The chemical basis of adenosine conservation throughout the *Tetrahymena* ribozyme

LORI ORTOLEVA-DONNELLY,¹ ALEXANDER A. SZEWCZAK,¹
ROBIN R. GUTELL,² and SCOTT A. STROBEL¹

¹Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA

²Department of Chemistry and Biochemistry, Campus Box 215, University of Colorado,
Boulder, Colorado 80309-0215, USA

ABSTRACT

Adenosines are present at a disproportionately high frequency within several RNA structural motifs. To explore the importance of individual adenosine functional groups for group I intron activity, we performed Nucleotide Analog Interference Mapping (NAIM) with a collection of adenosine analogues. This paper reports the synthesis, transcriptional incorporation, and the observed interference pattern throughout the *Tetrahymena* group I intron for eight adenosine derivatives tagged with an α -phosphorothioate linkage for use in NAIM. All of the analogues were accurately incorporated into the transcript as an A. The sites that interfere with the 3'-exon ligation reaction of the *Tetrahymena* intron are coincident with the sites of phylogenetic conservation, yet the interference patterns for each analogue are different. These interference data provide several biochemical constraints that improve our understanding of the *Tetrahymena* ribozyme structure. For example, the data support an essential A-platform within the J6/6a region, major groove packing of the P3 and P7 helices, minor groove packing of the P3 and J4/5 helices, and an axial model for binding of the guanosine cofactor. The data also identify several essential functional groups within a highly conserved single-stranded region in the core of the intron (J8/7). At four sites in the intron, interference was observed with 2'-fluoro A, but not with 2'-deoxy A. Based upon comparison with the P4-P6 crystal structure, this may provide a biochemical signature for nucleotide positions where the ribose sugar adopts an essential C2'-endo conformation. In other cases where there is interference with 2'-deoxy A, the presence or absence of 2'-fluoro A interference helps to establish whether the 2'-OH acts as a hydrogen bond donor or acceptor. Mapping of the *Tetrahymena* intron establishes a basis set of information that will allow these reagents to be used with confidence in systems that are less well understood.

Keywords: A-platform; G binding; Group I intron; interference mapping; RNA structure

INTRODUCTION

The *Tetrahymena* group I intron is one of several large catalytically active RNAs that folds into a compact globular structure (Fig. 1; Cech, 1993). The intron catalyzes two transesterification reactions in the course of RNA self-splicing (Cech et al., 1992). The first step consists of nucleophilic attack at the 5' splice site by the 3'-OH of an exogenous guanosine cofactor. In the second step, the 5'-exon attacks the 3' splice site to produce ligated exons. Under the appropriate conditions, the intron can also catalyze the reverse of either of these two reactions, resulting in exon ligation back onto the intron (Fig. 2A; Woodson & Cech, 1989; Green et al., 1990; Beaudry & Joyce, 1992). To understand the

reaction specificity and catalytic rate enhancement achieved by this ribozyme, it is necessary to understand the structural basis of intron function. Improved atomic resolution biochemical methods are needed to identify the specific chemical groups within the intron that are essential to its activity and, as a consequence, perhaps identify unique biochemical signatures for specific RNA structural motifs.

The recently reported crystal structure of the P4-P6 domain of the *Tetrahymena* group I intron has provided tremendous insight into the secondary and tertiary structures involved in RNA folding (Cate & Doudna, 1996; Cate et al., 1996a, 1996b; Cate et al., 1997). Many of the structural motifs important for domain folding involve adenosine-rich sequences. These motifs are likely to be present in a wide variety of RNA structures. One example is the adenosine-platform, which is observed at three different positions within the domain (Fig. 1;

Reprint requests to: Scott Strobel, Department of Molecular Biophysics and Biochemistry, 260 Whitney Ave., Yale University, New Haven, Connecticut 06520, USA; e-mail: strobel@csb.yale.edu.

FIGURE 1. Primary sequence and secondary structure of the L-21 G414 version of the *Tetrahymena* group I intron (Cech et al., 1994). This ribozyme binds the oligonucleotide CCCUCdTAAAA and transfers the AAAAA onto the 3' end of the intron in a reaction analogous to the reverse of the second step of splicing. Numbering of the nucleotides discussed in the text is shown, as are the names of the helical (P1–P9) and single-stranded regions (J6/6a, J8/7, etc.) of the RNA. The three A-platforms in P4–P6 are shown as adjacent A's with a heavy underline. The long thin lines indicate regions known to make tertiary interactions within the three-dimensional structure. Thick lines designate connectivity of the RNA strand.

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 (Beaudry & Joyce, 1992; Mei & Herschlag, 1996). 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'-terminus onto the 3' end of the intron. The equilibrium constant for the chemical step of this reversible reaction is approximately 1 (Mei & Herschlag, 1996). **B:** Scheme for the identification of the chemical groups important for RNA activity by NAIM (Strobel & Shetty, 1997). The phosphorothioate-tagged nucleotide analogue (indicated as $\delta\alpha$ S) is randomly incorporated into the transcript in place of A. If the analogue 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 the analogue disrupts activity (right side), then the subset of ribozymes that have the analogue incorporated at that 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 produces a sequencing ladder with gaps that correspond to sites intolerant of analogue substitution. A α S serves as a control to insure 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 analogue incorporation at a given site (not shown).

Cate et al., 1996b). This motif involves a side by side alignment of two consecutive A's to form a pseudo-base pair that serves as a platform for tertiary stacking interactions. In one of the three occurrences, the

A-platform makes tertiary interactions with another example of an A-rich motif, the GAAA tetraloop frequently found at the end of RNA hairpin loops (Woese et al., 1990). The A's in this and related GNRA tetraloops

appear to be widely utilized in tertiary interactions (Murphy & Cech, 1994; Pley et al., 1994; Costa & Michel, 1995; Costa et al., 1997). A third example of an A-rich motif in the P4-P6 structure is the asymmetric A-rich bulge (AAUAA; positions 183–187; Fig. 1), which makes extensive tertiary interactions with the minor groove of the P4 helix (Cate et al., 1996a). The A-rich bulge also has a corkscrew turn in the phosphate backbone that forms the binding site for two divalent metal ions (Cate et al., 1997).

The importance of A nucleotides among this wide variety of structural motifs explains an observation made several years ago about the conservation and sequence distribution of A's within ribosomal RNA phylogeny. A simple analysis of the 16S rRNA secondary structure revealed that there is a bias for unpaired A's (Gutell et al., 1985). More than 60% of the A's are unpaired, whereas only 30% of the G, C, and U nucleotides were unpaired. Among the unpaired A's, there is also a preference for the A's to occur in adjacent positions. Furthermore, of the universally conserved nucleotides, more than 80% of the A's were unpaired, whereas only 50% of the U, G, and C's were unpaired. At the time of this analysis (1985), there were no structural or functional explanations for this strong bias in favor of unpaired A's. We now know that there are a number of structural motifs that utilize A's and require this pattern of sequence conservation.

Several chemical reagents are available that covalently modify functional groups on A and other nucleotides (Stern et al., 1989). For example, dimethylsulfate (DMS) is used to explore the role of the N1 of A (von Ahsen & Noller, 1993) and diethylpyrocarbonate (DEPC) modifies A at the N7 position (Peattie, 1979). These reagents have proven valuable in biochemical structural and functional studies of RNA (Inoue & Cech, 1985; Moazed et al., 1986; von Ahsen & Noller, 1993; Butcher & Burke, 1994; Beattie et al., 1995). However, these chemical reagents have some severe limitations. They leave several functional groups unmodified, and therefore untested. They also probe the RNA by adding steric bulk to the nucleotide, which may not correlate with involvement of the unmodified functional group in the RNA structure.

Given the importance of A-rich motifs within the structure of the *Tetrahymena* intron and the prevalence of A residues in the unpaired regions of several different RNA molecules, we focused on developing improved biochemical methods to analyze these sequences. Nucleotide Analog Interference Mapping (NAIM) is an efficient biochemical approach to identify individual functional groups and tertiary hydrogen bonds essential for RNA activity (Fig. 2B; Strobel & Shetty, 1997). In this approach, a nucleotide analogue that contains an alteration of one functional group is covalently tagged with a 5'-phosphorothioate linkage and randomly incorporated into an RNA transcript by T7 RNA polymer-

ase at a level of approximately 5%. The RNA transcripts are then separated into active and inactive fractions based upon their ability to perform a defined function. For the group I intron experiments described here, this function is the ability to perform the 3'-exon ligation reaction, which is analogous to the reverse of the second step of splicing (Fig. 2A; Beaudry & Joyce, 1992; Mei & Herschlag, 1996). Cleavage of the phosphorothioate linkages with iodine (Gish & Eckstein, 1988) and resolution of the cleavage products by PAGE yields a sequencing ladder with gaps that correspond to sites where analogue substitution is detrimental to activity. This interference approach makes it possible to simultaneously, yet individually, test the importance of specific functional groups at every position within an RNA.

Our initial efforts with this chemogenetic approach tested the role of the N2 amine of G using the phosphorothioate-tagged derivative of inosine as the nucleotide analogue (Strobel & Shetty, 1997). Other groups have tested the importance of the 2'-OH using 2'-deoxy and 2'-O-methoxy derivatives of A, C, G, and U (Gaur & Krupp, 1993; Conrad et al., 1995; Hardt et al., 1996). In principle, NAIM can be extended to any nucleotide analogue that can be incorporated into an RNA by *in vitro* transcription. In this paper, we utilize eight derivatives of adenosine that each make an incremental change in either a nucleobase or ribosyl functional group (Fig. 3). We report the synthesis, incorporation, and sites of interference for each analogue within the *Tetrahymena* ribozyme. These interference data provide several biochemical constraints for modeling the intron active site. They also confirm the biochemical relevance of some of the A-rich motifs within the P4-P6 domain (Cate et al., 1996a, 1996b). The data indicate that some of the tertiary interactions proposed within the phylogenetic model of the group I intron require revision (Michel & Westhof, 1990; Lehnert et al., 1996). Given the importance and diversity of A-rich folding motifs, these A analogues are likely to be valuable reagents for investigating the relationship of structure and function within a variety of RNAs.

RESULTS AND DISCUSSION

Eight nucleotide analogues of adenosine were selected for use in NAIM (Fig. 3). In this collection, three analogues (^{NMe}AαS, PurαS, and 2APαS; abbreviations defined in Fig. 3) were used to examine the role of the N6 amino group. These nucleotides make it possible to explore the requirement for the N6 position by deleting the amine (2APαS and PurαS) or by replacing one proton of the amine with a methyl group (^{NMe}AαS). Interpretation of ^{NMe}AαS interference is somewhat complicated because the methyl group can adopt either the *s-cis* or *s-trans* rotamer. In the context of a nucleoside free in solution, the N6 methyl group prefers the *s-cis* rotamer 20:1 (Engel & von Hippel, 1974); however, in

FIGURE 3. Eight nucleotide analogues used for NAIM in this study. Each analogue was synthesized as the 5'-O-(1-thio)nucleoside triphosphate (Eckstein, 1985; Arabshahi & Frey, 1994). Each analogue is shown as the monophosphate derivative, the form in which it is incorporated during transcription. Functional group(s) modified relative to the parental A nucleotide (box) are identified by a shadowed box. Numbering of the adenosine rings is indicated on the parental nucleotide, and the abbreviation used in the text for each analogue is listed in parentheses.

the context of a DNA duplex, the methyl group is *s-trans* and occupies the major groove face of the helix (Fazakerley et al., 1985; Lingbeck et al., 1996). Despite the *s-cis* rotameric preference, ^{NMe}A substitution destabilizes a duplex by about 1 kcal·mol⁻¹ (Engel & von Hippel, 1978; Guo et al., 1995), which makes it an analogue suitable for probing the helical major groove for tertiary interactions.

Two of the analogues (DAPαS and 2APαS) were used to determine the compatibility of an additional N2 amino group on the minor groove face of the nucleotide. This is a functional group present only on G, and therefore interference with these two nucleotides indicates sites that should be incompatible with G mutation. A fifth analogue (7dAαS) was designed to test the role of the N7 imino group of A in the major groove by changing the N7 from a hydrogen bond acceptor to C-H. We also attempted to prepare 3-deaza-adenosine, but several attempts to synthesize the triphosphate from commercially available unprotected nucleoside were unsuccessful.

Three analogues (^{OMe}AαS, ^FAαS, and dAαS) were used to differentiate the contribution made by the 2'-OH. dAαS and ^FAαS both delete the 2'-OH, but replace it with atoms vastly different in electronegativity. dAαS

replaces the 2'-OH with a proton, whereas ^FAαS has a fluorine at the 2'-position. dAαS can neither donate nor accept a hydrogen bond, but the 2'-fluoro group of ^FAαS can still act as a hydrogen bond acceptor (Withers et al., 1988; Herschlag et al., 1993a). ^{OMe}AαS can also act as a hydrogen bond acceptor, but only if there is sufficient space to accommodate the additional steric bulk of the methyl group. In addition to the effects on the hydrogen bonding character of the 2'-position, these analogues have an indirect effect on the sugar pucker of the ribose ring (Guschlbauer & Jankowski, 1980). The more electronegative the substituent, the more the C3'-endo sugar conformation is preferred (Uesugi et al., 1979). The propensity toward C3'-endo sugar pucker is ^FAαS > AαS > dAαS.

Analogue incorporation: Efficiency and accuracy

Each of the analogues was synthesized as the 5'-O-(1-thio)nucleoside triphosphate derivative using the procedure described by Arabshahi and Frey (1994). The triphosphate facilitates incorporation of the analogue into a growing RNA transcript, and the phosphorothio-

ate serves as a chemical tag to identify the sites of analogue incorporation.

Prior to their use in NAIM, we determined the efficiency and accuracy of analogue incorporation throughout the RNA transcript. Each of the triphosphates was randomly incorporated into the L-21 G414 form of the group I intron (Table 1; Strobel & Shetty, 1997). The resulting RNAs were 5' end-labeled and the phosphorothioate linkages were cut with iodine (Fig. 4A). All the analogues were incorporated exclusively as an A, however, the relative incorporation efficiency at each position varied somewhat between the analogues. Efficient analogue incorporation was observed for the nucleotides DAP α S, 7dA α S, and ^{NM}eA α S. The intensity of analogue incorporation at each site throughout the molecule was approximately equivalent to A α S (Fig. 4A, lanes 1–4). These three analogues do not alter any of the functional groups necessary for Watson–Crick base pairing. The low concentration of DAPTP α S needed to achieve 5% incorporation (25 μ M DAPTP α S and 1 mM ATP, Table 1) suggests that the additional N2 amino group makes a positive contribution by hydrogen bonding to the O2 of T during transcription (Strobel et al., 1994). The fact that ^{NM}eA α S is incorporated efficiently throughout the RNA transcript provides additional evidence that the N6 methyl group can occupy the major groove face of the nucleotide.

Somewhat surprisingly, Pur α S was also accurately incorporated into the RNA, although the efficiency of incorporation was quite uneven at several positions in the transcript (Fig. 4A, lane 5). Pur α S lacks the N6 amino group and can form only one hydrogen bond to the opposing T during transcription. As a result of the poor hydrogen bonding of Pur α S with T, high ratios of PurTP α S to ATP (2:0.5 mM) were required to achieve efficient incorporation and, even then, Pur α S was not

incorporated at all positions within consecutive runs of A's. At four positions within the transcript (A103–A105, A171–A173, A218–A219, A268–A270), Pur α S was incorporated efficiently at the first (or 5'-most) A in the series, but it was not incorporated at the subsequent A's (Fig. 4A, lane 5). However, this was not the case at all consecutive A's because normal incorporation levels were observed at each position within several other runs of A (e.g., A64–A66, A87–A90, and A151–A153). It is unclear how sequence context affects incorporation at these sites. Independent of the cause, seven positions within the L-21 G414 sequence (A104, A105, A172, A173, A219, A269, and A270) were uninformative for Pur α S due to lack of incorporation.

2AP α S, like Pur α S, omits the N6 amine found on A; however, 2AP α S has an additional N2 amino group that might hydrogen bond to the O2 of T during transcription. We found that 2AP α S incorporated slightly better than Pur α S. Weak, but detectable incorporation was found at all seven of the sites that were uninformative for Pur α S (Fig. 4A, lane 12). Thus, it was possible to obtain information about the N6 amine at all the resolvable sites within the molecule. Interpretation of the 2AP α S interference data is complicated by the simultaneous changes at both the N2 and N6 positions within the nucleotide. In order to make conclusions about the importance of the N6 amine, effects from 2AP α S must be compared to both A α S (phosphorothioate effects) and DAP α S (N2 amine effects) controls. Experiments with 2AP α S were further complicated because RNAs containing this analogue were unstable and had a half-life of only a few days in the freezer, which made it necessary to transcribe the RNA containing 2AP α S prior to each experiment.

Use of the T7 RNA polymerase mutant Y639F for efficient analogue incorporation

Previous work with 2'-deoxy and 2'-methoxy derivatives reported reasonably efficient analogue incorporation into tRNA using the wild-type version of T7 RNA polymerase with high analogue to NTP ratios, and transcription buffers containing Mn²⁺ (Conrad et al., 1995). Unfortunately, we were unable to incorporate efficiently (<1%) dA α S, ^{OMe}eA α S, or ^FA α S into the L-21 G414 RNA using these transcription conditions (data not shown). Given that the 2'-hydroxyl is likely to be used widely in RNA folding and recognition, it is imperative that a procedure be developed for the efficient incorporation of 2'-substituted analogues. Sousa and Padilla (1995) have reported that a Y639F point mutant of the T7 RNA polymerase efficiently incorporates 2'-deoxynucleotides into what would otherwise be an RNA polymer. Subsequent work demonstrated that this polymerase can incorporate 2'-methoxy analogues in the presence of Mn²⁺, and that it can also incorporate 2'-deoxy-2'-fluoro-NTPs, and 2'-deoxy-2'-thio-CTP (Huang

TABLE 1. Concentration of ATP and analogue used to achieve approximately 5% analogue incorporation during RNA transcription.^a

Analogue δ TP α S	[δ TP α S] (mM)	[ATP] (mM)	Polymerase (wt or Y639F)
A α S (S _P isomer)	0.05	1.0	wt
7dA α S	1.0	1.0	wt
^{NM} eA α S	0.4	1.0	wt
DAP α S	0.025	1.0	wt
Pur α S	2.0	0.5	wt
2AP α S	2.0	0.5	wt
dA α S	1.5	1.0	Y639F
^{OMe} eA α S	2.0	0.1	Y639F
^F A α S	1.0	1.0	Y639F

^aIn all cases, 1 mM CTP, UTP, and GTP, 40 mM Tris-HCl, pH 7.5, 4 mM spermidine, 10 mM DTT, 15 mM MgCl₂, 0.05% Triton X-100, and 0.05 μ g/ μ L DNA template were included in the reaction. The A α S incorporation used the S_P isomer of ATP α S, whereas all other analogues were transcribed from the diastereomeric mixture of the S_P and R_P isomers. The polymerase used was either the wild-type version of T7 RNA polymerase or a version of the polymerase containing the Y639F point mutation (Sousa & Padilla, 1995).

FIGURE 4. Analogue incorporation and interference reactions. **A:** The L-21 G414 5' end-labeled control showing the extent and positions of analogue incorporation throughout the intron. The I_2 -treated $A\alpha S$ standard is shown in lanes 1 and 11. The phosphorothioate-tagged analogue incorporated into the other RNAs is listed above the lane numbers. Nucleotide numbers corresponding to several of the bands are marked to the left of each gel. Addition (lanes 1–5, 11–15) or omission (lanes 6–10, 16–20) of iodine is indicated. This particular gel was electrophoresed at 75 watts for 1.25 h. Longer electrophoretic times were used to improve the signal resolution of the nucleotides toward the 3' end of the RNA (not shown). **B:** 3'-Exon ligation reaction of L-21 G414 RNA with dT(–1)S. This autoradiogram reveals the sites of analogue interference throughout the intron. A complete description of the reaction conditions is included in Materials and Methods. Figure legends are the same as those in Figure 4A. This particular gel was electrophoresed at 75 watts for 2.25 h. It provides maximal resolution of the J8/7 region of the intron (nt 308–299). Sites of strong interference within this region are indicated with an asterisk. A302 and A306 were uninformative under these assay conditions due to a strong phosphorothioate effect at both positions. Longer electrophoretic times were used to resolve the cleavage products toward the 5' end of the intron (not shown). Interference results for 2AP αS are not shown on this autoradiogram, and the no-iodine controls (lanes 9–13) are only shown for a subset of the nucleotide analogues. The no-iodine controls for the remaining nucleotides were essentially identical. **C:** 3'-Exon ligation reaction of L-21 G414 RNA with rT(–1)S. Only the cleavage products from nucleotides surrounding the J8/7 region of this autoradiogram are shown. Unlike the reaction conditions in Figure 4B, the use of rT(–1)S in the presence of Mn^{2+} made it possible to gain information about nt A302 and A306. Sites of interference are marked with an asterisk. 2AP αS interference was not measured at these two sites because DAP αS and Pur αS were fully informative.

et al., 1997a, 1997b; Raines & Gottlieb, 1998). Given these interesting properties, we overexpressed and purified the Y639F polymerase for use in NAIM.

The mutant polymerase was sufficiently accurate that RNAs transcribed with Y639F were as active as those transcribed with the wild-type form of the polymerase (data not shown). Furthermore, all three of the 2'-substituted nucleotides (dA αS , $^{OMe}A\alpha S$, and $^FA\alpha S$) were accurately and efficiently incorporated into the L-21 G414 as defined by 5' end-labeling and analysis of the

transcripts (Fig. 4A, lanes 11, 13–15). No additional bands were detected in the phosphorothioate sequencing analysis. This confirms that, although the Y639F polymerase lacks fidelity with regard to the identity of the ribose sugar, it is not excessively prone to introducing mutations during transcription (Sousa & Padilla, 1995).

Using this polymerase, dA αS and $^FA\alpha S$ were evenly incorporated at modest ratios of phosphorothioate to ATP (Table 1). However, even with this mutant poly-

merase, high ratios of $^{OMe}ATP_{\alpha S}$ to ATP were necessary to obtain close to 5% analogue incorporation (Table 1). The efficiency of $^{OMe}A_{\alpha S}$ incorporation was uneven throughout the transcript, but there was at least some incorporation at every A position (Fig. 4A, lane 14). In addition to the 2'-substituted analogues, the Y639F polymerase incorporates other nucleotides that contain modifications in the minor groove (L. Ortoleva-Donnelly & S.A. Strobel, unpubl. obs.), which makes it a valuable reagent for NAIM.

Phosphorothioate interference

Prior to performing NAIM with a complete series of analogues, it was necessary to identify a reaction condition that was selective, but had a minimum number of uninformative sites due to strong phosphorothioate interference. Previous experiments with the *Tetrahymena* group I intron have shown that $A_{\alpha S}$ incorporation can inhibit splicing activity (Deeney et al., 1987), although the number and location of the detrimental sites varied with the experimental conditions (e.g., temperature, salt concentrations, incubation time), and the category of ribozyme reaction being studied (3' splice site hydrolysis, 3' splicing by CU addition, or 5'-exon cleavage by G; Waring, 1989; Christian & Yarus, 1992, 1993). Some of the apparent differences between these experimental results can be explained by experiments that allowed the splicing reaction to proceed too far to completion, which reduced the experimental signal (Waring, 1989), or by experiments that used primer extension to identify the sites of interference, which introduced excessive background noise into the data (Christian & Yarus, 1993). Nevertheless, there are some real, and possibly significant, differences in the phosphorothioate interference patterns observed for the first versus the second step of splicing, although it is still difficult to make conclusions about the importance of these differences.

For ease and efficiency of experimental analysis, we elected to study the 3'-exon ligation reaction (Beaudry & Joyce, 1992; Mei & Herschlag, 1996). This reaction, wherein the 3'-OH of the terminal G (G414) nucleophilically attacks an oligonucleotide substrate that mimics the 5'-3' ligated exons, is analogous to the reverse of the second step of splicing. The reaction transfers the 3'-exon onto the 3' end of the RNA, and presents three very important advantages over previous efforts to map sites of phosphorothioate interference in the *Tetrahymena* group I intron. (1) Using a 3'-radiolabeled substrate, the active molecules in the ribozyme population become radioactively labeled during the ligation reaction. No additional purification of the RNA is necessary prior to iodine treatment and gel electrophoresis, which makes it possible to directly visualize the interference pattern without using reverse transcriptase. (2) Ease of the reaction makes it feasible to com-

pare interference patterns under a variety of reaction conditions. (3) Unlike the self-splicing reactions studied previously, the substrate in the reaction is a synthetic oligonucleotide that can be altered chemically to adjust the selectivity of the reaction. We found the reaction to be maximally informative using a substrate with a 2'-deoxy substitution at the cleavage site [dT(-1)S: CCCUCdTAAAAA] in a reaction buffer containing 3 mM $MgCl_2$ and 1 mM $Mn(OAc)_2$. The 2'-deoxy substitution reduces the rate of chemistry by more than 1,000-fold (Herschlag et al., 1993b), which slows the reaction sufficiently that more subtle effects on activity can be detected. The low metal concentration partially destabilizes the structure of the ribozyme, but the presence of a small amount of the thiophilic manganese ion minimizes the phosphorothioate effects that are present at several positions throughout the intron (Christian & Yarus, 1993).

Under these conditions, there are 10 sites of $A_{\alpha S}$ interference in the *Tetrahymena* group I intron for the 3'-exon ligation reaction (Fig. 4B, lane 1). Most of these map within the conserved catalytic core of the RNA. Moderate interference was observed at 5 of the 10 sites (A57, A97, A207, A210, and A301), strong interference was detected at 3 sites (A263, A304, and A308), and complete interference was seen at 2 sites (A302 and A306; Fig. 4B, lane 1). Interference at A302 and A306 could be partially rescued using an oligonucleotide substrate with a ribose at the cleavage site [rT(-1)S: CCCUCUAAAAA], but only in the presence of Mn^{2+} (Fig. 4C, lane 1). Complete interference was observed at both of these sites using a buffer containing 4 mM $MgCl_2$ even with a ribose substrate (data not shown). The presence of 1 mM Mn^{2+} also partially or completely rescued the phosphorothioate effects at A57, A97, A113, A114, A115, A206, A207, and A210 that were observed using a 4 mM Mg^{2+} buffer. Strong phosphorothioate effects that can be rescued by Mn^{2+} have been interpreted to be sites of divalent metal ion binding within the intron (Christian & Yarus, 1993). Of particular importance to this experiment is that every A within the molecule has an $A_{\alpha S}$ cleavage signal, so all resolvable sites are informative for NAIM.

Nucleotide analogue interference quantitation

The interference pattern of each of the eight nucleotide analogues was determined for 3'-exon ligation (Fig. 4B). For technical reasons, only 105 of the 115 A's within the L-21 G414 sequence were informative in this assay. Nonspecific cleavage was observed consistently at A290 in the control lane lacking iodine (Fig. 4B). This nucleotide is in the P8 hairpin loop and is not essential to ribozyme activity (Doudna et al., 1991; Nakamura et al., 1995). In addition, positions close to the 5' and 3' ends of L-21 G414 could not be analyzed because the cleavage products were not sufficiently

resolved from the full-length intron. This included seven nucleotides at the 5' end and two nucleotides at the 3' end of the intron. Because of phosphorothioate effects, A302 and A306 were only informative using the rT(−1)S substrate in the presence of Mn^{2+} , so interference at these two sites was measured separately from the rest of the intron (Fig. 4C).

Based upon the band intensities in the 3'-exon ligation and 5' end-labeled control experiments, an interference κ value was calculated for each A position in the intron (see Materials and Methods). 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. As might be expected, most positions did not show any effect upon analogue substitution. Greater than 90% of the interference κ values were between 0.67 and 1.5. This data range provides a conservative estimate of the experimental noise in the system. In the data range from 1.5 to 2.0 (or, for the case of enhancement, from 0.67 to 0.5), there were subtle but reproducible effects. We have chosen to be conservative in our interpretation of the data and will consider only interference values greater than 2 (or less than 0.5) to be significant. Three-dimensional histograms plotting the magnitude of the interference κ values for each analogue at each position are shown in Figure 5. Every one of the analogues has a unique interference pattern throughout the intron with regard to both the distribution and the intensity of interference.

Sites of interference are largely coincident with sites of phylogenetic conservation

The interference pattern for this series of eight A analogues provides valuable biochemical information about the structure and function of the group I catalytic RNA. A composite view of the eight interference patterns reveals that the sites of interference map primarily onto two regions at the core of the ribozyme, P7-P3-P8 and

helices P4 and P6 (Fig. 6A). Strong interference was also detected at the junction between the P2 and P2.1 helices, which is known to be essential for tethering the P1 helix into the active site of the *Tetrahymena* intron (Downs & Cech, 1990, 1994). The sites of analogue interference correlate almost exactly with the most highly conserved A nucleotide positions among the group IC1 and IC2 introns (Fig. 6B; Michel & Westhof, 1990; Damberger & Gutell, 1994). Of the 21 conserved A nucleotides in the core region of the intron, 19 show interference with at least one of the analogues. This includes several positions within the P4-P6 and P7-P3-P8 helices that were demonstrated by mutagenesis to be essential for catalytic function (Couture et al., 1990; Pyle et al., 1992). In contrast, there is no interference within the P8 helix and the P9 extension. Although both of these structural elements are important for intron stability, there is very little primary sequence conservation within these regions of the intron (Fig. 6B). Because high sequence conservation implies that the nucleotide is structurally or functionally important, the coincidence between sites of interference and phylogenetic conservation provides strong validation for NAIM as a method for the biochemical characterization of RNA.

Conserved positions that do not show interference

There were a few exceptions to the correlation between phylogeny and interference. Most of the divergence occurred within the P5abc subdomain, where the only site of even modest interference was from $^{OMe}A\alpha S$ substitution at A183 (Fig. 5). Interference at this site is in agreement with a tertiary hydrogen bond observed between the 2'-OH of A183 and the 2'-OH of G110 in the P4-P6 crystal structure (Cate et al., 1996a). It is noteworthy that P5abc is not conserved among all the subclasses of group I introns (the data plotted in Fig. 6B is only for the IC1 and IC2 introns; Michel & Westhof, 1990; Damberger & Gutell, 1994), and muta-

FIGURE 5. Individual histograms plotting the magnitude of the interference κ value versus nucleotide position superimposed on the intron secondary structure for $A\alpha S$ and each of the eight analogues tested. Interference κ values ≥ 2.0 are shown as gray bars. Values greater than 6 are assigned a magnitude of 6 within this graph. A white bar (of which there is only one example at A207 for $^FA\alpha S$) indicates that κ was ≤ 0.5 . In this single case, the magnitude of the bar corresponds to $1/\kappa$, and indicates that there is enhancement of activity due to analogue substitution at that site. White boxes indicate positions that were uninformative in the assay because they are a nucleotide other than A. Gray boxes indicate A sites that were not informative in the assay due to incomplete resolution of the cleavage products on the sequencing gel (all analogues were uninformative at A24, A28, A29, A30, A31, A35, A38, A407, and A410), lack of analogue incorporation at that site ($Pur\alpha S$ was uninformative at A104, A105, A172, A173, A219, A269, and A270), degradation in the no iodine control (all analogues were uninformative at A290), or not measuring interference at that site ($2AP\alpha S$ was not tested at A302 and A306). The nucleotide number within the *Tetrahymena* sequence for each site of interference is shown adjacent to the bar. The error in the κ value at each position is $\leq 20\%$. The value is the average of at least two and as many as eight independent experimental measurements. A302 and A306 were assayed under conditions different from the rest of the intron due to complete phosphorothioate inhibition at these two sites under the standard reaction conditions.

FIGURE 5. (Legend on facing page.)

FIGURE 6. **A:** Composite histogram of the interference pattern throughout the *Tetrahymena* group I intron using the data presented in Figure 5. At each nucleotide position, the maximum interference observed from any of the nine analogues is shown as a bar whose height corresponds to the magnitude of the interference κ value. The symbols in the figure are the same as in Figure 5. This graph demonstrates that nearly all the interference sites cluster into the P4-P6 helices or the P7-P3-P8 subdomain. **B:** Composite histogram showing the relationship between the sites of nucleotide analogue interference and the sites of phylogenetic sequence conservation. The height of the bar at each position is proportional to the conservation value calculated from the modified Shannon equation (Damberger & Gutell, 1994; R.R. Gutell, unpubl. results) using a sequence alignment of 131 group IC1 and IC2 introns. In this calculation, a value of 2.0 indicates a site that is completely invariant, a value of 1.5 is approximately 90% conserved, a value near 1.0 is about 60–80% conserved, and a value approaching 0.5 is about 40–50% conserved. Negative values indicate a complete lack of conservation at a particular position. Only conservation values between 2.0 and 0.5 for A nucleotides within the sequence are plotted to show the positions that are most conserved among this class of introns. Values less than 0.5 are shown as black boxes. White boxes are nucleotides in the sequence other than A's. The color of the bars indicate the extent of interference at a given position. A black bar indicates that at least one of the nine analogues had an interference κ value greater than 3.5, a gray bar indicates an interference κ value between 1.8 and 3.5. A white bar indicates that none of the analogues had an interference κ value above 1.8. Asterisks show the three sites that are not conserved but still demonstrated interference. This graph demonstrates that there is a strong correlation between the sites of interference and the sites of sequence conservation. The most notable exception is within the P5abc domain, where there is moderate conservation, but where no interference was detected.

tions that disrupt P4-P6 domain folding have no apparent effect upon folding of the intact intron (Laggerbauer et al., 1994). In fact, several of these mutations actually improved the intron folding rates (D.K. Treiber & J.R. Williamson, unpubl. results). Furthermore, the J6a/6b region (A225, A226, and A248), which participates in an extensive tertiary interaction with the GAAA tetraloop of P5b (Cate et al., 1996a, 1996b), was unaffected by analogue substitution. Apparently, the P4-P6 domain is sufficiently stable in the context of the complete intron that single functional group modifications are not enough to disturb activity. By contrast, we have observed strong interference at several of these positions using a gel shift assay for folding of the P4-P6 domain (S. Basu & S.A. Strobel, unpubl. results; Murphy & Cech, 1993).

Two conserved positions within the intron core region (A214 and A268) did not show interference with at least one of the eight analogues included in this study (Fig. 6B). A214 and A268 are both at the ends of each of their helices (P4 and P7, respectively) and both are base paired to a conserved U (U107 and U307, respectively). It is possible that this collection of analogues did not modify the chemical group important for function at these two potentially homologous positions (for example, the N1 and N3 groups were not altered in this set of analogues).

Interference at nonconserved sites

Three positions within the core of the intron are not conserved phylogenetically, but do demonstrate interference (marked with an asterisks in Fig. 6B). The primary divergence from phylogeny is at A256, which showed interference from a wide variety of analogues. This position will be discussed below. The other two examples are A94 and A210. Interference at these two sites was primarily from analogues that add steric bulk to the nucleotide (^{OMe}A α S for A94 and ^{NMe}A α S for A210). Interference from bulky analogues such as these is not necessarily expected to correlate with the mutability of a given site. For example, A210 is a bulged nucleotide within the P4 helix, in a segment of the active site known to be densely packed (Michel & Westhof, 1990; Tanner & Cech, 1997). A210 is found only among the six sequenced species of *Tetrahymena*'s LSU rRNA and in the LSU rRNA intron of *Physarum polycephalum* (Damberger & Gutell, 1994). Given the close packing in this region, it is reasonable to expect that a bulky group on the amine would not be tolerated.

A biochemical signature for essential C2'-endo sugar pucker

A210 provides another example of how this collection of analogues can be used to probe the structure of

RNA. In addition to the $^{\text{NMe}}\text{A}\alpha\text{S}$ effect, A210 also showed interference with $^{\text{F}}\text{A}\alpha\text{S}$, but there was no interference from $\text{dA}\alpha\text{S}$ or $^{\text{OMe}}\text{A}\alpha\text{S}$. The lack of interference with $\text{dA}\alpha\text{S}$ strongly suggests that the 2'-OH does not make a direct contribution to activity, but interference with $^{\text{F}}\text{A}\alpha\text{S}$ argues in favor of an indirect contribution by the 2'-OH. A210 is one of only a few examples within the P4-P6 crystal structure where the ribose adopts a C2'-endo sugar pucker (Cate et al., 1996a). The unusual conformation of the A210 sugar allows the base to be flipped out of the helix without disrupting the helical continuity of P4. $^{\text{F}}\text{A}\alpha\text{S}$ substitution may disrupt the required C2'-endo sugar pucker because the highly electronegative 2'-fluoro group strongly favors the C3'-endo conformation (Uesugi et al., 1979). In contrast, the 2'-deoxy ribose could more easily adopt either the C2'-endo or C3'-endo conformation.

Interference with $^{\text{F}}\text{A}\alpha\text{S}$ coupled with tolerance for $\text{dA}\alpha\text{S}$ substitution might be predictive of C2'-endo sugar conformations within RNA. This pattern was seen at three other A positions within the *Tetrahymena* intron, A218, A256, and A304. One of these, A218, is also present within the P4-P6 crystal structure and it is also in a C2'-endo sugar pucker (Cate et al., 1996a). It is possible that A256 and A304 also adopt this alternative conformation. Each of these three positions will be discussed in detail below. Clearly, the 2'-fluoro and 2'-deoxy phosphorothioate derivatives of the complete series of nucleotides (i.e., A, G, C, and U) are potentially valuable reagents to identify positions where alterations in sugar pucker affect RNA folding.

Interference at the junction of helices P2 and P2.1

The interference pattern explains the chemical basis of A conservation throughout the intron active site. We will discuss each of the regions in the molecule that show interference and outline structural predictions for the active site based upon the interference pattern.

The junction of helices P2 and P2.1 is comprised entirely of A's, including nucleotides A28–A31, A57, A94, and A95. A57 and A95 are efficiently photo-crosslinked upon exposure of the intron to ultraviolet light, and the interaction between these two nucleotides is important for properly tethering the P1 helix into the active site (Downs & Cech, 1990, 1994). Point mutations at either position promote miscleavage at two sites other than the normal 5' splice site (Downs & Cech, 1994). A57 is highly conserved (98%) among the IC1 introns that have the P2.1 helix, but it is missing in all other classes of group I introns. A95 is also well conserved among the IC1-2 introns (94%), but less well conserved in the other subgroups.

Within P2 and P2.1, interference was only detected at the interface between the two helices, specifically at A57, A94, and A95. A57 only showed interference with

$^{\text{NMe}}\text{A}\alpha\text{S}$ and A94 only with $^{\text{OMe}}\text{A}\alpha\text{S}$. A95 showed interference with all the analogues that modify the N7 or the N6 positions ($7\text{dA}\alpha\text{S}$, $^{\text{NMe}}\text{A}\alpha\text{S}$, $\text{Pur}\alpha\text{S}$, and $2\text{AP}\alpha\text{S}$), but it was tolerant of functional group modification on the minor groove face, including addition of an N2 exocyclic amine and modification of the 2'-OH. An interference pattern involving this set of analogues strongly suggests interaction with the major groove, or Hoogsteen face of the nucleotide. A similar pattern is seen at several other positions within the ribozyme.

The revised model of the *Tetrahymena* intron group I intron has the ribose sugar of A95 positioned against the minor groove of the U59·G92 wobble pair in the P2.1 helix (Lehnert et al., 1996). Lack of interference at A95 from analogues that modify the ribose sugar argues against such an orientation. In the revised Michel and Westhof model, the O2' of A94 is proposed to interact with the N7 of A95, which agrees with the interference data, but there are no tertiary contacts to the N6 amino groups of either A57 or A95, as the data also predict.

Major and minor groove recognition of the P3 helix

The *Tetrahymena thermophila* P3 helix contains a single A nucleotide, A97, that is almost 90% conserved in IC1 and IC2 introns. Although only this single site was informative within the helix, the interference data provide valuable information about P3 helix packing within the active site. As was observed at A95, the interference pattern at A97 is characteristic of hydrogen bonding to the Hoogsteen face of the base (interference with $7\text{dA}\alpha\text{S}$, $^{\text{NMe}}\text{A}\alpha\text{S}$, $\text{Pur}\alpha\text{S}$, and $2\text{AP}\alpha\text{S}$). This implies that the major groove face of the P3 helix is involved in tertiary structure formation. However, unlike the pattern at A95, analogues that modify the minor groove functional groups also caused interference at A97. Interference with $\text{dA}\alpha\text{S}$ and $^{\text{OMe}}\text{A}\alpha\text{S}$ suggest that there is close approach in the P3 minor groove involving the 2'-OH. $^{\text{F}}\text{A}\alpha\text{S}$ substitution at A97 did not effect activity, which is an interference pattern exactly opposite to that seen at A210. In this case, deleting the 2'-OH is detrimental, but replacing it with the highly electronegative fluoro group had no effect, suggesting that the 2'-OH of A97 acts as a hydrogen bond acceptor, although its hydrogen bonding partner is unknown. A similar pattern was observed at A207, where the 2'-OH accepts a hydrogen bond from the N2 amine of G22 (Strobel et al., 1998).

The interference pattern at A97 suggests that two different structural elements converge at the P3 helix. The likely candidates are the J8/7 single-stranded region and the P2–P2.1 helical junction. The major groove interaction is consistent with the structure model proposed originally by Michel and Westhof (1990), where the Hoogsteen face of the A97–U277 pair makes a base triple with U300 (Fig. 7A). Further evidence in

FIGURE 7. One proposed base triple between U300 and the A97–U277 base pair. **A:** Wild-type U·A-U triple expected to form in almost 90% of all group IC1 and IC2 introns. The number of the nucleotide is shown within the ring, the functional groups of A97 that show interference are indicated with a shadowed box, and the occurrence of these nucleotides among 141 examples is shown. **B:** C+G·C triple expected to form in the introns where U300 is mutated to a C. **C:** One possible U·C-G triple predicted for the four introns that have a C97–G277 base pair. All three of the triples conserve the hydrogen bond to the 4 position of the pyrimidine at position 300.

support of this tertiary interaction comes from a larger group I intron sequence alignment (Damberger & Gutell, 1994; R.R. Gutell, unpublished results). Approximately 126 (90%) of the known IC1 and IC2 sequences are U·A-U at positions 300–97–277. In each of the six sequences where U300 is changed to a C, the 97–277 base pair changes to a G·C pair. Given that the C+G·C triple (where the N3 of C is protonated) is isosteric with U·A-U (Fig. 7A,B), this is the expected covariation for a conserved interaction between these three nucleotides. Furthermore, there are four introns where U300 is conserved, but 97–277 is changed to a C·G pair.

Although not completely isosteric, a U·C-G triple could retain one of the two hydrogen bonds present in the U·A-U triple (Fig. 7C). While additional experiments will be necessary to determine if these potential tertiary hydrogen bonds form in the folded structure, the interference and comparative sequence data are consistent with an interaction between J8/7 and the major groove of P3 at U300. The interaction in the minor groove of P3 still needs to be explored.

A G·U wobble receptor in J4/5

The interference pattern within the J4/5 region (nt A113, A114, A206, and A207) was reported previously for this set of analogues (Strobel et al., 1998). The data led us to conclude that the consecutive sheared A·A pairs within J4/5 act as a receptor for the universally conserved G·U wobble pair at the cleavage site of the intron (Fig. 8). The exocyclic amine of G forms two hydrogen bonds with the minor groove face of A207, and the 2'-OH of G forms two hydrogen bonds with the minor groove face of A114. In addition, there is a fifth hydrogen bond predicted by modeling of this region that was not identified in the original report of the helix-packing motif (Fig. 8). The 2'-OH of G23 is within convenient hydrogen bonding distance of the 2'-OH of C208. Previous thermodynamic analysis of a single 2'-deoxy substitution at G23 demonstrated that the 2'-OH contributes about $0.8 \text{ kcal} \cdot \text{mol}^{-1}$ to tertiary binding (Strobel & Cech, 1993; Narlikar et al., 1997), which is consistent with a single hydrogen bond to G23. Substitution of 2'-O-methylguanosine at G23 had no effect on P1 helix docking (Strobel & Cech, 1993), which suggests that the 2'-OH of G23 acts as a hydrogen bond acceptor.

Evidence in support of an A-platform in J6/6a

Another critical region with an interesting interference pattern is J6/6a, a symmetric three-nucleotide internal loop that includes nucleotides A218, A219, and A256. A218 and A219 are nearly invariant (>95%) among the IA and IC introns that have this segment. In the P4-P6 crystal structure, A218 and A219 are aligned side by side in a pseudo-base pairing arrangement termed an A-platform (Fig. 9; Cate et al., 1996b). They provide an interface for intermolecular association between J6/6a and L5c of two different P4-P6 molecules within the crystal lattice. The A218–A219 platform is not involved in the intramolecular folding of the P4-P6 domain, although it might play a role in folding the intact intron (Cate et al., 1996b). For the A's to adopt this side-by-side alignment, the ribose sugar of A218 adopts a C2'-endo conformation. The only possible hydrogen bonding interaction between the nucleobases is between the N3 of A218 and the N6 of A219 (Fig. 9), so the stability of the pair appears to be derived primarily from stacking interactions (Cate et al., 1996b).

FIGURE 8. G·U wobble receptor. Interactions between the highly conserved G22·U-1 pair in P1 and the consecutively stacked sheared A·A pairs in J4/5 were described previously (Strobel et al., 1998). The model suggests that there may also be an interaction between the 2'-OH of G23 and the 2'-OH of C208 (hydrogen bond at the bottom of the figure). This interaction was present, but not reported in the previous discussion of the model. This fifth hydrogen bond is supported by thermodynamic evidence using single functional group substitutions of the G23 2'-OH (Strobel & Cech, 1993; Narlikar et al., 1997). The structure of the J4/5 region is from the P4-P6 crystal structure (Cate et al., 1996a).

NAIM analysis confirms that the A nucleotides within J6/6a are important for activity. The data are consistent with an A·A pseudo-base pair, although the data are not sufficiently transparent that a pseudo-pair could have

FIGURE 9. A218·A219 pseudo-base pair observed within the J6/6a A-platform of the P4-P6 structure (Cate et al., 1996a). The J6a/6 segment is also shown. The ribose sugar of A218 is in a C2'-endo conformation, and its 2'-OH appears to contact electron density attributable to a metal (not shown) located below the A·A pseudo pair. The A's are approximately coplanar with G254. Nucleotides G254–A256 on the opposite strand of the internal loop are likely to be distorted away from the pseudo-pair because an unnatural G in the transcript (not shown) occupies the position expected for A256. If there is a tertiary interaction within this region, it is most likely formed with G254 and C255.

been predicted a priori. Interference was observed at A218 with both analogues (DAP α S and 2AP α S) that introduce an additional N2 amine. Interpreted in light of the pseudo-pair, an N2 amine at A218 would clash with the N6 amine of A219. The lack of interference at A219 upon deletion of the N6 amine (Pur α S) indicates that the single potential hydrogen bond between the bases is dispensable for activity under these conditions.

The most informative data in this region came from the 2'-OH analogues. Interference was observed with F A α S at A218, but not with dA α S. This is similar to the pattern seen at A210 and argues that A218 adopts a C2'-endo sugar pucker. This conclusion is in full agreement with the crystal structure where A218 is in a C2'-endo conformation that allows the bases to form the pseudo-pair. Strong interference also occurred with OMe A α S at A218, which suggests that there is close approach to the 2'-OH group. Within the P4-P6 structure, there is electron density in the major groove immediately below the A·A pair, which may correspond to a metal binding site (S. Basu, R. Rambo, J.H. Cate, S.A. Strobel, & J.A. Doudna, unpubl. results). The 2'-OH of A218 points directly toward this electron density. OMe A α S substitution may prevent metal binding due to steric occlusion, which provides indirect evidence that this metal is essential for intron activity.

Intermediate levels of interference were observed with several other analogues that are not necessarily expected to show interference based simply upon an A·A pseudo-pair. The data indicate that the N6 of A218 is important (Pur α S, 2AP α S, and NMe A α S interference) and that an N2 amino group or methylation of the N6 at A219 is not tolerated. Although these might simply be stacking effects, the data suggest that the tertiary structure near J6/6a may involve close approach to the Watson–Crick faces of both nucleotides.

A256 in the J6a/6 loop shows a complicated interference pattern. The identity of this base is not conserved, but there is strong phylogenetic evidence to suggest that it base pairs with the opposing nucleotide

in the loop, position 217. In the IC3 introns, this is a C-G pair in 99% of the sequences and, among the ID introns, it is usually (94%) a G-U pair. The *Tetrahymena* intron is an exception, with a C217·A256 juxtaposition. Thus, there is significant evolutionary pressure for base pairing at A256, but the identity of the base pair is not a universal element of group I introns. Nevertheless, A256 showed interference from a wide diversity of analogues. It was affected by modifications on the Hoogsteen face (7dA α S and Pur α S), and by modifications in the minor groove (DAP α S and ^{OMe}A α S). Interference with ^FA α S, but not with dA α S, suggests that this base may also adopt a C2'-endo sugar pucker, although the electron density at A256 was too distorted to confirm this possibility.

In all three examples of A-platforms within the P4-P6 structure, the A·A pseudo-pair mediated tertiary structure formation. Because the J6/6a A-platform in the crystal structure was involved in an intermolecular contact that does not occur within the full-length ribozyme, it is unclear if a tertiary interaction is made between J6/6a and the rest of the intron. In the crystal structure, the A218–A219 pseudo-pair is stacked upon P6 and P6a (Cate et al., 1996a), and it serves as an intramolecular continuation of the P4-P6 helix. In so doing, the A-platform frees the opposing J6a/6 strand of the internal loop (nt G254 and C255) to make a two-base pair intermolecular pseudoknot with L5c. If the P4-P6 structure is an accurate predictor of the intronic structure, the most likely tertiary contacts in this region are made by G254 and C255. The biological relevance of the P4-P6 structure for J6a/6 is questionable because A256 was displaced from C217 by the unnatural 5'-terminal G that was introduced for transcriptional initiation (Cate et al., 1996b). This G also displaced C255 and G254 from their presumed location within the intron structure (Fig. 9). Base pairing of A256 with C217 and close packing of C255 and G254 against the pseudo-pair would explain the interference data observed on the Watson–Crick faces of A218 and A219.

If G254 and C255 do not make an intermolecular contact to L5c within the intron structure, do they make an intramolecular contact somewhere else within the intron? Although A218 and A219 are not depicted as an A-platform within the Michel and Westhof model, it is interesting that J6a/6 is quite close to the P3 helix. Further experiments are necessary to determine if G254 and/or C255 make transdomain tertiary contacts to the bases or ribose backbone of this helical element.

Interference within the G cofactor binding site

The G binding site is located within the P7 helix and is centered around the G264–C311 pair (Michel et al., 1989). Two A's (A263 and A265) within the P7 helix flank this essential base pair and both have been implicated in G binding (Yarus et al., 1991a). The G bind-

ing site was identified originally by Michel et al. (1989), who demonstrated a change in substrate specificity from G to 2AP when the G-C pair was mutated to A-U. This was evidence for a direct hydrogen bond between the H1 of G and the O6 of G264. They also proposed a second hydrogen bond between the N2 amine of G and the N7 of G264, which defines an equatorial alignment of G with the G264–C311 pair (Fig. 10A). DMS interference experiments on the *sunY* intron showed that methylation at the N7 of G264 (G96 in the *sunY* numbering system) blocked G-mediated splicing and that G binding protected the N7 of G264 from DMS methylation (von Ahsen & Noller, 1993). This confirms that G is close to G264, but does not demand that the N7 of G264 hydrogen bond to the N2 of the G cofactor. Subsequent work by Yarus et al. (1991b) showed that the flanking base pair A265–U310 also contributed directly to G binding. They showed a change in substrate specificity to DAP when A265–U310 was mutated to G-C. This suggests that there is an additional hydrogen bond between the O6 of G and the N6 of A265. The restraint of a second hydrogen bond with the base pair below G264–C311 argued that the G is out of plane from the G-C pair and in an axial position (Yarus et al., 1991b; Fig. 10B).

Yarus et al. (1991a) used these constraints to generate an energy-minimized model of the G binding site that they termed axial III. This structure is a hybrid between the axial and equatorial models of the G binding site, although it is largely equatorial. The model includes a base triple between A263 and the minor groove face of the G264–C311 base pair. Upon inspection of their model, the geometry and distance of the A265 amine and the G O6 do not seem to be consistent with a hydrogen bond between these groups. Furthermore, the helix appears to be distorted from an A-form geometry and the purine and pyrimidine bases throughout the structure do not appear to be planar. The interference and mutational data suggest that axial III is unlikely to accurately represent the G binding site.

The NAIM data support a central tenant of the original axial model (Fig. 10B), which is that the amine of A265 hydrogen bonds to the O6 of G (Yarus et al., 1991a). ^{NMe}A α S substitution at A265 significantly impairs ribozyme function, however, both analogues that delete the N6 amine (2AP α S and Pur α S) did not interfere with activity. Although it is possible that ^{NMe}A α S substitution prevented P7 helix formation, our preferred interpretation is that G414 could not occupy its binding site when the Hoogsteen face of A265 was sterically blocked by a methyl group. Nevertheless, the effective molar concentration of G414 was sufficient to overcome the loss of a single hydrogen bond due to deletion of the A265 amine.

A263 is also adjacent to the G binding site, it is semi-conserved as either an A or a C, and it is always un-

FIGURE 10. An equatorial (A) and an axial (B) model for G cofactor binding to the G binding site (Michel et al., 1989; Yarus et al., 1991b). In both models, the G is in an anti configuration (Lin et al., 1994), and there is a hydrogen bond between the H1 of G and the O6 of G264. The equatorial model has a second hydrogen bond involving the N2 of G and the N7 of G264, whereas the axial model has a second hydrogen bond involving the O6 of G and the N6 of A265. Biochemical evidence exists to support the central tenets of both models, including interference data reported in this paper, which support the axial model. Although the hybrid model termed axial III by Yarus et al. (1991a; not shown) is unlikely to represent the structure of the bound G cofactor, it remains uncertain as to how these two models can be reconciled, or if a version of both of them might not occur at some point along the pathway toward G binding (Profenno et al., 1997).

paired. A263 covaries with the C262–G312 base pair in a way that avoids base pairing of A263 with G312 (Michel et al., 1989; Gautheret et al., 1995). Mutations at A263 had no effect on intron splicing (Yarus et al., 1991a), and no interference was detected with any of the analogues in this study. Lack of interference and lack of an effect upon mutation argues against the G262–C312·A263 base triple proposed in the axial III model (Yarus et al., 1991a). Instead, it suggests that there is simply a requirement for a bulged nucleotide, and perhaps an unusual position of the phosphate–ribose backbone is important for G binding. Consistent with this possibility, there was a phosphorothioate effect at A263 that could not be rescued with Mn^{2+} . Our data suggest that if A263 does make a direct contribution to G binding, it does so primarily via its phosphate. If G binds in an axial orientation, then the phosphate of A263 is closest to the 2'-OH of G, whereas if G binds in an equatorial alignment, then the phosphate is closest to the N2 amine (Fig. 10). Both the N2 and 2'-OH of G are known to contribute significantly to G binding (Bass & Cech, 1984; McConnell & Cech, 1995; Li & Turner, 1997; Profenno et al., 1997).

Another unpaired nucleotide that may participate in G binding is A261, which is one of the most highly conserved nucleotides among all classes of group I introns, with only a few exceptions among nearly 500 sequences. This is the level of conservation only seen among the most essential nucleotides in the intron,

which would be consistent with a role in G binding. The interference pattern at this site suggests that several functional groups of A261 are essential for activity, including the 2'-OH of the sugar (interference with dA α S, $^{\text{OMe}}$ A α S, and $^{\text{F}}$ A α S) and the Hoogsteen face of the base (interference with 7dA α S, $^{\text{NMe}}$ A α S, 2AP α S).

A major issue that is not at all resolved in an axial model for G binding (Fig. 10B) is what interacts with the exocyclic amine of G, and why interference is observed at the N7 of G264 with DMS (Bass & Cech, 1984; von Ahsen & Noller, 1993). The N2 amino group of G contributes 3.0 kcal·mol $^{-1}$ to binding (McConnell & Cech, 1995). Furthermore, N^2 -methylguanosine binds poorly, which suggests that both amine protons are involved in cofactor recognition (Bass & Cech, 1984). Two potential hydrogen bonding partners are the N7 of G264 and the R_P phosphate oxygen of A263. Both of these hydrogen bonds require an equatorial alignment of G, which is incompatible with the A265 interaction. Fewer obvious options exist for the amine in the axial orientation, although A261 may participate.

Thus, although the axial III hybrid model of Yarus et al. (1991a) seems unlikely, there are data to support both the axial and the equatorial models for G binding. Profenno et al. (1997) have recently shown that G binds to the intron in at least two steps. This includes a bimolecular step followed by at least one conformational change. An explanation that could reconcile the interference and mutational data is if the G binding pathway

includes one step with an axial and another step with an equatorial alignment of the base. The change in the alignment of G might be necessary to bring the 3'-OH nucleophile into position for in-line attack at the scissile phosphate (McSwiggen & Cech, 1989; Rajagopal et al., 1989). In this situation, interference might not result from a single static structure, but rather from disruption of one structure along a dynamic pathway of structural conformations.

Additional major groove contacts in P7

There is one additional A in the P7 helix (A308) that was informative in our assay. A308 is base paired with U267. The U267–A308 base pair is 95% conserved among all of the group I intron sequences. A308 is three base pairs removed from the G binding site, so it is unlikely to participate directly in G binding, although it could make an indirect contribution. A308 had an interference pattern indicative of Hoogsteen pairing in the major groove of P7 (interference with 7dA α S, ^{NMe}A α S, 2AP α S, and Pur α S), and there was no interference from any of the minor groove modified analogues. This suggests that an essential tertiary contact is made in the major groove of P7 immediately below the G binding site. Nothing is included in this region of the molecule within the Michel and Westhof model (Michel & Westhof, 1990; Lehnert et al., 1996). Possible candidates for this major groove interaction include the 3' end of J8/7 or the J3/4 linker segment. The nucleotides in both of these regions are very highly conserved. It might also be an essential metal binding site. Whatever makes contacts with the P7 helix at A308, it is sufficiently close to G264 that it may participate in G binding.

Interference within the J8/7 region

A final region with a particularly striking interference pattern is J8/7, a single-stranded segment between helices P8 and P7. Previous studies have implicated J8/7 in binding of the P1 substrate helix (Pyle et al., 1992; Strobel & Cech, 1993; Michel & Westhof, 1994), interaction with the P4 helix (Tanner et al., 1997; Tanner & Cech, 1997), proximity to the G binding site (Wang & Cech, 1992), and formation of the catalytic active site (Michel & Westhof, 1990; Christian & Yarus, 1993). It is an essential and highly conserved segment of the group I intron. Four A's exist in this seven-nucleotide segment, A301, A302, A304, and A306. Each displays at least some level of interference from this set of analogues, but each position is susceptible to analogue substitution in a different way.

Nucleotides A301 and A302 have been proposed to orient the substrate helix into the ribozyme active site by making direct tertiary contact with P1 (Pyle et al., 1992; Strobel & Cech, 1993; Michel & Westhof, 1994). Both nucleotides are highly conserved (>99% among

IC1-2 introns). A hydrogen bond between the N1 of A302 and the 2'-OH of U-3 within the P1 helix was demonstrated through DMS footprinting and mutagenesis (Pyle et al., 1992). Michel and Westhof (1994) proposed a similar interaction between A301 and G25.

NAIM analysis at A301 and A302 suggests that both nucleotides are involved in a collection of important interactions within the active site. Both sites show complete interference with DAP α S, which suggests that there is close approach to the minor groove edge of each nucleotide. The inability to accommodate the extra N2 amine on DAP α S explains why these positions are almost never mutated to G (only one exception at each position among 131 IC1-2 sequences). Both sites show interference with dA α S and ^{OMe}A α S, which implies that both of the 2'-hydroxyls are important for function. Both positions also show interference with Pur α S and from DMS modification at the N1 position (Pyle et al., 1992).

Although similar, the two sites are not equivalent. ^FA α S substitution at A301 interferes with activity, which suggests that the A301 2'-OH is an essential hydrogen bond donor. Activity is not affected by ^FA α S substitution at A302, where it might act as a hydrogen bond acceptor. 7dA α S interferes with activity when incorporated at A301, but not A302, whereas A302, but not A301, has a strong phosphorothioate effect that can be partially rescued with Mn²⁺. Although the interference patterns differ in detail, they strongly imply that both the minor and major faces of A301 and A302 are involved in forming the intron active site. One face of each nucleotide interacts with P1, whereas the other face might contact P3.

C2'-endo conformation at A304

Nucleotides at the 3' end of the J8/7 region are the most likely positions to participate directly in the chemical mechanism of group I intron catalysis. These include positions G303, A304, and A306. U305 forms an interdomain base triple in the major groove of P4 (Tanner et al., 1997; Tanner & Cech, 1997), so it is unlikely to participate directly in substrate activation. Because the 3'-exon ligation assay does not distinguish between effects on substrate binding and chemistry, interference within this region is only suggestive of functional groups that might directly participate in the chemical transition state.

Of the three positions, two (A304 and A306) were informative in this assay. Among the IC1 and IC2 introns, A304 is predominantly an A, and almost always a purine (77% A and 22% G). However, none of the base-modified analogues used in this study affected activity when incorporated at A304. This includes 7dA α S, which modifies one of the two base functional groups conserved between A and G. A304 did show interference from analogues that disrupted the ribose sugar,

although, again, no specific functional groups were implicated directly in activity. $^{\text{F}}\text{A}\alpha\text{S}$ substitution caused interference, but $\text{dA}\alpha\text{S}$ did not. This is the same pattern that was observed at A210 and A218, and suggests that the ribose ring adopts an essential C2'-endo sugar pucker. This is a particularly intriguing site for such a conformation. The switch from C3'-endo to C2'-endo lengthens the phosphate-to-phosphate distance from approximately 5.9 Å to 7.0 Å, and extends the ribose backbone (Saenger, 1984). A C2'-endo conformation is often observed in hairpin loops or bulged residues where a large distance must be transversed in the space of one nucleotide (Wyatt & Tinoco, 1993). On the 5'-side of A304, G303 makes a tertiary contact with the P1 helix (Strobel & Shetty, 1997; L. Ortoleva-Donnelly, A.A. Szewczak, & S.A. Strobel, unpubl. results), whereas on the 3'-side, U305 is docked into the major groove of P4 (Tanner et al., 1997; Tanner & Cech, 1997). Although P1 and P4 are fairly close in tertiary space within this region of the molecule (Wang et al., 1993; Strobel et al., 1998), a C2'-endo conformation at A304 might be necessary for the nucleotide to bridge the gap between these two helices.

An essential 2'-OH at A306

Similar to A304, A306 is usually an A and is almost always a purine (86% A and 13% G for A306). As was true for A304, none of the base-modified analogues had a strong affect upon activity when incorporated at A306. The only exception was $7\text{dA}\alpha\text{S}$, which had a slight effect that might have been artificially amplified by the strong phosphorothioate effect present at this site. Although there might be some involvement by the N7 imine, the data argue against a significant contribution by the base functional groups of A306. In contrast, strong interference was detected from $\text{dA}\alpha\text{S}$ and $^{\text{F}}\text{A}\alpha\text{S}$, which argues that the 2'-OH of A306 makes an essential contribution as a hydrogen donor. Given its proximity to the scissile phosphate, the 2'-OH of A306 is a good candidate for a catalytically important functional group within the intron active site.

None of the A304 or A306 base functional groups tested in this experiment contributed significantly to activity, yet both residues are conserved as an A or a G. Given that purines form more stable stacking interactions than pyrimidines, we propose that A304 and A306 may be stacked within the active site. Stacking of A304 and A306 places U305 into a "bulged" conformation relative to J8/7. This would allow it to form the base triple demonstrated between U305 and the P4 helix (Tanner et al., 1997; Tanner & Cech, 1997). A304, U305, and A306 all show strong phosphorothioate effects that can be at least partially rescued with Mn^{2+} . Divalent metal ions coordinated to these nucleotides might mediate the close approach of phosphates in J8/7 with functional groups in the P4 or P7 helices (Streicher

et al., 1996). These phosphates may also participate in metal ion coordination to the P1 helix, including coordination of the two catalytic metals within the active site (Piccirilli et al., 1993; Streicher et al., 1996; McConnell et al., 1997; Weinstein et al., 1997).

Interference data and structural models

Nucleotide analogue interference mapping provides high-resolution biochemical information about the structure and function of RNA. Some of the constraints suggested by the interference data are not satisfied within the original or modified version of the Michel and Westhof *Tetrahymena* group I intron model (Michel & Westhof, 1990, 1994; Lehnert et al., 1996). Their model has served as a predictive and fairly accurate indicator of helical positioning within the active site, but, at several positions, it does not appear to be accurate at the biochemical resolution detected by NAIM. Revisions to the model that incorporate these data appear to be warranted. Although we have not completed the effort, we have generated a model for the interactions between P1, J8/7, and P4-P6 (partially depicted in Fig. 8; Strobel et al., 1998) that is fully compatible with the interference and mutagenesis data (L. Ortoleva-Donnelly, A.A. Szewczak, & S.A. Strobel, unpubl. results). Using this central structure as the starting point, the interference data provide several additional constraints that must be met as the model is expanded to include other regions of the active site.

In this report, we have focused exclusively upon A's within the catalytic core of the *Tetrahymena* intron. The experiments reveal an important relationship between the conservation of a nucleotide at a given position and the utilization of the functional groups of that nucleotide within the RNA structure. We expect a similar pattern will occur within other RNAs. The methods and reagents reported here can be generalized to study the chemical basis of adenosine conservation in almost any RNA, including the extensive array of conserved A's within RNaseP, the group II intron, and the ribosome. Based upon the extensive involvement of A's within tertiary structural motifs, these phosphorothioate-tagged nucleotides are likely to be powerful tools for RNA structure-function analysis. It should also be possible to expand the collection of nucleotides to include analogues of C, G, and U. The development of a complete collection of nucleotide reagents for NAIM should make it possible to identify interference fingerprints for a full catalogue of RNA structural motifs.

MATERIALS AND METHODS

Synthesis of phosphorothioate-tagged nucleotide triphosphates

Diaminopurine nucleoside (2-amino adenosine), 2'-deoxy-2'-fluoro adenosine, and 2'-deoxy-2'-methoxy adenosine were

gifts from Schering Pharmaceutical Research, Isis Pharmaceuticals, and Ribozyme Pharmaceuticals Inc., respectively. 2-Aminopurine was synthesized in one step from 6-mercaptoguanosine (Sigma) as described previously (Fox et al., 1957). Purine riboside, *N*²-methyladenosine, 7-deaza-adenosine (tubercidin), and 2'-deoxyadenosine were purchased from Sigma. Synthesis was performed using reagent-grade chemicals as received unless otherwise stated. Interference experiments with dATP α S were performed with material purchased from Amersham. Synthesis of DAPTP α S was reported previously (Strobel & Shetty, 1997).

Each nucleoside (50–100-mg scale) was converted to the 5'-*O*-(1-thio)-nucleoside triphosphate using a procedure analogous to that reported for the synthesis of 5'-*O*-(1-thio)-2'-deoxyadenosine triphosphate (Arabshahi & Frey, 1994). The nucleoside (100 mg, 0.37 mmol) was dried under vacuum at 110 °C for 12 h and dissolved in triethylphosphate (5 mL). The nucleotide was reacted with thiophosphorylchloride (42 μ L, 0.41 mmol, 1.1 equivalents) in the presence of triethylamine (180 μ L, 0.41 mmol, 1.1 equivalents) at room temperature for 30 min to form the 5'-*O*-(1-thio, 1,1-dichloro)phosphoryl derivative of the nucleoside. Formation of this product was monitored by cellulose TLC using 0.5 M LiCl (aq) as the solvent system. This intermediate was converted directly to the triphosphate by addition of tributylammonium pyrophosphate (0.5 g, 1.0 mmol) in triethylphosphate (3 mL) and stirring at room temperature for an additional 30 min. Formation of the cyclic triphosphate was monitored by silica TLC using 6:3:1 *n*-propanol, ammonium hydroxide, water as the solvent system. The triphosphate typically had an *R*_f between 0.15 and 0.3, compared with 0.4–0.6 for the monophosphate and 0.8–0.9 for the nucleoside. The triphosphate was precipitated by the addition of excess triethylamine (2.5 mL), centrifuged, decanted, and the residue dissolved in aqueous triethylammonium bicarbonate (10 mL, 50 mM pH 7.5). The crude product was left at room temperature overnight to achieve ring opening of the cyclic triphosphate. Purification by DEAE-A25 Sephadex chromatography using a linear triethylammonium bicarbonate (TEAB) buffer gradient (0.05–0.8 M) afforded each of the 5'-*O*-(1-thio)nucleoside triphosphates as a diastereomeric mixture in 10–40% yield. Each eluted from the gradient at approximately 0.6 M TEAB. The triphosphates were characterized by ³¹P NMR and UV absorbance. The concentrations were determined using the absorbance and extinction coefficient for each nucleotide at pH 7.5. PurTP α S, ³¹P NMR (H₂O): 43.67 (m), –6.27 (d), –22.91 (t); λ_{max} : 262 nm; ϵ_{262} : 5700. 2APTP α S, ³¹P NMR (H₂O): 43.93 (m), –10.27 (d), –23.59 (t); λ_{max} : 304 nm; ϵ_{304} : 6000. NMeATP α S, ³¹P NMR (H₂O): 43.23 (m), –5.97 (d), –22.87 (t); λ_{max} : 263 nm; ϵ_{263} : 16,000. 7dATP α S, ³¹P NMR (H₂O): 43.041 (m), –6.071 (d), –22.870 (t); λ_{max} : 271 nm; ϵ_{270} : 12,000. OMeATP α S, ³¹P NMR (H₂O): 43.984 (m), –10.266 (d), –23.502 (t); λ_{max} : 260 nm; ϵ_{260} : 15,000. F¹⁸ATP α S, ³¹P NMR (H₂O): 44.045 (m), –9.995 (d), –23.402 (t); λ_{max} : 259 nm; ϵ_{260} : 15,000.

RNA transcription and analogue incorporation

Analogues were randomly incorporated into the L-21 G414 form of the *Tetrahymena* group I intron by T7 RNA polymerase transcription from the *Ear* I-digested plasmid pUCL-21G414 (Strobel & Shetty, 1997). RNAs were transcribed in

40 mM Tris-HCl, pH 7.5, 4 mM spermidine, 10 mM DTT, 15 mM MgCl₂, 0.05% Triton X-100, and 0.05 μ g/ μ L DNA template. A level of approximately 5% analogue incorporation was achieved by normalizing iodine cleavage intensities to that of a transcript made from 50 μ M ATP α S (*S*_P diastereomer only) and 1 mM of each NTP (Christian & Yarus, 1993). Each of the analogues were incorporated by including the analogue in the transcription mixture using the concentrations of analogue and ATP outlined in Table 1. In all cases, 1 mM of CTP, UTP, and GTP were also included in the transcription reaction. All of the RNAs were purified by PAGE (6% denaturing), eluted into 10 mM Tris, pH 7.5, 0.1 mM EDTA (TE), precipitated with ethanol, resuspended in TE, and stored at –20 °C.

Interference reactions

The L-21 G414 RNA (50 nM) (Strobel & Shetty, 1997) was pre-incubated in reaction buffer [50 mM HEPES, pH 7.0, 3 mM MgCl₂, and 1 mM Mn(OAc)₂] at 50 °C for 10 min. dT(–1)S (1 nM) radiolabeled at its 3' end with poly(A) polymerase and [α -³²P]cordycepin (Lingner & Keller, 1993) was dissolved in the same reaction buffer, added to the ribozyme solution, and the mixture was incubated at 50 °C for 10 min. The reaction was quenched by the addition of two volumes of stop solution (8 M urea, 50 mM EDTA, 0.01% bromophenol blue, 0.01% xylene cyanol), and split into two tubes. One-tenth volume of 100 mM iodine in ethanol was added to one of the tubes to cleave the phosphorothioate bonds (Gish & Eckstein, 1988). The solutions were heated to 90 °C for 2 min and the cleavage products were resolved by electrophoresis on a denaturing 5% polyacrylamide gel. The second portion of 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 5' end-labeled L-21 G414 RNAs were used to control for variability in analogue incorporation at each position. The L-21 G414 RNAs were treated with alkaline phosphatase and 5' end-labeled with T4 polynucleotide kinase and [γ -³²P]ATP. The radiolabeled RNAs were purified by PAGE (6% denaturing) and eluted into 0.1% SDS in TE overnight. RNAs were extracted with 1 volume of phenol/chloroform (1:1) to remove the SDS, ethanol precipitated, and redissolved in TE. RNAs were treated with iodine and the cleavage products were resolved in the same manner as for the reacted RNAs. Because the ligation reaction radiolabels the 3' end of the RNA and the kinase reaction labels the 5' end, the sequence orientation is reversed between the two gels. Several loadings of the same reaction samples were electrophoresed for variable amounts of time (from 1 to 5 h at 75 watts) to maximally resolve each region of sequence. Nucleotides very close to the 5' end were difficult to resolve for the 3'-exon ligation reaction, whereas nucleotides close to the 3' end were difficult to resolve for the 5'-labeled control. This is particularly true of consecutive A runs, of which there are several at the 5' end of the intron (A103–A105; A94–A95; A87–A90; A64–A66).

For comparison of the reactions in the presence or absence of Mn²⁺ (primary data not shown), the incubation time of the enzyme with the substrate dT(–1)S was adjusted to normalize for the extent of reaction. Reactions containing

4 mM MgCl₂ were incubated for 10 min, whereas reactions containing 3 mM MgCl₂ and 1 mM Mn(OAc)₂ were incubated for 2 min. Information was gained about A302 and A306 by incubating the ribozymes with the all ribose substrate rT(−1)S in buffer containing 3 mM MgCl₂ and 1 mM Mn(OAc)₂ for 5 min. No information could be obtained at these two positions using dT(−1)S because of a complete loss of function due to the phosphorothioate effects at these two sites.

Interference quantitation

Peak intensities for both the parental nucleotide (AαS) and the nucleotide analogue (δAαS) were quantitated by PhosphorImager analysis at each position in the 3′-ligation experiment 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:

$$\text{Interference} = \frac{\frac{\text{A}\alpha\text{S 3'-Ligation Reaction}}{\delta\text{A}\alpha\text{S 3'-Ligation Reaction}}}{\frac{\text{A}\alpha\text{S 5'-Labeled Control}}{\delta\text{A}\alpha\text{S 5'-Labeled Control}}} \quad (1)$$

The resulting value normalizes for phosphorothioate effects and variability in the extent of analogue incorporation at each position. All the interference values were further normalized to account for differences in loading and extent of reaction between lanes. The fully normalized interference values (κ) plotted in Figure 5 were obtained by calculating the average interference value at all positions in the RNA that were within two standard deviations from the mean and then dividing each individual interference value by the normalized average (the averages ranged from 0.8 to 1.2). This normalized all the data to a scale where a κ value of 1 indicated no interference. The interference κ values for A302 and A306 were calculated independently of those for the rest of the intron using 5′-control data quantitated from Figure 4A and interference data quantitated from Figure 4C. At some positions, analogue substitution resulted in complete loss of band intensity in the 3′-ligation experiment. In these cases, it was difficult to measure accurately the experimental baseline due to incomplete resolution of the missing band from the neighboring bands on the sequencing gel. Therefore, any effects that were quantitated as being greater than six-fold were defined as having a maximum magnitude of 6. All the κ values greater than 2 or less than 0.5 are plotted in Figure 5.

Phylogenetic comparisons

Group I intron sequences were identified and extracted from GenBank (<http://www2.ncbi.nlm.nih.gov/genbank/query form.html>) and aligned on a SUN Workstation with the program AE2 developed by T. Macke and available from the Ribosomal RNA Database Project (<http://rdp.life.uiuc.edu/>) (Maidak et al., 1997). Analysis of the group I intron sequences was performed on the sequence alignments and secondary structure diagrams as described (Damberger & Gutell, 1994; Gautheret et al., 1995; R.R. Gutell, unpublished results).

ACKNOWLEDGMENTS

We thank Rui Sousa for the gift of the Y639F mutant T7 polymerase clone, Leo Beigelman and Jasenka Adamic for advice and assistance with phosphorothioate synthesis, and S. Ryder and L. Weinstein for critical comments during the preparation of this manuscript. This work was supported by NIH grant GM48207 (to R.R.G.), and a Beckman Young Investigator Award, an American Cancer Society Junior Faculty Research Award, and NIH grant GM54839 (to S.A.S.).

Received January 14, 1998; returned for revision February 12, 1998; revised manuscript received February 19, 1998

REFERENCES

- Arabshahi A, Frey PA. 1994. A simplified procedure for synthesizing nucleoside 1-thiotriphosphates: dATPαS, dGTPαS, UTPαS, and dTTPαS. *Biochem Biophys Res Commun* 204:150–155.
- Bass BL, Cech TR. 1984. Specific interaction between the self-splicing RNA of *Tetrahymena* and its guanosine substrate: Implications for biological catalysis by RNA. *Nature* 308:820–826.
- Beattie TL, Olive JE, Collins RA. 1995. A secondary-structure model for the self-cleaving region of *Neurospora* VS RNA. *Proc Natl Acad Sci USA* 92:4686–4690.
- Beaudry AA, Joyce GF. 1992. Directed evolution of an RNA enzyme. *Science* 257:635–641.
- Butcher SE, Burke JM. 1994. Structure-mapping of the hairpin ribozyme. *J Mol Biol* 244:52–63.
- Cate JH, Doudna JA. 1996. Metal-binding sites in the major groove of a large ribozyme domain. *Structure* 4:1221–1229.
- Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA. 1996a. Crystal structure of a group I ribozyme domain: Principles of RNA packing. *Science* 273:1678–1685.
- Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Szewczak AA, Kundrot CE, Cech TR, Doudna JA. 1996b. RNA tertiary structure mediation by adenosine platforms. *Science* 273:1696–1699.
- Cate JH, Hanna RL, Doudna JA. 1997. A magnesium ion core at the heart of a ribozyme domain. *Nature Struct Biol* 4:553–558.
- Cech TR. 1993. Structure and mechanism of the large catalytic RNAs: Group I and group II introns and ribonuclease P. In: Gesteland RF, Atkins JF, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 239–269.
- Cech TR, Damberger SH, Gutell RR. 1994. Representation of the secondary and tertiary structure of group I introns. *Nature Struct Biol* 1:273–280.
- Cech TR, Herschlag D, Piccirilli JA, Pyle AM. 1992. RNA catalysis by a group I ribozyme. *J Biol Chem* 267:17479–17482.
- Christian EL, Yarus M. 1992. Analysis of the role of phosphate oxygens in the group I intron from *Tetrahymena*. *J Mol Biol* 228:743–758.
- Christian EL, Yarus M. 1993. Metal coordination sites that contribute to structure and catalysis in the group I intron from *Tetrahymena*. *Biochemistry* 32:4475–4480.
- Conrad F, Hanne A, Gaur RK, Krupp G. 1995. Enzymatic synthesis of 2′-modified nucleic acids: Identification of important phosphate and ribose moieties in RNase P substrates. *Nucleic Acids Res* 23:1845–1853.
- Costa M, Deme E, Jacquier A, Michel F. 1997. Multiple tertiary interactions involving domain II of group II self-splicing introns. *J Mol Biol* 270:520–536.
- Costa M, Michel F. 1995. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J* 14:1276–1285.
- Couture S, Ellington AD, Gerber AS, Cherry JM, Doudna JA, Green R, Hanna M, Pace U, Rajagopal J, Szostak JW. 1990. Mutational analysis of conserved nucleotides in a self-splicing group I intron. *J Mol Biol* 215:345–358.
- Damberger SH, Gutell RR. 1994. A comparative database of group I intron structures. *Nucleic Acids Res* 22:3508–3510.
- Deeney CMM, Eperon IC, Potter BVL. 1987. Self-splicing of *Tetrahymena* ribosomal RNA can proceed with phosphorothioate substitution at the splice sites. *Nucleic Acids Res Symp Ser* 18:277–280.

- Doudna JA, Couture S, Szostak JW. 1991. A multisubunit ribozyme that is a catalyst of and template for complementary strand RNA synthesis. *Science* 251:1605–1608.
- Downs WD, Cech TR. 1990. An ultraviolet-inducible adenosine–adenosine cross-link reflects the catalytic structure of the *Tetrahymena* ribozyme. *Biochemistry* 29:5605–5613.
- Downs WD, Cech TR. 1994. A tertiary interaction in the *Tetrahymena* intron contributes to selection of the 5' splice site. *Genes & Dev* 8:1198–1211.
- Eckstein F. 1985. Nucleoside phosphorothioates. *Annu Rev Biochem* 54:367–402.
- Engel JD, von Hippel PH. 1974. Effects of methylation on the stability of nucleic acid conformations: Studies at the monomer level. *Biochemistry* 13:4143–4158.
- Engel JD, von Hippel PH. 1978. Effects of methylation on the stability of nucleic acid conformations. *J Biol Chem* 253:927–934.
- Fazakerley GV, Teoule R, Guy A, Fritzsche H, Guschlbauer W. 1985. NMR Studies on oligodeoxyribonucleotides containing the dam methylation site GATC. Comparison between d(GGATCC) and d(GGm6ATCC). *Biochemistry* 24:4540–4548.
- Fox JJ, Wempen I, Hampton A, Doerr IL. 1957. Thiation of nucleosides. I. Synthesis of 2-amino-6-mercapto-9- β -D-ribofuranosyl-purine ("thioguanosine") and related purine nucleosides. *J Org Chem* 80:1669–1675.
- Gaur RK, Krupp G. 1993. Modification interference approach to detect ribose moieties important for the optimal activity of a ribozyme. *Nucleic Acids Res* 21:21–26.
- Gautheret D, Damberger SH, Gutell RR. 1995. Identification of base-triples in RNA using comparative sequence analysis. *J Mol Biol* 248:27–43.
- Gish G, Eckstein F. 1988. DNA and RNA sequence determination based on phosphorothioate chemistry. *Science* 240:1520–1522.
- Green R, Ellington AD, Szostak JW. 1990. In vitro genetic analysis of the *Tetrahymena* self-splicing intron. *Nature* 347:406–408.
- Guo Q, Lu M, Kallenbach NR. 1995. Effect of hemimethylation and methylation of adenine on the structure and stability of model DNA duplexes. *Biochemistry* 34:16359–16364.
- Guschlbauer W, Jankowski K. 1980. Nucleoside conformation is determined by the electronegativity of the sugar substituent. *Nucleic Acids Res* 8:1421–1433.
- Gutell RR, Weiser B, Woese CR, Noller HF. 1985. Comparative anatomy of 16S-like ribosomal RNA. *Prog Nucleic Acid Res Mol Biol* 32:155–216.
- Hardt WD, Erdmann VA, Hartmann RL. 1996. R_P-deoxy phosphorothioate modification interference experiments identify 2'-OH groups in RNase P RNA that are crucial to tRNA binding. *RNA* 2:1189–1198.
- Herschlag D, Eckstein F, Cech TR. 1993a. Contributions of 2'-hydroxyl groups of the RNA substrate to binding and catalysis by the *Tetrahymena* ribozyme. An energetic picture of an active site composed of RNA. *Biochemistry* 32:8299–8311.
- Herschlag D, Eckstein F, Cech TR. 1993b. The importance of being ribose at the cleavage site in the *Tetrahymena* ribozyme reaction. *Biochemistry* 32:8312–8321.
- Huang Y, Beaudry A, McSwiggen J, Sousa R. 1997a. Determinants of ribose specificity in RNA polymerization: Effects of Mn²⁺ and dNMP incorporation into transcripts. *Biochemistry*. Forthcoming.
- Huang Y, Eckstein F, Padilla R, Sousa R. 1997b. Mechanism of ribose 2'-group discrimination by an RNA polymerase. *Biochemistry* 36:8231–8242.
- Inoue T, Cech TR. 1985. Secondary structure of the circular form of the *Tetrahymena* rRNA intervening sequence: A technique for RNA structure analysis using chemical probes and reverse transcriptase. *Proc Natl Acad Sci USA* 82:648–652.
- Laggerbauer B, Murphy FL, Cech TR. 1994. Two major tertiary folding transitions of the *Tetrahymena* catalytic RNA. *EMBO J* 13:2669–2676.
- Lehnert V, Jaeger L, Michel F, Westhof E. 1996. New loop–loop tertiary interactions in self-splicing introns of subgroup IC and ID: A complete 3D model of the *Tetrahymena thermophila* ribozyme. *Chem & Biol* 3:993–1009.
- Li Y, Turner DH. 1997. Effects of Mg²⁺ and the 2'-OH of guanosine on steps required for substrate binding and reactivity with the *Tetrahymena* ribozyme reveal several local folding transitions. *Biochemistry* 36:11131–11139.
- Lin CW, Hanna M, Szostak JW. 1994. Evidence that the guanosine substrate of the *Tetrahymena* ribozyme is bound in the anti conformation and that N7 contributes to binding. *Biochemistry* 33:2703–2707.
- Lingbeck J, Kubinec MG, Miller J, Reid BR, Drobny GP, Kennedy MA. 1996. Effect of adenine methylation on the structure and dynamics of TpA steps in DNA: NMR structure determination of [d(CGAGGTTTAAACCTCG)]₂ and its A9-methylated derivative at 750 Mhz. *Biochemistry* 35:719–734.
- Lingner J, Keller W. 1993. 3' End labeling of RNA with recombinant yeast poly(A) polymerase. *Nucleic Acids Res* 21:2917–2920.
- Maidak BL, Olsen GJ, Larsen N, Overbeek R, McCaughey MJ, Woese CR. 1997. The RDP (Ribosomal Database Project). *Nucleic Acids Res* 25:109–111.
- McConnell TS, Cech TR. 1995. A positive entropy change for guanosine binding and for the chemical step in the *Tetrahymena* ribozyme reaction. *Biochemistry* 34:4056–4067.
- McConnell TS, Herschlag D, Cech TR. 1997. Effects of divalent metal ions on individual steps of the *Tetrahymena* ribozyme reaction. *Biochemistry* 36:8293–8303.
- McSwiggen JA, Cech TR. 1989. Stereochemistry of RNA cleavage by the *Tetrahymena* ribozyme and evidence that the chemical step is not rate-limiting. *Science* 244:679–683.
- Mei R, Herschlag D. 1996. Mechanistic investigations of a ribozyme derived from the *Tetrahymena* group I intron. Insights into catalysis and the second step of self-splicing. *Biochemistry* 35:5796–5809.
- Michel F, Hanna M, Green R, Bartel DP, Szostak JW. 1989. The guanosine binding site of the *Tetrahymena* ribozyme. *Nature* 342:391–395.
- Michel F, Westhof E. 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol* 216:585–610.
- Michel F, Westhof E. 1994. Slippery substrates. *Nature Struct Biol* 1:5–7.
- Moazed D, Stern S, Noller H. 1986. Rapid chemical probing of conformation in 16S ribosomal RNA and 30S ribosomal subunits using primer extension. *J Mol Biol* 187:399–416.
- Murphy FL, Cech TR. 1993. An independently folding domain of RNA tertiary structure within the *Tetrahymena* ribozyme. *Biochemistry* 32:5291–5300.
- Murphy FL, Cech TR. 1994. GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. *J Mol Biol* 236:49–63.
- Nakamura TM, Wang VH, Zaug AJ, Griffith JD, Cech TR. 1995. Relative orientation of RNA helices in a group I ribozyme determined by helix extension electron microscopy. *EMBO J* 14:4849–4859.
- Narlikar GJ, Khosla M, Usman N, Herschlag D. 1997. Quantitating tertiary binding energies of 2'-OH groups on the P1 duplex of the *Tetrahymena* ribozyme: Intrinsic binding energy in an RNA enzyme. *Biochemistry* 36:2465–2477.
- Peattie DA. 1979. Direct chemical method for sequencing RNA. *Proc Natl Acad Sci USA* 76:1760.
- Piccirilli JA, Vyle JS, Caruthers MH, Cech TR. 1993. Metal ion catalysis in the *Tetrahymena* ribozyme reaction. *Nature* 362:85–88.
- Pley HW, Flaherty KM, McKay DB. 1994. Model for an RNA tertiary interaction from the structure of an intermolecular complex between a GAAA tetraloop and an RNA helix. *Nature* 372:111–113.
- Proffen LA, Kierzek R, Testa SM, Turner DH. 1997. Guanosine binds to the *Tetrahymena* ribozyme in more than one step, and its 2'-OH and the nonbridging pro-S_P phosphoryl oxygen at the cleavage site are required for productive docking. *Biochemistry* 36:12477–12485.
- Pyle AM, Murphy FL, Cech TR. 1992. RNA substrate binding site in the catalytic core of the *Tetrahymena* ribozyme. *Nature* 358:123–128.
- Raines K, Gottlieb PA. 1998. Enzymatic incorporation of 2'-thio-CTP into the HDV ribozyme. *RNA* 4:340–345.
- Rajagopal J, Doudna JA, Szostak JW. 1989. Stereochemical course of catalysis by the *Tetrahymena* ribozyme. *Science* 244:692–694.
- Saenger W. 1984. *Principles of nucleic acid structure*. New York: Springer-Verlag.
- Sousa R, Padilla R. 1995. A mutant T7 RNA polymerase as a DNA polymerase. *EMBO J* 14:4609–4621.

- Stern S, Moazed D, Noller HF. 1989. Analysis of RNA structure using chemical and enzymatic probing monitored by primer extension. *Methods Enzymol* 164:481–489.
- Streicher B, Westhof E, Schroeder R. 1996. The environment of two metal ions surrounding the splice site of a group I intron. *EMBO J* 15:2556–2564.
- Strobel SA, Cech TR. 1993. Tertiary interactions with the internal guide sequence mediate docking of the P1 helix into the catalytic core of the *Tetrahymena* ribozyme. *Biochemistry* 32:13593–13604.
- Strobel SA, Cech TR, Usman N, Beigelman L. 1994. The 2,6-diaminopurine riboside·5-methylisocytidine wobble base pair: An isoenergetic substitution for the study of G·U pairs in RNA. *Biochemistry* 33:13824–13853.
- Strobel SA, Ortoleva-Donnelly L, Ryder SP, Cate JH, Moncoeur E. 1998. Complementary sets of noncanonical base pairs mediate RNA helix packing in the group I intron active site. *Nature Struct Biol* 5:60–66.
- Strobel SA, Shetty K. 1997. Defining the chemical groups essential for *Tetrahymena* group I intron function by nucleotide analog interference mapping. *Proc Natl Acad Sci USA* 94:2903–2908.
- Tanner MA, Anderson EM, Gutell RR, Cech TR. 1997. Mutagenesis and comparative sequence analysis of a base triple joining the two domains of group I ribozymes. *RNA* 3:1037–1051.
- Tanner MA, Cech TR. 1997. Joining the two domains of a group I ribozyme to form the catalytic core. *Science* 275:847–849.
- Uesugi S, Miki H, Ikehara M, Iwahashi H, Kyogoku Y. 1979. A linear relationship between electronegativity of 2'-substituents and conformation of adenine nucleosides. *Tet Lett* 42:4073–4076.
- von Ahsen U, Noller HF. 1993. Methylation interference experiments identify bases that are essential for distinct catalytic functions of a group I ribozyme. *EMBO J* 12:4747–4754.
- Wang JF, Cech TR. 1992. Tertiary structure around the guanosine-binding site of the *Tetrahymena* ribozyme. *Science* 256:526–529.
- Wang JF, Downs WD, Cech TR. 1993. Movement of the guide sequence during RNA catalysis by a group I ribozyme. *Science* 260:504–508.
- Waring RB. 1989. Identification of phosphate groups important to self-splicing of the *Tetrahymena* rRNA intron as determined by phosphorothioate substitution. *Nucleic Acids Res* 17:10281–10293.
- Weinstein LB, Jones BCNM, Cosstick R, Cech TR. 1997. A second catalytic metal ion in a group I ribozyme. *Nature* 388:805–808.
- Withers SG, Street IP, Percival MD. 1988. Fluorinated carbohydrates as probes of enzyme specificity and mechanism. *Fluorinated carbohydrates. Chemical and biochemical aspects*. Washington, DC: American Chemical Society. pp 59–77.
- Woese CR, Winker S, Gutell RR. 1990. Architecture of ribosomal RNA: Constraints on the sequence of "tetraloops." *Proc Natl Acad Sci USA* 87:8467–8471.
- Woodson SA, Cech TR. 1989. Reverse self-splicing of the *Tetrahymena* group I intron: Implication for the directionality of splicing and for intron transposition. *Cell* 57:335–345.
- Wyatt JR, Tinoco I. 1993. RNA structural elements and RNA function. In: Gesteland RF, Atkins JF, eds. *The RNA world*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 465–496.
- Yarus M, Illangesekare M, Christian E. 1991a. An axial binding site in the *Tetrahymena* precursor RNA. *J Mol Biol* 222:995–1012.
- Yarus M, Illangesekare M, Christian E. 1991b. Selection of small molecules by the *Tetrahymena* catalytic center. *Nucleic Acids Res* 19:1297–1304.