Identifying RNA Minor Groove Tertiary Contacts by Nucleotide Analogue Interference Mapping with N^2-Methylguanosine†

Lori Ortoleva-Donnelly, Matthew Kronman, and Scott A. Strobel*

Department of Molecular Biophysics and Biochemistry, Yale University, 260 Whitney Avenue, New Haven, Connecticut 06520

Received March 30, 1998; Revised Manuscript Received July 2, 1998

ABSTRACT: Nucleotide analogue interference mapping (NAIM) is a general biochemical method that rapidly identifies the chemical groups important for RNA function. In principle, NAIM can be extended to any nucleotide that can be incorporated into an in vitro transcript by an RNA polymerase. Here we report the synthesis of 5′-*O*-(1-thio)-N^2-methylguanosine triphosphate ($m^2G\alpha S$) and its incorporation into two reverse splicing forms of the *Tetrahymena* group I intron using a mutant form of T7 RNA polymerase. This analogue replaces one proton of the N2 exocyclic amine with a methyl group, but is as stable as guanosine (G) for secondary structure formation. We have identified three sites of $m^2G\alpha S$ interference within the *Tetrahymena* intron: G22, G212, and G303. All three of these guanosine residues are known to utilize their exocyclic amino groups to participate in tertiary hydrogen bonds within the ribozyme structure. Unlike the interference pattern with the phosphorothioate of inosine (IαS, an analogue that deletes the N2 amine of G), $m^2G\alpha S$ substitution did not cause interference at positions attributable to secondary structural stability effects. Given that the RNA minor groove is likely to be widely used for helix packing, $m^2G\alpha S$ provides an especially valuable reagent to identify RNA minor groove tertiary contacts in less well-characterized RNAs.

The wide and shallow minor groove of the RNA A-form double helix appears to be commonly used for helix packing interactions that are necessary for the formation of RNA tertiary structure (*1*). Structural motifs involving the RNA minor groove include the "ribose zipper" and the "wobble receptor", both of which employ nucleotide functional groups unique to the minor groove surface (*2, 3*). Unfortunately, the minor groove is difficult to analyze biochemically because the chemical reagents typically used in footprinting and interference studies are not informative for the minor groove functional groups. One reagent that is reactive with these groups is kethoxal, which forms a covalent bridge between the N2 exocyclic amine and N1 groups of G when both functional groups are accessible (*4, 5*). While this reagent is valuable for identifying unpaired Gs within a sequence, kethoxal cannot be used to identify tertiary interactions within helical segments of the RNA. Thus, despite the importance of the minor groove in RNA helix packing, there are significant deficiencies in our ability to explore the minor groove face of the helix with the probing reagents that are currently available.

Nucleotide analogue interference mapping (NAIM)[1] is an efficient method to define the chemical basis of RNA function (*6*−*9*). In principle, the method is generalizable to a wide variety of nucleotide derivatives. In this approach, the nucleotide analogue is chemically tagged with a phosphorothioate linkage and randomly incorporated into the RNA transcript. The sites of phosphorothioate incorporation are detected by cleavage of the linkage with iodine (*10*). This makes it possible to simultaneously, yet individually, quantitate every position where the nucleotide was incorporated into the RNA with an interference assay that is as simple as RNA sequencing. This approach has been used with 2′-deoxy and 2′-methoxy analogues to identify the 2′-OH groups in tRNA and RNase P that are essential for binding (*6, 7, 11*). It has also been employed with inosine and a series of eight adenosine derivatives including 7-deazaadenosine, purine riboside, diaminopurine riboside, and N^6-methyladenosine to explore group I intron catalysis (*3, 8, 9*).

The *Tetrahymena* ribozyme provides an ideal system to further develop NAIM methodology (Figure 1). The intron catalyzes two consecutive transesterification reactions in the course of RNA self-splicing (*12*). It can also catalyze the reverse of these two reactions, resulting in ligation of the exon back onto the intron (Figure 2A,B) (*13*−*16*). By using a radiolabeled oligonucleotide analogue of the exons, the active ribozymes in the population become site-specifically-labeled upon exon ligation, which provides a remarkably simple assay to probe for sites of interference (Figure 2C) (*8*). Furthermore, the *Tetrahymena* ribozyme ranks as possibly the best characterized large RNA, because a crystal structure of the P4−P6 domain of the intron (160 of the 414

† This work was supported by NSF Grant CHE-9701787, a Beckman Young Investigator Award, a Searle Scholar Award, and a Junior Faculty Research Award from the American Cancer Society.

* To whom correspondence should be addressed. Phone: (203)-432-9772. FAX: (203)432-5175. Email address: strobel@csb.yale.edu.

[1] Abbreviations: m^2G, N^2-methylguanosine; $m^2G\alpha S$, 5′-*O*-(1-thio)-N^2-methylguanosine monophosphate; $m^2GTP\alpha S$, 5′-*O*-(1-thio)-N^2-methylguanosine triphosphate; GαS, 5′-*O*-(1-thio)guanosine; GTPαS, 5′-*O*-(1-thio)guanosine triphosphate; IαS, 5′-*O*-(1-thio)inosine monophosphate; NAIM, nucleotide analogue interference mapping; dT, thymidine; rT, 5-methyluridine; dT(-1)S, CCCUC(dT)AAAAA; dT(-1)P, CCCUC(dT); rT(-1)P, CCCUC(rT); SDS, sodium dodecyl sulfate; IGS, internal guide sequence.

S0006-2960(98)00723-5 CCC: $15.00
Published on Web 08/28/1998

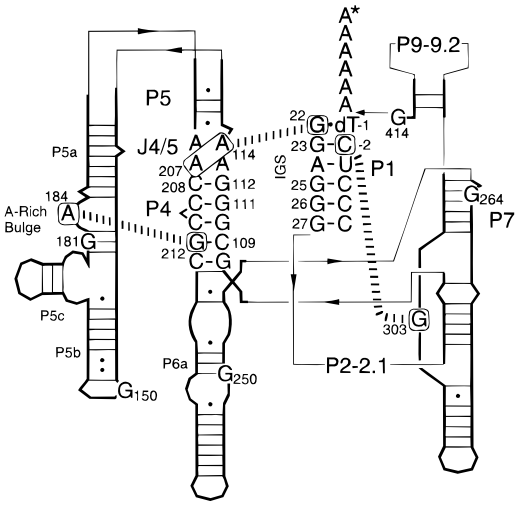

FIGURE 1: Secondary structure of the L-21 G414 form of the *Tetrahymena* group I intron. Nucleotides discussed in the text are shown, as are the names of the helical and single-stranded regions of the RNA. Other nucleotides are depicted as heavy lines. Connectivity within the ribozyme sequence is shown as thin lines, and the tertiary hydrogen bonds formed by the three Gs that display $m^2G\alpha S$ interference are shown as dashed lines. This ribozyme binds the oligonucleotide dT(-1)S, CCCUC(dT)AAAAA, and transfers the AAAAA onto G414 at the 3′-end of the intron in a reaction analogous to the reverse of the second step of splicing (*15, 16*).

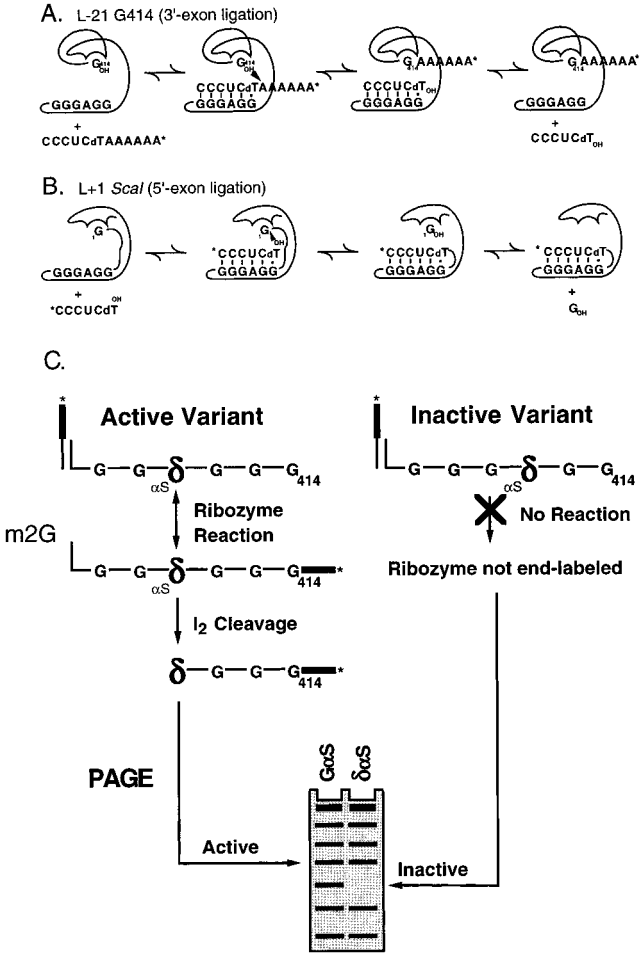

FIGURE 2: (A) Scheme for the reaction of the L-21 G414 ribozyme with oligonucleotide substrate. This reaction is analogous to the reverse of the second step of splicing (*13, 15, 16*). The ribozyme binds the substrate to form the P1 helix, which docks into the active site. The terminal guanosine (G414) nucleophilically attacks the substrate and transfers the 3′-end of the oligonucleotide onto the 3′-end of the intron. The equilibrium constant for the chemical step of this reversible reaction is approximately 1 (*16*). This reaction selectively 3′-end-labels the active ribozymes in the population when a 3′-end-labeled substrate (∗) is used. (B) Scheme for L+1 *Sca*I ribozyme reaction analogous to the reverse of the first step of splicing (*14*). The ribozyme binds the 5′-exon oligonucleotide analogue, where the 3′-OH of the exon attacks the 3′-phosphate of G1, releasing G1 as a free nucleotide and adding the 5′-exon onto the ribozyme. This reaction selectively 5′-end-labels the active ribozymes in the population when a 5′-end-labeled (∗) oligonucleotide is used. (C) Scheme for the identification of the chemical groups important for RNA activity by NAIM (*8*). The phosphorothioate-tagged nucleotide analogue (indicated as $\delta\alpha S$) is randomly incorporated into the transcript in place of G. If $m^2G\alpha S$ does not interfere with function at a particular position (left side), then ribozymes with the analogue at that site perform the ligation reaction and become radiolabeled. If $m^2G\alpha S$ disrupts activity (right side), then the subset of ribozymes that have $m^2G\alpha S$ incorporated at the susceptible site do not perform the ligation reaction and are not radiolabeled. Cleavage of the phosphorothioate linkages by treatment with iodine and resolution of the cleavage products by PAGE produce a sequencing ladder with gaps that correspond to sites intolerant of $m^2G\alpha S$ substitution. $G\alpha S$ serves as a control to ensure that loss of activity is not due to the phosphorothioate group. Unreacted RNA is also 5′-end-labeled to ensure that the gap in the sequencing ladder is not due to lack of $m^2G\alpha S$ incorporation at a given site (not shown).

nucleotides) is available, and three-dimensional models of the rest of the intron have been proposed based upon phylogenetic and other biochemical experiments (*2, 17*−*20*). Thus, interference in this ribozyme system provides a basis set to calibrate NAIM methodology prior to its application on less well-characterized RNAs. Yet, even within this RNA, the interference results can help to refine and improve our understanding of group I intron structure and catalysis.

Previous mapping experiments of the *Tetrahymena* group I intron with the phosphorothioate of inosine (IαS), a G analogue that replaces the N2 exocyclic amine with a proton (Figure 3), revealed several sites that interfered with ribozyme activity (*8*). Interference at many of these sites is unlikely to indicate direct participation of the amine in tertiary hydrogen bonding, but rather reflects a loss of duplex stability. In hopes of developing a better analogue to analyze the role of the minor groove exocyclic amine in tertiary structure formation, we synthesized the 5′-*O*-(1-thio)-N^2-methylguanosine triphosphate ($m^2GTP\alpha S$) and utilized it in NAIM. Instead of deleting the amine, this analogue replaces one of the amino protons with a methyl group (Figure 3). We find that the sites of interference throughout the *Tetrahymena* group I intron are exclusively at positions where the exocyclic amine of G is known to participate in long-range tertiary hydrogen bonds. Thus, $m^2G\alpha S$ provides a means to identify tertiary interactions within the RNA minor groove, a region that is relatively uninformative using previously available biochemical methodologies.

METHODS

Synthesis of $m^2GTP\alpha S$. O^6-[2-(Nitrophenyl)ethyl]-N^2-methylguanosine was prepared as previously described (*21*−*23*). N^2-Methylguanosine was synthesized by treating O^6-[2-(nitrophenyl)ethyl]-N^2-methylguanosine (1.0 g, 2.3 mmol) with 1,8-diazabicyclo[5.4.0]undec-7-ene (5 mL) at room temperature overnight. Water (50 mL) was added to the reaction, and the aqueous phase was extracted with CH_2Cl_2

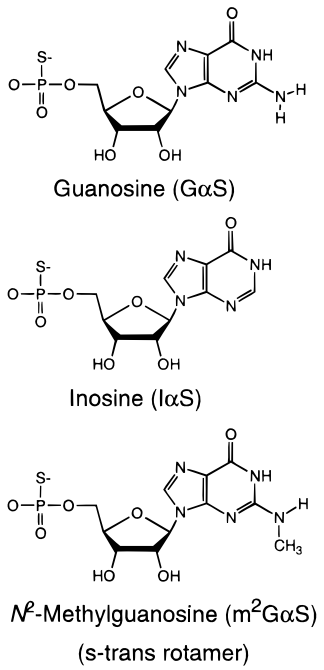

FIGURE 3: Parental G nucleotide and the two nucleotide analogues used for NAIM in this work. Each nucleotide is shown as the monophosphate derivative, the form in which it is incorporated during transcription. N^2-Methylguanosine can adopt either the s-trans (shown) or the s-cis rotamer.

(3 × 100 mL) and with ether (3 × 100 mL). The water was removed by rotary evaporation and the reaction dried overnight in vacuo. Authentic N^2-methylguanosine (*24*) was obtained as an off-white solid (340 mg, 50% yield) by recrystallization of the residue with hot methanol.

Synthesis of the 5′-*O*-(1-thio)-N^2-methylguanosine triphosphate followed the general procedure outlined by Arabshahi and Frey (*25*). N^2-Methylguanosine (50 mg, 0.17 mmol) was dried under vacuum at 110 °C for 16 h and dissolved in triethyl phosphate (2 mL). Trioctylamine (83 μL, 0.19 mmol, 1.1 equiv) and thiophosphoryl chloride (20 μL, 0.19 mmol, 1.1 equiv) were added to the reaction and stirred under argon at room temperature for 30 min to form 5′-*O*-(1-thio-1,1-dichloro)phosphoryl-N^2-methylguanosine. The reaction was about 40% complete based upon thin-layer chromatography (TLC) on cellulose plates using 0.5 M LiCl (aq) as the solvent system. This was converted directly to the triphosphate by addition of tributylammonium pyrophosphate (105 mg, 0.34 mmol, 2.0 equiv) in triethyl phosphate (3 mL) and stirring at room temperature for an additional 30 min. Formation of the N^2-methylguanosine 5′-*O*-(1-thio)-*cyclo*-triphosphate was monitored by silica TLC using 6:3:1 1-propanol, ammonium hydroxide, water as the solvent system. In this system, the triphosphate had an R_f of 0.2 compared to 0.6 for the monophosphate and 0.8 for the free nucleoside. The triphosphate was precipitated by addition of excess triethylamine (2.5 mL), centrifuged, and decanted, and the residue was dissolved in aqueous triethylammonium bicarbonate (TEAB) (50 mM, pH 7.5, 10 mL). The crude product was left at room temperature overnight to achieve hydrolytic ring opening of the cyclic triphosphate. Purification by DEAE-A25 Sephadex chromatography using a linear gradient of 0.05−0.8 M TEAB afforded 5′-*O*-(1-thio)-N^2-methylguanosine triphosphate (m^2GTPαS) as a diastereomeric mixture in 22% yield. The triphosphate eluted at approximately 0.6 M TEAB. ^{31}P NMR (H_2O): 43.37 (m), −6.87 (d), −23.00 (t); λ_{max} 254 nm; ϵ_{254} 13 000.

Transcriptional Incorporation of m^2GαS. Plasmid templates for RNA transcription were prepared by ionic exchange chromatography (Qiagen), digested with the appropriate restriction enzyme, phenol extracted, and ethanol precipitated. pUCL-21G414 was cut with *Ear*I, and pUCL+1 was cut with *Sca*I (*8*). m^2GαS was randomly incorporated into the L-21 G414 or L+1 *Sca*I forms of the intron by in vitro transcription using the wild-type or Y639F mutant form of T7 RNA polymerase (*26*). RNAs were transcribed in 40 mM Tris-HCl, pH 7.5, 4 mM spermidine, 10 mM DTT, 15 mM $MgCl_2$, 0.05% Triton X-100, 0.05 μg/μL DNA template, and 1 mM CTP, UTP, and ATP. Using the pUCL-21G414 plasmid, various concentration ratios of m^2GTPαS (0.1, 0.5, 1.0, and 2.0 mM) and GTP (0.1, 0.5, and 1.0 mM) were tested to identify a transcription condition that gave approximately 5% m^2GαS incorporation as defined by comparison to a transcript made with 50 μM GTPαS (S_P diastereomer only) and 1 mM GTP (*27*). Following this determination, the L+1 *Sca*I RNA was transcribed using 1.0 mM m^2GTPαS, 0.5 mM GTP, and the Y639F polymerase. All the RNAs were purified by PAGE (6% denaturing), eluted into 10 mM Tris, pH 7.5, 0.1 mM EDTA (TE) overnight at 4 °C, precipitated with NaCl and ethanol (−80 °C for 2 h), resuspended in TE, and stored at −20 °C.

5′- and 3′-End-Labeling of RNA. Three oligonucleotides were utilized in the interference mapping experiments. dT(-1)S, CCCUC(dT)AAAAA (20 pmol), was radiolabeled at the 3′-end with [α-^{32}P]cordycepin by yeast poly(A) polymerase (Amersham) (*28*) and used as a substrate for the 3′-exon ligation experiments. dT(-1)P, CCCUC(dT), and rT(-1)P, CCCUC(rT), were each radiolabeled at the 5′-end with [γ-^{32}P]ATP by T4 polynucleotide kinase and used as substrates for the 5′-exon ligation experiments. Both RNAs were purified by PAGE (10% nondenaturing), eluted in TE, and used in the ligation assays without further treatment.

L-21 G414 RNAs (2.5 pmol) were treated with calf intestinal alkaline phosphatase (2 units, 30 min, 37 °C) to remove the 5′-phosphate and heated to 85 °C for 15 min to inactivate the phosphatase. The RNAs were 5′-end-labeled using T4 polynucleotide kinase (5 units) and [γ-^{32}P]ATP at 37 °C for 30 min. The radiolabeled RNAs were purified by PAGE (6% denaturing) and eluted into 0.1% SDS in TE overnight. The SDS was removed by extraction with 1 volume of phenol/chloroform (1:1), and the RNAs were precipitated with NaCl and ethanol, centrifuged, decanted, and resuspended in 50 μL of TE. The specific activities of the 5′-end-labeled RNAs (cpm/μL) were normalized by scintillation counting. The RNAs were cleaved by the addition of 0.1 volume of 100 mM iodine in ethanol and heated to 90 °C for 1 min, and the cleavage products were resolved by PAGE on a 6% or a 5% denaturing gel. Several loadings of the same reaction samples were electrophoresed for variable amounts of time (from 1 to 5 h) to maximally resolve each region of the sequence.

Despite several attempts, we were unable to obtain clean sequence information by 5′-end-labeling the L+1 *Sca*I RNA due to some heterogeneity at the 5′-end of the L+1 *Sca*I RNA. To obtain information for the positions near the 5′-end of the transcript that is comparable to that derived from the 5′-end-labeled control, the L+1 *Sca*I RNA was incubated

with a highly reactive substrate, rT(-1)P, under permissive reaction conditions (10 mM $MgCl_2$, 50 °C, 50 mM HEPES, pH 7.0) for 10 min. The faster reacting ribose substrate reduces the amount of interference observed throughout the length of the RNA and provides a minimum estimate as to the extent of analogue incorporation at each site. This version of the "5′-end-labeled control" was used in the L+1 *Sca*I interference calculations for positions in the internal guide sequence (IGS) and P2 helix (*8*).

m^2GαS Interference Mapping of the Tetrahymena Ribozyme. Interference mapping of m^2GαS was performed using 50 nM ribozyme and 10 nM oligonucleotide substrate in a buffer containing 4 mM $MgCl_2$, 50 mM HEPES, pH 7.0, at 50 °C for 10 min. The L-21 G414 ribozyme was incubated with 3′-end-labeled dT(-1)S, and the L+1 *Sca*I ribozyme was incubated with 5′-end-labeled dT(-1)P. The ligation reactions were stopped by the addition of 1 volume of urea loading buffer (8 M urea, 50 mM EDTA, 0.01% bromophenol blue, 0.01% xylene cyanol) and worked up with iodine as described above. The reaction containing no iodine was run in parallel to confirm that the cleavage pattern was specific to the iodine treatment and not due to nonspecific degradation. The transcriptional efficiencies of m^2GαS incorporation were determined using the 3′-exon ligation reaction of L-21 G414 RNAs with dT(-1)S. Relative efficiencies were calculated by comparing the intensity of the cleavage products throughout the length of the intron to those of the GαS control (5% incorporation standard).

Interference Quantitation. Peak intensities for both the parental nucleotide (GαS) and the nucleotide analogue (m^2GαS) were quantitated by PhosphorImager analysis at each position for the 3′-exon ligation or 5′-exon ligation experiments and the 5′-end-labeled control. The extent of interference at each position was calculated by substituting the band intensities at each nucleotide position into the equation:

$$\text{Interference} = \frac{\text{GαS ligation reaction}/\text{m}^2\text{GαS ligation reaction}}{\text{GαS labeled control}/\text{m}^2\text{GαS labeled control}} \quad (1)$$

The resulting interference value normalizes for phosphorothioate effects assumed to be equivalent for both GαS and m^2GαS. It also controls for variability in the extent of analogue incorporation or reactivity with iodine at each position. All the interference values were further normalized to account for differences in loading and extent of reaction between lanes by calculating the average interference value at all positions in the RNA that were within two standard deviations from the mean and dividing each individual interference value by the normalized average (the averages ranged from 0.8 to 1.2). This resulted in an interference κ value for each position in both the 3′-exon and 5′-exon ligation reactions. A κ value of 1 indicates that there is no effect of substituting the analogue at that site, a value greater than 1 indicates inhibition of activity, and a value less than 1 indicates that activity is enhanced by analogue substitution at that site.

RESULTS

Transcriptional Incorporation of m^2GαS. The α-phosphorothioate-tagged triphosphate of m^2G was synthesized for use in NAIM. To be useful in this method, it must be efficiently and accurately incorporated into an RNA by in vitro transcription. Initial efforts to incorporate m^2GαS were unsuccessful using T7 RNA polymerase, probably because the methyl group occupies a prominent position in the minor groove that is likely to be involved in the error-reading mechanism of the polymerase (*29*). No incorporation was observed even at high ratios of m^2GTPαS to GTP (data not shown). We had previously obtained and overexpressed a Y639F point mutant of T7 RNA polymerase to perform NAIM with 2′-deoxynucleotides. This mutation was reported to cause reduced selectivity for the 2′-position of the nucleotide during transcription (*26*). The mutant polymerase efficiently incorporates 2′-deoxy, 2′-methoxy, 2′-fluoro, and 2′-thio nucleotide triphosphates into RNA transcripts (*9*, *30*−*32*). It seemed reasonable to expect that this polymerase might also have enhanced tolerance for an additional methyl group in the helical minor groove.

We repeated the transcriptions using the Y639F polymerase and found that m^2GαS was efficiently incorporated into the L-21 G414 RNA. An incorporation level of approximately 5% was achieved using a ratio of 1.5 mM m^2GTPαS to 0.5 mM GTP. Based upon iodine treatment of the 5′-end-labeled L-21 G414 transcript, m^2GαS was incorporated at every G position within the RNA, and no significant incorporation was detected at any non-G sites within the sequence (Figure 4A, lanes 2 and 3). Furthermore, the efficiency of incorporation at each G was generally equivalent to that seen for the GαS standard.

m^2GαS Interference Mapping. The sites of m^2GαS incorporation that interfere with intron activity were mapped using the L-21 G414 ribozyme (*8*). pUCL-21G414 encodes a form of the group I intron that includes the terminal G414, but lacks the first 21 nucleotides of the intron. In the presence of an oligonucleotide substrate analogous to the 5′-3′ ligated exon product, dT(-1)S, this intron can perform the reverse of the second step of splicing by using the 3′-OH of the terminal G414 as the nucleophile to attack the splice site between the exons (Figure 2A) (*15*, *16*). This reaction transfers the 3′-exon onto the 3′-end of the intron, which radiolabels the active molecules in the population if the oligonucleotide substrate is 3′-end-labeled. Following the ligation reaction, the RNAs are digested with iodine and the cleavage products resolved by PAGE (Figure 4A). The intensities of individual bands in the m^2GαS and GαS cleavage ladders were compared to the 5′-end-labeled controls to identify the sites of interference. Although the majority of the 107 Gs within the ribozyme were informative, complete data could not be obtained for 10 sites due to the inability to separate the cleavage products at the positions furthest from the location of the radiolabel (G22−G32 for the 3′-exon ligation reaction, and G405−G414 for the 5′-end-labeled control). As might be expected, most positions did not show any effect upon analogue substitution. Greater than 95% of the interference κ values were between 0.67 and 1.5.

Two sites, G212 and G303, showed complete interference with m^2GαS in the 3′-exon ligation reaction (Figure 4A, lanes 8 and 9; Figure 5). Both of these positions are completely conserved among the 131 examples of IC1 and IC2 introns, of which the *Tetrahymena* intron is a member (*33*). The only other Gs that are conserved to this level among IC1

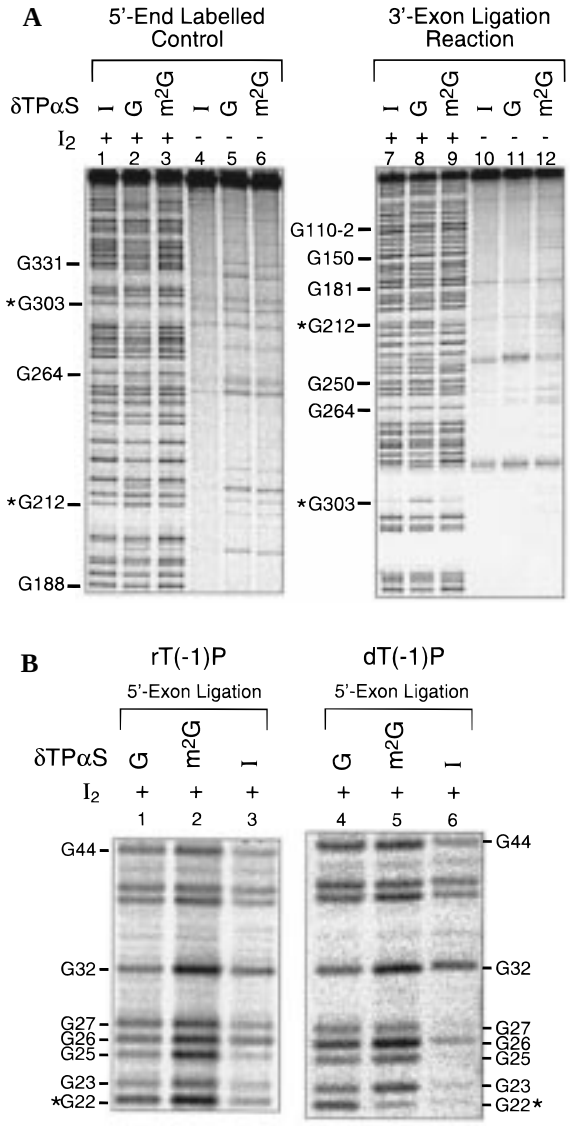

FIGURE 4: Analogue incorporation and interference reactions for 3′- and 5′-exon ligation. (A) Interference mapping for the 3′-exon ligation reaction. The phosphorothioate-tagged analogue incorporated into each RNA is listed above the lane numbers. The nucleotide number corresponding to several of the bands is marked to the left of each gel. The addition (lanes 1−3, 7−9) or omission (lanes 4−6, 10−12) of iodine is indicated. 5′-End Labeled Control: The L-21 G414 5′-end-labeled control showing the extent and positions of analogue incorporation throughout the intron. Cleavage products at positions G212 and G303 (asterisks) show m^2GαS was incorporated at these sites (lane 3). This particular gel was electrophoresed at 75 W for 2.5 h. Longer electrophoretic times were used to improve the signal resolution of the nucleotides toward the 3′-end of the RNA (not shown). 3′-Exon Ligation Reaction: The 3′-exon ligation reaction of L-21 G414 RNA with dT(-1)S. This autoradiogram reveals the sites of analogue interference throughout the intron. The m^2GαS cleavage products at G303 and G212 (asterisks) are of substantially lower intensity than in the 5′-end-labeled control (lane 9). This particular gel was electrophoresed at 75 W for 1.75 h. Longer electrophoretic times were used to resolve the cleavage products toward the 5′-end of the intron (not shown). (B) The 5′-exon ligation reaction of L+1 *Sca*I RNA with rT(-1)P or dT(-1)P in 10 or 4 mM $MgCl_2$, respectively. This autoradiogram shows the interference results for nucleotide positions within the IGS. The asterisk is to call attention to the m^2GαS interference seen at G22. m^2GαS interference was also observed at G212 (not shown). Interference was observed with IαS at all the positions within the IGS (compare lanes 3 and 6). The no-iodine control lanes are omitted, but they were equivalent to those shown in panel A.

and IC2 introns are G22 and G264. G22 was not informative in this assay. G264 is known to be essential to intron function because it forms part of the G binding site, but it does this using functional groups in the major groove of the P7 helix that would not be affected by m^2GαS substitution (*34*). No other sites showed detrimental effects due to m^2GαS substitution, including G111 and G112 where modest interference was previously observed with IαS substitution (*8*). G303 also showed interference with IαS, though there was no IαS interference at G212.

One region of particular interest within the ribozyme is the internal guide sequence (IGS) strand of the P1 helix (Figure 1). Unfortunately, the five Gs within the IGS were uninformative because the cleavage products could not be resolved in the 3′-exon ligation reaction. To gain information about these nucleotides, we used the L+1 *Sca*I ribozyme, which lacks the last 3 nucleotides at the 3′-end of the intron (including G414), but includes the first 21 nucleotides at the 5′-end (*8*). This form of the intron includes a G as the first base of the intron that is equivalent to the exogenous G added to the 5′-end of the intron after the first step of splicing (*35*). In the presence of a 5′-exon oligonucleotide substrate, this intron performs a reaction that is analogous to the reverse of the first step of splicing, wherein it transfers the 5′-exon onto the 5′-end of the intron with concomitant release of the terminal G (*13, 14*) (Figure 2B). This reaction places the radiolabel at the 5′-end of the intron where the IGS nucleotides can be readily resolved. Using this reaction, all five of the Gs within the P1 helix showed interference with IαS, though only G22 is conserved (*8, 19*).

In contrast to the results with IαS interference, G22 was the only position in the IGS that showed interference from m^2GαS substitution (Figure 4B). No m^2GαS interference was detected at G23, G25, G26, or G27. G22 is universally conserved among all known group I introns and is essential for defining the 5′-splice site (*33, 36*−*38*). In the 5′-exon ligation reaction, strong interference was also detected at G212, though there was no interference at G303 (data not shown). Thus, using two G analogues that alter the exocyclic amine, there are sites where interference was observed with both analogues (G22 and G303), as well as sites where only one of the analogues interfered with activity; G212 only showed interference with m^2GαS, while G23, G25, G26, G27, G111, and G112 only showed interference with IαS.

Mutagenesis of the Closing Base Pair in the P4 Helix. Given the reduced stabilities of duplexes containing inosine compared to m^2G (*39*−*41*), the lack of m^2GαS interference at some of the IαS interference sites suggests that reduced duplex stability is the likely cause of IαS interference at these positions. This interpretation implies that the stability of the closing G−C base pair in the P4 helix might be important for 3′-exon ligation activity. To test this possibility directly in the L-21 G414 ribozyme, we mutated the G112−C208 pair to an A112−U208 pair and measured the rates of 3′-exon ligation with the dT(-1)S substrate under the same conditions as those used for the interference experiments (*16*). k_{cat} for 3′-exon ligation was unaffected by the base pair mutation (0.015 ± 0.002 and 0.014 ± 0.001 min^{-1} for G112−C208 and A112−U208, respectively), but K_m increased by approximately 2.5-fold (54 and 140 nM, respectively).

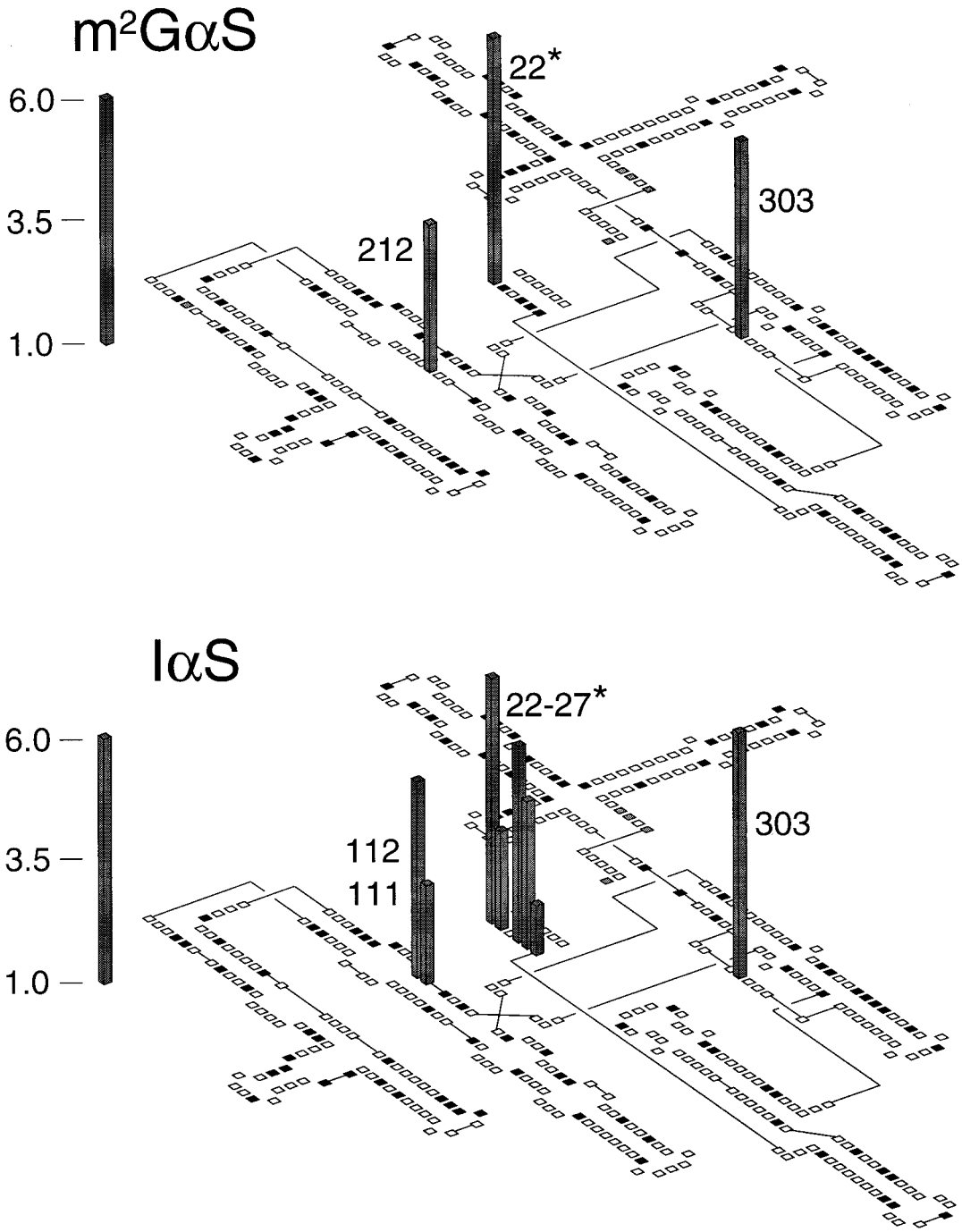

FIGURE 5: Individual histograms for both of the G analogues plotting the magnitude of the interference κ value versus nucleotide position superimposed on the intron secondary structure. Interference κ values ≥ 2.0 are shown as gray bars with the nucleotide number listed adjacent to the bar. Values greater than 6 are assigned a magnitude of 6 for this graph. Black boxes indicate G sites where the κ value was less than 2 (meaning no interference), and white boxes mark nucleotide positions other than G. Gray boxes indicate G sites that were not informative in the assay due to incomplete resolution of the cleavage products on the sequencing gel (G400−G414). The 5′-exon ligation reaction with L+1 *Sca*I RNA was used to obtain values within the IGS (marked with an asterisk). All other values in the histogram were for the 3′-exon ligation reaction using the L-21 G414 RNA. The error in the κ value at each position is $\leq 20\%$. Plotted values are the average of at least two, and as many as eight, independent experimental measurements.

DISCUSSION

Nucleotide analogue interference mapping (NAIM) is a general biochemical method to identify the chemical groups that are important for RNA function (Figure 2C). In a previous report, we used IαS to identify the G positions in the *Tetrahymena* intron where the exocyclic amine is important for 3′-exon and 5′-exon ligation (*8*). These included all five of the Gs within the internal guide sequence (IGS), the closing two G−C base pairs at the top of the P4 helix, and G303 which is located within the J8/7 single-stranded segment that threads through the center of the catalytic core (Figure 5). Based upon sequence conservation and mutagenesis experiments, not all the sites of IαS interference are expected to be involved in tertiary contacts within the ribozyme fold (*33*, *42*).

Inosine substitution, which deletes the N2 exocyclic amine of G, can affect at least two aspects of RNA folding, duplex

stability and tertiary structure formation. Thermodynamic characterization of model RNA duplexes containing inosine substitutions shows that an I−C pair is between 1 and 2 kcal·mol^{-1} less stable than a G−C pair (*39, 40*). Yet, while inosine substitution can destabilize RNA duplexes, it can also have dramatic effects on ribozyme activity if the N2 amine participates in tertiary hydrogen bonding, such as seen at position G22 within the *Tetrahymena* intron (*43*). Because inosine affects both secondary and tertiary stability, it is not possible with IαS interference mapping alone to distinguish between these two contributions to RNA structure.

N^2-Methylguanosine (m^{2}G) is a second analogue that modifies the exocyclic amine of G, but instead of deleting the functional group, it substitutes one proton of the amine with a methyl group (Figure 3). Duplex studies with m^{2}G have shown that it can form base pairs with C, U, and A that are at least as stable as those formed by G, which indicates that the methyl group is equally stable on either the s-trans or the s-cis face of the base (*41, 44*). We reasoned that this property might make it possible to use the phosphorothioate of m^{2}G to probe specifically for essential minor groove tertiary contacts in RNA without encountering the secondary structural effects that are observed with IαS.

Toward this objective, we synthesized the phosphorothioate of *N*^{2}-methylguanosine (m^{2}GαS) and randomly incorporated it into two reverse splicing forms of the *Tetrahymena* group I intron using a mutant form of the T7 RNA polymerase. We observed three sites of m^{2}GαS interference within the *Tetrahymena* intron, G22, G212, and G303 (Figure 5). These sites are located in the P1 helix, P4 helix, and J8/7 single-stranded region, respectively (Figure 1). All three of the Gs are completely conserved among the IC1−IC2 introns (which includes the *Tetrahymena* intron) (*33*), and all three Gs are known to participate in minor groove tertiary hydrogen bonding between helical elements of the ribozyme structure (*2, 3, 45*) (Figure 6).

Interference of P1 Helix Docking into J4/5. G22 forms a wobble pair with U-1 at the 5′-splice site of the intron (*46*), and these two bases are the only conserved residues within the P1 helix (Figure 1) (*19, 42*). While U-1 is essential for holding the G in a wobble configuration, U-1 does not participate directly in ground-state tertiary interactions with the intron core (*43*). Instead, the G utilizes its amine and its 2′-OH to dock into a wobble receptor located within the J4/5 region (*3, 47, 48*). Within a G·U wobble pair, neither of the amino protons participate in duplex formation; however, both protons of the G22 amino group participate in tertiary hydrogen bonds with the two consecutively stacked sheared A·A pairs that constitute the wobble receptor (*3*). One proton donates a hydrogen bond to the N3 of A207 while the second proton donates a hydrogen bond to the 2′-OH of A207 (Figure 6A). Interference with m^{2}GαS and IαS is in full agreement with this model for P1 helix docking, because placement of the m^{2}GαS methyl group onto either the s-cis or the s-trans face of the nucleotide would disrupt a hydrogen bond essential for docking the 5′-exon into the active site.

Interference of the P4 Interaction with the A-Rich Bulge. G212 base-pairs with C109 in the P4 helix. A G−C pair utilizes one of the exocyclic amino protons for duplex formation, but leaves the second proton unpaired in the minor groove. The crystal structure of the P4−P6 domain demonstrated that G212, and nucleotides surrounding it, makes extensive minor groove tertiary interactions with the A-rich bulge of the P5abc extension (Figure 1) (*2, 49*). These contacts are largely sequence-nonspecific hydrogen bonds to 2′-OHs on the P4 helix. The single exception is a tertiary hydrogen bond between the N2 amine of G212 and the N3 of A184 (Figure 6B). Strong interference at G212 with m^{2}GαS suggests that the interaction between P4 and the A-rich bulge is important for activity, though the absence of IαS interference indicates that deletion of the hydrogen bond is not sufficient to affect intron splicing. In this closely packed region of the RNA, the additional steric bulk of the N2-methyl group in the P4 minor groove is likely to be more destabilizing than deletion of the amine.

m^{2}GαS interference at G212 is the first example of significant loss of intron splicing activity resulting from a

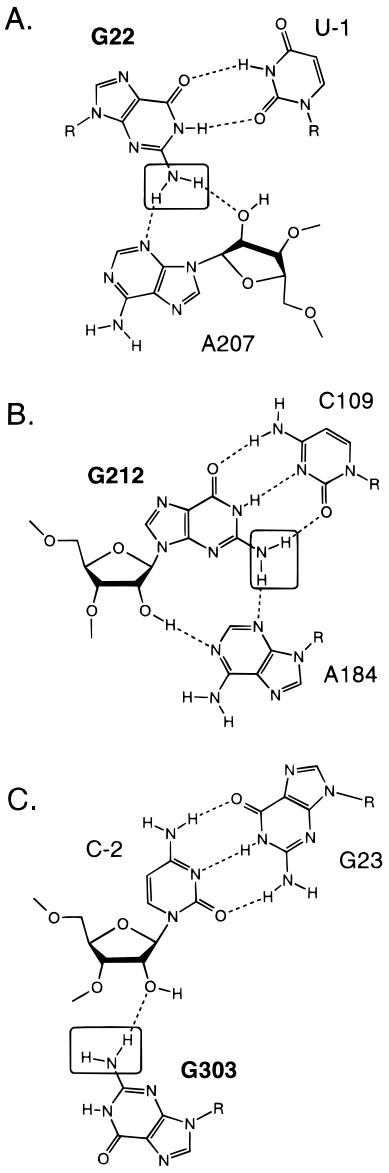

FIGURE 6: Tertiary hydrogen bonding interactions postulated for the three Gs (bold) that show m^{2}GαS interference. In each case, the amino group that shows interference is highlighted with a box. (A) G22 forms a wobble pair with U-1 and makes a tertiary interaction with A207 (*3*). (B) G212 forms a Watson−Crick pair with C109 and makes a tertiary interaction with A184 (*2*). (C) G303 is in a single-stranded region of the ribozyme, but makes a tertiary interaction with C-2 (*45*).

single point mutation or functional group change in the interface between P5abc and the P4−P6 helical stack (*9*, *50*). Apparently, there is sufficient energetic redundancy to overcome most minor structural alterations in the context of the intact intron (*50*). The P4−P6 crystal structure suggests three other sites within the domain that are strong candidates for $m^2G\alpha S$ interference (*2*). The N2 amino groups of G150 and G250 participate in the tetraloop−tetraloop receptor interaction between P5b and P6a, and the amine of G181 helps form the substructure of the A-rich bulge; however, interference was not observed at any of these three sites in either ligation assay (Figure 5). $m^2G\alpha S$ interference at G212, but not at G150 or G250, suggests that the A-rich bulge interaction with P4 may be more important for intron function than the tetraloop/tetraloop receptor interaction.

Interference from P1 Helix Docking into J8/7. Unlike the other two sites that demonstrated $m^2G\alpha S$ interference, G303 is in a single-stranded segment of the intron, J8/7 (Figure 1), which forms an extended triple helix complex with the minor groove of the P1 helix (*45*). Interference suppression experiments using a substrate with a 2′-deoxy substitution at C-2 demonstrated that G303 forms a tertiary hydrogen bond between its exocyclic amine and the 2′-OH of C-2 in the P1 helix (Figures 1 and 6C) (*45*). However, the interaction proposed to form between these two residues only utilizes one of the G303 amino protons. Presumably $m^2G\alpha S$ could adopt the s-cis rotamer to facilitate hydrogen bonding by the s-trans proton of the amine. The fact that interference is observed with both $I\alpha S$ and $m^2G\alpha S$ suggests either that the s-cis proton is involved in an additional hydrogen bond or that there is insufficient space in the P1−J8/7 packing interface to accommodate an s-cis methyl group.

It is somewhat curious that $I\alpha S$ and $m^2G\alpha S$ interference is observed at G303 for the 3′-exon but not the 5′-exon ligation reaction. While this could reflect an important conformational difference between the first and second steps of splicing, the more likely explanation is that there is additional energetic redundancy for P1 helix docking within the 5′-exon ligation reaction. In the L+1 *Sca*I construct, the reactive G is covalently connected to the P1 helix. Additional stabilization energy for P1 docking is provided by G binding into the P7 helix, which would make the G303 interaction with C-2 less critical. In the L-21 G414 ribozyme, the reactive G is at the 3′-end of the RNA where it does not contribute directly to P1 helix binding. In this arrangement, the interaction between G303 and C-2 is necessary for alignment of the substrate into the active site.

Interference with $I\alpha S$ but Not $m^2G\alpha S$ Suggests Duplex Stability Is Important. There are six sites that showed interference with $I\alpha S$ that were not sensitive to $m^2G\alpha S$ substitution. All the sites are in double-stranded regions of the intron, and they cluster into two helices. Four of the sites (G23, G25, G26, and G27) are in the P1 helix, and two of the sites (G111 and G112) are at the top of the P4 helix just below the sheared A·A pairs in J4/5.

The only phylogenetically conserved sequence in the P1 helix is the G22·U-1 base pair at the cleavage site. The remaining positions maintain complementarity between the 5′-exon and the IGS, but the primary sequence is not conserved and can be mutated as long as base-pairing is retained (*42*). This lack of sequence conservation suggests that base-specific functional groups do not participate directly in tertiary interactions with the ribozyme catalytic core, which implies that $I\alpha S$ interference at these sites results from reduced secondary structural stability rather than the loss of a tertiary hydrogen bond to an N2-exocyclic amine. In a competitive environment where multiple ribozymes (50 nM) are vying for a slow reacting substrate ($k_{cat} = 0.015$ min^{-1}, $k_{off} > 0.1$ min^{-1}) that is present in limiting quantity (10 nM, 0.2 equiv), the inability of some ribozymes to form a stable duplex between the IGS and the substrate confers a significant disadvantage compared to other variants in the population. In contrast, interference was not observed with $m^2G\alpha S$ at the nonconserved positions in the P1 helix, because this analogue does not affect duplex stability (*41*). Therefore, interference from $I\alpha S$, but not from $m^2G\alpha S$, defines a biochemical signature to identify positions where secondary structural stability is critical for RNA function.

This pattern of interference is also seen within the P4 helix where nucleotides G111 and G112 showed weak interference with $I\alpha S$. This is a challenging region of the sequence to analyze experimentally by NAIM because it is difficult to resolve the cleavage products for the 3 consecutive Gs (G110−G112) that are located more than 300 nucleotides from the radiolabel at the 3′-terminus. Furthermore, interference at these sites appears to be highly dependent upon the reaction conditions. For example, $I\alpha S$ interference at G111 and G112 was not observed under slightly modified divalent metal conditions (such as 3 mM Mg^{2+} and 1 mM Mn^{2+}; M. Kronman and S. A. Strobel, unpublished results) nor was it observed in the 5′-exon ligation reaction (*8*). Nevertheless, the interference data with $I\alpha S$ and $m^2G\alpha S$ at 4 mM Mg^{2+} imply that ribozymes with a stable G112−C208 closing base pair in the P4 helix have a modest selective advantage over those with an $I\alpha S$112−C208 pair.

To directly test the importance of the P4 closing base pair, we mutated the G112−C208 pair to A−U in the context of the L-21 G414 ribozyme and measured the second-order rate constant for 3′-exon ligation under the same conditions as those used for the interference experiments (*16*). k_{cat} for the mutant enzyme was the same as the wild-type, but the K_m was 2.5-fold higher. The magnitude of the $I\alpha S$ interference suggested that the effect would be larger, but it is possible that the A−U mutation is not as destabilizing as the $I\alpha S$−C pair that occurs in the interference experiment. Thermodynamic measurements of model duplexes have shown that an I−C pair is in fact slightly less stable (0.3 kcal·mol^{-1}) than an A−U pair (*39*).

The G112−C208 pair is immediately below two consecutive sheared A·A pairs that act as a wobble receptor for the P1 helix (Figure 1) (*3*). Thermodynamic measurements of model duplexes containing consecutive A·A or G·A mismatches have demonstrated that the mispairs show a net stabilization from a G−C closing pair, but a net destabilization from an A−U closing pair (*51*, *52*). One dramatic consequence of this destabilization is that mutation of the closing base pair converts the G·A mispair from a sheared conformation to a imino hydrogen-bonded conformation (*53*, *54*). While the A·A pair cannot undergo this secondary structural transition, there is a precedent to argue that the stability of the P4 closing pair is important to stabilize the consecutively stacked sheared A·A pairs located immediately above it.

m^2GαS Interference Only at Sites of Direct Tertiary Hydrogen Bonding. Immediately adjacent to the sites of m^2GαS interference at G22 and G212 within the P1 and P4 helices, respectively, there are G−C pairs that do not show interference (Figures 1 and 5). These Gs participate in tertiary hydrogen bonds, but utilize functional groups other than their N2 amines. For example the 2′-OHs of the G23−C-2 pair each are thought to participate in a direct hydrogen bond; the 2′-OH of G23 bonds with the 2′-OH of C208 (*9*), and the 2′-OH of C-2 bonds with the N2 amine of G303 (*45*). Despite tremendous close packing within the minor groove face of this region of the molecule, interference was not observed with m^2GαS at G23. A similar pattern was observed in the P4 helix, where the 2′-OHs of G110, G212, and C109 all participate in hydrogen bonds to groups within the A-rich bulge. Nevertheless, m^2GαS interference was not observed in the G110−C211 base pair above the C109−G212 pair, nor was it observed in the G108−C213 base pair below G212. Thus, within the *Tetrahymena* ribozyme, m^2GαS interference is only observed at Gs that participate directly in tertiary hydrogen bonds via their exocyclic amines, not just at Gs that are closely packed within the tertiary structure.

All three of the sites of m^2GαS interference within the *Tetrahymena* intron are highly conserved and participate in tertiary hydrogen bonding interactions via their N2 amino groups. The sites of interference within this intron occurred at a G−C pair, a G·U pair, and an unpaired G. Given that the wide and shallow RNA minor groove is likely to be a common interface for RNA helix packing, and that there are relatively few reagents that probe this face of the helix, m^2GαS provides a valuable tool for RNA structure/function mapping in less well-characterized RNAs. This is particularly true if m^2GαS is used in combination with IαS, because together these analogues make it possible to differentiate sites of essential secondary structural stability from sites of tertiary hydrogen bonding.

ACKNOWLEDGMENT

We thank Charles Cheng for assistance with nucleotide synthesis, Rui Sousa for the gift of the Y639F mutant T7 polymerase clone, and Alexander Szewczak, Sean P. Ryder, and Juliane Strauss-Soukkup for critical comments on the manuscript.

REFERENCES

1. Strobel, S. A., and Doudna, J. A. (1997) *Trends Biochem. Sci. 22*, 262−266.
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. Strobel, S. A., Ortoleva-Donnelly, L., Ryder, S. P., Cate, J. H., and Moncoeur, E. (1998) *Nat. Struct. Biol. 5*, 60−66.
4. Litt, M., and Hancock, V. (1967) *Biochemistry 6*, 1848−1854.
5. Noller, H. F. (1974) *Biochemistry 13*, 4694−4703.
6. Gaur, R. K., and Krupp, G. (1993) *Nucleic Acids Res. 21*, 21−26.
7. Hardt, W. D., Erdmann, V. A., and Hartmann, R. K. (1996) *RNA 2*, 1189−1198.
8. Strobel, S. A., and Shetty, K. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 2903−2908.
9. Ortoleva-Donnelly, L., Szewczak, A. A., Gutell, R. R., and Strobel, S. A. (1998) *RNA 4*, 498−519.
10. Gish, G., and Eckstein, F. (1988) *Science 240*, 1520−1522.
11. Conrad, F., Hanne, A., Gaur, R. K., and Krupp, G. (1995) *Nucleic Acids Res. 23*, 1845−1853.
12. Cech, T. R. (1990) *Annu. Rev. Biochem. 59*, 543−568.
13. Woodson, S. A., and Cech, T. R. (1989) *Cell 57*, 335−345.
14. Green, R., Ellington, A. D., and Szostak, J. W. (1990) *Nature 347*, 406−408.
15. Beaudry, A. A., and Joyce, G. F. (1992) *Science 257*, 635−641.
16. Mei, R., and Herschlag, D. (1996) *Biochemistry 35*, 5796−5809.
17. 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.
18. Cate, J. H., Hanna, R. L., and Doudna, J. A. (1997) *Nat. Struct. Biol. 4*, 553−558.
19. Michel, F., and Westhof, E. (1990) *J. Mol. Biol. 216*, 585−610.
20. Lehnert, V., Jaeger, L., Michel, F., and Westhof, E. (1996) *Chem. Biol. 3*, 993−1009.
21. Robins, M. J., and Uznanski, B. (1981) *Can. J. Chem. 59*, 2608−2611.
22. Avino, A. M., Mayordomo, A., Espuny, R., Bach, M., and Eritja, R. (1995) *Nucleotides Nucleosides 14*, 1613−1617.
23. Rife, J., Cheng, C., Moore, P. B., and Strobel, S. A. (1998) *Nucleotides Nucleosides* (in press).
24. Sekine, M., and Satoh, T. (1991) *J. Org. Chem. 56*, 1224−1227.
25. Arabshahi, A., and Frey, P. A. (1994) *Biochem. Biophys. Res. Commun. 204*, 150−155.
26. Sousa, R., and Padilla, R. (1995) *EMBO J. 14*, 4609−4621.
27. Christian, E. L., and Yarus, M. (1992) *J. Mol. Biol. 228*, 743−758.
28. Lingner, J., and Keller, W. (1993) *Nucleic Acids Res. 21*, 2917−2920.
29. Steitz, T. A., Beese, L., Freemont, P. S., Friedman, J. M., and Sanderson, M. R. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 465−471.
30. Huang, Y., Eckstein, F., Padilla, R., and Sousa, R. (1997) *Biochemistry 36*, 8231−8242.
31. Huang, Y., Beaudry, A., McSwiggen, J., and Sousa, R. (1997) *Biochemistry 36*, 13718−13728.
32. Raines, K., and Gottlieb, P. A. (1998) *RNA 4*, 340−345.
33. Damberger, S. H., and Gutell, R. R. (1994) *Nucleic Acids Res. 22*, 3508−3510.
34. Michel, F., Hanna, M., Green, R., Bartel, D. P., and Szostak, J. W. (1989) *Nature 342*, 391−395.
35. Grabowski, P. J., Zaug, A. J., and Cech, T. R. (1981) *Cell 23*, 467−476.
36. Doudna, J. A., Cormack, B. P., and Szostak, J. W. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 7402−7406.
37. Barfod, E. T., and Cech, T. R. (1989) *Mol. Cell. Biol. 9*, 3657−3666.
38. Strobel, S. A., and Cech, T. R. (1996) *Biochemistry 35*, 1201−1211.
39. Turner, D. H., Sugimoto, N., Kierzek, R., and Dreiker, S. D. (1987) *J. Am. Chem. Soc. 109*, 3783−3785.
40. Strobel, S. A., Cech, T. R., Usman, N., and Beigelman, L. (1994) *Biochemistry 33*, 13824−13853.
41. Rife, J., Cheng, C., Moore, P. B., and Strobel, S. A. (1998) *Nucleic Acids Res.*, In press.
42. Murphy, F. L., and Cech, T. R. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9218−9222.
43. Strobel, S. A., and Cech, T. R. (1995) *Science 267*, 675−679.
44. Engel, J. D., and von Hippel, P. H. (1974) *Biochemistry 13*, 4143−4158.
45. Szewczak, A. A., Ortoleva-Donnelly, L., Ryder, S. P., and Strobel, S. A. (1998) (submitted for publication).
46. Waring, R. B., Towner, P., Minter, S. J., and Davies, R. W. (1986) *Nature 321*, 133−139.

47. Strobel, S. A., and Cech, T. R. (1993) *Biochemistry 32*, 13593−13604.
48. Wang, J.-F., Downs, W. D., and Cech, T. R. (1993) *Science 260*, 504−508.
49. Flor, P. J., Flanegan, J. B., and Cech, T. R. (1989) *EMBO J. 8*, 3391−3399.
50. Laggerbauer, B., Murphy, F. L., and Cech, T. R. (1994) *EMBO J. 13*, 2669−2676.
51. Walter, A. E., Wu, M., and Turner, D. H. (1994) *Biochemistry 33*, 11349−11354.
52. Wu, M., McDowell, J. A., and Turner, D. H. (1995) *Biochemistry 34*, 3204−3211.
53. SantaLucia, J., and Turner, D. H. (1993) *Biochemistry 32*, 12612−12623.
54. Wu, M., and Turner, D. H. (1996) *Biochemistry 35*, 9677−9689.

BI980723J