Molecular Recognition by the *Candida albicans* Group I Intron: Tertiary Interactions with an Imino G·A Pair Facilitate Binding of the 5′ Exon and Lower the K_M for Guanosine†

Matthew D. Disney‡,§ and Douglas H. Turner*,‡,§,||

Departments of Chemistry and Pediatrics and Center for Human Genetics and Molecular Pediatric Disease, University of Rochester, Rochester, New York 14627-0216

Received February 5, 2002; Revised Manuscript Received April 22, 2002

ABSTRACT: A G·A pair at position −5 in the P1 helix of the *Candida albicans* ribozyme contributes to tertiary binding of the 5′ exon substrate [Disney, M. D., Haidaris, C. G., and Turner, D. H. (2001) *Biochemistry 40*, 6507−6519]. Here, the G in the G·A pair is replaced with inosine (I) in both semisynthetic ribozymes and oligonucleotide mimics of the internal guide sequence. Comparisons of oligonucleotide binding affinity for these and other sequences indicate that the G·A pair is in an imino conformation where the exocyclic amine of G contributes ∼1.4 kcal/mol to tertiary interactions that help dock the ribozyme's P1 helix. Furthermore, replacement of the G·A pair with a G-C pair produces less favorable interactions with the 2′-hydroxyl group at the −3 position and a less favorable K_M for pG in a ribozyme-catalyzed transesterification reaction. These results are also consistent with the G·A pair promoting docking of the P1 helix into the catalytic core. Evidently, tertiary interactions with the exocyclic amino group of a G in a single G·A pair can increase the equilibrium constant for tertiary folding of RNA by roughly 10-fold at 37 °C. Results with a G·U or G·G pair replacing the G·A pair at the −5 position suggest similar tertiary interactions with these pairs.

Functions of RNAs and RNA−protein complexes often depend on folding into proper three-dimensional structures. The motifs and interactions that contribute to tertiary folding are currently being discovered (*1*). Recent publication of several high-resolution crystal structures of the ribosome and other structured RNAs has greatly increased the amount of information available on RNA tertiary folding motifs (*2*−*11*). Although there is a large database of information on tertiary motifs, little is known about their energetics.

Stability parameters for tertiary interactions coupled with a better understanding of sequences that can form these interactions could facilitate prediction of RNA tertiary structure from sequence. A database of tertiary structure energetics is beginning to emerge. These include tertiary contacts to 2′-hydroxyl groups (*12*−*16*), the ribose zipper motif (*17*, *18*), the exocyclic amine of G in a terminal G·U pair (*19*−*22*), various base triple interactions (*23*−*25*), and a single G·A pair in a helix (*26*). In the latter case, tertiary interactions were more than 2 kcal/mol less favorable when the G·A pair at position −5 in the P1 helix of the *Candida albicans* group I ribozyme was replaced with a G-C pair, but were 0.8 kcal/mol more favorable when the G·A pair was replaced with a G·U pair. This suggested that tertiary contacts are made with the exocyclic amino group placed in the minor groove by an imino conformation G·A pair or a wobble conformation G·U pair. This hypothesis is tested here by substituting inosine (I)[1] for guanine in various contexts. Results indicate that tertiary contacts with the exocyclic amine of G in internal non-Watson−Crick pairs can contribute ∼10-fold to RNA tertiary folding at 37 °C.

MATERIALS AND METHODS

Buffers and General Protocols. H10Mg buffer contains 135 mM KCl, 50 mM Hepes (25 mM Na^+ Hepes), and 10 mM $MgCl_2$ (pH 7.5) M10Mg buffer replaces Hepes with Mes (pH 6.5). Stop buffer contains 10 M urea and 15 mM EDTA, in TBE buffer [100 mM Tris, 90 mM boric acid, and 1 mM EDTA (pH 8.4)]. The T7 RNA polymerase, containing an N-terminal seven-His tag, was overexpressed in *Escherichia coli* and purified using Ni−NTA−agarose resin (Qiagen) (*27*).

Oligonucleotide Synthesis. Oligonucleotides were synthesized and purified as previously described (*28*). Inosine phosphoramidite was from ChemGenes; all other monomers were from Glen Research (Baltimore, MD). Oligonucleotide purity was checked by HPLC as described previously (*28*); all oligonucleotides were at least 95% pure. The C-10/1x ribozyme was synthesized by T7 transcription, purified on a 5% polyacrylamide denaturing gel, isolated from the gel by electroelution with pH 7.0 TBE buffer, ethanol precipitated, resuspended in sterile water, and stored at −20 °C.

† This work was supported by NIH Grant GM22939. M.D.D. was partially supported by Sherman-Clarke, Elon Huntington Hooker, and Arnold Weissberger Memorial Fellowships.

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

‡ Department of Chemistry.

§ Center for Human Genetics and Molecular Pediatric Disease.

|| Department of Pediatrics.

[1] Abbreviations: C-10/1x, *C. albicans* group I ribozyme; Hepes, *N*-(2-hydroxyethyl)piperazine-*N*′-2-ethanesulfonic acid; I, inosine; IGS, internal guide sequence; K_M, Michaelis−Menten constant; Mes, 2-(*N*-morpholino)ethanesulfonic acid.

10.1021/bi020102x CCC: $22.00
Published on Web 05/29/2002

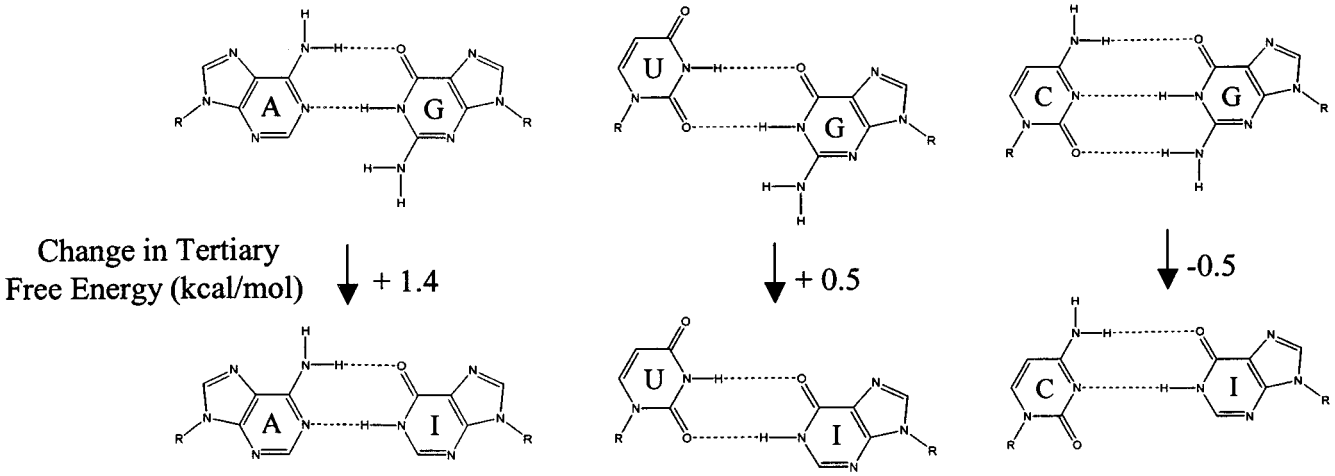

FIGURE 1: Proposed structure of base pairs at position −5 and the penalty in tertiary binding observed for replacing the exocyclic amino group of G with a hydrogen.

Plasmid Synthesis. Truncated ribozyme, C-25/1x, was PCR amplified from the *C. albicans* precursor plasmid, C-h (*26*), and ligated into a pGEM-3zf(+) vector into the *Sfi*I and *Xba*I sites as described for the C-10/1x ribozyme (*26*). The transcription start site was chosen to maximize the number of initial guanosines, which have been shown to be the best sequences for starting transcription (*29*). The plasmid was sequenced in both directions (CORE sequencing Lab, University of Rochester) to confirm its identity.

Construction of Semisynthetic Ribozymes. The protocol of Strobel and Cech (*14*) was used, with minor changes described below. Transcription reactions were performed with a GMP to GTP ratio of 20:1. After transcription, the reaction mixture was extracted with 1 volume of a phenol/chloroform/isoamyl alcohol (24:25:1) mixture and ethanol precipitated with 3 volumes of ethanol. To remove the large amount of salt in the pellet, it was resuspended and dialyzed against sterile water overnight. After dialysis, RNA was ethanol precipitated, purified on a 5% polyacrylamide denaturing gel, and isolated from the gel as described above. Ligation and tertiary disruption procedures were performed as described previously (*14*) except that the ligation reaction mixtures were incubated at 37 °C for ∼4 h. The ligation oligonucleotide was d(CATGGCGTCCCTACTTTTGCC), and the tertiary disruption oligonucleotide was d(CCAGCCCATACCTTTCCGTGCTCTACGACGGCC).

Optical Melting and Gel Binding Experiments. Optical melting experiments were performed in H10Mg buffer as described previously (*28*) with a Gilford 250 UV−vis spectrophotometer equipped with a temperature programmer. Binding of various 5′ exon substrates was assessed with direct gel binding assays in H10Mg buffer as described previously (*26*); all ribozymes used in gel binding experiments were synthesized via splint ligation.

The 5′GACUCU/3′C**I**GAGG, 5′GACUCU/3′CGGAG**I**, and 5′GACUC**C**/3′CGGAG**I** duplexes were too unstable for the thermodynamics to be measured experimentally. Therefore, stabilities were estimated by assessing a related duplex and correcting for the expected difference in stability on the basis of nearest-neighbor parameters (*30*−*32*); a description of the calculations for each of these duplexes is given below.

The duplex stability of 5′GACUCU/3′C**I**GAGG was estimated from the measured stability of 5′GACUC**C**/3′C**I**GAGG by correcting for the 1.6 kcal/mol difference in G-C versus G·U pairs (*30*−*32*) and for adding a terminal G·U pair to the duplex (*31*, *32*). The duplex stability of 5′GACUCU/3′CGGAG**I** was estimated from the measured stability of 5′G**C**CUCU/3′CGGAG**I** by correcting for the 4.3 kcal/mol difference in the G-C versus G·A pair in this context (*26*). Stabilities of duplexes 5′GAC**dU**CU/3′CGGAGG and 5′GAC**dU**CU/3′C**I**GAGG were estimated from the stabilities of 5′GACUCU/3′CGGAGG and 5′GACUCU/3′C**I**GAGG by correcting with the 0.5 kcal/mol difference for introducing a deoxy substitution into a related duplex in the same position, 5′GACUCU/3′CUGAGG versus 5′GAC**dU**CU/3′CUGAGG (*26*).

Kinetics. Experiments with the RNA substrates were performed with ∼40 pM substrate and 400 nM saturating ribozyme. Experiments with the weakly binding RNA−DNA chimera substrate, d(GACUC)Ud(CAA), were performed with ∼40 pM substrate and 200 nM ribozyme. Wild-type ribozyme used in kinetic experiments was prepared by transcription from the C10/1x plasmid (*26*), and mutant ribozymes were prepared by splint ligation. All ribozymes were refolded by incubation in H10Mg buffer at 50 °C for 10 min and slow cooling to 37 °C in a circulating water bath. Reactions were started by adding a solution containing 5′ end-labeled oligonucleotide and pG to a solution containing ribozyme. Data analysis and general methods were those described by McConnell et al. (*33*).

RESULTS

Binding of Oligonucleotides to Ribozymes and Internal Guide Sequence (IGS) Mimics. To determine the contribution of the guanine exocyclic amino groups at positions −1 and −5 in the P1 helix to binding, these G's were replaced with I in both ribozymes and oligonucleotide mimics of the IGS (Figures 1 and 2). Binding of oligonucleotides to ribozyme and to IGS mimics was assessed by gel retardation and optical melting experiments, respectively (Figure 3 and Tables 1 and 2).

The free energy of tertiary interactions is calculated by subtracting the free energy of base pairing from the free energy of binding to ribozyme, where the free energy of base pairing is taken to be the free energy of forming the isolated

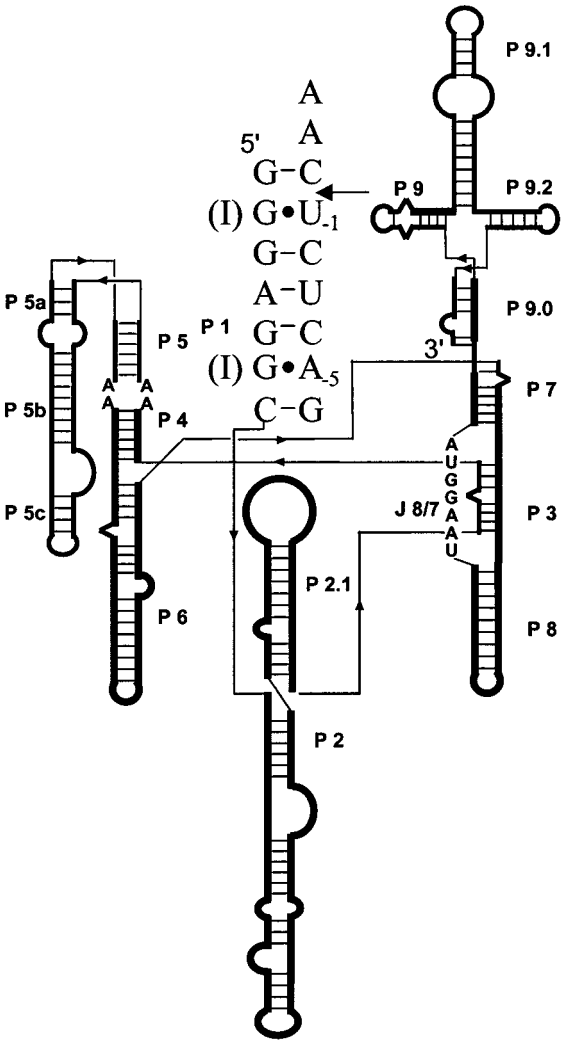

FIGURE 2: Schematic of the secondary structure of the *C. albicans* ribozyme bound to the substrate GACUCUCAA. The splice site is indicated by an arrow. Inosine substitutions are indicated (I), and termed positions −1 and −5.

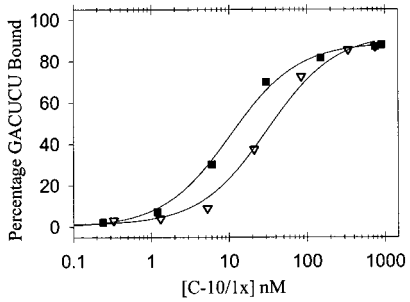

FIGURE 3: Plots of data for binding of GACUCU to wild-type (■) and position −5 I mutant ribozymes (▽). Data were fit to the equation $\theta = [\text{C-10/1x}]/([\text{C-10/1x}] + K_d) + c$ (*14*, *54*), where Θ is the percent 5′ exon mimic bound, K_d is the dissociation constant, [C-10/1x] is the ribozyme concentration, and *c* is a constant.

P1 duplex without the unpaired 5′G present in the ribozyme (Tables 1 and 2). For the wild-type ribozyme, the most favorable tertiary free energies were measured for oligonucleotides with a noncanonical pair at the −5 position. The GACUCU, G**G**CUCU, and G**U**CUCU oligonucleotides bind with tertiary free energies ranging from −7.1 to −7.9 kcal/mol, whereas G**C**CUCU only exhibits −4.8 kcal/mol of tertiary free energy. Tertiary interactions become less favorable when the G in a noncanonical pair at position −5 is changed to I, but become slightly more favorable when a G-C pair at the −5 position is changed to an I·C pair.

Base pairing of 5′GACUCU and 5′G**G**CUCU to the IGS mimic 5′GGAG**I**C is more favorable than that to 5′GGAGGC by ∼0.7 kcal/mol (Tables 1 and 2). This suggests that the hydrogen bonding patterns in the G·A and G·G pairs do not involve the exocyclic amine of the IGS G5, thus making this functional group available for tertiary interactions. Binding of 5′G**C**CUCU and 5′G**U**CUCU is less favorable to 5′GGAG**I**C than to 5′GGAGGC. For 5′G**C**CUCU, the free energy is less favorable by 1.6 kcal/mol, which is the expected increment for the loss of a hydrogen bond (*30*, *34*). For 5′G**U**CUCU, where a hydrogen bond is not lost, base pairing is only 0.6 kcal/mol less favorable. This could reflect a reduced level of stacking interactions (*35*), since the amino group of the G is expected to overlap the carbonyl of the adjacent C (see Figure 7b of ref *36*).

Binding of oligonucleotides with 2′-deoxy substitutions at position −3 was studied because this 2′-hydroxyl group contributes to tertiary interactions (Table 2) (*26*). For the wild-type ribozyme, the effect of removing this 2′-hydroxyl group is similar to that previously determined (*26*); for G·A and G-C pairs at the −5 position, the hydroxyl group at the −3 position makes tertiary interactions more favorable by 2.0 and 0.2 kcal/mol, respectively. Changing the pair at the −5 position from G·A to I·A has essentially no effect on the tertiary free energy contribution of the 2′-hydroxyl group at the −3 position. Changing the G-C pair to an I·C pair at the −5 position, however, makes the tertiary interaction of the 2′-hydroxyl at the −3 position more favorable by 1.7 kcal/mol (Table 2). This suggests that the ribozyme binds differently to the substrate with a −5 position I·C pair than the one with a −5 position G-C pair.

K_M for pG Reacting with Various Bound 5′ Exon Mimics. The enhanced free energies observed with noncanonical pairs at position −5 presumably reflect enhanced docking of the P1 helix into the catalytic site of the ribozyme. To test for altered P1 helix docking, K_Ms for pG were measured for the ribozyme bound to various 5′ exon mimics (Table 3 and Figure 4). These experiments were based on the observation that the K_M for binding of G to the *Tetrahymena thermophila* ribozyme is coupled to binding of substrate (*33*, *37*); the K_M of G in a ribozyme-catalyzed transesterification reaction is more favorable when the P1 helix is correctly docked by forming its full complement of tertiary contacts than when some of these tertiary contacts are not formed (*20*, *33*, *38*). For example, the K_M for pG in a reaction catalyzed by the *T. thermophila* ribozyme is 4-fold more favorable with an RNA substrate, CCUCUA, than with an RNA−DNA chimeric substrate, d(CCUC)Ud(AAAA) (*33*). This is due to the less favorable docking of the P1 helix with d(CCUC)-Ud(AAAA) because 2′-hydroxyl groups at the −2 and −3 positions contribute to docking (*12*, *15*, *16*, *33*) and likely to proper orienting of substrate.

The K_M for pG was measured when the pair at the −5 position of P1 was a G·A, G·U, G-C, I·A, or I·C pair, and also when the pair at the −1 position was a G·U or G-C pair with the G·A pair at the −5 position. Measurements were also made with d(GACUC)Ud(CAA), a substrate that mimics the weakly docked substrate used by McConnell et al. (*33*) (Table 3). Ribozymes with substrates that form either

Table 1: Thermodynamics for Binding to IGS Mimics in H10Mg Buffer[a]

	thermodynamics from $1/T_M$ vs $\ln(C_T/4)$ plots				thermodynamics from the average curve fits			
oligonucleotide	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	T_m (°C)	$-\Delta G°_{37}$ (kcal/mol)	$-\Delta H°$ (kcal/mol)	$-\Delta S°$ (eu)	T_m (°C)
binding to 5′GGAGGC								
5′GACUCU[b]	4.2	–	–	–	–	–	–	–
5′G**C**CUCU[b]	8.5 ± 0.1	59.3 ± 2.1	163.6 ± 6.7	47.7	8.8 ± 0.5	66.4 ± 12.1	185.6 ± 37.6	47.9
5′G**G**CUCU[b]	4.1 ± 0.5	54.2 ± 9.9	161.6 ± 32.1	23.7	4.1 ± 0.6	52.4 ± 9.9	155.6 ± 33.7	23.1
5′G**U**CUCU[b]	4.9 ± 0.2	52.1 ± 3.6	152.3 ± 12.0	27.6	4.9 ± 0.3	54.0 ± 6.6	158.5 ± 22.1	27.7
5′G**C**CUC**C**[b]	10.7 ± 0.6	77.9 ± 10.6	216.6 ± 23.6	54.5	10.7 ± 0.6	77.0 ± 9.6	213.8 ± 29.2	54.7
5′G**C**C**dU**C**U**[c]	(8.2 ± 0.1)	(73.3 ± 4.9)	(209.9 ± 15.7)	(44.2)	(8.5 ± 0.4)	(87.2 ± 11.1)	(253.8 ± 34.7)	(44.0)
binding to 5′GGAG**I**C								
5′GACUC**C**	6.3 ± 0.3	45.6 ± 6.3	126.8 ± 20.3	35.5	6.4 ± 0.1	57.7 ± 10.8	149.3 ± 34.7	36.3
5′G**C**CUC**U**	6.9 ± 0.1	62.8 ± 4.7	180.1 ± 15.3	39.1	7.0 ± 0.2	64.5 ± 8.4	185.4 ± 27.0	39.3
5′G**G**CUCU	4.9 ± 0.4	35.2 ± 5.5	97.7 ± 18.3	23.5	4.9 ± 0.3	39.4 ± 7.0	111.4 ± 22.5	24.4
5′G**U**CUCU	4.3 ± 0.1	51.1 ± 2.5	150.7 ± 8.3	24.1	4.2 ± 0.5	50.1 ± 9.6	147.8 ± 32.5	23.4
5′G**C**C**dU**CU	6.1 ± 0.2	52.9 ± 5.9	150.9 ± 19.3	34.3	6.1 ± 0.2	51.9 ± 5.5	147.6 ± 17.8	34.5
binding to 5′**I**GAGGC								
5′G**C**CUCU	8.7 ± 0.2	55.5 ± 5.6	151.0 ± 17.4	49.3	8.9 ± 0.5	62.7 ± 8.7	173.6 ± 26.7	49.2
5′G**C**CUC**C**	9.1 ± 0.1	64.9 ± 2.9	179.7 ± 9.1	49.9	9.0 ± 0.2	63.3 ± 5.3	175.1 ± 16.4	49.8

[a] Thermodynamic parameters are reported with extra significant figures to allow for accurate calculation of T_m. [b] Values were previously reported (*26*). [c] Values in parentheses are estimates because the $\Delta H°$ values differ by more than 15%, indicating non-two-state melting behavior.

Table 2: Binding of 5′ Exon Mimics to Ribozymes at 37 °C

P1 duplex	K_d (nM)	$-\Delta G°_{37,total}$ (kcal/mol)	$-\Delta G°_{37,base\ pairing}$ (kcal/mol)	$-\Delta G°_{37,tertiary\ binding}$ (kcal/mol)[a]	$-\Delta\Delta G°_{37,x\rightarrow y}$ (kcal/mol)[b]
5′GACUCU 3′CGGAGG	10 ± 1	11.3	4.2	7.1	–
5′GACUCU 3′C**I**GAGG	45 ± 9	10.4	4.7[e]	5.7	−1.4
5′GACUCU 3′CGGAG**I**	800 ± 150	8.6	4.4[e]	∼4.2	−2.9
5′GACUC**A** 3′CGGAGG	>4000	–	–	–	–
5′G**C**CUCU 3′CGGAGG	0.40 ± 0.02	13.3	8.5	4.8	–
5′G**C**CUCU 3′C**I**GAGG	2.0 ± 0.2	12.3	6.9	5.4	0.5
5′G**U**CUCU 3′CGGAGG	1.0 ± 0.2	12.8	4.9	7.9	–
5′G**U**CUCU 3′C**I**GAGG	6.0 ± 0.9	11.7	4.3	7.4	−0.5
5′G**G**CUCU 3′CGGAGG	7.3 ± 1	11.5	4.1	7.4	–
5′G**G**CUCU 3′C**I**GAGG	122 ± 19	9.8	4.9	4.9	−2.5
5′GACUCUCAA[c] 3′CGGAGGG	10 ± 2	11.3	–	–	–
5′GACUC**A**CAA[c] 3′CGGAGGG	>4000	<7.7	–	–	–
5′GACUCUC[c,d] 3′CGGAGGG	9.2 ± 4	11.4	5.7	5.7	–
		2′-Hydroxyl Substitutions (2′-OH → 2′-H)			
5′GAC**dU**CU 3′CGGAGG	142 ± 13	8.7	3.7[e]	5.1	−2.0
5′G**C**C**dU**CU 3′CGGAGG	0.9 ± 0.1	12.8	8.2	4.6	−0.2
5′GAC**dU**CU[a] 3′C**I**GAGG	∼2000	∼8.1	4.2[e]	∼4.0	∼−1.7
5′G**C**C**dU**CU 3′C**I**GAGG	200 ± 5	9.5	6.1	3.4	−1.9

[a] The $-\Delta G°_{37,tertiary\ binding}$ values are determined by subtracting the free energy of base pairing from the free energy of overall binding. [b] $\Delta\Delta G°_{37,X\rightarrow Y}$ is the tertiary free energy change either when G is changed to I or when the 2′-hydroxyl group is removed. These values were calculated before rounding off $\Delta G°_{37,tertiary\ binding}$. [c] Binding to ribozyme was assessed in M10Mg buffer at pH 6.3 to prevent hydrolysis of the 5′ exon mimic. [d] From ref *26*. [e] Estimated as described in Materials and Methods.

a G·A pair, GACUCUCAA, or a G·U pair, G**U**CUCUCAA, have similar K_Ms for pG of 8 and 13 μM, respectively. Less favorable K_Ms for pG are observed with G**C**CUCUCAA and GACUC**C**CAA as substrates, 45 and 61 μM, respectively. The RNA−DNA chimeric substrate, d(GACUC)Ud(CAA), which cannot form tertiary contacts with the position −3 and −2 2′-hydroxyl groups, also has a less favorable K_M for pG of 57 μM.

Table 3: Addition of pG to Various Substrates at 37 °C

P1 duplex	K_M(pG) (μM)	k_{cat} (min^{-1})	k_{cat}/K_M (M^{-1} min^{-1})
5′GACUCUCAA 3′CGGAGGG	8 ± 3	5.5 ± 0.9	150000
5′G**U**CUCUCAA 3′CGGAGGG	13 ± 3	3.1 ± 0.8	128000
5′G**C**CUCUCAA 3′CGGAGGG	45 ± 10	0.85 ± 0.2	20000
5′GACUC***C***CAA 3′CGGAGGG	61 ± 11	0.27 ± 0.05	1500
5′**d**(GACUC)U**d**(CAA) 3′CGGAGGG	57 ± 9	0.027 ± 0.003	56000
5′GACUCUCAA 3′C**I**GAGGG	23 ± 6	3.6 ± 0.4	70000
5′G**C**CUCUCAA 3′C**I**GAGGG	15 ± 5	1.8 ± 0.1	78000

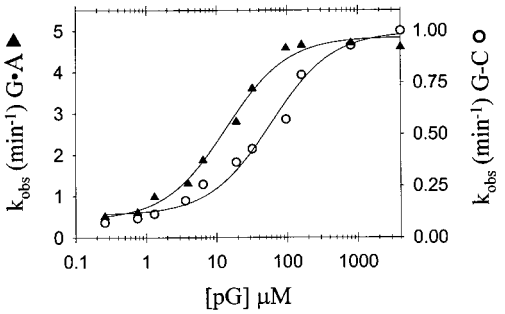

FIGURE 4: Plots of data for reaction rate vs pG concentration for GACUCUCAA (▲) and G**C**CUCUCAA (○) cleavage by the wild-type ribozyme. Data were fit to the equation $k_{obs} = k_c[pG]/(K_M + [pG]) + c$ (*20*, *33*, *38*), where k_{obs} is the rate of the reaction, k_c is the maximum rate of the reaction, K_M is the Michealis−Menten constant for binding of pG, and c is a constant.

These data suggest that substituting the G-C pair for the native G·A pair at the position −5 causes the P1 helix to dock into an orientation that is less favorable for reaction. Furthermore, the G·A pair can be substituted with the G·U pair without significantly affecting the K_M for pG (Table 3), indicating that these pairs promote optimal docking of the P1 helix.

The K_M of pG with the position −5 I ribozyme was also measured. Results show that K_Ms of pG are similar for ribozymes binding to substrates with either position −5 I·A or position −5 I-C pairs (Table 3). These K_Ms, however, are ∼2.5-fold less favorable than for the ribozyme bound to a substrate that forms a G·A pair, suggesting that in both cases the P1 helix is docked into an orientation that is not optimal for reaction.

DISCUSSION

Previously, the G·A pair at position −5 of P1 in the *C. albicans* group I ribozyme was implicated in contributing to tertiary interactions that dock P1 into the catalytic core (*26*). Here, the contribution of the exocyclic amino group of G to tertiary interactions was determined by measuring the stabilities of ribozyme−substrate complexes and short oligonucleotide duplexes with either a G·A or I·A pair at the −5 position (Tables 1 and 2).

Evidence for an Imino G·A Pair at the −5 Position. Stabilities of short duplexes indicate that the amino group of the IGS G at the −5 position is not involved in hydrogen bonding in a G·A pair (Tables 1 and 2). For example, the duplexes 5′GACUCU/3′CGGAGG and 5′GACUCU/3′C**I**GAGG have $\Delta G°_{37}$ values of −4.2 and −4.7 kcal/mol, respectively. No loss in stability suggests an imino G·A pair (Figure 1) since the exocyclic amine of guanine is involved in a hydrogen bond in other structures of G·A pairs with at least two hydrogen bonds; for other structures of G·A pairs that form two hydrogen bonds, see ref *39*. These results are consistent with a molecular modeling study of G·A pairs, which shows that a single sheared G·A pair cannot be inserted between two Watson−Crick pairs because of the backbone distortions introduced by a sheared G·A pair (*40*).

Binding to wild-type and mutant ribozymes shows that tertiary interactions are 1.4 kcal/mol less favorable when the substrate GACUCU is bound to the position −5 I ribozyme than to the wild type (Table 2 and Figure 3). Evidently, the exocyclic amine of G in the G·A pair contributes to tertiary interactions. Tertiary interactions are 0.5 kcal/mol less favorable when the substrate G**U**CUCU is bound to the position −5 I ribozyme, forming an I·U pair, rather than to the wild-type ribozyme, forming a G·U pair (Table 2). In contrast, tertiary interactions are 0.5 kcal/mol more favorable when the substrate G**C**CUCU forms an I·C pair rather than a G-C pair at position −5. The results suggest both G·A and G·U pairs accommodate a similar tertiary contact at position −5. Structures of the imino G·A and wobble G·U pairs place a free amino group of G in the minor groove, suggesting that this functional group forms tertiary contacts (Figure 1).

To test if the ability of G·A and G·U pairs to substitute for each other is a general phenomenon, the G·U pair at the splice site was substituted with a G·A pair (Table 2). The 5′ exon mimic GACUC**A** binds the ribozyme very weakly [K_d > 4000 nM (Table 2)]. If the G·A pair acts as a substitute for the G·U pair, then GACUC**A** should bind ∼3-fold more weakly than GACUCU, since this is the predicted difference in base pairing of these two oligonucleotides to the IGS (*30*). Previous experiments with tandem G·A pairs have shown that the environment around these pairs affects their structure (*41*, *42*). To test if the difference in binding of GACUCU and GACUC**A** is due to having G·U or G·A pairs at the end of a helix rather than at an internal position, GACUC**A**CAA and GACUCUCAA were studied (Table 2). The results with extended substrates mirror those observed with GACUC**X** and further suggest that the ability of G·A and G·U pairs to substitute for each other is context-dependent.

To determine if the loss in the level of tertiary binding that occurs when the exocyclic amine of G is removed is due to loss of a specific tertiary contact and not global rearrangement of the P1 helix docking, the tertiary free energy contribution of the 2′-hydroxyl group at position −3 was measured for substrates that form a G·A or G·I pair at position −5 (Table 2). For both the position −5 I and wild-type ribozymes, the tertiary contribution of the 2′-hydroxyl at position −3 is similar, −2.0 kcal/mol for the G·A pair and −1.7 kcal/mol for the I·A pair. Thus, the 1.4 kcal/mol less favorable tertiary binding due to replacing the G·A pair with a G·I pair is due to the specific loss of the tertiary contact to the exocyclic amine of G (Figure 1).

Evidence That G·G Pairs Also Allow Formation of Tertiary Interactions with an Exocyclic Amino Group. The largest change in tertiary binding with a G to I substitution at position −5 in P1 is the −2.5 kcal/mol observed for the

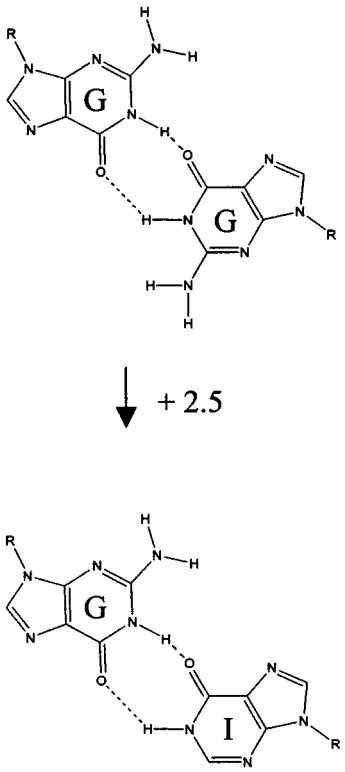

FIGURE 5: Proposed structure of the G·G pair at position −5 and the penalty in tertiary binding observed for replacing the exocyclic amino group of the IGS G with a hydrogen. Note that one of the G's would be in a syn conformation.

G·G pair. G·G pairs with two hydrogen bonds can also place an exocyclic amino group in the minor groove (*39*). For example, the G·G N1-carbonyl, symmetric structure (Figure 5) places a free amino group at about the same position as in an imino G·A pair or a wobble G·U pair. Studies of the effects of I and 7-deazaguanosine substitutions on duplex stabilities suggest the N1-carbonyl, symmetric structure is not stable in the middle of a helix but can be stable in the penultimate position of a helix (*43*). Thus, the data suggest that G·G pairs in the penultimate and possibly terminal positions of helices can also use an exocyclic amino group to form tertiary contacts.

Evidence for Misaligned Docking of the P1 Helix When the Pair at Position −5 Is a G-C Pair. Several pieces of evidence point to misaligned docking of the P1 helix when the pair at position −5 is a G-C pair compared to either a G·A or G·U pair. First, tertiary binding free energy contributed by the 2′-hydroxyl group at position −3 is negligible when the pair at position −5 is a G-C pair (Table 2). Second, the K_M of pG in the presence of the substrate G**C**CUCUCAA is ∼6-fold weaker than with the G·A pair and is close to the K_M observed with the RNA−DNA chimeric substrate that is a mimic of the undocked substrate used with the *T. thermophila* ribozyme (*33*) (Table 3 and Figure 4).

The above effects are decreased in magnitude when the G-C pair is changed to an I·C pair (Tables 2 and 3). The tertiary contribution of the 2′-hydroxyl group at position −3 is more favorable by 1.7 kcal/mol when the pair at position −5 is an I·C pair than when it is a G-C pair. Also, the K_M of pG decreases when the G-C pair is replaced with an I·C pair [45 and 15 μM, respectively (Table 3)]. Thus, removing one of the hydrogen bonds in the G-C pair allows the P1 helix to adjust itself, thereby making the docking equilibrium more favorable by forming tertiary contacts to the 2′-hydroxyl group at position −3. Furthermore, the 0.5 kcal/mol more favorable tertiary interactions when the G-C pair is substituted with an I·C pair at position −5 could be due to the increased contribution of this 2′-hydroxyl tertiary contact. This enhancement may be smaller in magnitude than the energy gained from formation of tertiary contacts to 2′-hydroxyl groups since there may be an energetic penalty if a structural rearrangement occurs to make the docking equilibrium more favorable.

Comparison to Results with the T. thermophila Ribozyme. In the *T. thermophila* ribozyme, several interactions have been shown to stabilize binding of the 5′ exon substrate. These include contacts to 2′-hydroxyl groups (*12*−*16*, *44*, *45*) and the splice site G·U pair (*19*, *20*, *22*, *38*, *46*). The involvement of the position −1 G·U pair in docking of the P1 helix in the *T. thermophila* ribozyme has been supported by several experiments. Replacement of this pair with an I·U pair makes tertiary interactions less favorable by 1.4 kcal/mol (*19*). The K_M of pG is 4-fold weaker with a substrate that forms a G-C pair rather than a G·U pair at position −1 (*38*). Also, the tertiary contribution of the 2′-hydroxyl group at position −3 is less favorable when the splice site G·U pair is replaced with a G-C pair (*38*). For the *C. albicans* ribozyme, the tertiary free energy contribution of the exocyclic amine of G in the splice site G·U pair is −2.9 kcal/mol (Table 1), which is 1.5 kcal/mol more favorable than for the *T. thermophila* ribozyme. Apparently, the contributions of hydrogen bonding groups to tertiary interactions are context-dependent, as also observed in secondary structures (*30*, *34*, *47*).

At position −5 in the *T. thermophila* intron, a G·U pair is formed with the natural substrate (*48*). Replacement of this pair with a G-C pair makes tertiary binding less favorable by ∼1.2 kcal/mol at 42 °C (*38*). At position −5 in the *C. albicans* ribozyme, however, replacement of the G·A pair with a G-C pair makes tertiary binding less favorable by ∼2.5 kcal/mol at 37 °C (Table 2) (*26*). Evidently, the position −5 G·A pair plays a larger role in docking of the P1 helix in *C. albicans* than the position −5 G·U pair in *T. thermophila* even though both present an exocyclic amine in the minor groove.

Inspection of an alignment of 87 group I introns (*49*) reveals that 13 of 87 secondary structures contain a non-Watson−Crick pair in the P1 helix at positions other than the splice site and these occurrences are distributed between the various group I intron classes. Of those 13, five have a non-Watson−Crick pair at the −5 position. Thus, a non-Watson−Crick pair at the −5 position is rare in group I introns, and therefore, the tertiary interaction described here is not used by the majority of group I introns to stabilize binding of the 5′ exon substrate.

Peripheral elements only conserved within a specific group I intron class have been shown to stabilize binding of the G substrate and facilitate acquisition of the intron's tertiary fold (*50*, *51*). Also, loop−loop interactions inferred from sequence comparisons have been shown to facilitate acquisition of the RNA's active state (*52*); for the 1CI subgroup of introns, these loop−loop interactions were inferred for less than 50% of the sequences that were studied. Mutation of these nonconserved elements, however, affects RNA splicing by

causing the RNA to be trapped in an inactive state, which is rescued at higher Mg^{2+} concentrations. Thus, nonconserved elements within group I introns can significantly contribute to acquisition of tertiary structure and binding of substrates.

Insights into Modeling Tertiary Structure. The G·U and G·A pairs are the most and second most common non-Watson−Crick pairs, respectively, in a database of RNA secondary structures (*53*). The data presented here suggest that both G·A and G·U pairs may appear more frequently than other non-Watson−Crick pairs because they can provide contacts for stabilizing tertiary folds. The data further suggest that G·G pairs in the penultimate or terminal positions of helices can also provide stabilizing tertiary interactions. The effects of G to I substitutions indicate that the exocyclic amino groups presented in the grooves of RNA helices by non-Watson−Crick pairs can enhance the equilibrium constant for tertiary folding by more than 10-fold.

ACKNOWLEDGMENT

We thank Mark Burkard, Jess Childs, and Brent Znosko for helpful advice.

REFERENCES

1. Batey, R. T., Rambo, R. P., and Doudna, J. A. (1999) *Angew. Chem., Int. Ed. Engl. 38*, 2326−2343.
2. 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.
3. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szewczak, A. A., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) *Science 273*, 1696−1699.
4. Golden, B. L., Gooding, A. R., Podell, E. R., and Cech, T. R. (1998) *Science 282*, 259−264.
5. Rupert, P. B., and Ferre-D'Amare, A. R. (2001) *Nature 410*, 780−786.
6. Batey, R. T., Rambo, R. P., Lucast, L., Rha, B., and Doudna, J. A. (2000) *Science 287*, 1232−1239.
7. Ferre-D'Amare, A. R., Zhou, K., and Doudna, J. A. (1998) *Nature 395*, 567−574.
8. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001) *Science 292*, 883−896.
9. Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A. (2000) *Science 289*, 905−920.
10. Ogle, J. M., Brodersen, D. E., Clemons, W. M., Jr., Tarry, M. J., Carter, A. P., and Ramakrishnan, V. (2001) *Science 292*, 897−902.
11. Carter, A. P., Clemons, W. M., Jr., Brodersen, D. E., Morgan-Warren, R. J., Hartsch, T., Wimberly, B. T., and Ramakrishnan, V. (2001) *Science 291*, 498−501.
12. Bevilacqua, P. C., and Turner, D. H. (1991) *Biochemistry 30*, 10632−10640.
13. Sugimoto, N., Tomka, M., Kierzek, R., Bevilacqua, P. C., and Turner, D. H. (1989) *Nucleic Acids Res. 17*, 355−371.
14. Strobel, S. A., and Cech, T. R. (1993) *Biochemistry 32*, 13593−13604.
15. Pyle, A. M., and Cech, T. R. (1991) *Nature 350*, 628−631.
16. Pyle, A. M., Murphy, F. L., and Cech, T. R. (1992) *Nature 358*, 123−128.
17. Silverman, S. K., and Cech, T. R. (1999) *Biochemistry 38*, 8691−8702.
18. Klostermeier, D., and Millar, D. P. (2001) *Biochemistry 40*, 11211−11218.
19. Strobel, S. A., and Cech, T. R. (1995) *Science 267*, 675−679.
20. Strobel, S. A., and Cech, T. R. (1996) *Biochemistry 35*, 1201−1211.
21. Strobel, S. A., and Shetty, K. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 2903−2908.
22. Strobel, S. A., and Ortoleva-Donnelly, L. (1999) *Chem. Biol. 6*, 153−165.
23. Doudna, J. A., and Cech, T. R. (1995) *RNA 1*, 36−45.
24. Szewczak, A. A., Ortoleva-Donnelly, L., Ryder, S. P., Moncoeur, E., and Strobel, S. A. (1998) *Nat. Struct. Biol. 5*, 1037−1042.
25. Szewczak, A. A., Ortoleva-Donnelly, L., Zivarts, M. V., Oyelere, A. K., Kazantsev, A. V., and Strobel, S. A. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 11183−11188.
26. Disney, M. D., Haidaris, C. G., and Turner, D. H. (2001) *Biochemistry 40*, 6507−6519.
27. Ichetovkin, I. E., Abramochkin, G., and Shrader, T. E. (1997) *J. Biol. Chem. 272*, 33009−33014.
28. Disney, M. D., Testa, S. M., and Turner, D. H. (2000) *Biochemistry 39*, 6991−7000.
29. Milligan, J. F., and Uhlenbeck, O. C. (1989) *Methods Enzymol. 180*, 51−62.
30. Turner, D. H. (2000) in *Nucleic Acids: Structures, Properties, and Functions* (Bloomfeild, V. A., Crothers, D. M., and Tinoco, I., Jr., Eds.) Chapter 8, University Science Books, Sausalito, CA.
31. Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999) *J. Mol. Biol. 288*, 911−940.
32. 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.
33. McConnell, T. S., Cech, T. R., and Herschlag, D. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 8362−8366.
34. Turner, D. H., Sugimoto, N., Kierzek, R., and Dreiker, S. D. (1987) *J. Am. Chem. Soc. 109*, 3783−3785.
35. Burkard, M. E., Kierzek, R., and Turner, D. H. (1999) *J. Mol. Biol. 290*, 967−982.
36. Chen, X., McDowell, J. A., Kierzek, R., Krugh, T. R., and Turner, D. H. (2000) *Biochemistry 39*, 8970−8982.
37. Bevilacqua, P. C., Johnson, K. A., and Turner, D. H. (1993) *Proc. Natl. Acad. Sci. U.S.A. 90*, 8357−8361.
38. Pyle, A. M., Moran, S., Strobel, S. A., Chapman, T., Turner, D. H., and Cech, T. R. (1994) *Biochemistry 33*, 13856−13863.
39. Burkard, M. B., Turner, D. H., and Tinoco, I., Jr. (1999) in *The RNA World, Appendix 1* (Gesteland, R. F., Cech, T. R., and Atkins, J. F., Eds.) 2nd ed., pp 675−680, Cold Spring Harbor Laboratory Press, Plainview, NY.
40. Gautheret, D., Konings, D., and Gutell, R. R. (1994) *J. Mol. Biol. 242*, 1−8.
41. Wu, M., and Turner, D. H. (1996) *Biochemistry 35*, 9677−9689.
42. SantaLucia, J., Jr., and Turner, D. H. (1993) *Biochemistry 32*, 12612−12623.
43. Burkard, M. E., and Turner, D. H. (2000) *Biochemistry 39*, 11748−11762.
44. Herschlag, D., and Cech, T. R. (1990) *Nature 344*, 405−409.
45. Herschlag, D., Eckstein, F., and Cech, T. R. (1993) *Biochemistry 32*, 8299−8311.
46. Knitt, D. S., Narlikar, G. J., and Herschlag, D. (1994) *Biochemistry 33*, 13864−13879.
47. SantaLucia, J., Jr., Kierzek, R., and Turner, D. H. (1992) *Science 256*, 217−219.
48. Cech, T. R., Damberger, S. H., and Gutell, R. R. (1994) *Nat. Struct. Biol. 1*, 273−280.
49. Michel, F., and Westhof, E. (1990) *J. Mol. Biol. 216*, 585−610.
50. Engelhardt, M. A., Doherty, E. A., Knitt, D. S., Doudna, J. A., and Herschlag, D. (2000) *Biochemistry 39*, 2639−2651.
51. Zarrinkar, P. P., and Williamson, J. R. (1996) *Nucleic Acids Res. 24*, 854−858.
52. Lehnert, V., Jaeger, L., Michel, F., and Westhof, E. (1996) *Chem. Biol. 3*, 993−1009.
53. Kierzek, R., Burkard, M. E., and Turner, D. H. (1999) *Biochemistry 38*, 14214−14223.
54. Testa, S. M., Haidaris, C. G., Gigliotti, F., and Turner, D. H. (1997) *Biochemistry 36*, 15303−15314.

BI020102X