a small intensity negative peak at ~245 nm and a large positive peak at ~265 nm. These spectral features are characteristic of a global A-form helical geometry.[2]

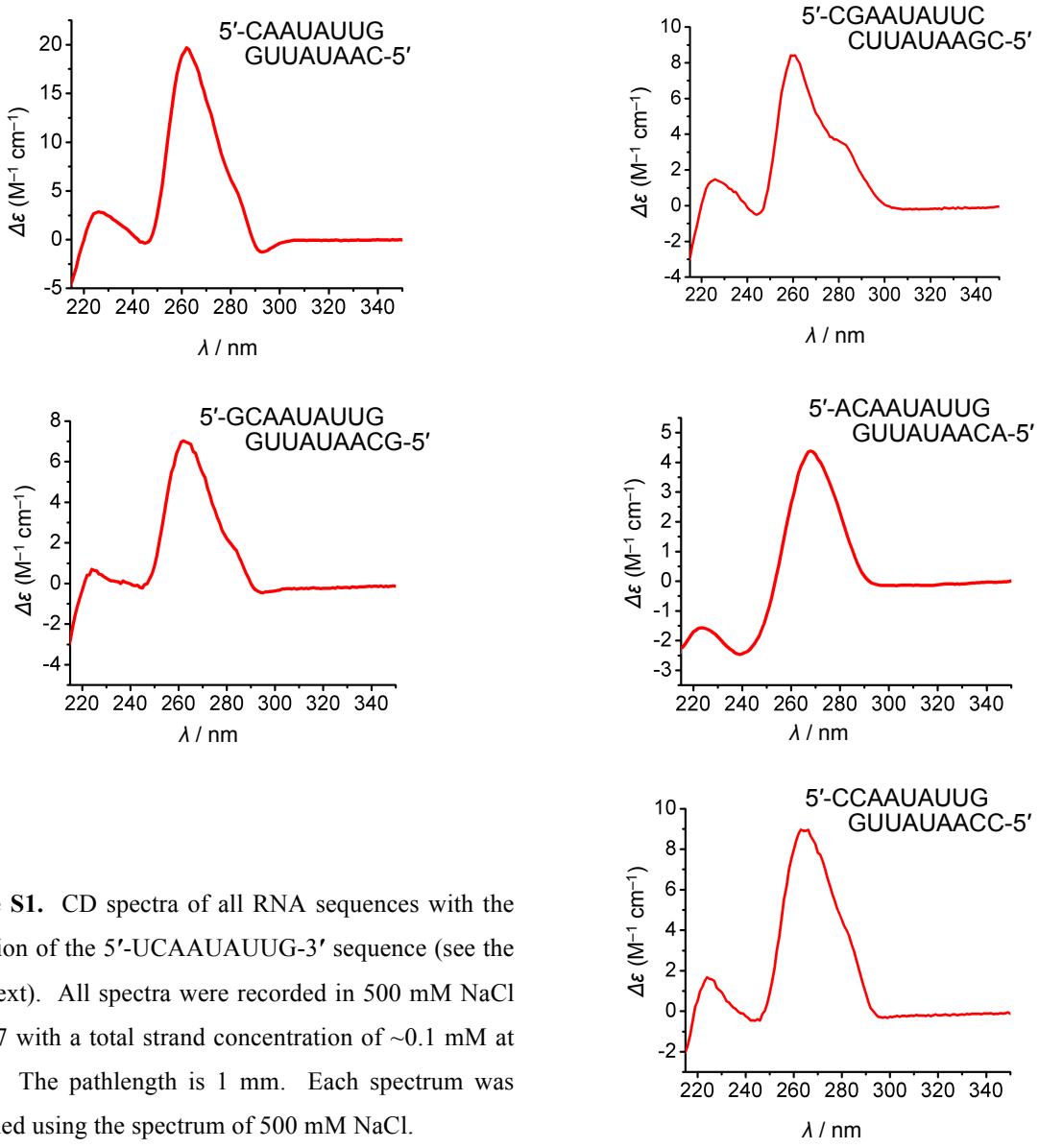


Figure S1. CD spectra of all RNA sequences with the exception of the 5′-UCAAUAUUG-3′ sequence (see the main text). All spectra were recorded in 500 mM NaCl at pH 7 with a total strand concentration of ~0.1 mM at 12 °C. The pathlength is 1 mm. Each spectrum was baselined using the spectrum of 500 mM NaCl.

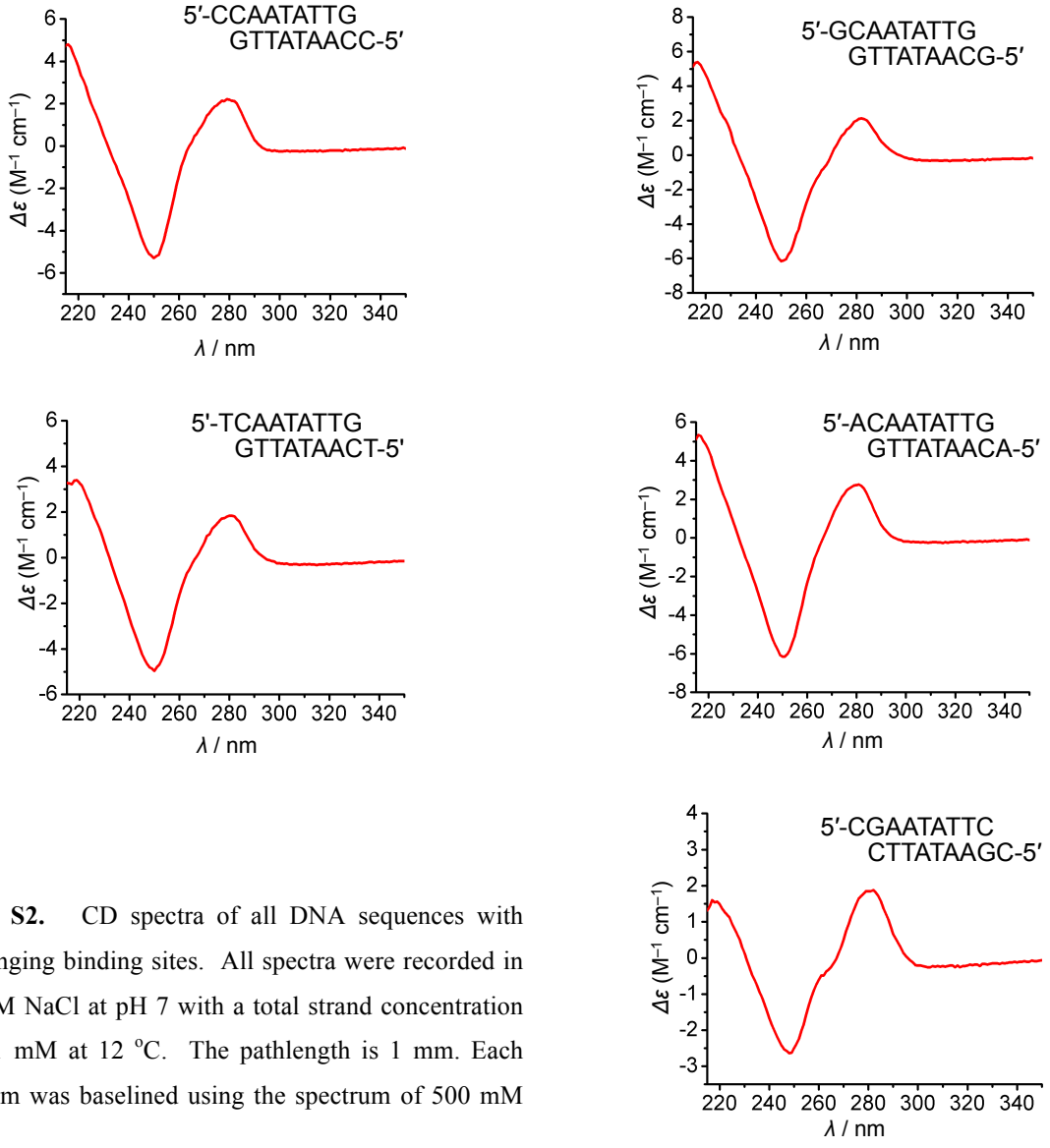


Figure S2. CD spectra of all DNA sequences with overhanging binding sites. All spectra were recorded in 500 mM NaCl at pH 7 with a total strand concentration of ~0.1 mM at 12 °C. The pathlength is 1 mm. Each spectrum was baselined using the spectrum of 500 mM NaCl.

All DNA duplexes (Fig. S2) exhibit roughly equally sized positive and negative peaks at ~280 and ~250 nm, respectively. These spectral features are characteristic of a global B-form helical geometry.[3]

4. Duplex Characterization by UV-Melting

We carried out a series of variable concentration UV-melting experiments on all RNA and DNA duplexes (Fig. S3a) in order to determine the thermodynamic parameters governing the formation of the double helix. From this data, the relative amounts of single and double stranded RNA or DNA can be calculated. As a representative example, we discuss the 5′-CAAUAUUG-3′ sequence, which has no overhang. It contains the same core sequence as nearly all of the duplexes studied in this paper. The melting points were determined at different concentrations, and from this data, a van't Hoff plot (Fig. S3b) was constructed. The van't Hoff analysis revealed that double helix formation is driven by an enthalpic contribution ΔH of –57.9 kcal mol^{-1} and an entropic cost ΔS of –166 cal mol^{-1} K^{-1}. From these values, we calculate that the concentration of single stranded RNA is less than 5 μM at 12 °C with a total single strand concentration of 3 mM, conditions that are the same as those under which all titration experiments were carried out. The concentration of single stranded RNA is low enough that we can justify safely excluding it from the mechanism of monomer binding. The van't Hoff analyses for the other RNA (Fig. S3b) and DNA (Fig. S4b) duplexes revealed similar values for the

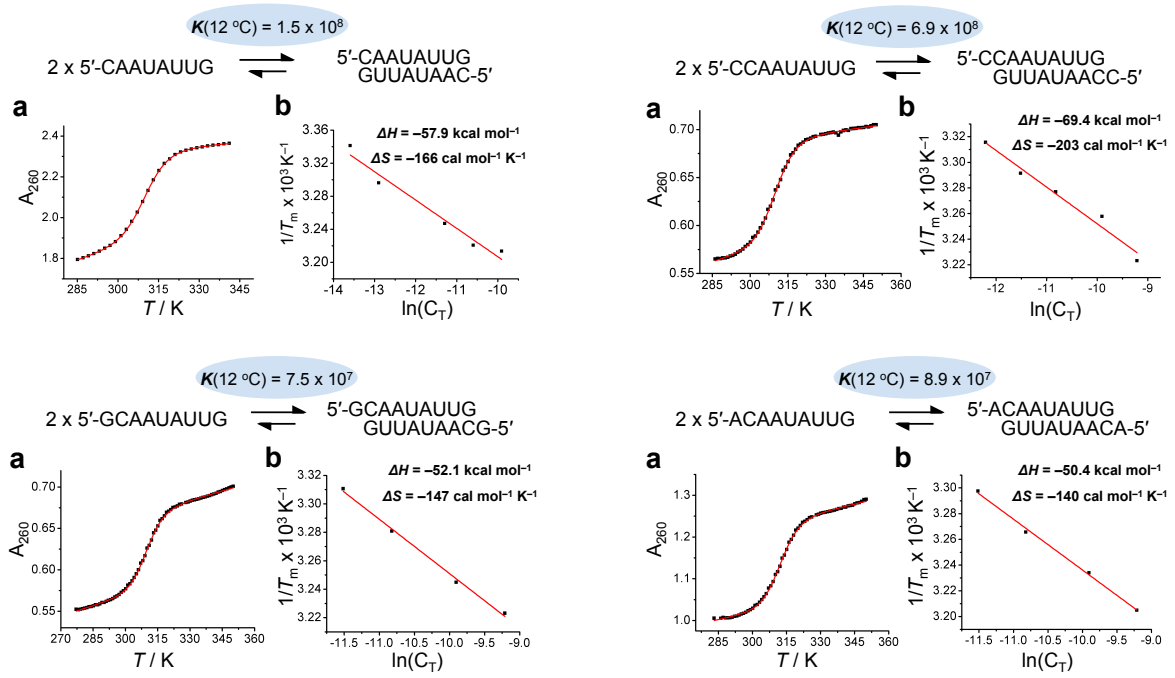


Figure S3. a. Thermal denaturation of RNA duplexes (0.1 mM) dissolved in H_2O (500 mM NaCl, pH 7). **b.** Van't Hoff plots for RNA sequences dissolved in H_2O (500 mM NaCl, pH 7). For 5′-UCAAUAUUG-3′, see the main text.

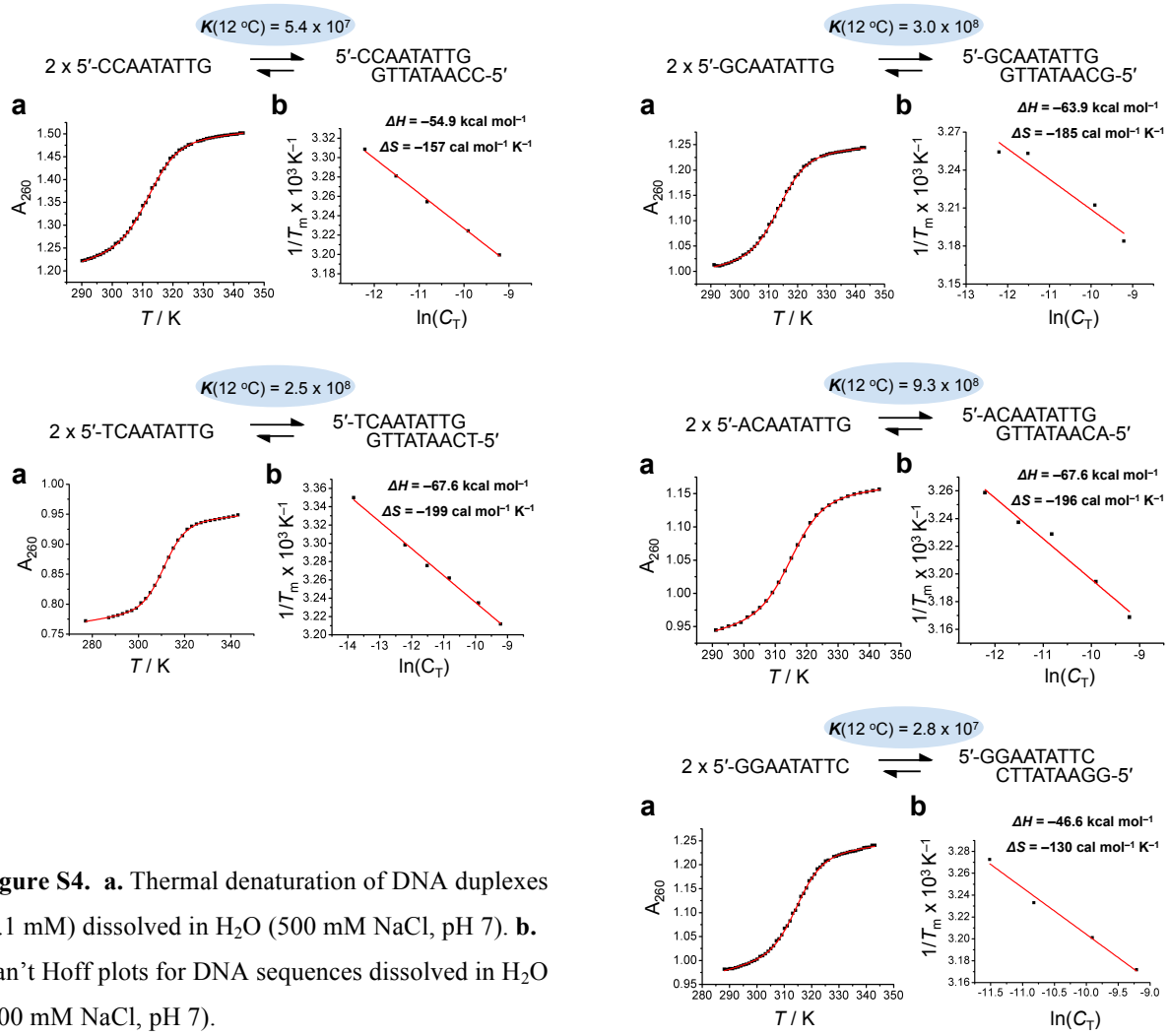


Figure S4. **a.** Thermal denaturation of DNA duplexes (0.1 mM) dissolved in H_2O (500 mM NaCl, pH 7). **b.** Van't Hoff plots for DNA sequences dissolved in H_2O (500 mM NaCl, pH 7).

concentration of single strands present in solution.

5. Assignment of Imino Resonances

Variable Temperature ^{1}H NMR and ^{1}H-^{1}H NOESY Experiments of 5′-CAAUAUUG-3′

We carried out a series of variable temperature ^{1}H NMR experiments as well as ^{1}H-^{1}H 2D NOESY experiment on the 5′-CAAUAUUG-3′ duplex in order to definitively assign the four imino proton resonances observed for all dangling end duplexes possessing the same central core. The variable temperature experiments also allowed us to evaluate the optimal temperature range at which to carry out the titration experiments. The variable temperature ^{1}H NMR experiments are shown in Fig. S5a from 0 to 36 °C recorded at a pH of 7. At 0 °C, all four imino resonances are clearly observed although they are relatively broad. Increasing the temperature

up to approximately 20 °C results in the progressive sharpening of the linewidths of all four resonances, after which temperature the peaks begin to broaden and are almost completely undetectable by 36 °C. The temperature at which each resonance broadens into the baseline is also informative as to their identities. This broadening is the result of an increase in the exchange rate with water. The protons closest to the termini exchange the fastest, while the protons nearest the center exchange the slowest. This exchange occurs when a base pair momentarily opens up and exposes the imino protons of both purine or pyrimidine components to the bulk solvent. Base pairs closest to the termini are thermodynamically most favored to open, while those in the center are the least. As expected, the resonance assigned to G8(H1) (red) broadens into the baseline at the lowest temperature. The green colored resonance assigned to U7(H3) broadens out next followed by the resonances assigned to U4(H3) (yellow) and U6(H3) (blue). The 2D NOESY spectrum (Fig. S5b) reveals a cross-peak between the U7(H3) and U6(H3) signals, and another one occurring between U6(H3) and U4(H3), data of which corroborates these assignments.

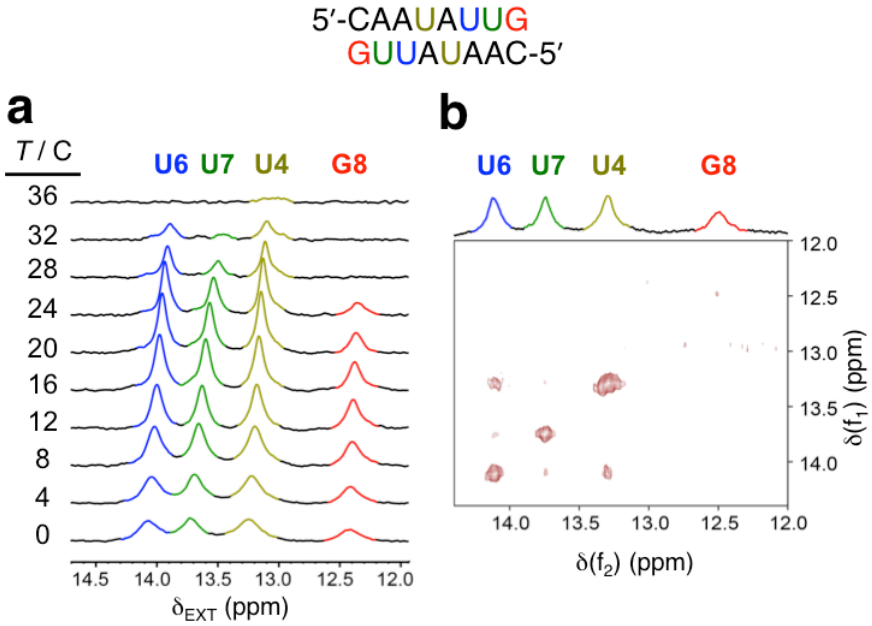


Figure S5. a. Variable temperature-^{1}H NMR and **b.** ^{1}H-^{1}H NOESY experiments of the 5′-CAAUAUUG-3′ RNA sequence containing no overhang. Each spectrum was recorded in the presence of 500 mM NaCl at pH 7 with a total strand concentration of 3 mM. The NOESY spectrum was recorded at 12 °C (200 ms mixing time). The cross-peaks represent the interaction between neighboring imino protons.

Variable pH-^{1}H NMR of 5′-CAAUAUUG-3′

We carried out a variable pH-^{1}H NMR experiment on the 5′-CAAUAUUG-3′ RNA duplex (Fig. S6) in order to evaluate the optimal pH range at which to carry out the titrations. All four imino proton signals are clearly observable from pH 4.7 to pH 8.0. At pH 8.5, the G8(H1) signal is no longer observable. All four imino resonances undergo only modest changes in their chemical shifts as a function pH.

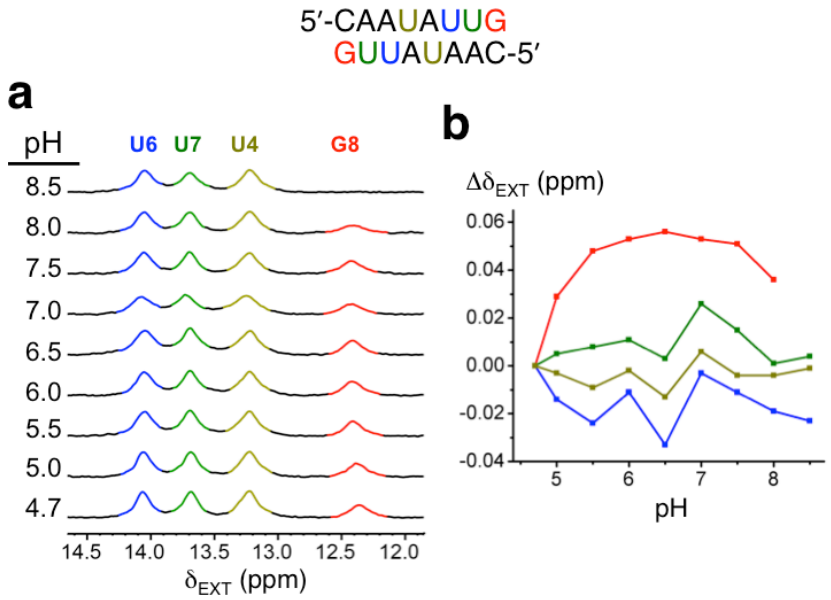
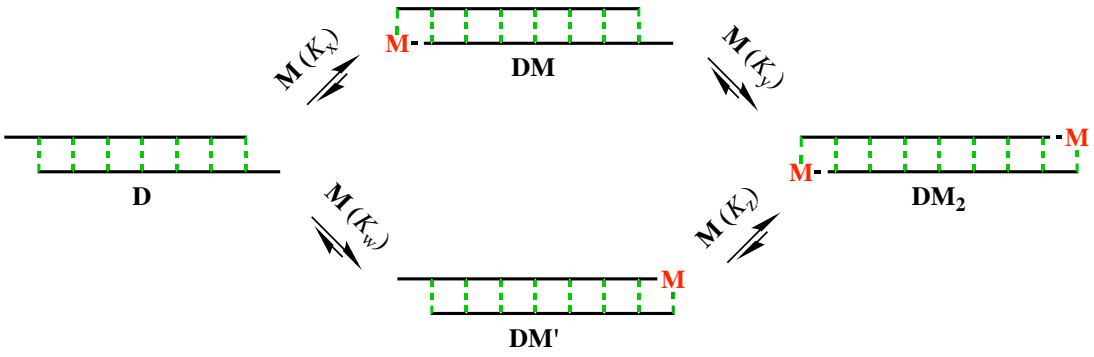


Figure S6. a. Variable pH spectra of the 5′-CAAUAUUG-3′ sequence with no overhanging binding site. Each spectrum was recorded in 500 mM NaCl at pH 7 with a total strand concentration of 3 mM at 12 °C. **b.** Plot of the change in the chemical shift for each imino resonance with respect to pH.

6. Derivation of the Binding Isotherms

We tested two binding isotherms[4] against all the titration data, one which assumes the overhangs are identical and noninteracting (statistical) and one which assumes that the binding of the first monomer can affect the binding constant of the second monomer (interacting). Consider the following equilibria between a duplex **D** and ribonucleotide monomer **M**.



Noncovalent binding equilibration is fast on the ^{1}H NMR timescale. Therefore, the observed chemical shift (δ_{OBS}) of the selected proton will be a weighted fraction of **D**, **DM**, **DM'** and **DM$_2$**, given by the relation:

$$\delta_{OBS} = f_D\delta_D + f_{DM}\delta_{DM} + f_{DM'}\delta_{DM'} + f_{DM2}\delta_{DM2} \quad (1)$$

where δ_D, δ_{DM}, $\delta_{DM'}$ and δ_{DM2} are the chemical shift values of the selected imino proton for **D**, **DM**, **DM'** and **DM$_2$**, respectively, and f_D, f_{DM}, $f_{DM'}$ and f_{DM2} are the molar fractions of **D**, **DM**, **DM'** and **DM$_2$** in solution at any given total concentration of monomer [M]. The sum of the mole fractions is equal to one. Therefore f_D is equal to $1 - (f_{DM} + f_{DM'} + f_{DM2})$, and substituting equation (1) with this expression yields:

$$\delta_{OBS} - [1 - (f_{DM} + f_{DM'} + f_{DM2})]\delta_D = f_{DM}\delta_{DM} + f_{DM'}\delta_{DM'} + f_{DM2}\delta_{DM2} \quad (2)$$

Defining $\Delta\delta$ as $\delta - \delta_D$, and rearranging the terms leads to:

$$\Delta\delta_{OBS} = f_{DM}\Delta\delta_{DM} + f_{DM'}\Delta\delta_{DM'} + f_{DM2}\Delta\delta_{DM2} \quad (3)$$

Rewriting in terms of [M] and microscopic equilibrium constants, Ks, yields:

$$\Delta\delta_{OBS} = (\Delta\delta_{DM}K_x[M] + \Delta\delta_{DM'}K_w[M] + \Delta\delta_{DM2}K_xK_y[M]^2) / (1 + K_x[M] + K_w[M] + K_xK_y[M]^2) \quad (4)$$

Assuming that $\Delta\delta_{DM}$ and $\Delta\delta_{DM'}$ are equal, then:

$$\Delta\delta_{OBS} = (\Delta\delta_{DM}[M](K_x + K_w) + \Delta\delta_{DM2}K_xK_y[M]^2) / (1 + [M](K_x + K_w) + K_xK_y[M]^2) \quad (5)$$

The supposition that the two binding sites are identical but the binding of the first monomer can influence the binding of the second can be expressed mathematically as:

$$K_x = K_w = K_1'; \; K_y = K_z = K_2' \quad (6a,b)$$

As a result of the binding sites being identical, the microscopic binding constants K_x and K_w leading to the singly bound DM and DM' duplexes are equal. This model can account for a mechanism in which the first bound monomer may induce conformational changes at the remaining binding site which affects the binding of the second monomer. By this reasoning, K_y and K_z are equal to a second, possibly different constant K_2'. Substituting equation (5) with these expressions leads to:

$$\Delta\delta_{OBS} = (\Delta\delta_{DM}[M](2K_1') + \Delta\delta_{DM2}K_1'K_2'[M]^2) / (1 + [M](2K_1') + K_1'K_2'[M]^2) \quad (7)$$

By letting $2K_1'$ equal K_1, the macroscopic binding constant governing the first binding event, and from the thermodynamic constraint that the product of K_1 and K_2 must be equal to that of K_1' and K_2' yields:

$$\Delta\delta_{OBS} = (\Delta\delta_{DM}K_1[M] + \Delta\delta_{DM2}K_1K_2[M]^2) / (1 + K_1[M] + K_1K_2[M]^2) \quad (8)$$

Assuming that $2\Delta\delta_{DM}$ is equal to $\Delta\delta_{DM2}$, which is also equal to $\Delta\delta_{MAX}$, then:

$$\Delta\delta_{OBS} = \Delta\delta_{MAX}(K_1[M] / 2 + K_1K_2[M]^2) / (1 + K_1[M] + K_1K_2[M]^2) \quad (9)$$

This equation is what we have called the interacting isotherm and assumes the binding of the first monomer can have an effect on the binding of the second. For binding sites which are both identical and noninteracting, where the microscopic binding constants are all equal, the macroscopic binding constants K_1 and K_2 are related to each other statistically,[4] such that $K_1 = 2K_a$ and $K_2 = {}^1/_2K_a$, where K_a is the intrinsic microscopic binding constant. Substituting these expressions into equation (9) leads to:

$$\Delta\delta_{OBS} = \{\Delta\delta_{MAX}K_a[M] (1 + K_a[M])\} / (1 + K_a[M])^2 = \Delta\delta_{MAX}K_a[M] / (1 + K_a[M]) \quad (10)$$

which has the identical form to a binding isotherm whose mechanism has one open binding site.

7. ^{1}H NMR Spectroscopic Titrations

RNA Duplexes

Titration of 5′-CCAAUAUUG-3′ with rGMP

Titration of rGMP into the 5′-CCAAUAUUG-3′ RNA double helix (Fig. S7) results in a large upfield shift of the G9(H1) resonance that occurs in a hyperbolic fashion. The U resonances also shift upfield, albeit to a much lesser extent, also in a fashion that appears hyperbolic. The modest shifts of U5(H3), U7(H3) and U8(H3) signals are likely the result of subtle changes in the double helix conformation upon binding (Fig. S27).

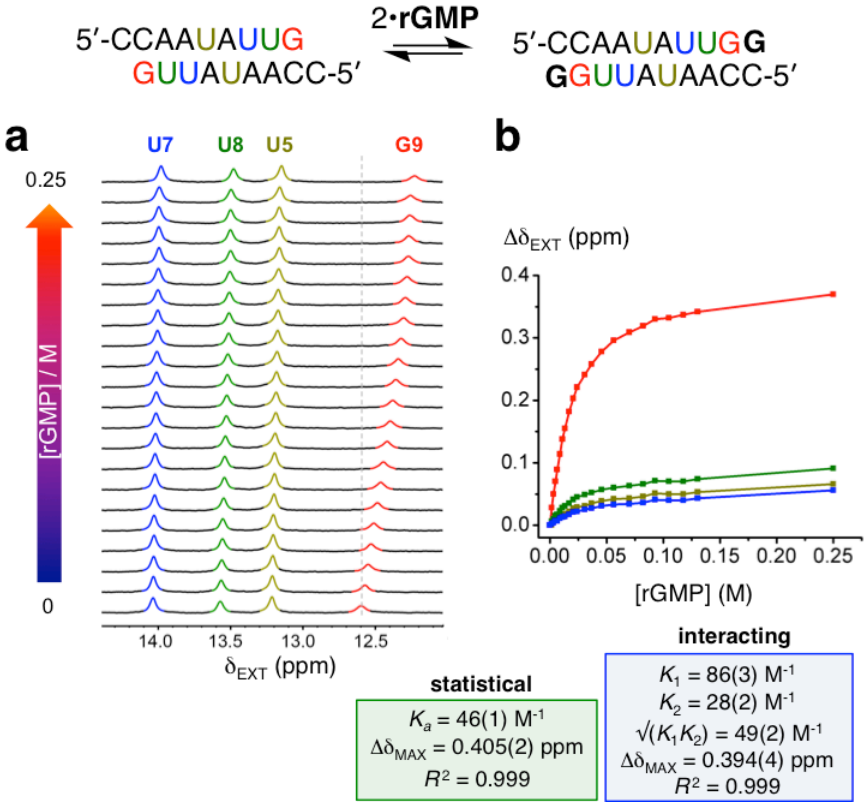


Figure S7. a. Titration of the 5′-CCAAUAUUG-3′ RNA duplex with rGMP while monitoring the imino region of the ^{1}H NMR spectra. Each spectrum was recorded at a total strand concentration of 3 mM at pH 7 at a Na^{+} concentration of 500 mM at 12 °C. **b.** Plot of the change in chemical shift of the imino resonances with respect to rGMP concentration. From these data, a fit to the statistical and interacting models was carried out.

Titration of 5′-GCAAUAUUG-3′ with rCMP

The titration of rCMP into the 5′-GCAAUAUUG-3′ RNA duplex (Fig. S8) reveals that the G9(H1) resonance shifts downfield as does the U8(H3) signal. The resonances of U5(H3) and U7(H3) both shift upfield. The maximum change of the G9(H1) chemical shift is much less compared to the titration of rAMP with the 5′-UCAAUAUUG-3′ duplex, while the change in chemical shift of U5(H3) for both titrations is approximately the same. The most internal resonances U7(H3) and U5(H3) undergo upfield shifts likely as a consequence of subtle changes

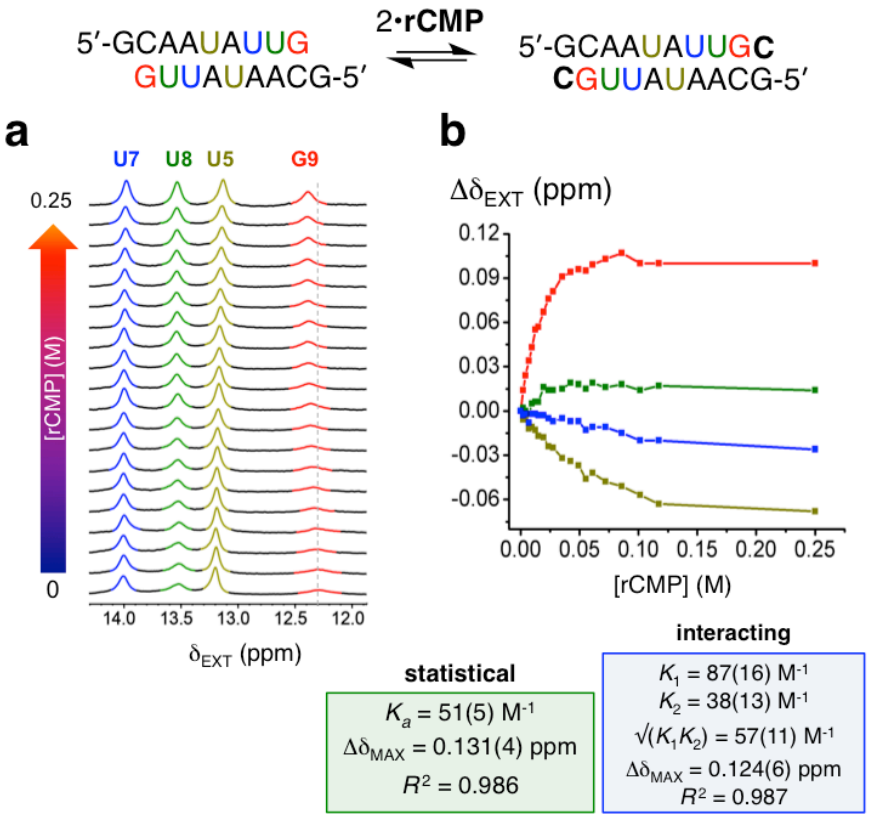


Figure S8. a. Titration of the 5′-GCAAUAUUG-3′ RNA duplex with rCMP while monitoring the imino region of the ^{1}H NMR spectra. Each spectrum was recorded at a total strand concentration of 3 mM at pH 7 at a Na^{+} concentration of 500 mM at 12 ^{o}C. **b.** Plot of the change in chemical shift of the imino resonances with respect to rCMP concentration. From these data, a fit to the statistical and interacting models was carried out.

in the conformation of the double helix upon binding of the monomer (Fig. S28). A global fit to all four resonances simultaneously using both the interacting and statistical models reveal that the binding is statistical within the error to the fits.

Titration of 5′-UCAAUAUUG-3′ with rAMP

We performed a titration on the 5′-UCAAUAUUG-3′ RNA duplex with rAMP from 0 to 250 mM (Fig. S9). The G9(H1) resonance is observed to shift the most by about 0.45 ppm upfield in a hyperbolic fashion. Like in the case of rGMP titration, G9(H1) signal shifts upfield, which is presumably due to the shielding effect of the purine ring orthogonal to G9(H1) (Fig. S29). All other U(H3) resonances also shift upfield, however, to much lesser extents ~0.03–0.06 ppm. We globally fit the change in chemical shift of all four imino proton resonances to both the statistical and interacting binding models. Comparing the values of the binding constants obtained between interacting and statistical models reveals the binding of the first and second monomer is purely statistical.

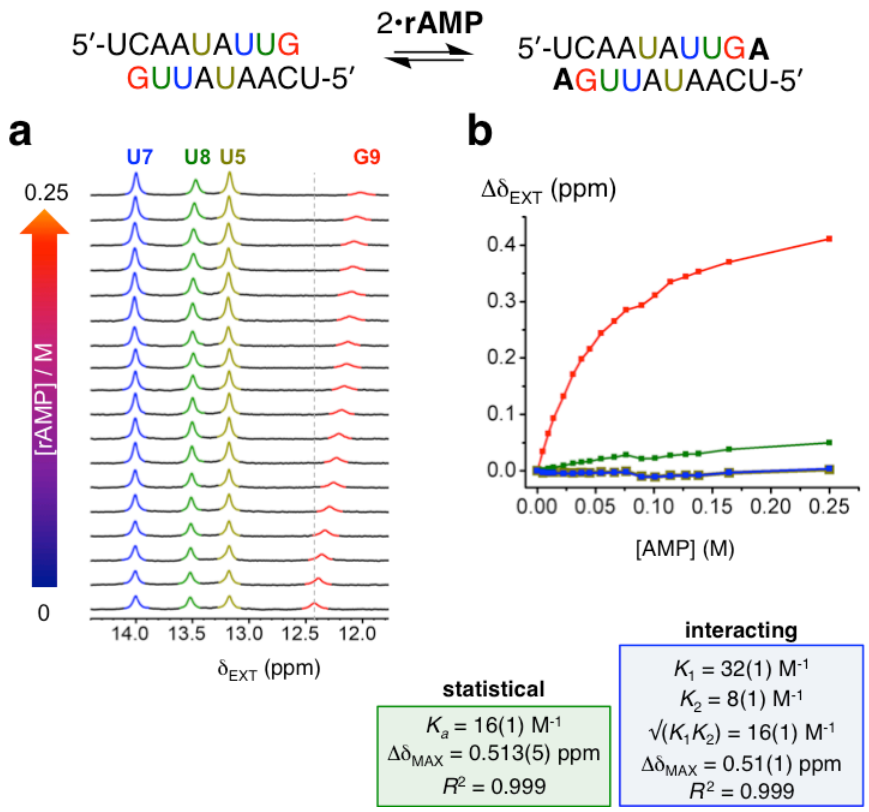


Figure S9. a. Titration of the 5′-UCAAUAUUG-3′ RNA duplex with rAMP while monitoring the imino region of the ^{1}H NMR spectra. Each spectrum was recorded at a total strand concentration of 3 mM at pH 7 at a Na^{+} concentration of 500 mM at 12 °C. **b.** Plot of the change in chemical shift of the imino resonances with respect to rAMP concentration. From these data, a fit to the statistical and interacting models was carried out.

Titration of 5′-ACAAUAUUG-3′ with UMP

The titration of the 5′-ACAAUAUUG-3′ duplex with UMP from 0 to 250 mM (Fig. S10) reveals

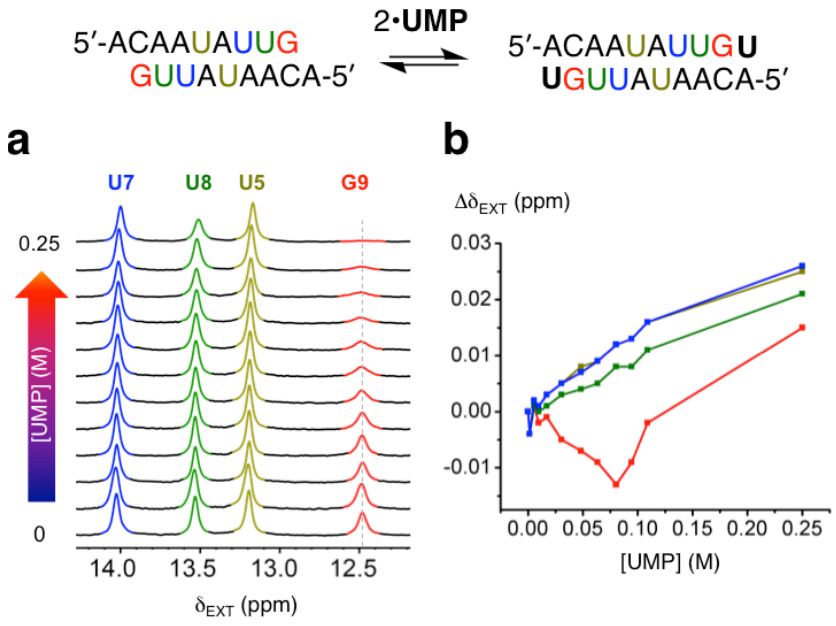


Figure S10. a. Titration of the 5′-ACAAUAUUG-3′ RNA duplex with UMP while monitoring the imino region of the ^{1}H NMR spectra. Each spectrum was recorded at a total strand concentration of 3 mM at pH 7 at a Na^{+} concentration of 500 mM at 12 °C. **b.** Plot of the change in chemical shift of the imino resonances with respect to UMP concentration.

that the binding is very weak. Although we cannot directly quantify the binding constant, because the change in $\Delta\delta_{G9(H1)}$ is too small, we estimate that the formation of a monomer-duplex complex is significantly less favored than the other RNA systems. All four imino resonances are relatively sharp at the beginning of the titration (0 M UMP concentration). Initially, $\delta_{G9(H1)}$ begins to shift downfield, which is what we expect for a pyrimidine base. This observation indicates that some amount of binding is actually occurring. However, after about 0.1 M UMP, the G9(H1) resonance reverse direction and starts to shift downfield in a fashion similar to that of the U resonances. We hypothesize that the binding of UMP is so weak, that the changes in chemical shift wrought by systematic changes in the physical environment outcompete those brought on by binding of UMP. Consistent with the hypothesis of weak binding, the G9(H1) and U8(H3) resonances both undergo significant line broadening. The likely explanation is that excess phosphate monoester results in catalysis of the exchange of G9(H1) and, to a lesser extent, U8(H3).

Titration of 5′-CCAAUAUUG-3′ with dGMP

The titration of dGMP into a solution containing the 5'-CCAAUAUUG-3' duplex from 0 to 250

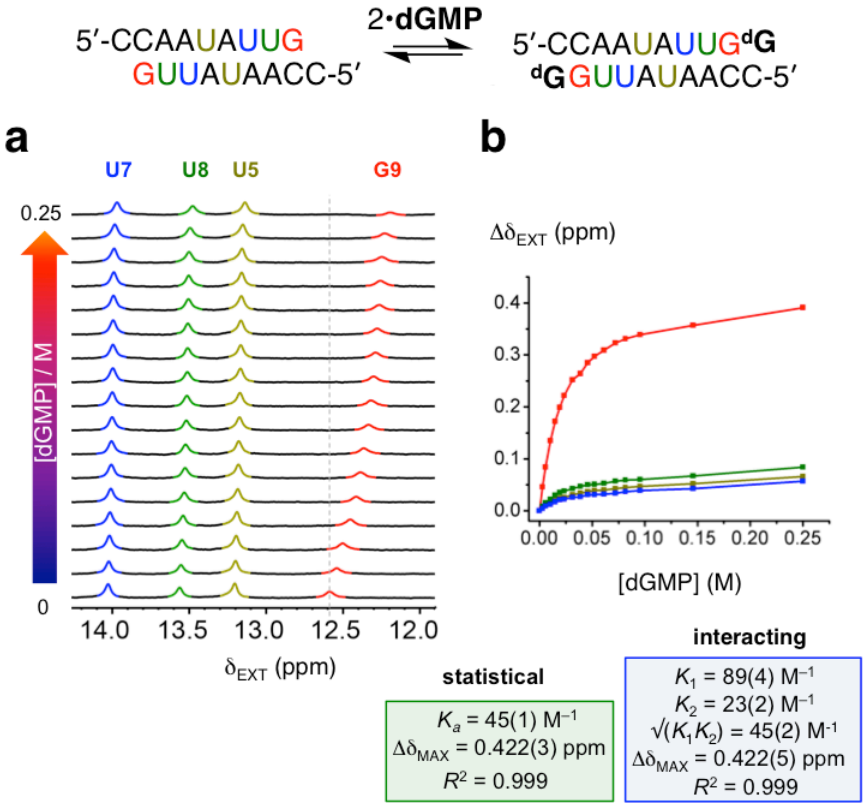


Figure S11. a. Titration of the 5′-CCAAUAUUG-3′ RNA duplex with dGMP while monitoring the imino region of the ^{1}H NMR spectra. Each spectrum was recorded at a total strand concentration of 3 mM at pH 7 at a Na^{+} concentration of 500 mM at 12 ^{o}C. **b.** Plot of the change in chemical shift of the imino resonances with respect to dGMP concentration. From these data, a fit to the statistical and interacting models was carried out.

mM is shown in Figure S11. All four resonances shift upfield as more dGMP is titrated into solution. The G9(H1) resonance undergoes the greatest amount of shifting. All four resonances were globally fit to the statistical and interacting binding models, respectively. The values of the binding constants obtained reveal that the binding of the first and second dGMP monomers follows a statistical mechanism.

Titration of 5′-GGAAUAUUC-3′ with rCMP

The titration of rCMP into the 5′-GGAAUAUUC-3′ duplex (Fig. S12), which has 3′-C instead of 3′-G adjacent to the binding site, was carried out in order to test the consequences of pyrimidine-pyrimidine over purine-pyrimidine base stacking. G2(H1) undergoes the largest change in frequency, shifting downfield by 0.25 ppm. The U8(H3) signal also shifts upfield, while the U7(H3) and U5(H3) signals both undergo modest downfield shifts. Comparing the values of the K_as obtained from the statistical and interacting models using a global fit of all four resonances reveals that the binding of the first and second monomers may be somewhat cooperative, with the binding of the first monomer assisting the binding of the second. We hypothesize that this

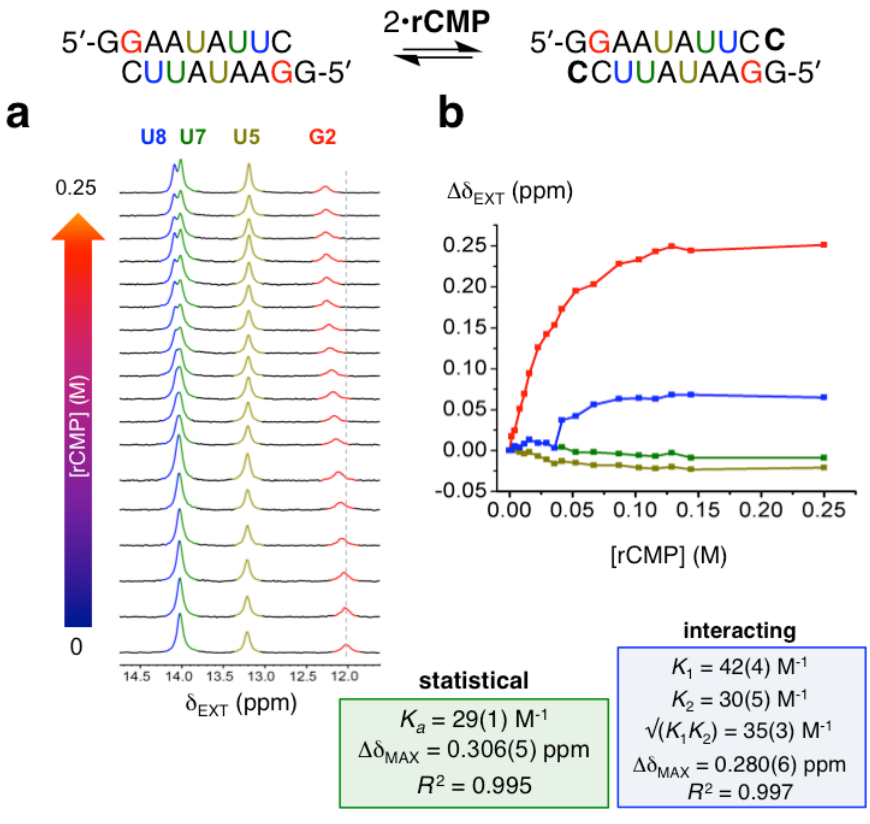


Figure S12. a. Titration of the 5′-GGAAUAUUC-3′ RNA duplex, possessing a C-primer nucleobase, with rCMP while monitoring the imino region of the ^{1}H NMR spectra. Each spectrum was recorded at a total strand concentration of 3 mM at pH 7 at a Na^+ concentration of 500 mM at 12 °C. **b.** Plot of the change in chemical shift of the imino resonances with respect to rCMP concentration. From these data, a fit to the statistical and interacting models was carried out.

observation might be the result of significant changes in conformation of the double helix that occur after binding of the first monomer that promote the binding of the second.

Titration of 5′-UCAAUAUUG-3′ with rGMP

We carried out a titration of rGMP into a solution of the 5′-UCAAUAUUG-3′ RNA double helix (Fig. S13) in order to evaluate the strength of a U:G mismatch. The G9(H1) resonance shifts upfield as the concentration of rGMP is increased. The U8(H3) resonance also shifts modestly upfield, while those of U5(H3) and U7(H3) shift modestly downfield. The G9(H1) signal starts out relatively broad at 0 mM rGMP, sharpens midway through the titration, and then broadens out again towards the end. Together, all these observations are consistent with specific, although weak, binding of rGMP to the 5′-U overhang of the 5'-UCAAUAUUG-3' RNA double helix. Binding of rGMP is sufficiently strong to slow down the exchange of the G9(H1) with water initially, but after about 50 mM, catalysis of the exchange process caused by excess rGMP begins to dominate, leading to line-broadening of the signal. This results in an increased

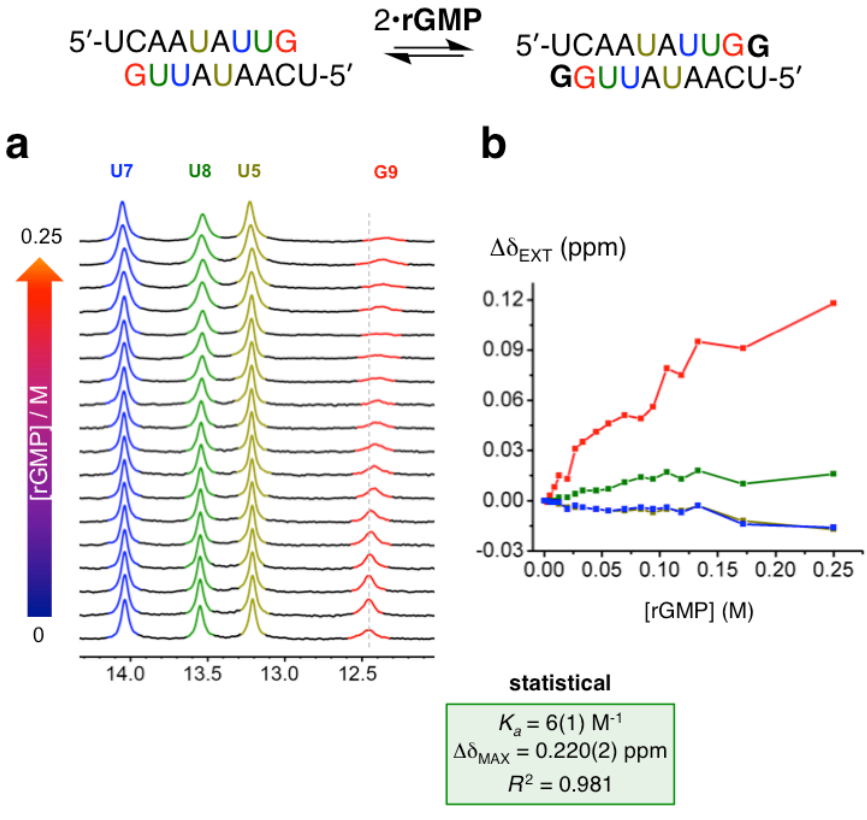


Figure S13 **a.** Titration of the 5′-UCAAUAUUG-3′ RNA duplex with rGMP while monitoring the imino region of the ^{1}H NMR spectra. Each spectrum was recorded at a total strand concentration of 3 mM at pH 7 at a Na^{+} concentration of 500 mM at 12 ^{o}C. **b.** Plot of the change in chemical shift of the imino resonances with respect to rGMP concentration. From these data, only a fit to the statistical was carried out. The fit to the interacting model produced values of binding constants with errors too large for confidence.

uncertainty of the peak maximum, which decreases the goodness of the fits obtained. The error obtained from the global fit to the interacting model is too large to conclude whether or not the binding is purely statistical.

Titration of 5′-CAAUAUUG-3′ with rCMP

As a control experiment, we carried out a titration on the 5′-CAAUAUUG-3′ RNA double helix, which possesses no 5′-overhang (Fig. S14). Only modest changes in the chemical shifts of all four resonances were observed ranging from approximately 0.01 to 0.02 ppm, an observation which indicates little to no interactions takes place between rCMP and the 5′-CAAUAUUG-3′ double helix. This magnitude of chemical shift, likely a consequence of systematic changes in the physical environment, is significantly smaller than the magnitude of chemical shift observed for the internal U resonances of the other RNA duplexes. The G8(H1) resonance undergoes significant line-broadening as the concentration of rCMP increases. This observation contrasts to the behavior of the analogous G9(H1) resonance when rCMP is titrated into a solution of the 5′-GCAAUAUUG-3′ double helix. This control experiment corroborates the hypothesis that excess monomer catalyzes exchange with H_2O, while monomer binding to an overhang slows it down.

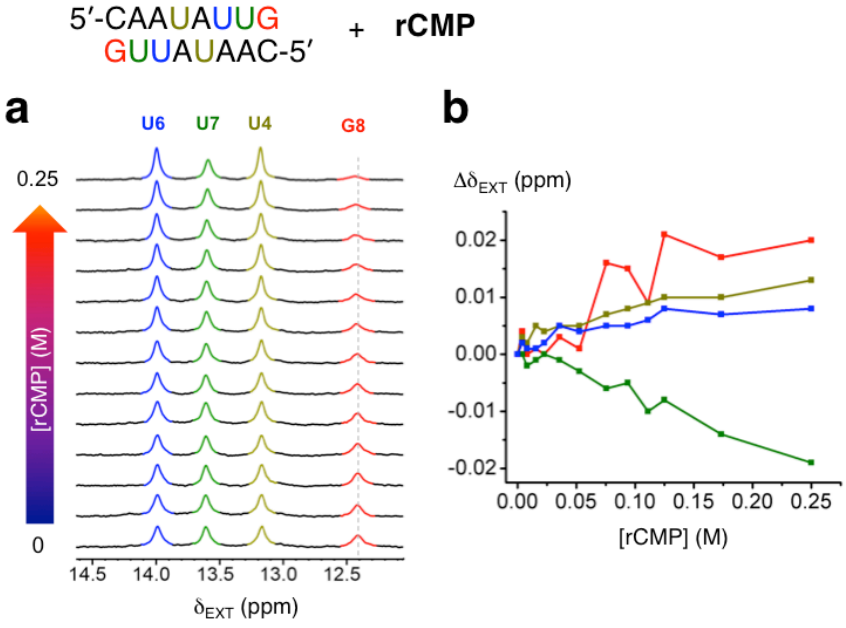
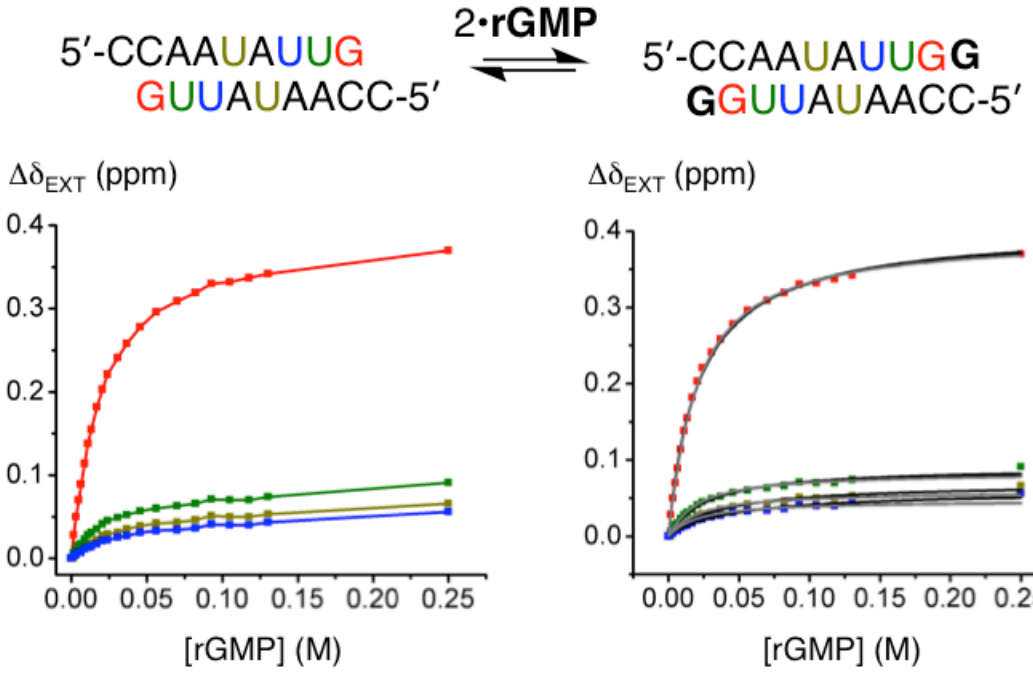
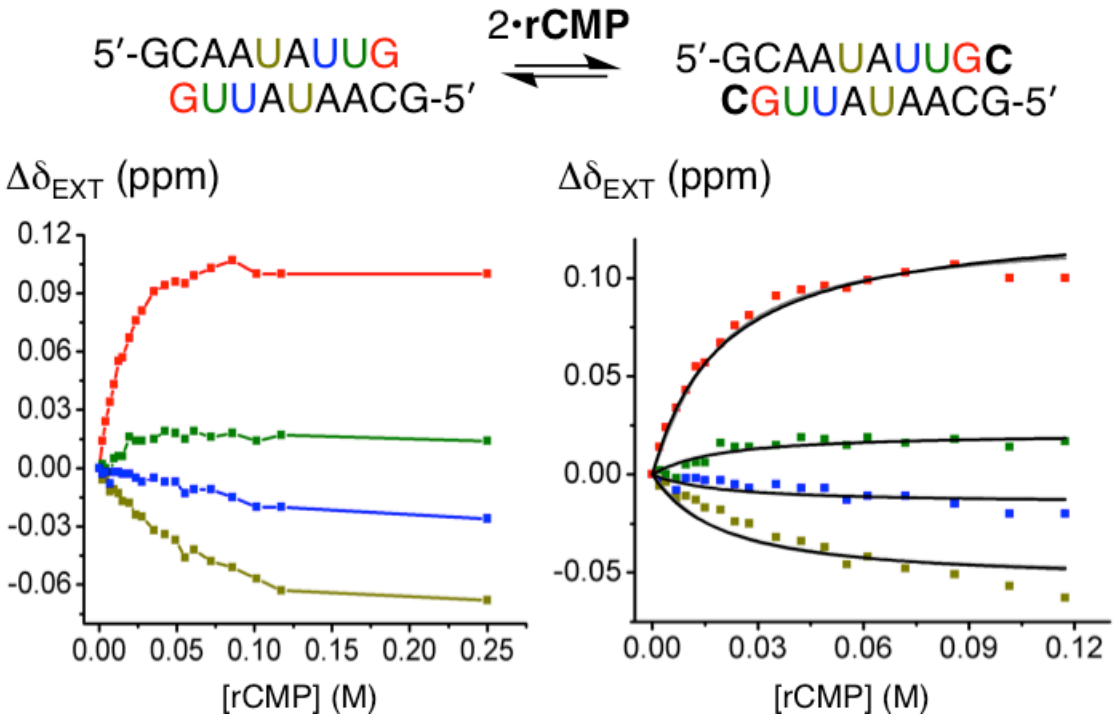
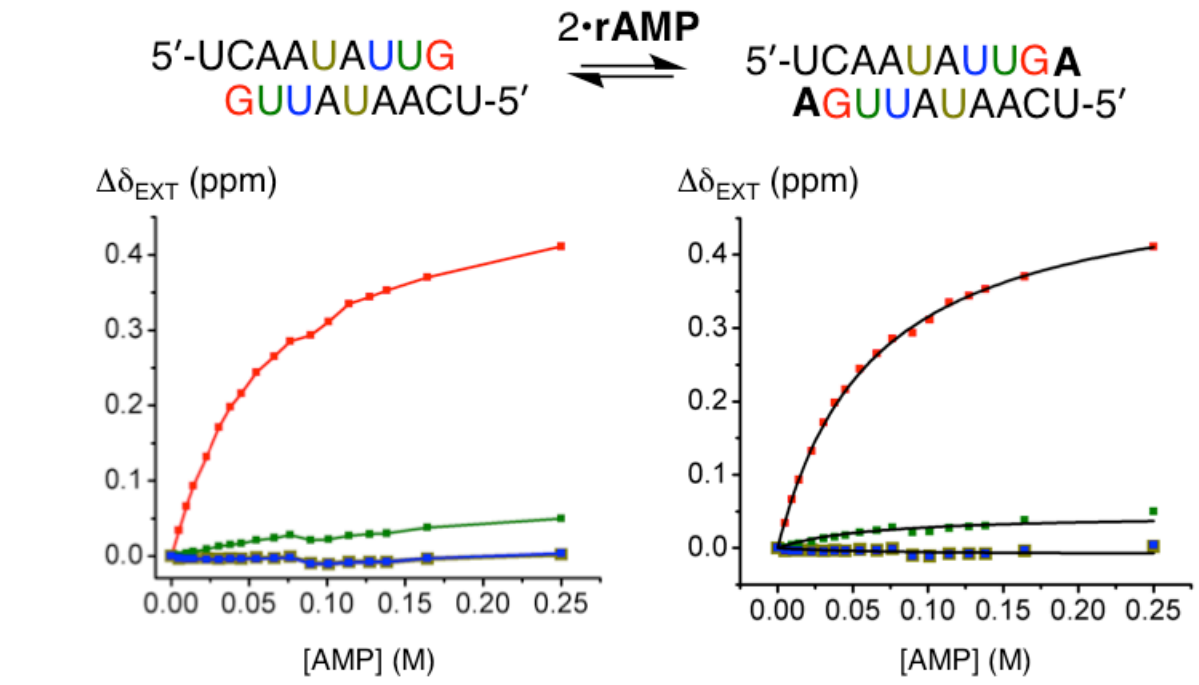


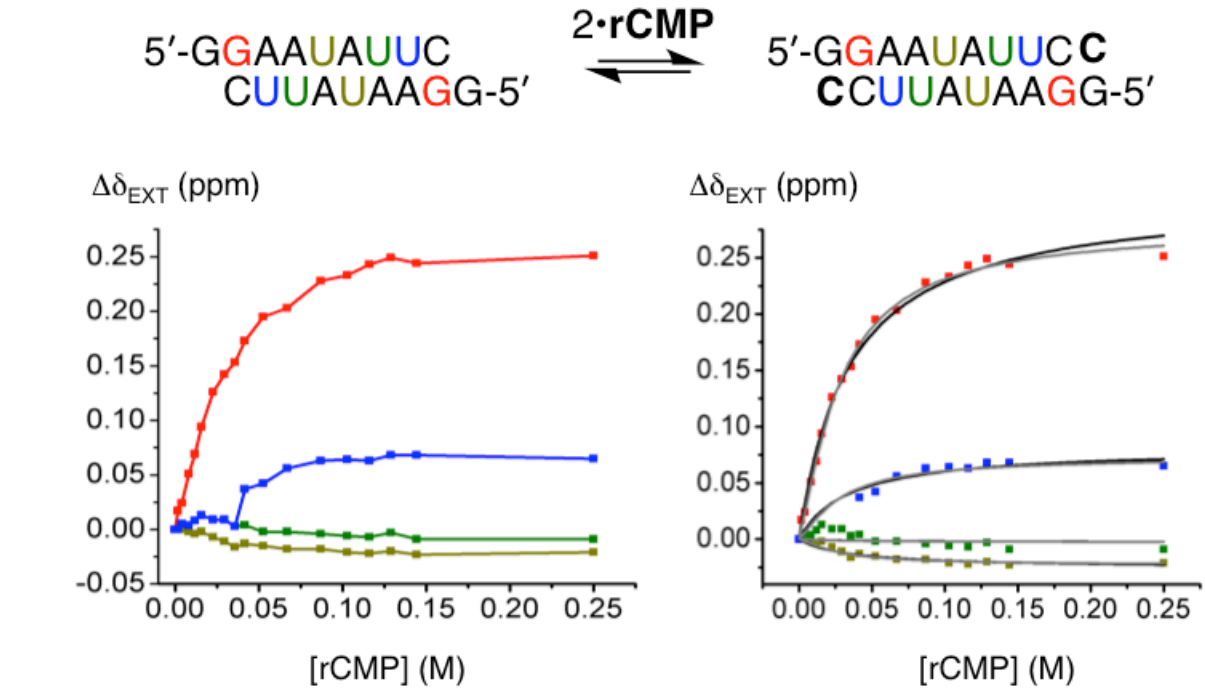
Figure S14. **a.** Titration of the 5′-CAAUAUUG-3′ duplex that possesses no overhanging binding site with rCMP while monitoring the imino region of the ^{1}H NMR spectra. Each spectrum was recorded at a total strand concentration of 3 mM at pH 7 at a Na^+ concentration of 500 mM at 12 ^{o}C. **b.** Plot of the change in chemical shift of the imino resonances with respect to rCMP concentration. Fits to the statistical and interacting models both yielded values with too much error to be confident in, a reflection of little-to-no binding.

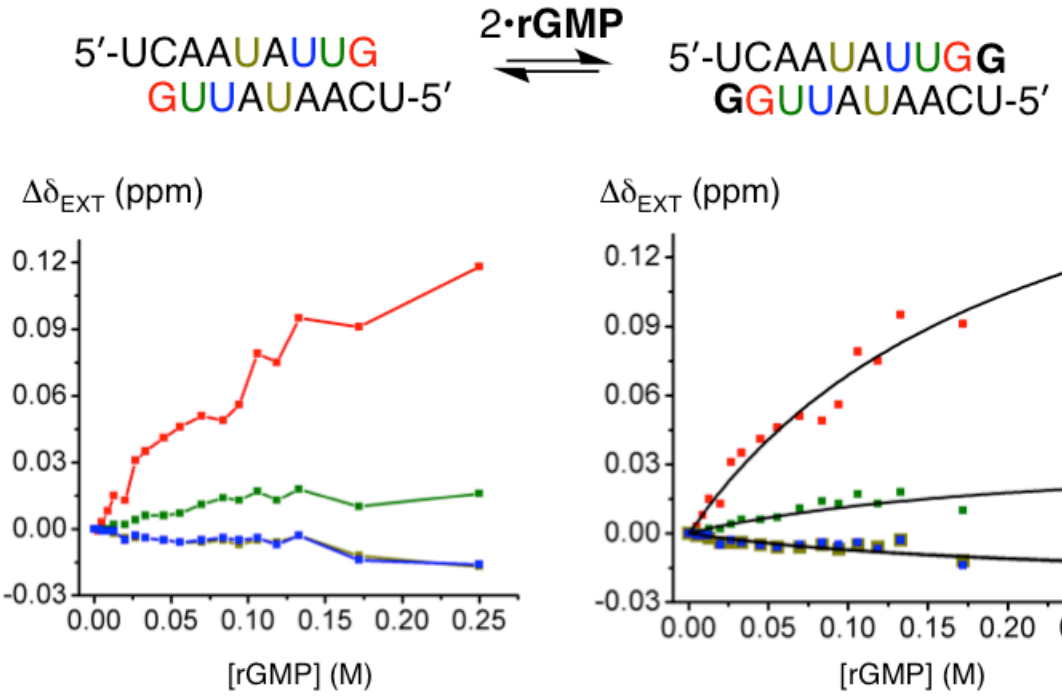
Below, in Fig. S15, we compare the best fit curves to the data presented for RNA titrations. Regarding the curves, the black lines are the statistical fits, while the gray ones are interacting fits. In most cases, these two lines overlap substantially.






5′-UCAAUAUUG
GUUAUAACU-5′
2•rAMP
5′-UCAAUAUUGA
AGUUAUAACU-5′
Δδ_EXT (ppm)
0.4
0.3
0.2
0.1
0.0
0.00 0.05 0.10 0.15 0.20 0.25
[AMP] (M)
Δδ_EXT (ppm)
0.4
0.3
0.2
0.1
0.0
0.00 0.05 0.10 0.15 0.20 0.25
[AMP] (M)


5′-GGAAUAUUC
CUUAUAAGG-5′
2•rCMP
5′-GGAAUAUUCC
CCUUAUAAGG-5′
Δδ_EXT (ppm)
0.25
0.20
0.15
0.10
0.05
0.00
-0.05
0.00 0.05 0.10 0.15 0.20 0.25
[rCMP] (M)
Δδ_EXT (ppm)
0.25
0.20
0.15
0.10
0.05
0.00
0.00 0.05 0.10 0.15 0.20 0.25
[rCMP] (M)



Curve fittings are based on the statistical binding model only.

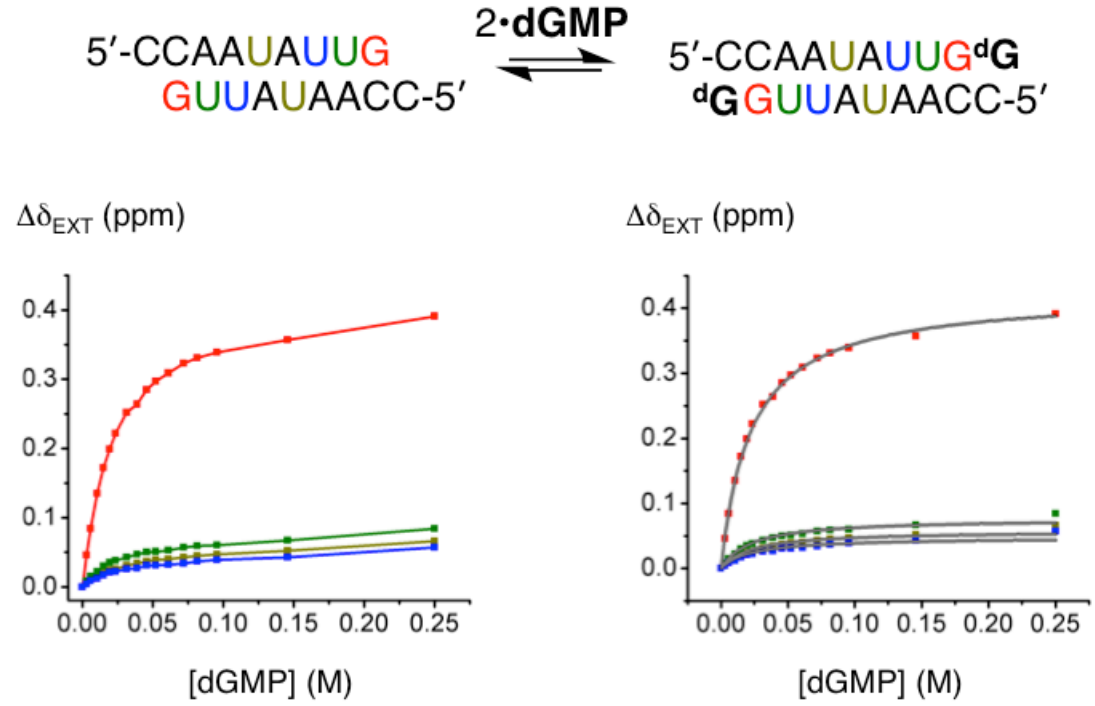


Fig. S15. Comparisons of curve fittings applied for selected RNA titrations.

DNA Duplexes

Titration of 5′-d(CCAATATTG)-3′ with rGMP

A titration of the 5′-d(CCAATATTG)-3′ duplex was performed with rGMP from 0 to 250 mM (Fig. S16). The G9(H1) resonance starts out broad, but quickly sharpens after the concentration of rGMP is raised. Towards the end of the titration, the signal becomes slightly broader again. All four imino proton resonances shift upfield as a function of rGMP concentration in a hyperbolic fashion. The G9(H1) signal shifts the most followed by the T8(H3), the T7(H3) and T5(H3) signals, respectively. However, unlike the RNA systems, a satisfactory rationale for which direction (upfield or downfield) that the DNA G9(H1) signal shifts is less obvious to us based on the MD minimized termini models (Fig. S32). Globally fitting the chemical shift data of all four resonances using the interacting and statistical models reveals that the binding of the first and second monomers is statistical.

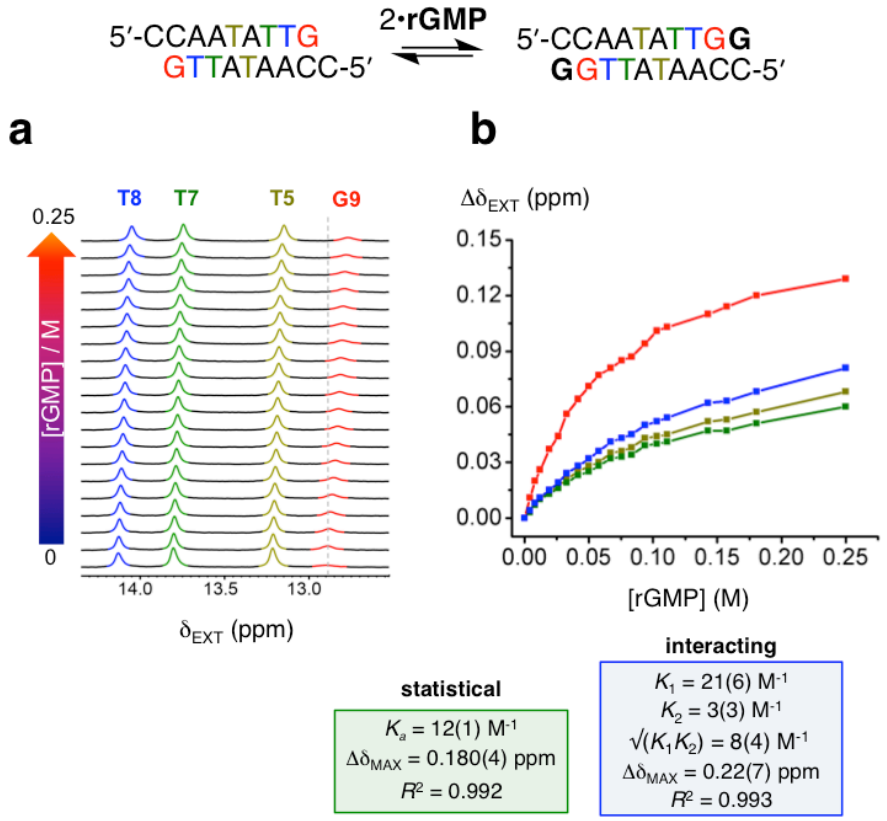


Figure S16. a. Titration of the 5′-d(CCAATATTG)-3′ duplex with rGMP while monitoring the imino region of the ^{1}H NMR spectra. Each spectrum was recorded at a total strand concentration of 3 mM at pH 7 at a Na^{+} concentration of 500 mM at 12 ^{o}C. **b.** Plot of the change in chemical shift of the imino resonances with respect to rGMP concentration. From these data, a fit to the statistical and interacting models was carried out.

Titration of 5′-d(GCAATATTG)-3′ with rCMP

A titration of the 5′-d(GCAATATTG)-3′ duplex with rCMP was carried out from 0 to 250 mM (Fig. S17). The G9(H1) resonance starts out broad, but sharpens as it shifts downfield with increasing rCMP concentrations. A global fit of the chemical shift data to the interacting and statistical models reveals that the binding may be slightly cooperative. This result could be an artifact of the relatively small change in chemical shift of the G9(H1) resonance (~0.06 ppm), a fact that leads to a larger effect of any systematic error in measurement of the peak positions.

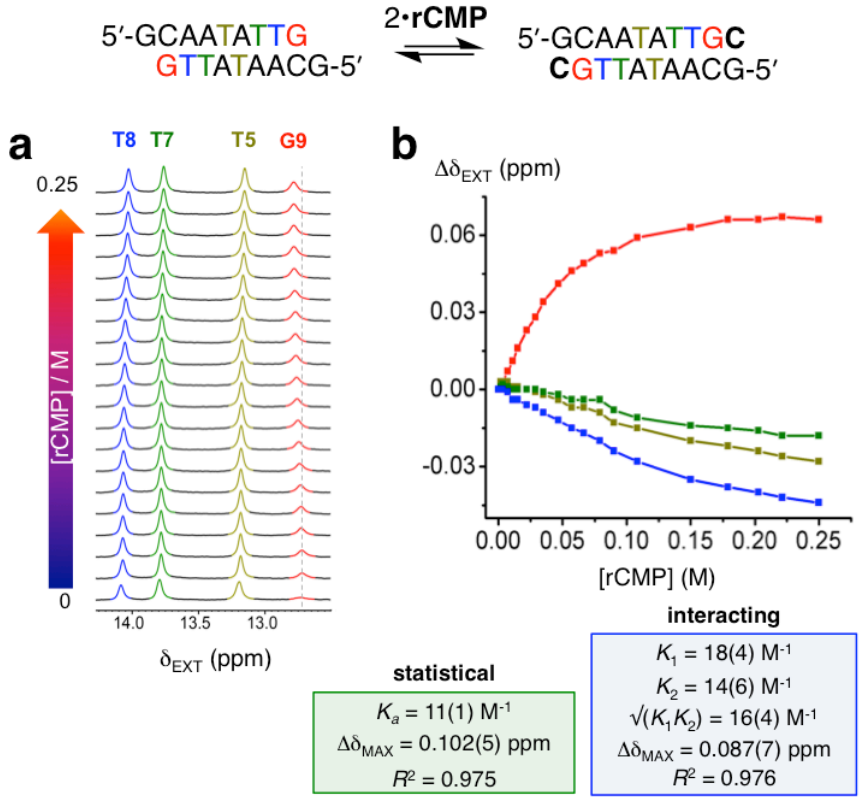


Figure S17. a. Titration of the 5′-d(GCAATATTG)-3′ duplex with rCMP while monitoring the imino region of the ^{1}H NMR spectra. Each spectrum was recorded at a total strand concentration of 3 mM at pH 7 at a Na^{+} concentration of 500 mM at 12 °C. **b.** Plot of the change in chemical shift of the imino resonances with respect to rCMP concentration. From these data, a fit to the statistical and interacting models was carried out.

Titration of 5′-d(TCAATATTG)-3′ with rAMP

The 5′-d(TCAATATTG)-3′ duplex was titrated with rAMP from 0 to 250 mM (Fig. S18). As the concentration of rAMP increases, the G9(H1) resonance begins to sharpen and grow in intensity; however, it gets broader after the rAMP concentration exceeded approximately 50 mM. This initial sharpening followed by broadening is consistent with the hypothesis of relatively weak binding of rAMP, followed by catalysis of proton exchange by excess rAMP at higher concentrations. The T imino resonances also undergo upfield shifting. The relatively small shifts of all three T imino protons display some degree of hyperbolic behavior. Comparison of the

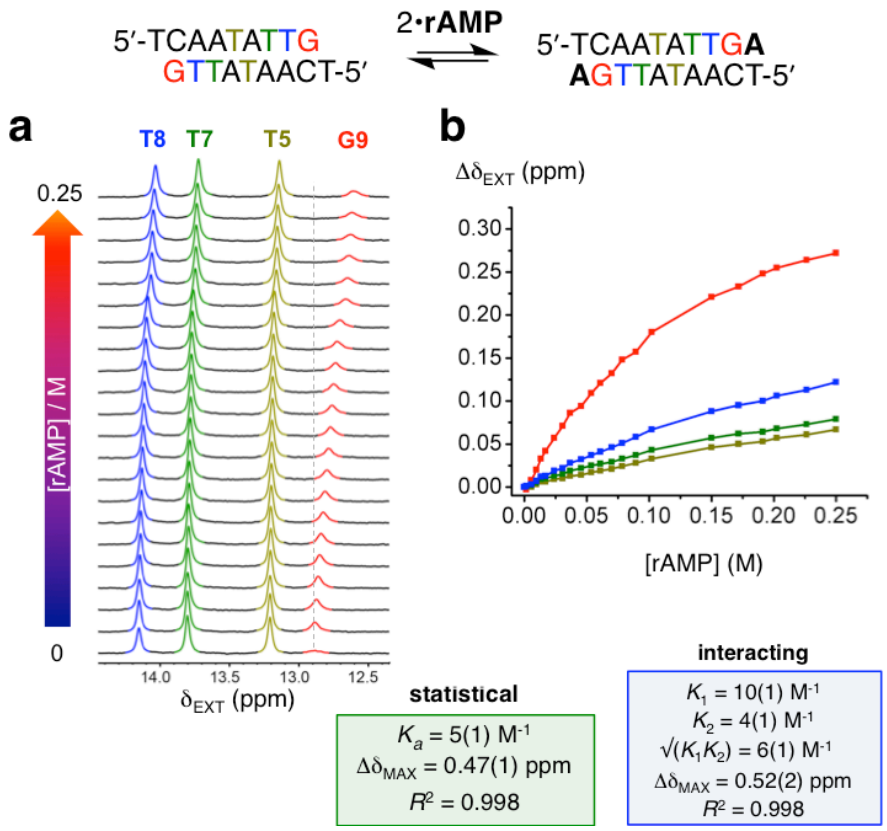


Figure S18. a. Titration of the 5′-d(TCAATATTG)-3′ duplex with rAMP while monitoring the imino region of the ^{1}H NMR spectra. Each spectrum was recorded at a total strand concentration of 3 mM at pH 7 at a Na^{+} concentration of 500 mM at 12 ^{o}C. **b.** Plot of the change in chemical shift of the imino resonances with respect to rAMP concentration. From these data, a fit to the statistical and interacting models was carried out.

values of the binding constants obtained using global fits with the isotherm models reveals that the binding of the first and second monomers are purely statistical.

Titration of 5′-d(ACAATATTG)-3′ with TMP

The titration of the 5′-d(ACAATATTG)-3′ duplex with TMP was performed from 0 to 250 mM (Fig. S19). The line width of the G9(H1) resonance starts out large, but gets smaller after the first addition of TMP, then becomes significantly large by 250 mM of TMP. Only modest downfield changes in the chemical shift of the G9(H1) signal are observed, which appear nearly linear. The other T signals all move upfield at increasing concentrations of TMP. These observations are consistent with binding of TMP, although it is weak. The observed changes in chemical shift, however, are insufficient for accurately measuring a binding constant, which is likely on the order of 1 M^{-1}.

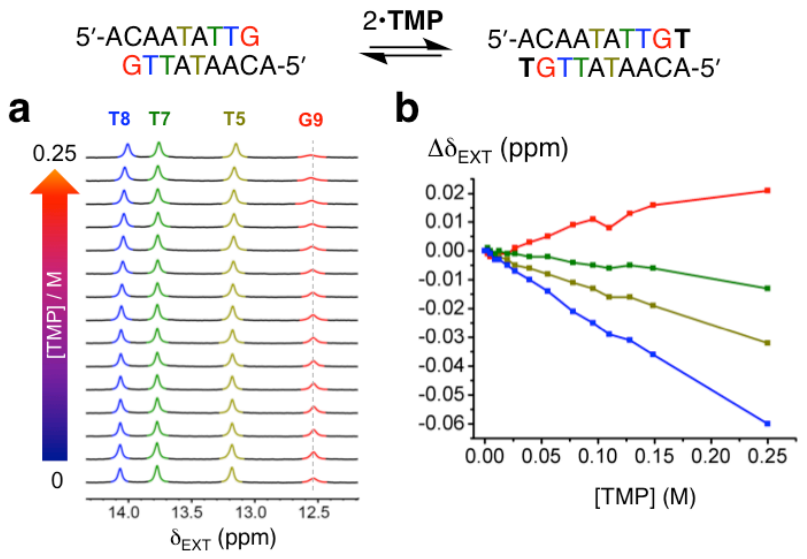


Figure S19. a. Titration of the 5′-d(ACAATATTG)-3′ duplex with TMP while monitoring the imino region of the ^{1}H NMR spectra. Each spectrum was recorded at a total strand concentration of 3 mM and Na^{+} concentration of 500 mM at pH 7 and 12 ^{o}C. **b.** Plot of the change in chemical shift of the imino proton resonances with respect to TMP concentration. Fits to the statistical and interacting models were carried out, but the values of K_a possessed errors too large to be confident in.

Titration of 5′-d(GGAATATTC)-3′ with rCMP

Effect of the nucleobase stacking on the noncovalent monomer binding thermodynamics was studied through titration of the 5′-d(GGAATATTC)-3′ duplex with rCMP (Fig. S20). Association constant, K_a for rCMP was too small to be measured due to the fact that no observable hyperbolic change occurred in the chemical shift of the $\delta_{G2(H1)}$.

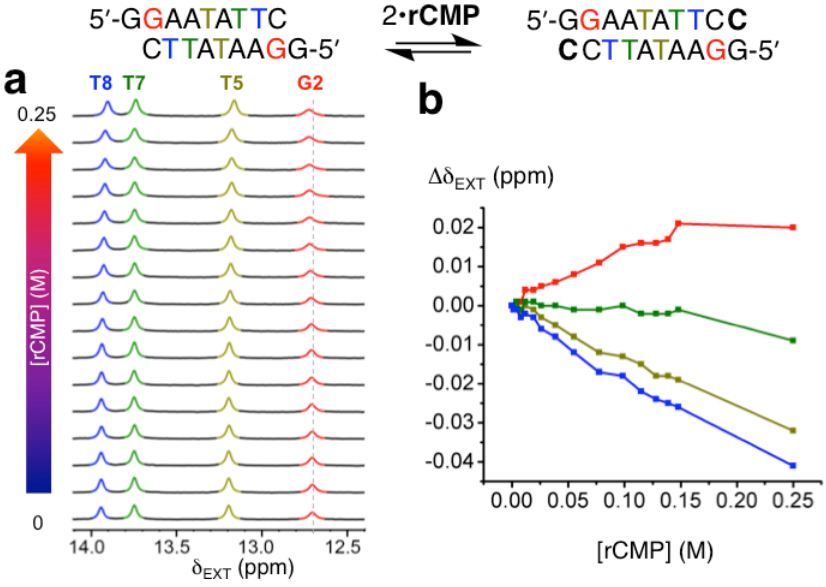


Figure S20. a. Titration of the 5′-d(GGAATATTC)-3′ duplex with rCMP while monitoring the imino region of the ^{1}H NMR spectra. Each spectrum was recorded at a total strand concentration of 3 mM and Na^{+} concentration of 500 mM at pH 7 and 12 ^{o}C. **b.** Plot of the change in chemical shift of the imino resonances with respect to rCMP concentration. Fits to the statistical and interacting models were carried out, but the values of K_a possessed errors too large to be confident in.

Titration of 5′-d(CCAATATTG)-3′ with dGMP

A comparative titration experiment involving a deoxyribonucleotide monomer was performed using the 5′-d(CCAATATTG)-3′ duplex and dGMP monomer (Fig. S21). The G9(H1) signal sharpens during the first half of the titration, as the concentration of rGMP is increased. Like in the case of rGMP titration (cf. Fig. 16), all four imino proton resonances shift upfield as a function of rGMP concentration in a hyperbolic fashion.

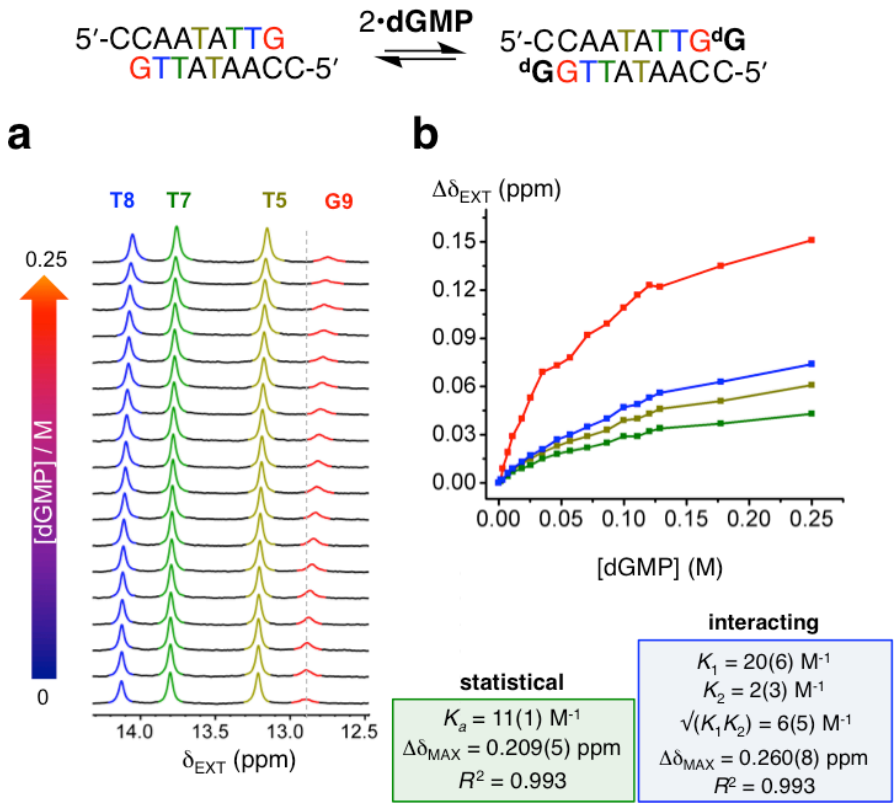
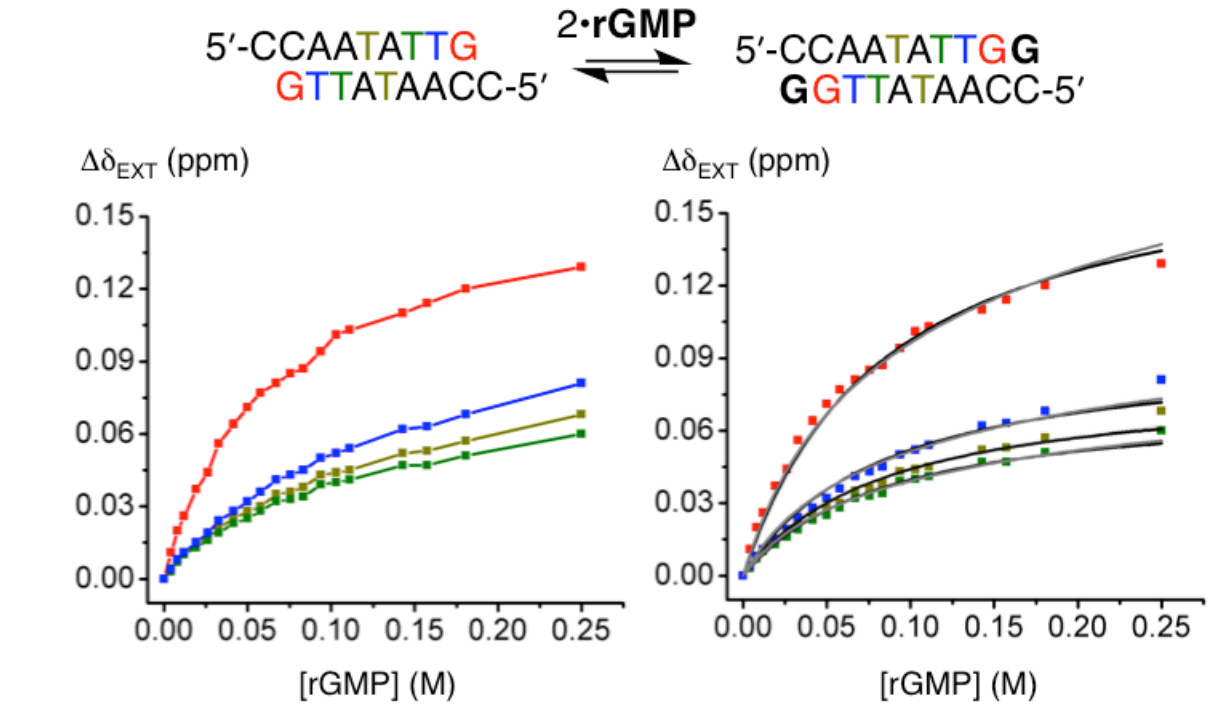
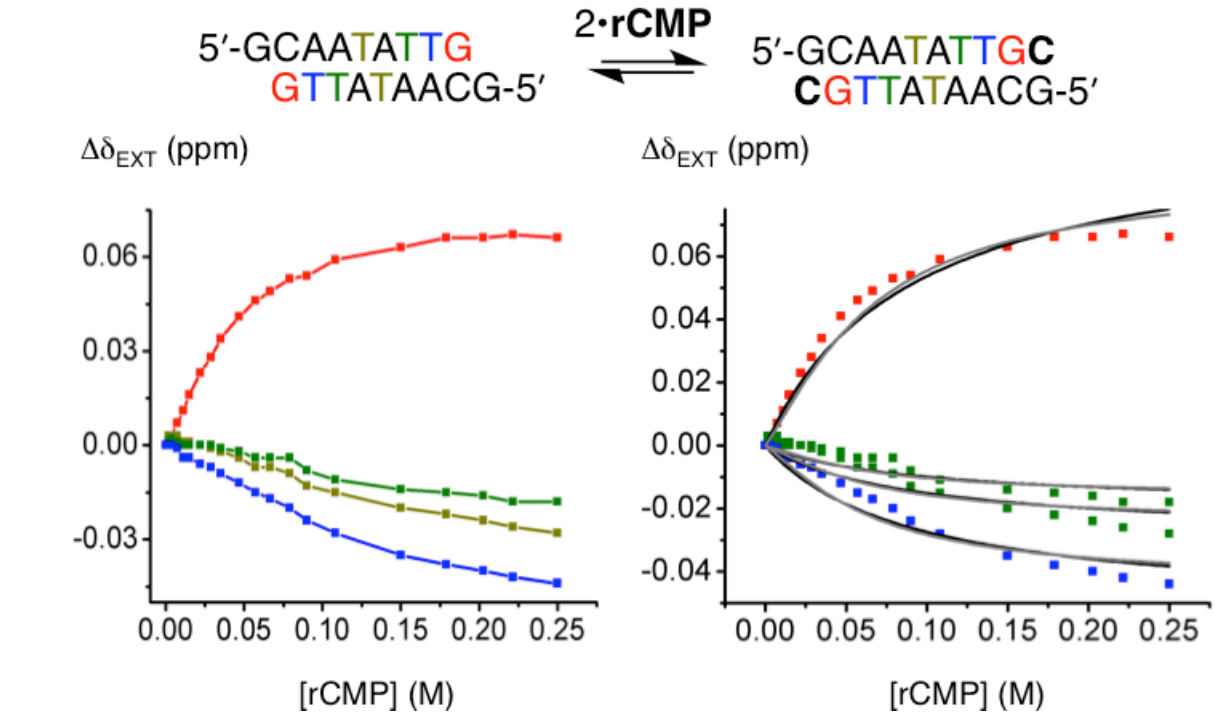


Figure S21. a. Titration of the 5′-d(CGAATATTC)-3′ duplex with dGMP while monitoring the imino region of the ^{1}H NMR spectra. Each spectrum was recorded at a total strand concentration of 3 mM and Na^{+} concentration of 500 mM at pH 7 and 12 ^{o}C. **b.** Plot of the change in chemical shift of the imino resonances with respect to dGMP concentration. From these data, a fit to the statistical and interacting models was carried out.

Below, in Fig. S22, we compare the best fit curves to the data presented for DNA titrations. Regarding the curves, the black lines are the statistical fits, while the gray ones are interacting fits. In most cases, these two lines overlap substantially.





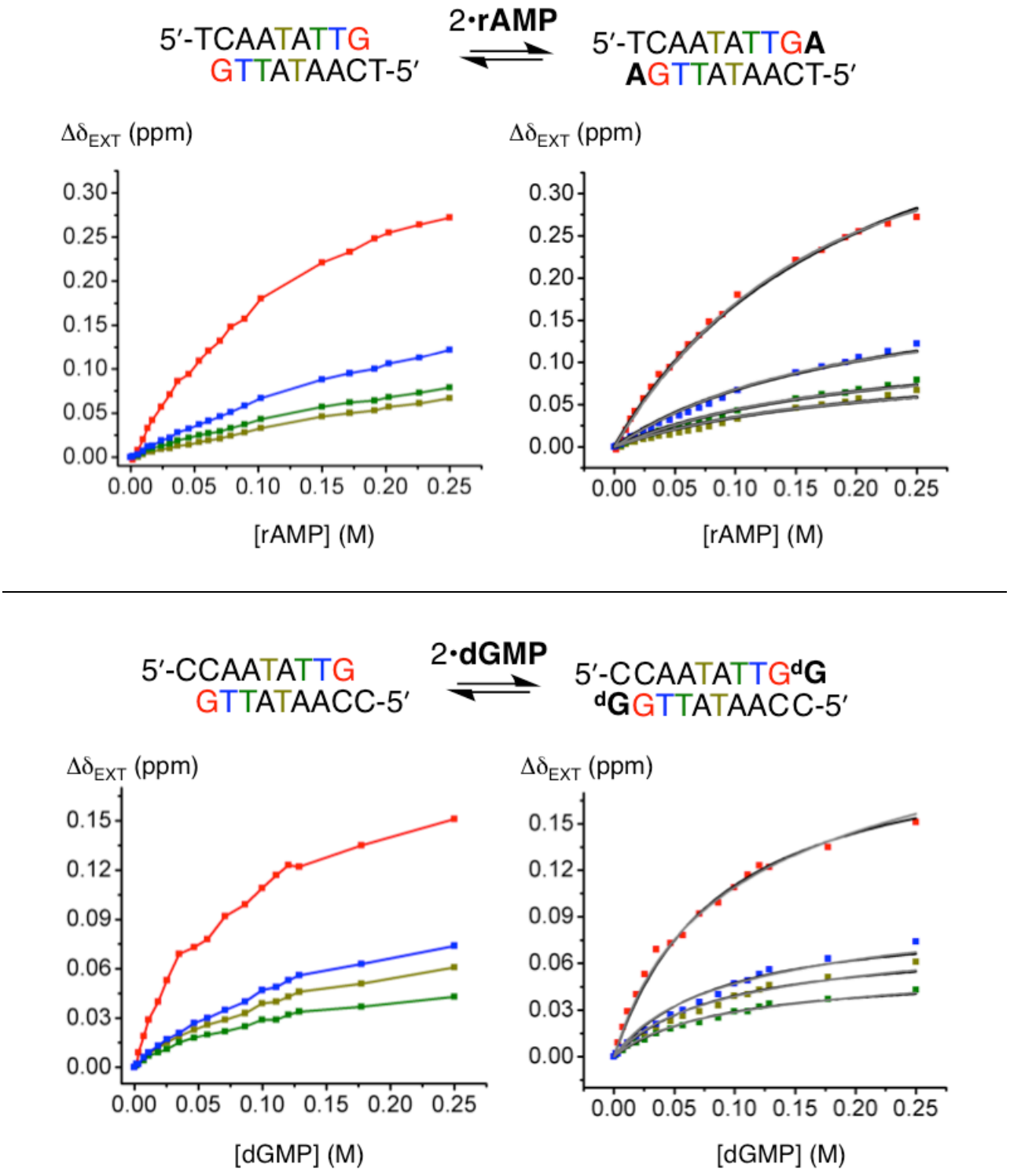


Fig. S22. Comparisons of curve fittings applied for selected DNA titrations.

8. Job Plot with 5′-d(TCAATATTG)-3′ and rAMP

In order to confirm the 2:1 stoichiometry of the binding event, we carried out a titration by the method of continuous variation on the 5′-d(TCAATATTG)-3′ duplex with rAMP from which we constructed a Job plot (Fig. S23). The total concentration of duplex and rAMP was kept constant at 20 mM as the mole fraction was varied between 0 and 1. Increasing the mole fraction of the duplex results in upfield shifts of all four imino signals. The most shifted resonance is G9(H1) while the least shifted is T5(H3). A Job plot of the change in chemical shift multiplied by the mol fraction of duplex against the mol fraction of duplex reveals a maximum around a mole fraction 0.35. This value is consistent with a 2:1 binding stoichiometry.

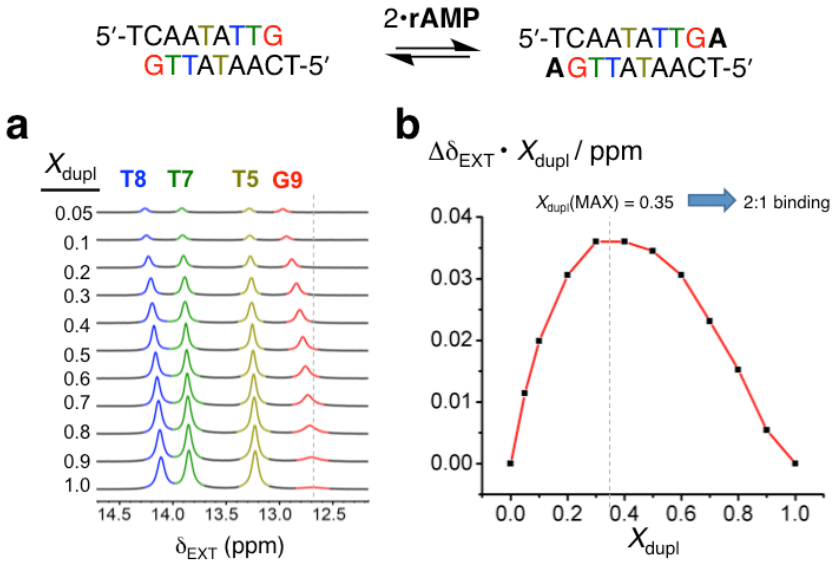


Figure S23. a. Titration by continuous variation of the concentration of 5′-d(TCAATATTG)-3′ duplex with rAMP while monitoring the imino region of the ^{1}H NMR spectra. The total concentration of duplex and rAMP was kept constant at 20 mM for each spectrum recorded. Each spectrum was recorded at a Na^{+} concentration of 500 mM at pH 7 and 12 ^{o}C. **b.** Job plot of the change in chemical shift of the imino resonances multiplied by the mole fraction of duplex ($\Delta\delta_{EXT} \bullet X_{dupl}$) versus the mole fraction of duplex (X_{dupl}). The Job plot is consistent with a 2 to 1 binding stoichiometry.

9. Nearest-Neighbor Analysis

All the nearest-neighbor parameters employed have been established by Turner et al.[5] We first calculate (Fig. S24) the enthalpic (ΔH_1) and entropic (ΔS_1) contributions for the full-length sequence. This full-length sequence serves as a model for the noncovalently bound rNMP-primer-template complex, where X is the 5′-overhang and Y is the complementary monomer. In the case of the nearest-neighbor analysis, the 3′-end Y is actually covalently bound to G, which will lead to larger predicted energies than what is observed. We next calculate the enthalpic (ΔH_2) and entropic (ΔS_2) contributions from the dangling-end duplexes, which omit Y. These dangling-end sequences are the exact same ones used in the experimental titrations. Finally, we subtract the enthalpic and entropic contributions of the dangling-end sequence from those of the full-length. Since the core sequences of both these duplexes are identical, the nearest neighbor

Full-length Sequence

5′-XCAAUAUUGY
YGUUAUAACX-5′

ΔH_1 = initiation + symmetry correction + 2(XC/GY + CA/UG + AA/UU + AU/AU) + UA/UA
ΔS_1 = initiation + symmetry correction + 2(XC/GY + CA/UG + AA/UU + AU/AU) + UA/UA

Dangling-end Sequence

5′-XCAAUAUUG
GUUAUAACX-5′

ΔH_1 = initiation + symmetry correction + 2(XC/G + CA/UG + AA/UU + AU/AU) + UA/UA
ΔS_1 = initiation + symmetry correction + 2(XC/G + CA/UG + AA/UU + AU/AU) + UA/UA

$\Delta H_1 - \Delta H_2 = \Delta\Delta H = 2(\text{XC/GY} - \text{XC/G})$
$\Delta S_1 - \Delta S_2 = \Delta\Delta S = 2(\text{XC/GY} - \text{XC/G})$
nearest-neighbor energy = $\Delta\Delta G/2$

Figure S24. Nearest-neighbor analysis. The propagation sequences are represented in 5′ to 3′ direction

parameters from these regions all cancel out. The only sequence parameters that do not cancel are those of the termini and dangling ends, respectively. The differences in enthalpy ($\Delta\Delta H$) and entropy ($\Delta\Delta S$) are used to calculate $\Delta\Delta G$ at 12 °C. As a consequence of the fact that the duplexes are palindromic, we normalize $\Delta\Delta G$ by a factor of two. This value is the nearest-neighbor predicted energy shown in Fig. 6 of the main text.

10. Statistical Correlation and Error Analysis

We have performed a statistical study of the correlation between $\Delta\delta_{MAX}$ and K_a (Fig. 25). We used all binding constants available from both RNA and DNA primer-template complexes. No correlation was found, i.e., Pearson's correlation coefficient was determined to be about –0.05. The correlation matrix scatter plots are shown below.

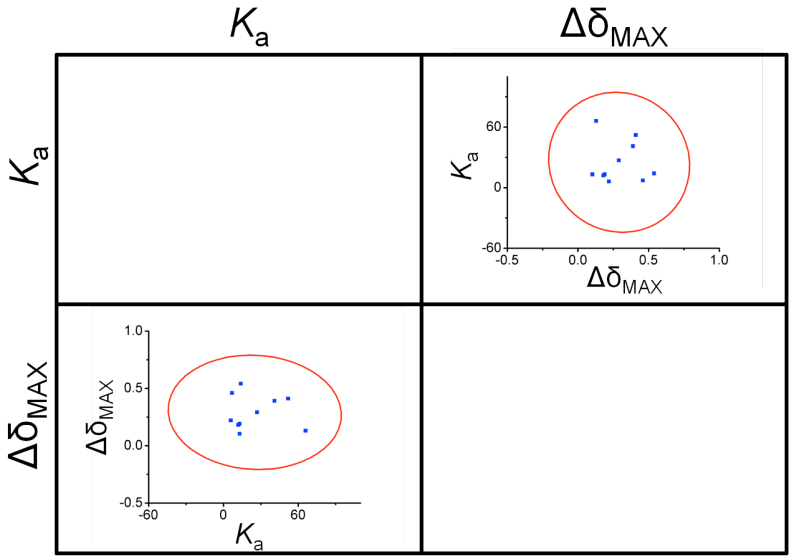


Fig. S25. The correlation matrix scatter plots.

This approximately null result is consistent with the hypothesis that the measured $\Delta\delta_{MAX}$ for the G9(H1) signal is mostly a result of ring-current effects, which are brought on by, and specific to the identity of each NMP monomer. The lack of correlation between $\Delta\delta_{MAX}$ and G9(H1) further strengthens the significance of the measured differences in binding energies between all the different monomers studied.

We have performed an error analysis of all RNA monomers using the two-tailed Student's *t* test (Fig. 26) in order to calculate the probability that the measured average values between any two monomers are truly different. The results are shown below.

two-tailed Student's *t* test

rGMP	X					
rCMP	0.800	X				
rAMP	>0.999	0.996	X			
UMP	NA	NA	NA	X		
dGMP	0.946	0.948	0.999	NA	X	
rCMP*	0.996	0.990	0.870	NA	0.981	X
	rGMP	**rCMP**	**rAMP**	**UMP**	**dGMP**	**rCMP***

Fig. S26. Error analysis performed by Student's *t* test. Here, rCMP* corresponds to the titration of rCMP to the 5′-GGAAUAUUC-3′ RNA duplex.

In each comparison, there is about a 95% chance or more that the measured binding constant values are truly different, with only two exceptions. The first exception is in the comparison of rGMP with rCMP, wherein the probability is at 80%. This slightly lower value is a consequence of the fact that the average values of K_a are not only relatively close to each other, but also the variance for the rCMP binding constant is relatively large. The reason for this large variance comes from the smaller $\Delta\delta_{MAX}$ (0.11 ppm) of the G9(H1), which makes the relative error for each individual chemical shift measurement larger compared to all other monomers leading to larger variance in the measured K_a values. The second exception involves the comparison of rCMP* with rAMP, which has an 87% probability. Again, this slightly lower percentage is mostly the result of the relatively large variance in the average value of K_a for rCMP*. Overall, however, the odds are in favor across the board that each measured average K_a is truly different from every other, which strengthens our justification for carrying out the nearest-neighbor analysis in particular.

11. Molecular Modeling

Below, selected minimized models of the rNMP monomers both bound and unbound to RNA and DNA duplexes. For each sequence, an ideal A-form 10-mer RNA and B-form 10-mer DNA duplex generated by COOT was used to model both the free and monomer-bound states of the 9-

mer duplex with one-nucleotide 5′-overhangs at both ends. See Methods in the main text for further details.

Binding of rGMP to the 5′-CCAAUAUUG-3′ duplex

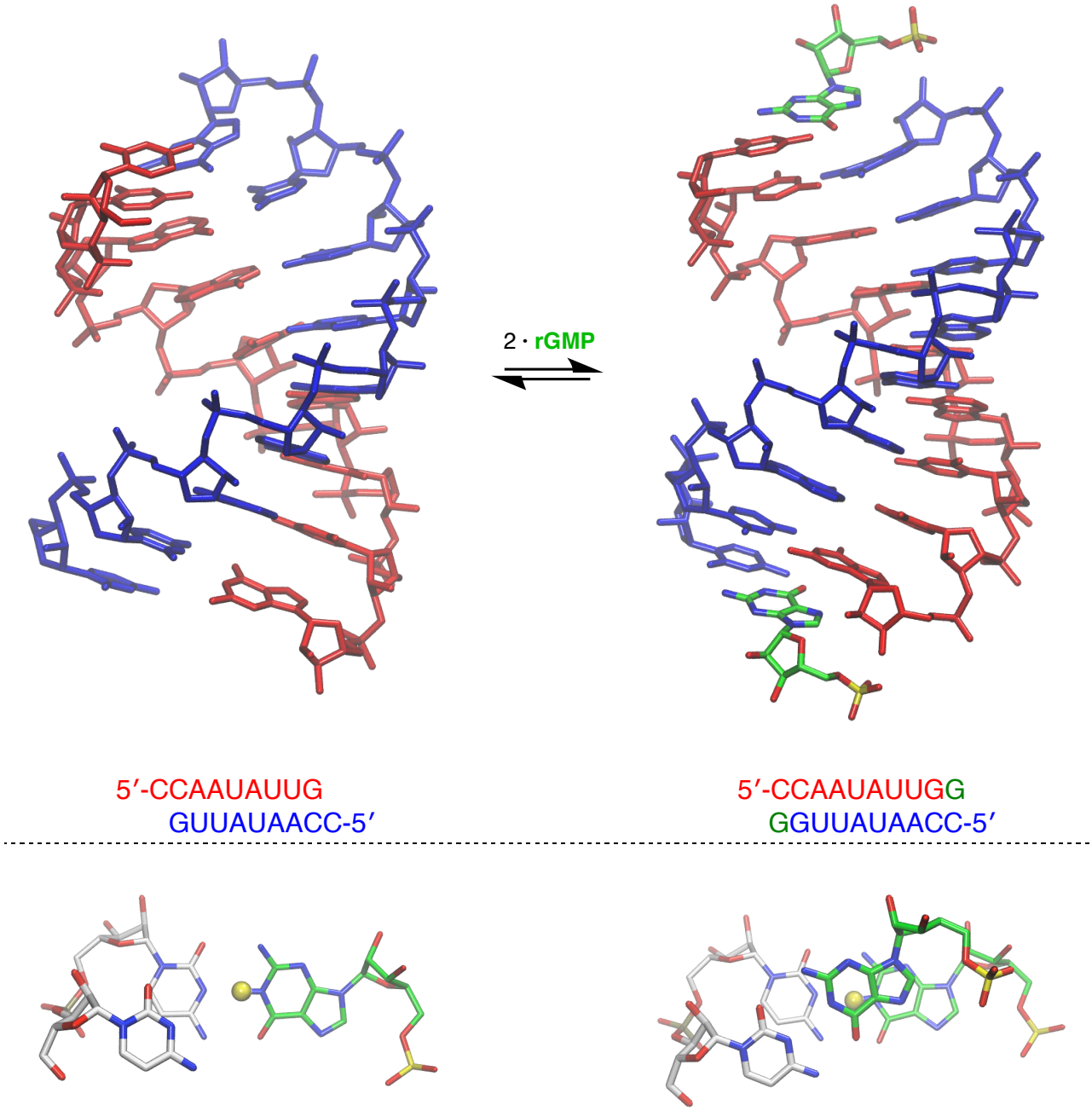


Figure S27. Top: Energy minimized models of A-form 5′-CCAAUAUUG-3′ RNA duplex in both free and rGMP-bound states. **Bottom:** Rotated views of the modeled duplex terminus.

Binding of rCMP to the 5′-GCAAUAUUG-3′ RNA duplex

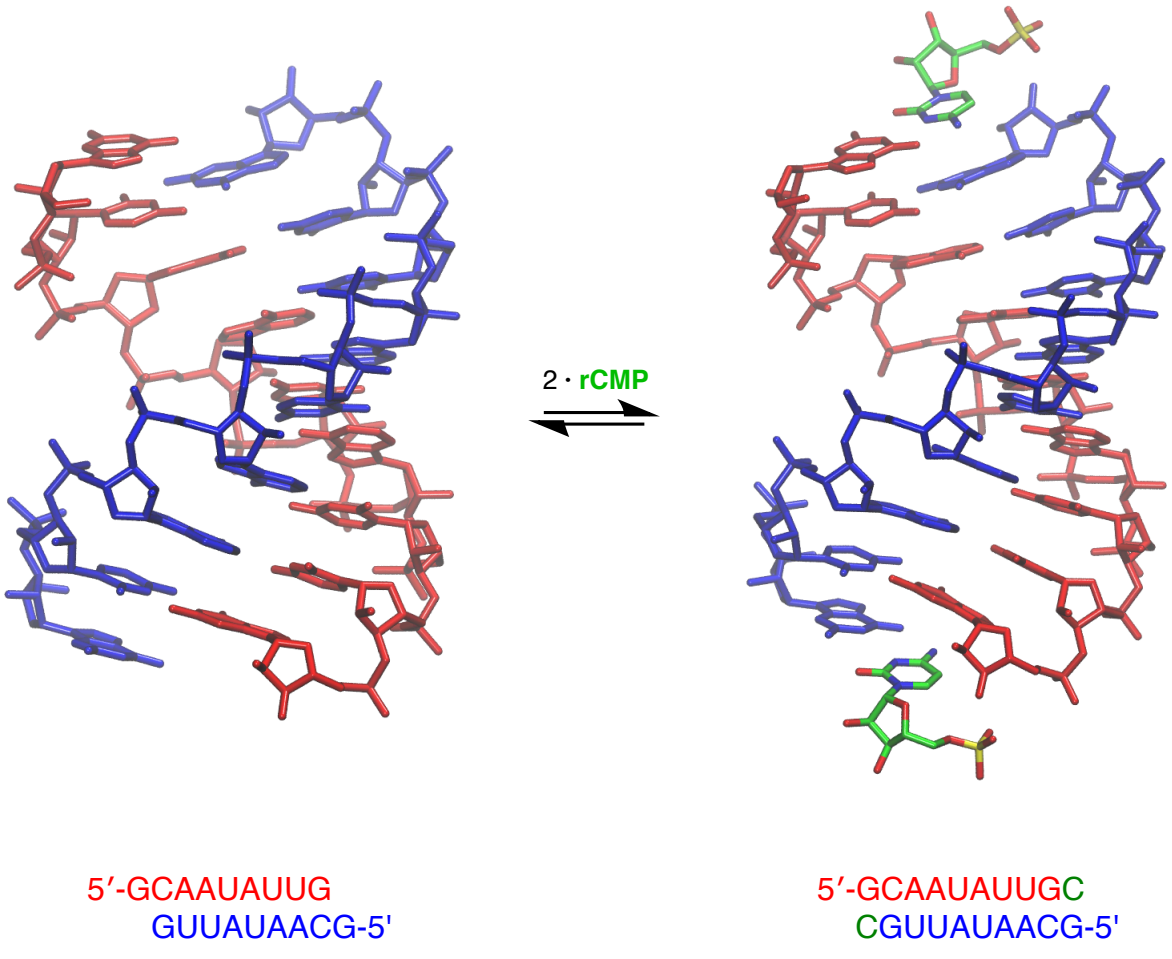


Figure S28. Top: Energy minimized models of A-form 5′-GCAAUAUUG-3′ RNA duplex in both free and rCMP-bound states. **Bottom:** Rotated views of the modeled duplex terminus.

Binding of rAMP to the 5′-UCAAUAUUG-3′ duplex

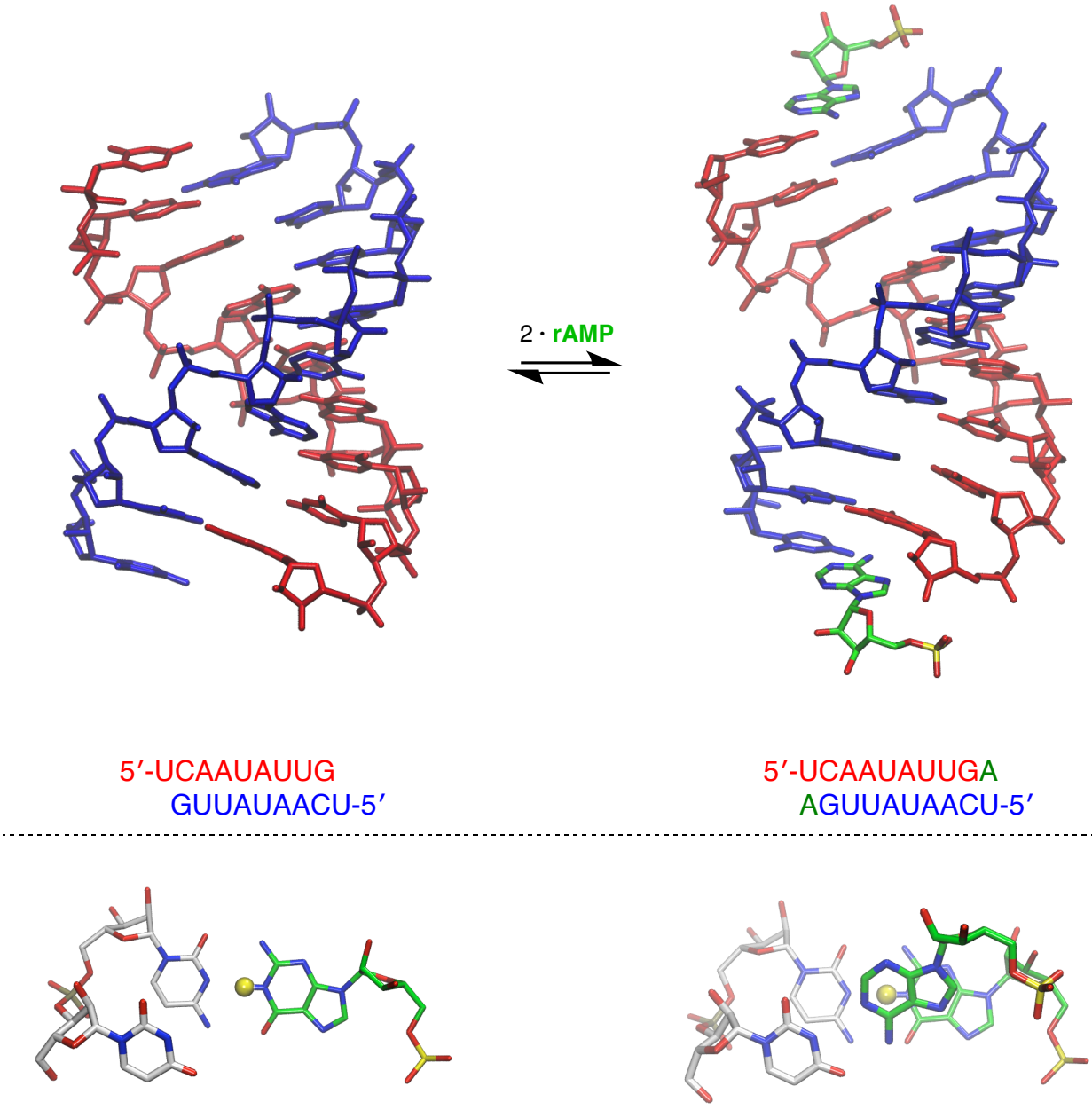


Figure S29. Top: Energy minimized models of A-form 5′-UCAAUAUUG-3′ RNA duplex in both free and rAMP-bound states. **Bottom:** Rotated views of the modeled duplex terminus.

Binding of rCMP to the 5′-GGAAUAUUC-3′ RNA duplex

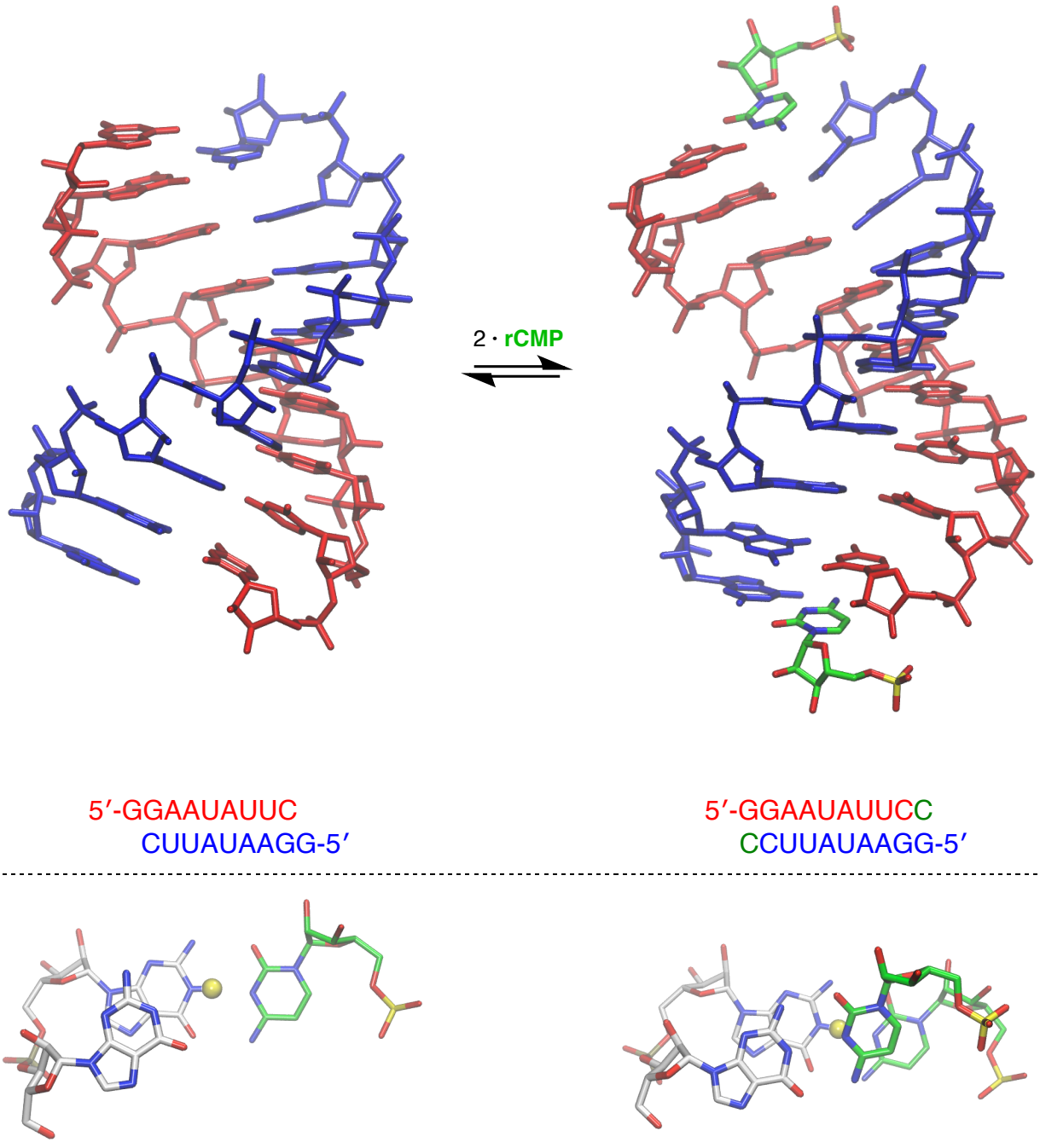


Figure S30. Top: Energy minimized models of A-form 5′-GGAAUAUUC-3′ RNA duplex in both free and rCMP-bound states. **Bottom:** Rotated views of the modeled duplex terminus.

Binding of rGMP to the 5′-UCAAUAUUG-3′ RNA duplex

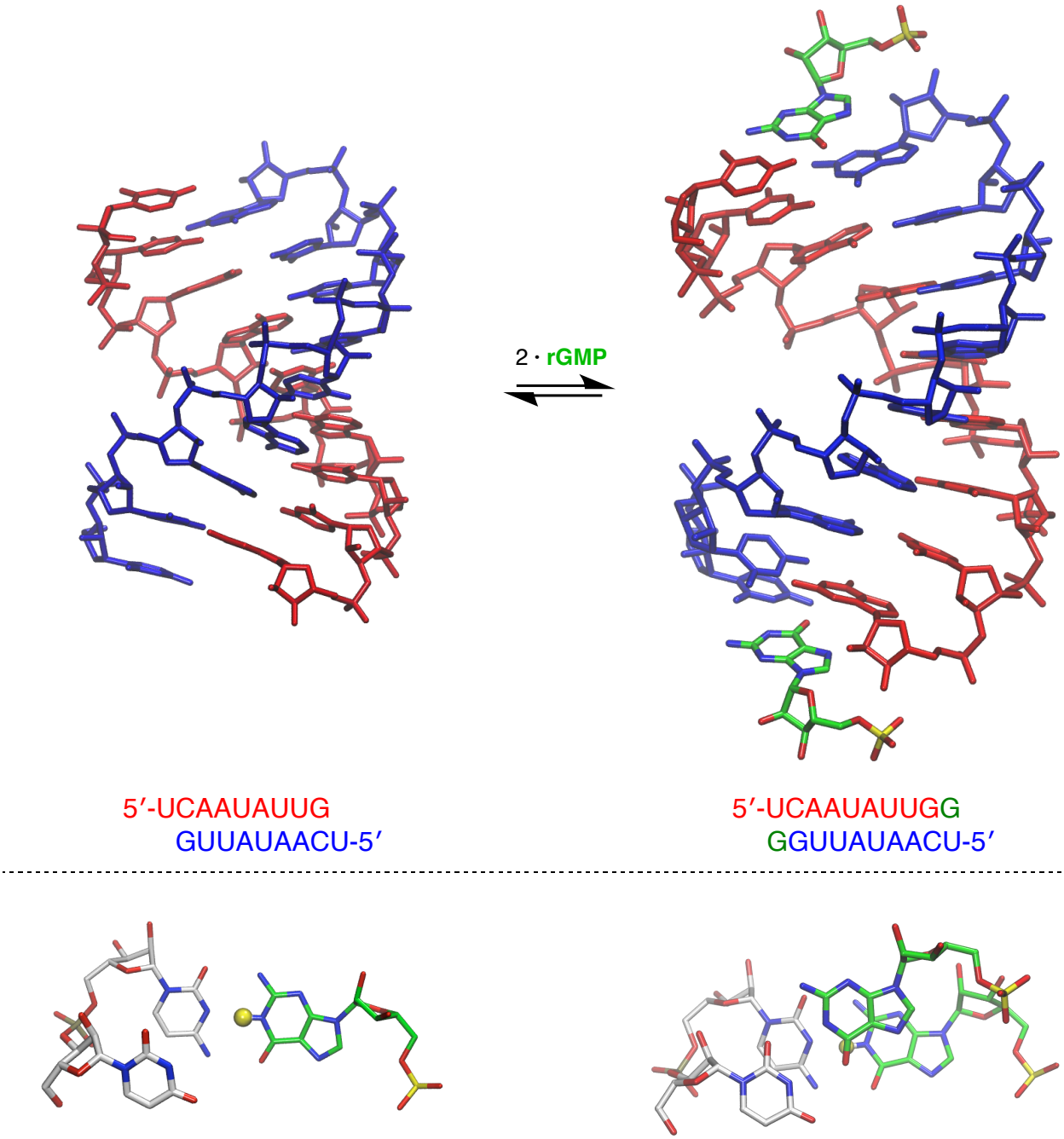


Figure S31. Top: Energy minimized models of A-form 5′-UCAAUAUUG-3′ RNA duplex in both free and rGMP-bound states. **Bottom:** Rotated views of the modeled duplex terminus.

Binding of rGMP to the 5′- d(CCAATATTG)-3′ DNA duplex

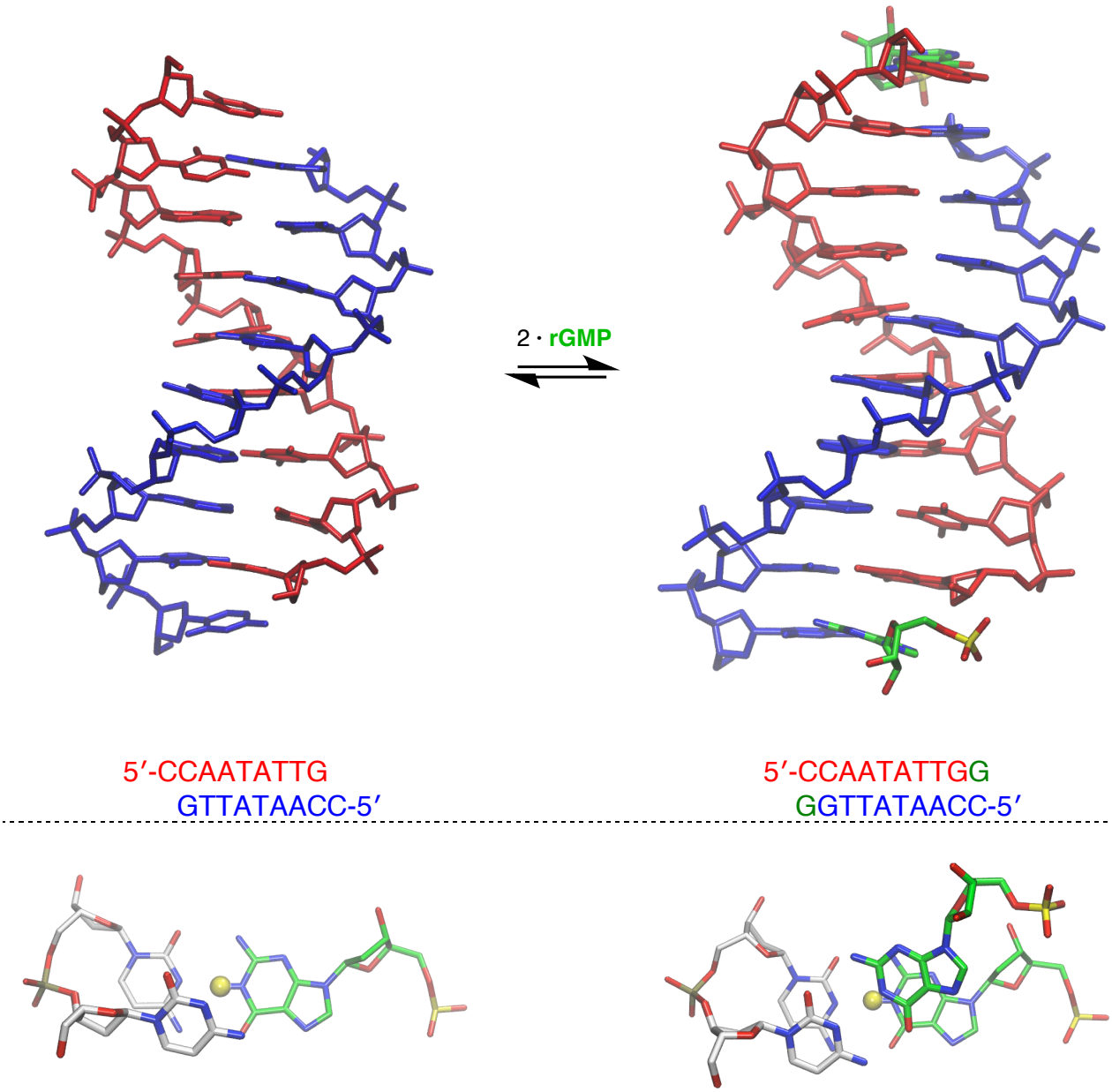


Figure S32. Top: Energy minimized models of B-form 5′-d(CCAATATTG)-3′ DNA duplex in both free and rGMP-bound states. **Bottom:** Rotated views of the modeled duplex terminus.

Binding of rCMP to the 5′- d(GCAATATTG)-3′ DNA duplex

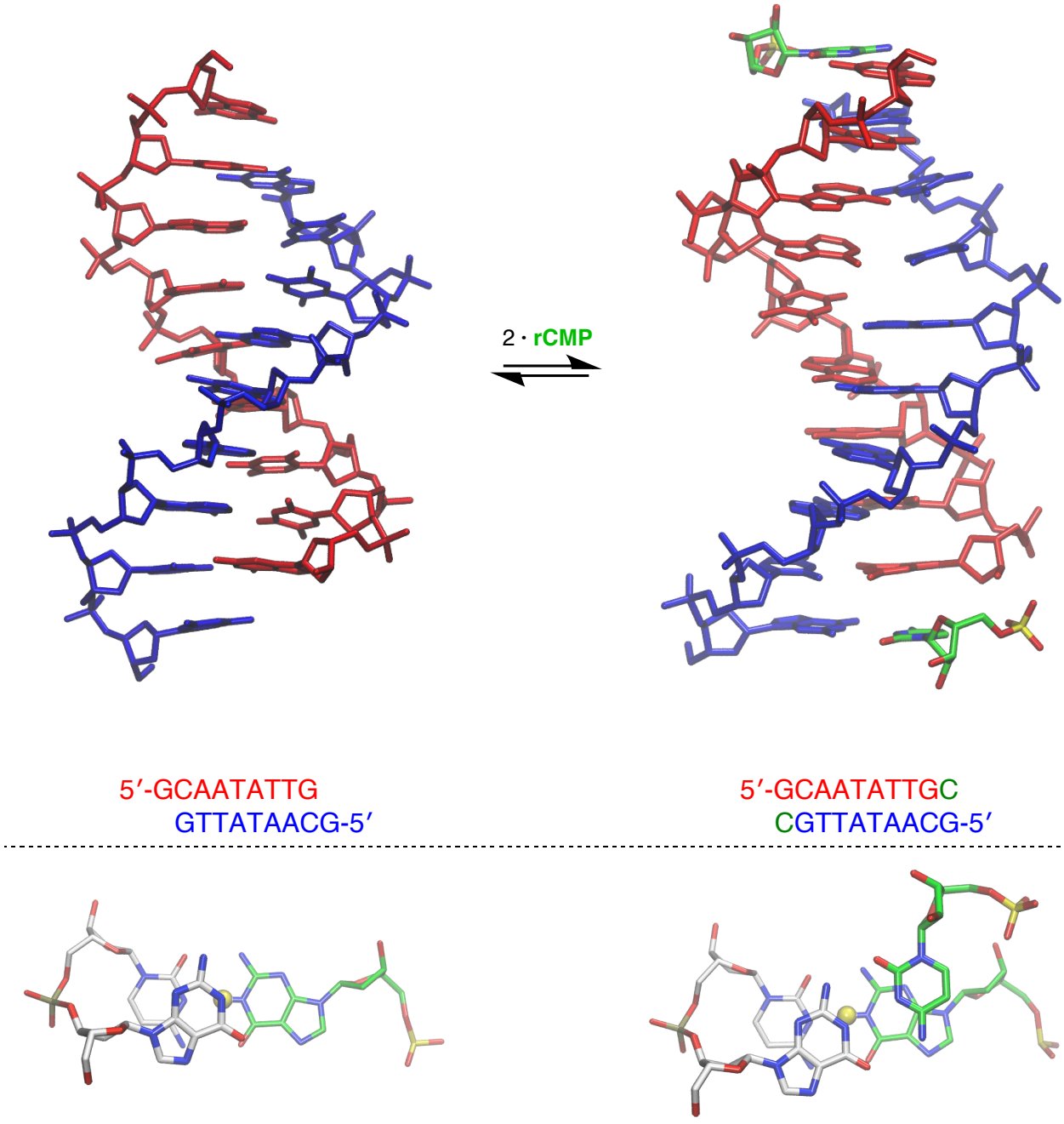


Figure S33. Top: Energy minimized models of B-form 5′-d(GCAATATTG)-3′ DNA duplex in both free and rCMP-bound states. **Bottom:** Rotated views of the modeled duplex terminus.

Binding of rAMP to the 5′-d(TCAATATTG)-3′ DNA duplex

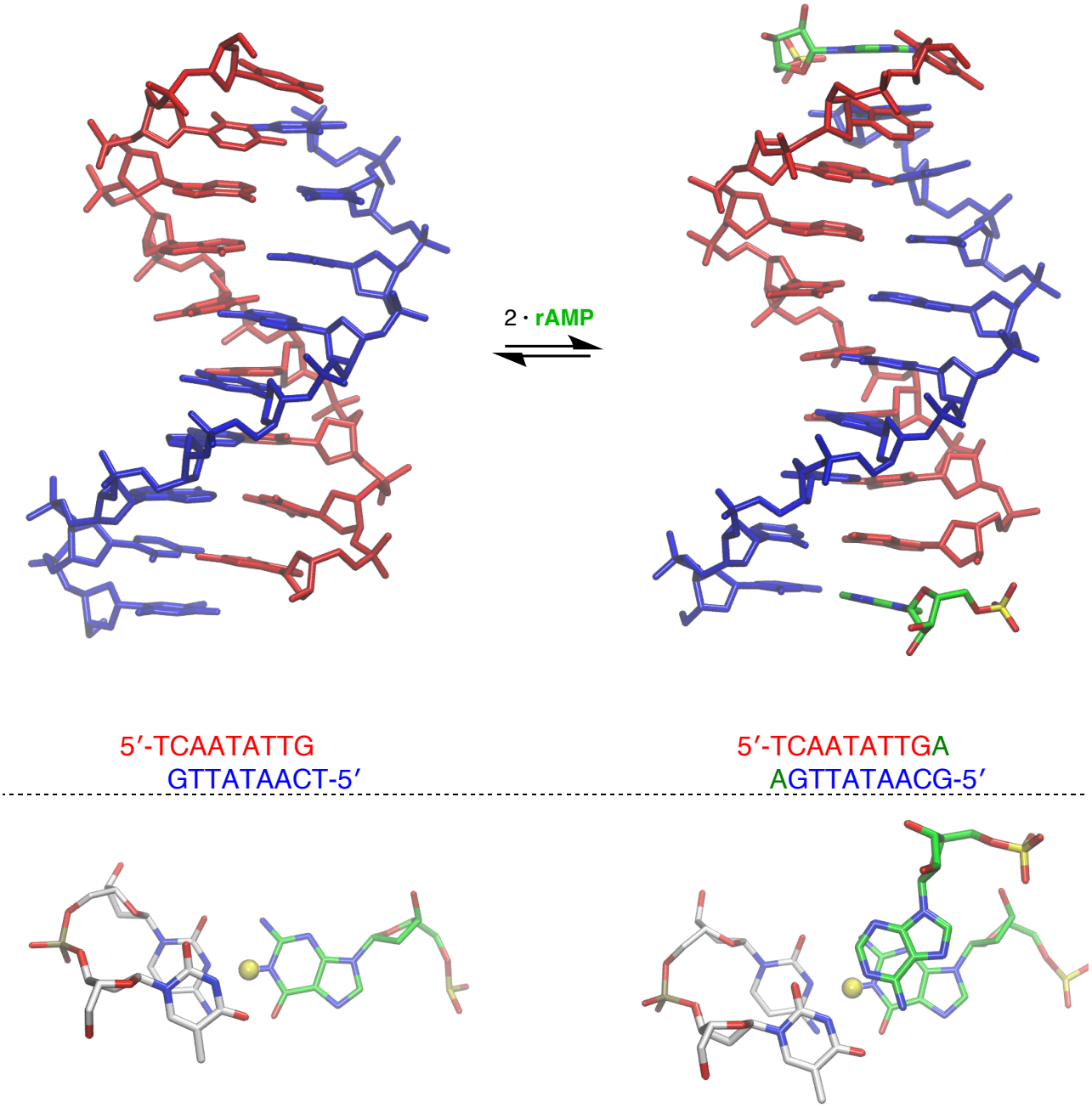


Figure S34. Top: Energy minimized models of B-form 5′-d(TCAATATTG)-3′ DNA duplex in both free and rAMP-bound states. **Bottom:** Rotated views of the modeled duplex terminus.

12. References

(1) Lippens, G.; Dhalluin, C.; Wieruszeski, J. M. *J. Biomol. NMR* **1995**, *5*, 327–331.

(2) Gratzer, W. B.; Richards, E. G. *Biopolymers* **1971**, *10*, 2607–2614.

(3) Ivanov, V. I.; Minchenkova, L. E.; Schyolkina, A. K.; Poletayev, A. I. *Biopolymers* **1973**, *12*, 89–110.

(4) Connors, K. A. Binding Constants: The Measure of Molecular Complex Stability, Wiley-Interscience; New York, **1987**.

(5) Turner, D. H.; Sugimoto, N.; Freier, S. M. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 167–192.