Molecular Recognition in Purine-Rich Internal Loops: Thermodynamic, Structural, and Dynamic Consequences of Purine for Adenine Substitutions in 5′(rGGCAAGCCU)$_2$[†,‡]

Brent M. Znosko,[§] Mark E. Burkard,[§] Thomas R. Krugh,[§] and Douglas H. Turner*[,§,||]

Department of Chemistry, University of Rochester, Rochester, New York 14627-0216, and Department of Pediatrics and Center for Human Genetics and Molecular Pediatric Disease, School of Medicine and Dentistry, University of Rochester, Rochester, New York 14642

Received April 30, 2002; Revised Manuscript Received September 30, 2002

ABSTRACT: The contribution of amino groups to the thermodynamics, structure, and dynamics of tandem A·A mismatches is investigated by substitution of purine (P) for adenine (A) within the RNA duplex, 5′(rGGCAAGCCU)$_2$, to give 5′(rGGCPAGCCU)$_2$, 5′(rGGCAPGCCU)$_2$, and 5′(rGGCPPGCCU)$_2$. The 5′(rGGCAAGCCU)$_2$ duplex has sheared A_{anti}·A_{anti} (A·A trans Hoogsteen/Sugar-edge) pairs in which the A5 amino group is involved in hydrogen bonds but the A4 amino group is not [Znosko, B. M., Burkard, M. E., Schroeder, S. J., Krugh, T. R., and Turner, D. H. (2002) *Biochemistry 41*, 14969−14977]. In comparison to 5′(rGGCAAGCCU)$_2$, replacing the amino group of A4 with a hydrogen stabilizes the duplex by 1.3 kcal/mol, replacement of the A5 amino group destabilizes the duplex by 0.6 kcal/mol, and replacement of both A4 and A5 amino groups destabilizes the duplex by 0.8 kcal/mol. In NMR structures, the P·A noncanonical pairs of the 5′(rGGCPAGCCU)$_2$ duplex have a sheared anti−anti structure (P·A trans Hoogsteen/Sugar-edge) with P4·A5 interstrand hydrogen bonding and A5 bases that interstrand stack, similar to the structure of 5′(rGGCAAGCCU)$_2$. In contrast, the A·P pairs of the 5′(rGGCAPGCCU)$_2$ duplex have a face-to-face conformation (A·P trans Watson−Crick/Watson−Crick) with intrastrand stacking resembling typical A-form geometry. Although the P5 bases in 5′(rGGCPPGCCU)$_2$ are involved in an interstrand stack, the loop region is largely undefined. The results illustrate that both hydrogen-bonded and non-hydrogen-bonded amino groups play important roles in determining the thermodynamic, structural, and dynamic characteristics of purine rich internal loops.

Little is known about the molecular interactions that determine the stabilities, shapes, and dynamics of internal loops in RNA (*1*, *2*). Knowledge of these interactions is important for modeling RNA structure from sequence (*3*, *4*). Here we provide a study of the effects of amino groups on the stability, structure, and dynamics of an internal loop with two sheared A·A pairs, i.e., with A·A trans Hoogsteen/Sugar-edge conformations (*5*, *6*).

Sheared purine·purine pairs are common in biology (*5*, *7*). Sheared A·A pairs have been observed in the group I intron (*8*, *9*), 30S ribosomal subunit (*10*), 50S ribosomal subunit (*11*), yeast tRNA (fmet) (*12*), tRNA aptamer (*13*), and loop B of the hairpin ribozyme (*14*). For example, tandem A·A pairs are a common feature of the J4/5 loop of group I introns (*15*−*17*).

The accompanying paper (*18*) reports the NMR structure of 5′(rGGCAAGCCU)$_2$, a duplex containing a tandem A·A motif that is found in several ribosomal RNAs. The predominant loop structure of this duplex has two interstrand hydrogen bonds involving the amino group of the A5 in each A·A pair: one to the N3 atom of A4 and one to the 2′ oxygen of the A4 sugar. The amino group of A4 is located in the major groove but is not involved in hydrogen bonds. Thus, the amino groups of A4 and A5 play very different roles in the structure. The NMR data suggest that A4 is more dynamic than A5.

Here, the contributions of amino groups to the thermodynamics, structure, and dynamics of the tandem A·A pairs in 5′(rGGCAAGCCU)$_2$ are investigated by replacing adenines with purines, thereby replacing the 6-amino group with a hydrogen. In particular, systematic substitutions in the 5′(rGGCAAGCCU)$_2$ duplex (referred to as the AA[1] duplex) result in 5′(rGGCPAGCCU)$_2$ (the PA duplex), 5′(rGGCAPGCCU)$_2$ (the AP duplex), and 5′(rGGCPPGCCU)$_2$ (the PP duplex) (Figure 1), where P represents a purine base.

[†] This publication was supported by NIH Grants GM 22939 (D.H.T.) and GM 53826 (T.R.K.).

[‡] Protein Data Bank entries 1MV1, 1MV2, and 1MV6.

* To whom correspondence should be addressed. Phone: (585) 275-3207. Fax: (585) 506-0205. E-mail: Turner@chem.rochester.edu.

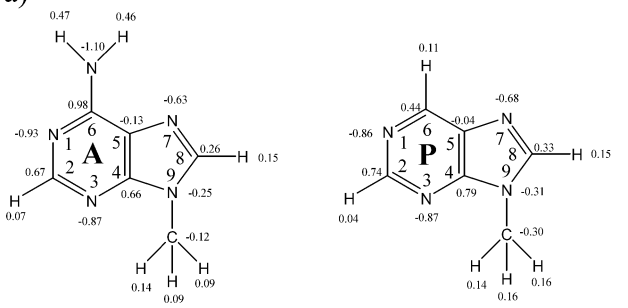
[§] Department of Chemistry.

[||] Department of Pediatrics and Center for Human Genetics and Molecular Pediatric Disease.

[1] Abbreviations: AA, 5′(rGGCAAGCCU)$_2$; AP, 5′(rGGCAPGCCU)$_2$; CPG, controlled pore glass; HETCOR, heteronuclear correlation spectroscopy; HMQC, heteronuclear multiple quantum correlation spectroscopy; P, purine; PA, 5′(rGGCPAGCCU)$_2$; PP, 5′(rGGCPPGCCU)$_2$; rMD, restrained molecular dynamics; RMSD, root-mean-squared deviation; T_M, melting temperature; TOCSY, total correlation spectroscopy.

10.1021/bi0203278 CCC: $22.00
Published on Web 11/21/2002

a)



b)

```
          123456789
PA        GGCPAGCCU3'
      3'UCCGAPCGG
        987654321

          123456789
AP        GGCAPGCCU3'
      3'UCCGPACGG
        987654321

          123456789
PP        GGCPPGCCU3'
      3'UCCGPPCGG
        987654321
```

FIGURE 1: (a) Structure, numbering scheme, and electrostatic potentials of 9-methyladenine and 9-methylpurine. The pK_a's for adenosine and purine riboside are 3.5 and 2.1, respectively. (b) Schematic representation of the duplexes studied by NMR.

These substitutions reveal the importance of the exocyclic amino groups in determining the properties of the sheared A_{anti}•A_{anti} noncanonical pairs.

MATERIALS AND METHODS

RNA Synthesis, Purification, and Optical Melting Experiments. The purine riboside 2′-*tert*-butyl dimethyl silyl phosphoramidites were acquired from ChemGenes (Ashland, MD). All other aspects of RNA synthesis, purification, and optical melting are described elsewhere (*18*).

Sample Preparation for NMR. The samples were lyophilized and redissolved in 250−350 μL of 80 mM NaCl, 3 mM KH_2PO_4, 7 mM K_2HPO_4, 0.5 mM Na_2EDTA, pH 6.5, and lyophilized. For the nonexchangeable proton samples, D_2O exchange was performed with lyophilization from 99.96% D_2O. The samples were redissolved in 250−350 μL of 99.996% D_2O (Cambridge Isotope Laboratories). The strand concentrations of the oligonucleotides for these samples were ∼2 mM. NMR spectra of the nonexchangeable protons were recorded at 30 °C, pH ∼7.7. NMR spectra of the exchangeable protons of ∼1 mM oligonucleotide were recorded at 0−45 °C in 80 mM NaCl, 3 mM KH_2PO_4, 7 mM K_2HPO_4, 0.5 mM Na_2EDTA, pH ∼6.5 in 90:10 H_2O/D_2O.

NMR Spectroscopy. All spectra were acquired on a Varian Inova 500 MHz spectrometer. The parameters and processing for one-dimensional imino proton spectra, one-dimensional NOE spectra, one-dimensional nonexchangeable proton spectra, one-dimensional phosphorus spectra, NOESY spectra, NOESY WATERGATE spectra, DQF−COSY spectra, ^{1}H-^{31}P HETCOR, TOCSY spectra, and T_1 relaxation experiments are described elsewhere (*18*).

Restraint Generation. The procedure for the generation of distance restraints is described elsewhere (*18*). For the PA duplex, a total of 86 NMR-derived interproton distance restraints per strand (50 intranucleotide, 33 internucleotide, and 3 interstrand) and 46 dihedral angle restraints per strand were included. For the AP duplex, a total of 71 NMR-derived interproton distance restraints per strand (44 intranucleotide, 25 internucleotide, and 2 interstrand) and 59 dihedral angle restraints per strand were included. For the PP duplex, a total of 63 NMR-derived interproton distance restraints per strand (45 intranucleotide, 13 internucleotide, and 5 interstrand) and 50 dihedral angle restraints per strand were included. For all duplexes, an additional 18 hydrogen bond restraints for six Watson−Crick G−C pairs (included as distance restraints of 1.8 to 2.5 Å) were used; hydrogen-bonding restraints were not used between A/P•A/P pairs. Because coupling in the DQF−COSY and HETCOR experiments suggested typical A-form geometry, parameters for the dihedral angles were loosely restrained to those observed in canonical A-form RNA duplexes for all residues except some angles of the A/P4 and A/P5 residues to allow for conformational freedom in the non-Watson−Crick region. Because the oligonucleotides form a self-complementary duplex, the same restraints were applied to both strands.

Simulated Annealing. Models of the duplexes consistent with NMR data were derived from restrained energy minimization and simulated annealing with the Discover 95 package on a Silicon Graphics computer. Five starting structures for each duplex were generated with Biosym Insight II software: standard A-form RNA, B-form RNA, and 3 A-form structures with varying axial rise and twist angle parameters. The molecular dynamics and minimization protocol is identical to that described elsewhere (*18*, *19*). In all cases, at least three of the final structures were derived from each of the helical starting structures. Molecular dynamic simulations with random structures as the starting structures also contain the same loop conformations with no violations in the loop regions. Due to greater convergence with A-form and B-form type starting structures, discussion will focus on the derived structures from these helical starting structures.

For each duplex, a theoretical NOESY spectrum using a spin matrix approach was calculated as a test of the final structure. In all cases, the theoretical and experimental spectra were almost identical, including the cross-peaks for the loop and stem-loop junction.

Electrostatic Potential Calculations. Electrostatic potentials were determined for 9-methyladenine and 9-methylpurine by ab initio calculations with the Gaussian 98 software package (*20*). The PRCG algorithm with the MM3 force field (*21*) was used to generate energy-minimized structures. PM3 (*22*, *23*) was used to optimize the geometry of the resulting conformations. Electrostatic calculations were performed using the UMP2 method (*24*) with a 6-31G basis set (*25*). Calculations are done in vacuo.

RESULTS

Electrostatic Calculations. Figure 1 shows electrostatic potentials calculated for 9-methyladenine and 9-methylpurine. As expected, the largest change occurs at C6.

Table 1: Thermodynamic Parameters for Duplex Formation[a]

	T_M^{-1} vs log C_T plots				average of curve fits			
	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	$-\Delta G°_{37}$ (kcal/mol)	T_M[b] (°C)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	$-\Delta G°_{37}$ (kcal/mol)	T_M[b] (°C)
(rGGCAAGCCU)$_2$[c]	77.2 ± 2.0	216.5 ± 6.0	10.04 ± 0.11	55.6	75.3 ± 6.2	210.5 ± 18.8	9.98 ± 0.40	55.8
(rGGCPAGCCU)$_2$	91.5 ± 8.2	258.5 ± 24.8	11.30 ± 0.50	57.3	87.1 ± 6.4	245.2 ± 19.4	11.04 ± 0.45	57.3
(rGGCAPGCCU)$_2$	80.9 ± 3.6	230.4 ± 11.0	9.44 ± 0.16	52.1	78.7 ± 9.9	223.5 ± 30.5	9.37 ± 0.45	52.3
(rGGCPPGCCU)$_2$	85.2 ± 2.4	244.8 ± 7.7	9.22 ± 0.07	50.4	73.6 ± 3.9	208.4 ± 12.2	8.91 ± 0.16	51.3

[a] Solutions are 1 M NaCl, 20 mM sodium cacodylate, 0.5 mM Na_2EDTA, pH 7. [b] Calculated at 10^{-4} M oligonucleotide concentration. [c] Ref *18*.

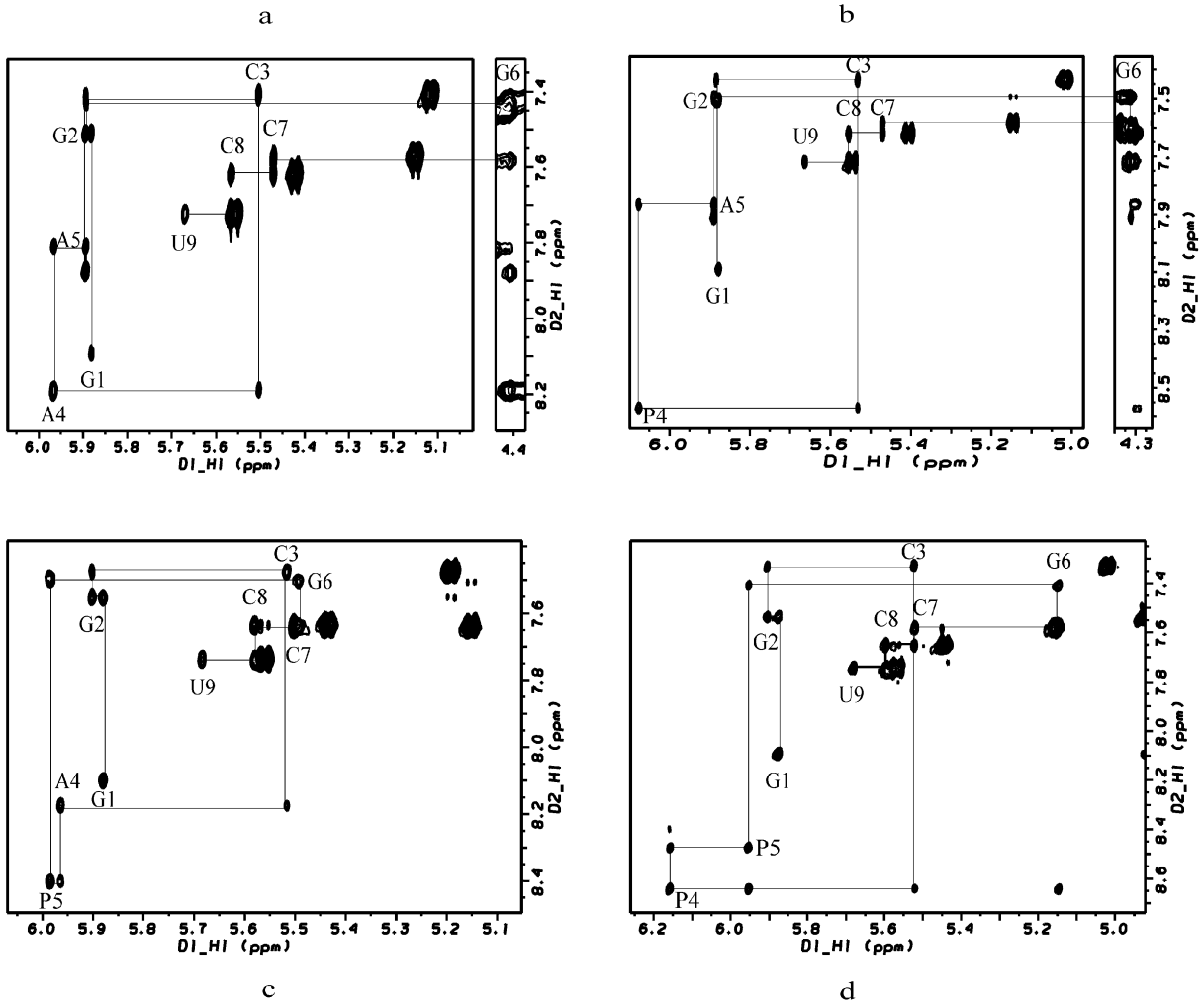


FIGURE 2: (H8/H6/H2)−(H1′/H5) region of the 400 ms NOESY spectra of (a) 5′(rGGCAAGCCU)$_2$ (*18*), (b) 5′(rGGCPAGCCU)$_2$, (c) 5′(rGGCAPGCCU)$_2$, and (d) 5′(rGGCPPGCCU)$_2$. Sequential assignments of base to H1′ protons are connected by lines. Intranucleotide H1′−H6/H8 cross-peaks are labeled.

Thermodynamic Parameters. Thermodynamic parameters for the duplexes are listed in Table 1. The parameters from the fits of melting curves and from T_M^{-1} versus log C_T plots agree within 15%, suggesting that the two-state model is a reasonable approximation for these transitions.

Chemical Shift Assignments. The NMR resonances of the duplexes were assigned essentially as described by Varani et al. (*26*−*28*) and as described for 5′(rGGCAAGCCU)$_2$ (*18*). Table S1 in Supporting Information summarizes chemical shift assignments at 30 °C. The chemical shifts of the PH6 protons are between 8.65 and 9.03 ppm. In all three duplexes, all eight ^{31}P resonances were observed, and all are in a 1 ppm range, suggesting a typical A-form geometry (*29*).

Exchangeable Proton Resonances. Typical imino−imino and imino−amino cross-peaks were observed and assigned for each of the three duplexes on the basis of 1D NOE difference and 2D NOESY WATERGATE spectra (Supporting Information). It is evident from the imino chemical shifts that three distinct G−C pairs are formed, as expected for each of the duplexes (Supporting Information). An imino resonance was not observed for the dangling U residue in any of the duplexes.

Nonexchangeable Proton Resonances. For each duplex, the (H8/H6/H2)−(H1′/H5) region of the 400 ms NOESY spectrum at 30 °C is shown in Figure 2. The assignments follow standard connectivity pathways from residues G1 to U9. One of the main differences in the NOESY walk region of the three duplexes is the chemical shift of the G6H1′ resonance. In the AA and PA duplexes, the G6H1′ cross-peaks are shifted upfield to ∼4.5 and ∼4.3 ppm, respectively.

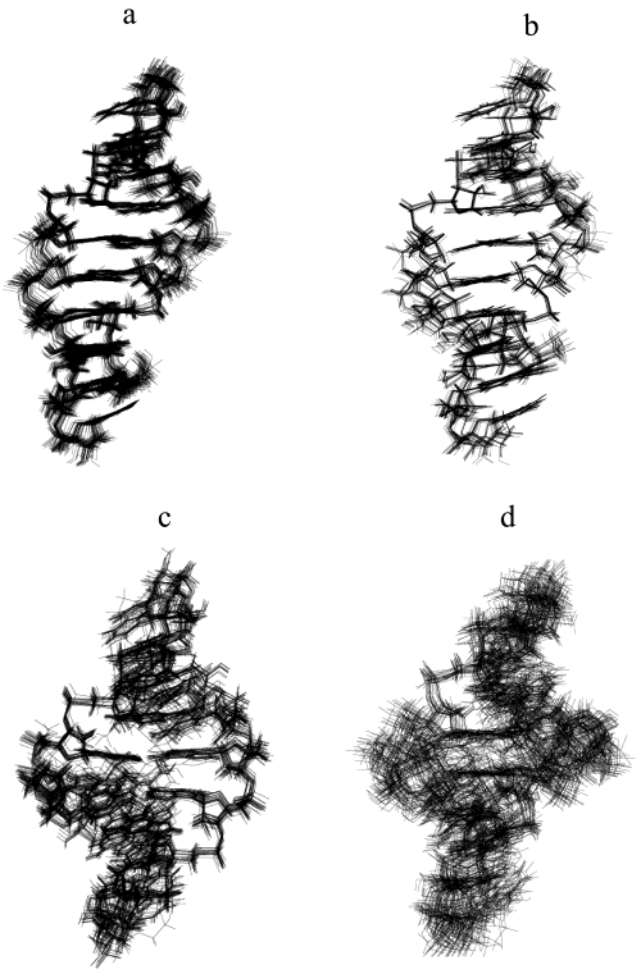


FIGURE 3: Superposition of (a) 32 low-energy structures derived from rMD of 5′(rGGCAAGCCU)$_2$ with RMSD = 0.42 Å (*18*), (b) 21 low-energy structures derived from rMD of 5′(rGGCPAGCCU)$_2$ with RMSD = 0.40 Å, (c) 20 low-energy structures derived from rMD of 5′(rGGCAPGCCU)$_2$ with RMSD = 0.98 Å, and (d) 33 low-energy structures derived from rMD of 5′(rGGCPPGCCU)$_2$ with RMSD = 0.99 Å.

In the AP duplex, the G6H1′ proton resonates near the other H1′ protons, at ~5.5 ppm. In the PP duplex, the G6H1′ proton resonates at ~5.2 ppm. An upfield shift of H1′ resonances, due to location above or below bases with strong ring currents, has been reported previously (*30*−*33*). Apparently, the stacked arrangement of A5 on G6H1′ is disrupted in the AP and PP duplexes.

Structure Determination. Distance and dihedral angle restraints for structure determination are listed in the Supporting Information. Structures derived in the absence of solvent and structures derived with water as the solvent both displayed similar structural features and overall geometry. Since there were no apparent differences, further discussion is based on the 50 structures derived in the absence of solvent.

Structures of 5′(rGGCPAGCCU)$_2$. Of the 50 structures generated by the rMD protocol described in Materials and Methods, the 21 with the lowest energies have an average simulated total energy of −399 ± 3 kcal/mol. Superposition of these 21 structures (Figure 3b) reveals that the overall structure and local features are well determined. The average RMSD for the all-atom pairwise superposition of these 21 structures is 0.40 Å. These 21 structures converged to satisfy all distance and dihedral angle restraints within 0.08 Å and 2.0°, respectively. (The minimized average structure has only four distance violations, with all less than 0.1 Å and having an average violation of 0.02 Å.) These structures contain sheared P_{anti}·A_{anti} hydrogen-bonded pairs (Figure 4b) similar to the A·A pairs in 5′(rGGCAAGCCU)$_2$ (*18*) (Figure 4a). A minor groove view of the P·A pairs and closing C−G pairs is shown in Figure 5b.

T_1 relaxation times are given in Supporting Information. The long relaxation times of the P4H2 (8.3 s), A5H2 (6.0 s), and P4H6 (9.3 s) are consistent with the observation that other protons are not close enough to generate cross-peaks in NOESY spectra.

Structures of 5′(rGGCAPGCCU)$_2$. Of the 50 structures generated by the rMD protocol described in Materials and Methods, the 20 with the lowest energies have an average simulated total energy of −382 ± 7 kcal/mol. The superposition reveals that the overall structure and local features (Figure 3c) are moderately well determined. The average RMSD for the all-atom pairwise superposition of these 20 structures is 0.98 Å. These 20 structures converged to satisfy all distance and dihedral angle restraints within 0.07 Å and 2.0 °, respectively. (The minimized average structure has only one distance violation, which is 0.03 Å.) Unlike the sheared A_{anti}·A_{anti} hydrogen bonded pairs of the AA duplex and the PA duplex (Figure 4a, b), the A·P pairs are in a face-to-face geometry (Figure 4c). A minor groove view of the A·P pairs and closing C−G pairs is shown in Figure 5c.

The T_1 relaxation times are consistent with the final structure (Supporting Information). The slow relaxation time (6.3 s) of P5H2 is typical. The P5H6 also relaxed slowly (7.7 s), as only one other proton is close enough to generate a cross-peak in any of the NOESY spectra.

Structures of 5′(rGGCPPGCCU)$_2$. Of the 50 structures generated by the rMD protocol described in Materials and Methods, 33 have an average simulated total energy of −365 ± 5 kcal/mol. The superposition reveals that the overall structure and local features are dynamic (Figure 3d). The average RMSD for the all-atom pairwise superposition of these 33 structures is 0.99 Å. These 33 structures converged to satisfy all distance and dihedral angle restraints within 0.07 Å and 3.5 °, respectively. (The minimized average structure has only five distance violations, with all less than 0.1 Å and having an average violation of 0.03 Å.) Similar to the sheared A_{anti}·A_{anti} hydrogen bonded pairs of the AA duplex and the PA duplex (Figure 4a, b), the P·P pairs are in a sheared conformation (Figure 4d). Many different conformations of the loop region satisfy the distance restraints, however, which results in the large RMSD in comparison to the AA and PA duplexes. A minor groove view of the P·P pairs and closing C−G pairs is shown in Figure 5d.

The T_1 relaxation times were measured. The uncertainty in the values is too large, however, to determine if any value is significantly different from any other (Supporting Information).

DISCUSSION

Thermodynamics. The melting temperatures of all duplexes are dependent on oligonucleotide concentration, showing that duplexes are formed rather than hairpins. Thermodynamic

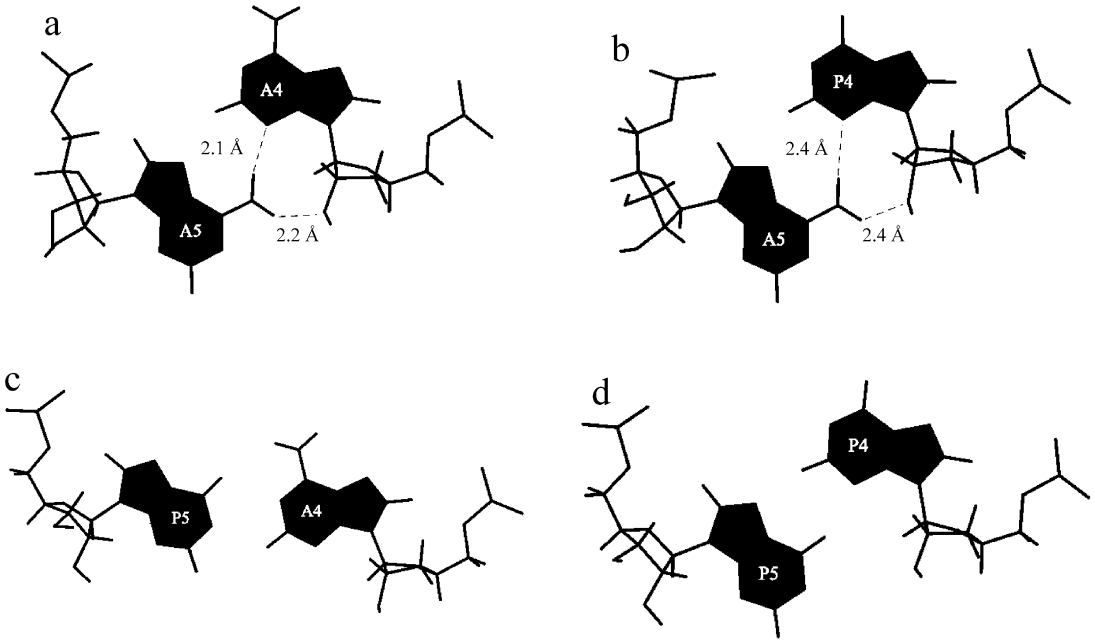


FIGURE 4: A/P·A/P pairs from the lowest energy structures of (a) 5′(rGGCAAGCCU)$_2$ (*18*), (b) 5′(rGGCPAGCCU)$_2$, (c) 5′(rGGCAPGCCU)$_2$, and (d) 5′(rGGCPPGCCU)$_2$. Dashed lines indicate functional groups that are close enough to form hydrogen bonds. The numbers indicate the average distance in angstroms between the hydrogen and the hydrogen-bond acceptor. A C−H hydrogen bond may be possible between the following: A4H2 and A5N7 of AA, with an average distance of 2.6 Å, P4H2 and A5N7 of PA, with an average distance of 2.5 Å, A4N1 and P5H6 of AP, with an average distance of 2.7 Å, and A4H2 and P5N1 of AP, with an average distance of 2.8 Å. Other low-energy structures for sequences a−c are similar to these lowest energy structures, but low-energy structures for sequence d are more diverse (see Figure 3). The figures were created with 3DNA (*45*).

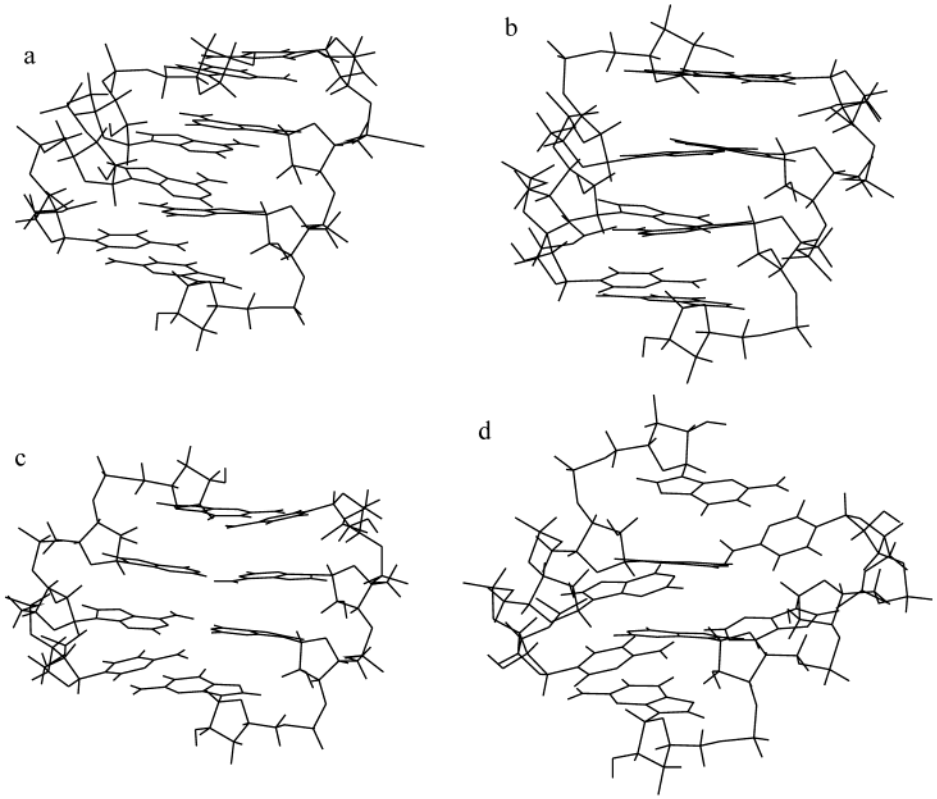


FIGURE 5: View of the minor groove of the lowest-energy structures of (a) 5′(rGGCAAGCCU)$_2$ (*18*), (b) 5′(rGGCPAGCCU)$_2$, (c) 5′-(rGGCAPGCCU)$_2$, and (d) 5′(rGGCPPGCCU)$_2$. Shown are the A/P·A/P pairs and the closing base pairs.

increments for the internal loops (Table 2) were calculated as previously described (*18*, *34*) by using thermodynamic parameters in Table 1 and previously reported parameters for motifs other than the internal loops (*35*−*38*). The 0.82 kcal/mol difference in ΔG°_{37} between the AA and PP systems is similar to the 1.07 kcal/mol difference observed earlier between 5′(rGCAAGCG)$_2$ and 5′(rGCPPGCG)$_2$ (*39*).

Comparing the thermodynamic parameters of the loops provides insight into the effects of adenosine amino groups on the stabilities of internal loops. Most striking is that the

Table 2: Thermodynamic Parameters of Loop Formation[a]

	$\Delta G°_{37}$ (kcal/mol)	$\Delta H°$ (kcal/mol)	$\Delta S°$ (eu)
5′(rGGCAAGCCU)$_2$[b]	1.13	−5.0	−20.4
5′(rGGCPAGCCU)$_2$	−0.13	−19.3	−62.4
5′(rGGCAPGCCU)$_2$	1.73	−8.7	−34.3
5′(rGGCPPGCCU)$_2$	1.95	−13.0	−48.7
5′(rCGCAGGCG)$_2$[c]	−1.00	−16.5	−49.9
5′(rCGCPGGCG)$_2$[c]	1.68	−12.8	−46.7
5′(rGGCGAGCC)$_2$[c]	−0.72	−8.9	−26.5
5′(rGGCGPGCC)$_2$[c]	2.25	−0.4	−6.0
5′(rGCPPGCG)$_2$[c]	2.16	−12.2	−46.4

[a] Calculated from T_M^{-1} vs log C_T parameters using equations similar to (*34*): $\Delta G°_{37,loop} = \Delta G°_{37}$(rGGCXYGCCU)$_2$ − $\Delta G°_{37}$(rGGCGCC)$_2$ + $\Delta G°_{37}$(CG)$_2$ − 2$\Delta G°_{37}$(CU), where $\Delta G°_{37}$(rGGCXYGCCU)$_2$ is the measured value of the duplex containing the internal loop, $\Delta G°_{37}$(rGGCGCC)$_2$ is the measured value of the duplex without the loop and without the dangling end (*35*), $\Delta G°_{37}$(CG)$_2$ is the free energy increment for the nearest-neighbor base pair interaction interrupted by the internal loop (*36*), and $\Delta G°_{37}$(CU) is the free energy increment of −1.1 kcal/mol measured for a dangling U on a C−G base pair (*37*, *38*). [b] Ref *18*. [c] Ref *39*.

PA duplex is ∼1.3 kcal/mol more stable than the AA duplex, even though the two structures are very similar. The only major structural difference is that in the AA duplex, the amino group of A4 protrudes into the major groove, whereas this group is replaced by a hydrogen in the PA duplex (Figure 4). Evidently, the A4 amino groups are destabilizing if they are not involved in a structural hydrogen bond. Perhaps there are unfavorable stacking interactions between the overlapped A4 amino groups and between the C3 and A4 amino groups (Figure 6a). The distances between the nitrogens involved in these two interactions are 3.24 and 3.45 Å, respectively. According to electrostatic potential calculations, the amino group of adenine withdraws electron density from the adenine ring system. Thus, the purine ring system is more electron-rich than the adenine ring system (Figure 1). This is supported by the observation that in the PA duplex, the P4H2 and H8 resonances are 0.78 and 0.38 ppm further downfield, respectively, when compared to the A4H2 and H8 resonances in the AA duplex (Supporting Information). As a result, the overlap of electron-rich P4 bases may be more favorable than A4 bases. In addition, the P4−P4 overlap in the PA duplex (Figure 6b) is larger than that in the AA duplex (Figure 6a). It does not appear that N3 of purine is a better hydrogen bond acceptor than N3 of adenine, since both nitrogens have essentially the same electrostatic potential (Figure 1). Another potential reason for enhanced stability upon removal of the A4 amino group is that solvation of the amino group may be more favorable in the single strand than in the duplex. A similar rationale could explain the enhanced stabilities of 1.1 and 3.0 kcal/mol for internal 2-thio U·A and 4-thio U·G pairs over U·A and U·G pairs, respectively (*40*). In those cases, a carbonyl group not involved in a structural hydrogen bond is replaced by a less electronegative sulfur atom. Alternative rationales, however, have been suggested for this thio effect (*40*).

In contrast to the 1.3 kcal/mol stabilization observed for an AA to PA substitution in 5′(rGGCAAGCCU)$_2$, an AG to PG substitution in 5′(rCGCAGGCG)$_2$ destabilizes the duplex by 2.7 kcal/mol (*39*) (Table 2). Presumably, this is because neither the AG nor PG motifs can form sheared pairs (*41*), and the PG substitution eliminates a hydrogen bond found in a GA imino pair.

Comparisons between the AA and PA duplexes, which have a total of four interstrand hydrogen bonds in noncanonical pairs, and the AP and PP duplexes, which have no interstrand hydrogen bonds in noncanonical pairs, provide insight into the energetics of hydrogen bonds. The easiest comparison is between the PA and PP duplexes since they have similar average structures and only differ in the replacement of two amino groups by two hydrogens. The difference in $\Delta G°_{37}$ for duplex formation is 2.1 kcal/mol or ∼0.5 kcal/mol of hydrogen bond. Although the PA and PP duplexes have similar average structures, the PP duplex appears to be less well defined (see Figure 3b,d). Thus, the total contribution of a hydrogen bond may be more than 0.5 kcal/mol since some of the free energy is expended to make the structure of PA more rigid than PP.

Comparisons with the AP duplex are not straightforward because the conformation differs from that of the AA and PA duplexes. For example, the face-to-face geometry (Figure 4c) induces less backbone distortion and results in different stacking patterns compared to the sheared geometry. Nevertheless, the 1.9 kcal/mol difference in $\Delta G°_{37}$ of duplex formation between the AP and PA duplexes indicates that each hydrogen bond in the noncanonical pairs of PA contributes at least 0.5 kcal/mol to duplex stability.

Structural Analysis. Exchangeable ^{1}H NMR spectra show that Watson−Crick base pairs are formed between the G1−C8, G2−C7, and C3−G6 bases for all duplexes (Supporting Information). The base to H1′ region of the NOESY spectra indicates that all bases are in the anti conformation. For all nonterminal Watson−Crick residues, weak or nonexistent cross-peaks in the DQF−COSY spectra indicate H1′−H2′ couplings of <2 Hz, which suggests predominantly C3′-endo sugar puckers. The C3′-endo sugar puckers are confirmed by the large H3′−H4′ couplings. Strong cross-peaks are observed in the 100 ms mixing time NOESY spectra from base H8/H6 protons to 5′ neighboring H2′ protons for G2, C3, C7, and C8 residues, consistent with typical A-form structure. The base protons of the A/P4, A/P5, and G6 residues show medium-strength NOEs to 5′ neighboring H2′ protons, suggesting slight deviations from A-form structure near the A/P·A/P pairs. All observed resonances in the 1D ^{31}P spectrum are within a 1 ppm range, suggesting that the A/P·A/P pairs do not significantly distort the backbone from typical A-form geometry. For all the duplexes, C8−U9 intrastrand restraints position the dangling U9 to stack on C8.

Interstrand NOEs. Interstrand NOEs are important in defining the structures of the internal loops, as was the case for the tandem A·A internal loop (*18*). The process for distinguishing between interstrand and intrastrand cross-peaks in the loop region was discussed previously (*18*). A comparison of measured distances with those expected for sheared and face-to-face conformations is presented in Table 3.

Interstrand NOEs in the PA Duplex. Three interstrand NOEs, A5H1′−A5H2, P4H2−A5H8, and P4H2−P4H2′, were observed. The A5H1′−A5H2 cross-peak resulted in an NMR-derived distance of 3.2 Å. Since the AH2 proton cannot be closer than 4.5 Å to its own H1′ proton (*42*), this is an interstrand cross-peak. In an A-form helix, this

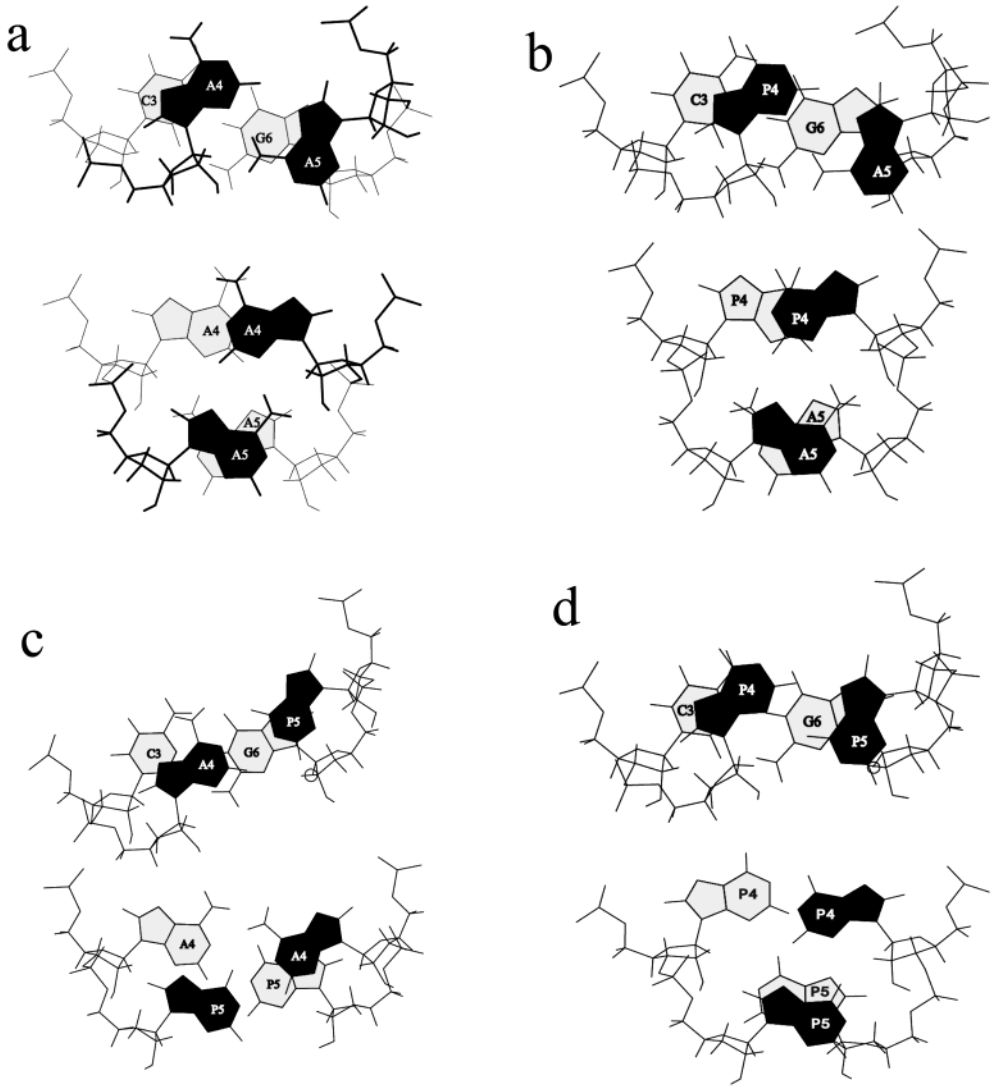


FIGURE 6: Base stacking between the A/P·A/P pair and adjacent C−G pair (top figure in each set) and adjacent A/P·A/P pairs (bottom figure in each set) for (a) 5′(rGGCAAGCCU)$_2$ (*18*), (b) 5′(rGGCPAGCCU)$_2$, (c) 5′(rGGCAPGCCU)$_2$, and (d) 5′(rGGCPPGCCU)$_2$. The darker residues are closer to the viewer. In the (CAAG)$_2$ loop, the amino nitrogens of the stacked A4 bases are separated by 3.2 Å and separated by 3.5 Å from the amino nitrogen of C3. Also, notice that in the AA and PA duplexes, G6H1′ is situated directly below the A5 ring. In the PP duplex, this proton is situated very close to the P5 ring, and in the AP duplex, this proton is rather far from the P5 ring, consistent with the chemical shift variations of this proton in different duplexes. Figures were created with 3DNA (*45*).

interstrand distance for equivalent protons is 4.7 Å, and thus the experimental distance of 3.2 Å suggests that the minor groove is narrower than found in an A-form duplex. A narrowing of the minor groove is consistent with formation of sheared P·A pairs. In addition, the A5H1′−A5H2 distance of 3.2 Å is consistent with an interstrand stack of A5 bases, as shown in Figure 6b. The P4H2−A5H8 cross-peak led to an NMR distance of 4.4 Å. The intrastrand P4H2−A5H8 distance in typical A-form geometry is >5.0 Å, and thus the P4H2−A5H8 distance of 4.4 Å must also be an interstrand distance, which is also consistent with the P4-A5 bases adopting a sheared conformation. A face-to-face geometry would result in an interstrand P4H2−A5H8 distance of ∼7 Å. The P4H2′−P4H2 cross-peak resulted in an NMR-derived distance of 4.1 Å. The typical A-form intraresidue distance for P4H2−P4H2′ is >5.0 Å. Therefore, this peak represents another interstrand cross-peak. These three interstrand distances are consistent with the sheared base conformation of the P4 and A5 bases in the PA duplex.

Interstrand NOEs in the AP Duplex. The A4H2−P5H2 and A4H2−P5H6 cross-peaks gave NMR-derived distances of 4.6 and 3.8 Å, respectively, which are consistent with the face-to-face base orientation (Figure 4c). Intrastrand distances between these protons are >5.0 Å in typical A-form geometry and thus would not be observed in short mixing time NOESY spectra. The interstrand cross-peaks unique to a sheared geometry are not observed even in the 400 ms mixing time NOESY spectrum of the AP duplex.

Interstrand NOEs in the PP Duplex. Five interstrand NOEs, P5H1′−P5H2, P4H1′−P5H6, P4H2′−P5H6, P5H6−P5H8, and P4H2−P5H6, are observed for the PP duplex. These are interstrand peaks since these proton pairs have intrastrand distances typically >5.0 Å. All of these are consistent with an interstrand stack. The interstrand P4H2−P5H6 cross-peak does not distinguish between a sheared or face-to-face geometry since this cross-peak would be expected in both conformations. Therefore, the NMR data are consistent with the PP duplex adopting an interstrand stack

Table 3: A Comparison of NMR Distances Derived from Interstrand NOE Cross-Peaks to Proton−Proton Distances in Sheared and Face-to-Face Conformations[a]

proton pair[b]	NMR distance	sheared intrastrand[c]	sheared interstrand[c]	face-to-face intrastrand[d]	face-to-face interstrand[d]	conformation[e]
4H1′−5H6	AA: n/a[f] PA: n/a AP: −[g] PP: 4.0	7.5	**2.7**[h]	7.3	7.6	sheared
4H2′−4H2	AA: − PA: 4.1 AP: − PP: −	5.1	**3.6**	5.4	8.9	sheared
4H2′−5H6	AA: n/a PA: n/a AP: − PP: 3.8	6.5	**3.6**	5.4	8.1	sheared
4H2−5H2	AA: − PA: − AP: 4.6 PP: −	7.8	7.3	4.3	**3.7**	face-to-face
4H2−5H6	AA: n/a PA: n/a AP: 3.8 PP: 3.1	4.7	3.8	6.2	**3.4**	sheared or face-to-face
4H2−5H8	AA: 4.4 PA: 4.4 AP: − PP: −	5.7	**3.5**	6.8	7.7	sheared
5H1′−5H2	AA: 3.5 PA: 3.2 AP: − PP: 3.7	4.7	**2.9**	4.5	4.3	sheared or face-to-face
5H6−5H8	AA: n/a PA: n/a AP: − PP: 4.5	5.2	**3.6**	5.1	10.7	sheared or face-to-face

[a] All distances are in Å. [b] Protons are represented by the residue number followed by the proton name. [c] Proton−proton distances in the sheared conformation were measured from the miminimized average structure of 5′(rGGCAAGCCU)$_2$. The sheared conformation is also known as A·A trans Hoogsteen/Sugar-edge (*6*). [d] Proton−proton distances in the face-to-face conformation were measured from a minimized standard A-form helix of 5′(rGGCAAGCCU)$_2$. The face-to-face conformation is also known as A·A cis Watson−Crick/Watson−Crick (*6*). [e] The most likely conformation if a cross-peak is observed for the proton−proton pair. [f] Not applicable due to unmodified residue lacking specified proton. [g] Dashes indicate the absence of an observable cross-peak in the 400 ms mixing time NOESY spectrum. [h] The shortest distance for each proton pair are shown in bold.

but no interstrand NOEs between the P4 and P5 bases provide direct information concerning the conformation of the pairs.

Hydrogen Bonding in the Loop Region of the PA Duplex. The NMR spectra do not provide any direct evidence for or against hydrogen bond formation of the loop bases, and thus no hydrogen bonding restraints were used for the loop bases. The 21 lowest-energy structures of the PA duplex, however, have functional groups that are close enough to form hydrogen bonds, similar to those observed in the AA duplex (*18*). The average hydrogen bond distances in the P·A pairs of the NMR structures are 2.36 Å for the interstrand P4N3−A5H6b atoms, 2.42 Å for the interstrand P4O2′−A5H6a atoms, and 2.16 Å for the intrastrand P4 2′ OH−A5O4′ atoms (Figure 4). In some of the structures, the P4 2′ OH hydrogen is close enough to form a bifurcated hydrogen bond with the A5O4′ and A5O5′ oxygens, with an average distance of 2.60 Å to the A5O5′ oxygen. The 21 lowest-energy structures all exhibit essentially the same hydrogen-bonding pattern. In addition, there is the possibility of small favorable electrostatic interactions from weak hydrogen bonds involving C−H groups in the loop: A5N7 is ∼2.5 Å from P4H2.

Hydrogen Bonding in the Loop Regions of the AP and PP Duplexes. Due to the lack of an amino group on P5, neither the AP nor PP duplexes has interstrand hydrogen bonding involving their opposing bases. Both the AP and PP duplexes, however, have intrastrand hydrogen bonds within the loop. In some of the structures, the A/P4 2′ OH hydrogen is close enough to form a hydrogen bond with the P5O5′ oxygen, with an average distance of 2.60 and 2.37 Å for the AP and PP duplexes, respectively. There is the possibility of small favorable electrostatic interactions from weak hydrogen bonds involving C−H groups in the loop region of AP: P5N7 is ∼2.8 Å from P4H2, and A4N1 is ∼2.7 Å from P5H6. Despite the paucity of hydrogen bonds in the loop, both the AP and PP duplexes appear to have predominant structures. Evidently, stacking, volume exclusion, backbone constraints, and perhaps solvation are sufficient to organize the residues of a small internal loop.

Stacking. In the AA, PA, and PP duplexes, the stacking of the mismatch on the adjacent C−G pair is similar (Figure 6). Stacking between mismatch pairs, however, differs significantly. In the AA and PA duplexes, there is moderate overlap of the A/P4 base and the interstrand A/P4 base. This stacking is not observed in the average PP duplex (Figure 6d), although the position of the P4 residues of the PP duplex is relatively undefined, presumably reflecting a lack of interstrand hydrogen bonding to stabilize the pairing. There

are only three interstrand NOEs involving the P4 residues of the PP duplex and a range of conformations of the P4 residue satisfy the NOE restraints. In the AA and PA duplexes, the A5 six-membered ring is directly above the G6H1′ proton, which is consistent with the significant upfield shift of the G6H1′ proton in the AA and PA duplexes (*43*). In the PP duplex, the G6H1′ proton is near the P5 six-membered ring, but not directly below it, which explains the reduced upfield shift of this proton. Thus, the chemical shifts for the G6H1′ proton are consistent with the varying degrees of overlap of G6H1′ with the ring of the A/P5 base in the final structures, providing additional confidence in the structures.

The sheared conformation of the mismatches in the PP duplex is surprising (Figure 6d). There are no hydrogen bonds that favor this conformation over the face-to-face conformation observed for the AP duplex. Presumably, the AP duplex prefers the face-to-face conformation because this requires less distortion in the sugar−phosphate backbone. There is no obvious interaction with the A4 amino group to favor the face-to-face conformation. Rationalizing this structural sensitivity to the presence of amino groups presents an interesting challenge to theoreticians.

Comparison of Dynamics. In addition to thermodynamic and structural differences, the removal of amino groups from the loop region also affects the dynamics of the loop. For the AA and PA duplexes, the A/P4H3′−A5P cross-peak in the HETCOR spectra is weak. Also, A/P4H1′−A/P4H2′ cross-peaks are observed in both the COSY and TOCSY spectra. All of these examples are suggestive of dynamics in the region of A/P4, including interconversion between C2′-endo and C3′-endo sugar puckers (*18*). Since there are possible hydrogen bonds between the A/P4 and A5 bases, it seems probable that, although somewhat dynamic, the A/P4 residue of the AA and PA duplexes spends a majority of the time in hydrogen-bonding range of the A5 residue. The direct hydrogen bonding between the A4 base and the A5 base in the AA duplex could possibly be replaced by a water-mediated hydrogen bond, causing the A4 base to swing out into the major groove, with only minor structural changes, as seen in GNRA tetraloops (*44*).

There is also NMR evidence of dynamics in the AP and PP duplexes. In both of these duplexes, the A/P4H3′−A/P5P cross-peak is less intense than the other HETCOR cross-peaks. In addition, A/P4H1′−A/P4H2′ and P5H1′−P5H2′ cross-peaks are observed in COSY and TOCSY spectra. This suggests that A/P4 and P5 are interconverting between C2′-endo and C3′-endo sugar puckers. Due to replacement of the P5 amino group with a hydrogen, the interstrand hydrogen bonding has been eliminated. Thus, it is not surprising that the P5 residues in these two duplexes exhibit greater dynamic character than in the AA and PA duplexes. Evidently, the 6-amino groups of adenosines in 5′(rGGCAAGCCU)$_2$ affect the thermodynamic, structural, and dynamic characteristics of this internal loop.

ACKNOWLEDGMENT

We thank Dr. S. J. Schroeder for NMR suggestions, Dr. M. D. Disney for helpful discussions, and S. Duan for help with the electrostatic potential calculations.

SUPPORTING INFORMATION AVAILABLE

Four tables listing proton and phosphorus chemical shifts, torsion angle restraints, distance restraints, and T_1 relaxation times are available. Six figures show 1D imino spectra, the imino−amino region of the NOESY WATERGATE spectra, the H1′-sugar region of the NOESY spectra, the base-sugar region of the NOESY spectra, HETCOR spectra, and the sugar−sugar region of the DQF−COSY spectra. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Burkard, M. E., Turner, D. H., and Tinoco, I., Jr. (1999) in *The RNA World* 2nd ed. (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) pp 233−264, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
2. Schroeder, S. J., Burkard, M. E., and Turner, D. H. (1999) *Biopolymers 52*, 157−167.
3. Major, F., Turcotte, M., Gautheret, D., Lapalme, G., Fillion, E., and Cedergren, R. (1991) *Science 253*, 1255−1260.
4. Pinard, R., Lambert, D., Heckman, J. E., Esteban, J. A., Gundlach, C. W., Hampel, K. J., Glick, G. D., Walter, N. G., Major, F., and Burke, J. M. (2001) *J. Mol. Biol. 307*, 51−65.
5. Leontis, N. B., and Westhof, E. (1998) *Q. Rev. Biophys. 31*, 399−455.
6. Leontis, N. B., Stombaugh, J., and Westhof, E. (2002) *Nucleic Acids Res. 30*, 3497−3531.
7. Chou, S. H., Zhu, L., and Reid, B. R. (1997) *J. Mol. Biol. 267*, 1055−1067.
8. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) *Science 273*, 1678−1685.
9. Juneau, K., Podell, E., Harrington, D. J., and Cech, T. R. (2001) *Structure 9*, 221−231.
10. Carter, A. P., Clemons, W. M., Brodersen, D. E., Morgan-Warren, R. J., Wimberly, B. T., and Ramakrishnan, V. (2000) *Nature 407*, 340−348.
11. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) *Science 289*, 905−920.
12. Basavappa, R., and Sigler, P. B. (1991) *EMBO J. 10*, 3105−3111.
13. Bullock, T. L., Sherlin, L. D., and Perona, J. J. (2000) *Nat. Struct. Biol. 7*, 497−504.
14. Butcher, S. E., Allain, F. H., and Feigon, J. (1999) *Nat. Struct. Biol. 6*, 212−216.
15. Michel, F., and Westhof, E. (1990) *J. Mol. Biol. 216*, 585−610.
16. Ortoleva-Donnelly, L., Szewczak, A. A., Gutell, R. R., and Strobel, S. A. (1998) *RNA 4*, 498−519.
17. Strauss-Soukup, J. K., and Strobel, S. A. (2000) *J. Mol. Biol. 302*, 339−358.
18. Znosko, B. M., Burkard, M. E., Schroeder, S. J., Krugh, T. R., and Turner, D. H. (2002) *Biochemistry 41*, 14969−14977.
19. Burkard, M. E., and Turner, D. H. (2000) *Biochemistry 39*, 11748−11762.
20. Gaussian 98, Frisch, M. J., Trucks, G. E., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., Montgomery, J. A., Jr., Stratmann, R. E., Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Kudin, K. N., Strain, M. C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G. A., Ayala, P. Y., Cui, Q., Morokuma, K., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Cioslowski, J., Ortiz, J. V., Baboul, A. G., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., Martin, R. L., Fox, D. J., Keith, T., Al-Laham, M. A., Peng, C. Y., Nanayakkara, A., Gonzalez, C., Challacombe, M., Gill, P. M. W., Johnson, B. G., Chen, W., Wong, M. W., Andres, J. L., Head-Gordon, M., Replogle, E. S., and Pople, J. A. (1998) *Gaussian98*, Gaussian, Inc., Pittsburgh, PA.
21. Allinger, N. L., Yuh, Y. H., and Lii, J. H. (1989) *J. Am. Chem. Soc. 111*, 8551−8566.
22. Stewart, J. J. P. (1989) *J. Comput. Chem. 10*, 209−220.
23. Stewart, J. J. P. (1989) *J. Comput. Chem. 10*, 221−264.
24. Harihara, P. C., and Pople, J. A. (1974) *Mol. Phys. 27*, 209−214.

25. Hehre, W. J., Ditchfield, R., and Pople, J. A. (1972) *J. Chem. Phys. 56*, 2257−2261.
26. Varani, G., and Tinoco, I. (1991) *Q. Rev. Biophys. 24*, 479−532.
27. Varani, G., Aboulela, F., and Allain, F. H. T. (1996) *Prog. Nucl. Magn. Reson. Spectrosc. 29*, 51−127.
28. Varani, G., Cheong, C., and Tinoco, I., Jr. (1991) *Biochemistry 30*, 3280−3289.
29. Gorenstein, D. (1984) *^{31}P NMR, Principles and Applications*, Academic Press, New York.
30. Chou, S. H., Zhu, L. M., and Reid, B. R. (1994) *J. Mol. Biol. 244*, 259−268.
31. Fountain, M. A., Serra, M. J., Krugh, T. R., and Turner, D. H. (1996) *Biochemistry 35*, 6539−6548.
32. Chou, S. H., and Chin, K. H. (2001) *J. Mol. Biol. 312*, 753−768.
33. SantaLucia, J., Jr., and Turner, D. H. (1993) *Biochemistry 32*, 12612−12623.
34. Gralla, J., and Crothers, D. M. (1973) *J. Mol. Biol. 78*, 301−319.
35. Freier, S. M., Sugimoto, N., Sinclair, A., Alkema, D., Neilson, T., Kierzek, R., Caruthers, M. H., and Turner, D. H. (1986) *Biochemistry 25*, 3214−3219.
36. Xia, T., SantaLucia, J., Jr., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998) *Biochemistry 37*, 14719−14735.
37. Turner, D. H., Sugimoto, N., and Freier, S. M. (1988) *Annu. Rev. Biophys. Biophys. Chem. 17*, 167−192.
38. Burkard, M. E., Kierzek, R., and Turner, D. H. (1999) *J. Mol. Biol. 290*, 967−982.
39. SantaLucia, J., Kierzek, R., and Turner, D. H. (1991) *J. Am. Chem. Soc. 113*, 4313−4322.
40. Testa, S. M., Disney, M. D., Turner, D. H., and Kierzek, R. (1999) *Biochemistry 38*, 16655−16662.
41. Gautheret, D., Konings, D., and Gutell, R. R. (1994) *J. Mol. Biol. 242*, 1−8.
42. Wutrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York.
43. Giessner-Prettre, C., Pullman, B., Borer, P. N., Kan, L. S., and Ts'o, P. O. (1976) *Biopolymers 15*, 2277−2286.
44. Jucker, F. M., Heus, H. A., Yip, P. F., Moors, E. H., and Pardi, A. (1996) *J. Mol. Biol. 264*, 968−980.
45. Lu, X. J., Shakked, Z., and Olson, W. K. (2000) *J. Mol. Biol. 300*, 819−840.
46. Saenger, W. (1989) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.

BI0203278