Identification of A-Minor Tertiary Interactions within a Bacterial Group I Intron Active Site by 3-Deazaadenosine Interference Mapping†

Juliane K. Soukup,‡,§ Noriaki Minakawa,|| Akira Matsuda,|| and Scott A. Strobel*,‡

Department of Molecular Biophysics and Biochemistry and Department of Chemistry, Yale University, 260 Whitney Avenue, New Haven, Connecticut 06520-8114, and Graduate School of Pharmaceutical Sciences, Hokkaido University, Kita-12, Nishi-6, Kita-ku, Sapporo 060-0812, Japan

Received April 8, 2002; Revised Manuscript Received June 20, 2002

ABSTRACT: The A-minor motifs appear to be the most ubiquitous helix packing elements within RNA tertiary structures. These motifs have been identified throughout the ribosome and almost every other tertiary-folded RNA for which structural information is available. These motifs utilize the packing of the donor adenosine's N1, N3, and/or 2′-OH against the 2′-OHs and minor groove edge of the acceptor base pair. The ability to identify biochemically which adenosines form A-minor motifs and which base pairs they contact is an important experimental objective. Toward this goal, we report the synthesis and transcriptional incorporation of 5′-*O*-(1-thio)-3-deazaadenosine triphosphate and its use in Nucleotide Analogue Interference Mapping (NAIM) and Nucleotide Analogue Interference Suppression (NAIS). This analogue makes it possible for the first time to explore the functional importance of the N3 imino group of adenosine in RNA polymers. Interference analysis of the group I self-splicing introns from *Tetrahymena* and *Azoarcus* indicates that A-minor motifs are integral to the helix packing interactions that define the 5′-splice site of the intron. Specifically, *Azoarcus* A58 in the J4/5 region contacts the G·U wobble pair at the cleavage site in the P1 helix, and *Azoarcus* A167 in the J8/7 region contacts the C13-G37 base pair in the P2 helix. Both of these structural features are conserved between the eukaryotic and bacterial introns. These results suggest that nucleotide analogue interference patterns can identify and distinguish A-minor interactions in RNA tertiary structure, particularly the most prevalent type I and type II varieties. Furthermore, clustering of 3-deazaadenosine interferences is suggestive of A patches, in which a series of consecutive A-minor motifs mediate helix packing. Biochemical identification of these interactions may provide valuable constraints for RNA structure prediction.

Structural evidence suggests that A-minor interactions constitute the most common tertiary motifs in RNA structure (*1, 2*). It has long been noted that adenosines are unusually abundant in the single-stranded or noncanonically paired regions of RNA secondary structure (*3, 4*). It now appears that these "unpaired" adenosines are the prime mediators of helix packing interactions in virtually every RNA for which structural information is available (*1*).

The motif is particularly abundant within the 16S and 23S rRNA tertiary folds of the ribosomal subunits (*5, 6*). For example, there are 170 A-minor motifs within the 50S ribosomal subunit, the vast majority of which are highly conserved across all kingdoms of life (*1*). These interactions also appear to play a major role in the RNA−RNA interactions at the interface of the 50S and 30S ribosomal subunits (*7*). The importance of the A-minor motif is prominently manifest in the role they play in mediating interactions of tRNA with the ribosome. The terminal adenosines of both the A-site and P-site tRNAs make A-minor interactions with 23S rRNA (*8*), and universally conserved adenosines in 16S rRNA decode the message by using consecutive A-minor motifs to contact the codon−anticodon helix in the ribosomal A-site (*9*). But the motif is not confined to the ribosome. Closer inspection of the P4−P6 crystal structure from the *Tetrahymena* group I intron revealed that A-minor interactions are at the heart of larger more specialized motifs, including the tetraloop−tetraloop receptor interaction and the A-rich bulge (*2, 10*). A-minor interactions are also responsible for helix packing interactions in the Hepatitis Delta Virus and hairpin ribozymes (*11, 12*).

There are four variants of the A-minor motif, termed types 0, I, II, and III (*1*) (Figure 1). In each case, the minor groove edge of the donor adenosine packs against the minor groove edge of the acceptor duplex. Each motif is classified by the way the donor adenosine interacts with the receptor base pair. The type I and type II motifs are highly specific for the adenine base (Figure 1) (*1*). The type 0 and type III interactions are substantially less adenosine-specific because the adenine base lies outside the minor groove edge of the base pair (Figure 1) (*1*). The donor adenosines are located either in single-stranded segments or within a variety of noncanonical pairings, including sheared or reverse Hoogsteen interactions (*1*). The adenosine strand runs antiparallel to the near strand in the acceptor base pair. The donor adenosine provides an extensive surface for van der Waals interactions, as a result of the nonfunctionalized C2 carbon of the heter-

† This work was supported by an NIH postdoctoral fellowship to J.K.S. and by NIH Grant GM54839 to S.A.S.

* Correspondence should be addressed to this author. Phone: (203)-432-9772, Fax: (203)432-5767, Email: strobel@csb.yale.edu.

‡ Yale University.

§ Current address: Department of Chemistry, Creighton University, 2500 California Plaza, Omaha, NE 68178.

|| Hokkaido University.

10.1021/bi020265l CCC: $22.00
Published on Web 07/23/2002

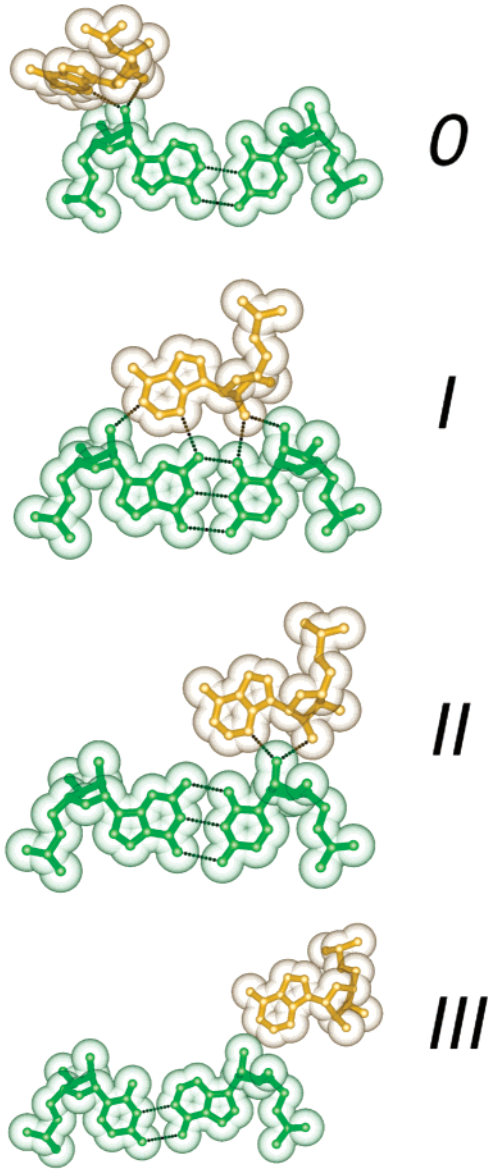

FIGURE 1: The four A-minor motifs identified within RNA tertiary structures. The motifs are shown in surface representation with the donor adenosine in yellow and the acceptor base pair in green. The motifs are distinguished by the position of the donor adenosine relative to the positions of the two 2′-OH groups of the receptor base pair. Types I and II show strong specificity for adenosine as the donor residue. Type I shows specificity for an acceptor C-G base pair (shown) in preference to the other three pairs (*1*). The hydrogen bonds between the nucleotides are indicated as dashed lines. Adapted from Nissen et al., 2001 (*1*).

ocyclic ring, and an ideal hydrogen bonding interface on the minor groove edge of the adenosine involving the N1 and N3 imino groups and the 2′-OH (*1*, *2*). Together these interactions provide a substantial quantity of tertiary stabilization energy (*2*).

Group I introns are proposed to utilize A-minor motifs for helix packing interactions within its active site (*13*, *14*) (Figure 2). While a low-resolution structure of a truncated variant of the *Tetrahymena* intron lacking the substrate and other helices has been reported (*15*), a high-resolution structure of a conformationally active group I intron has remained elusive. As a result, the molecular interactions responsible for 5′-splice site selection and helix packing in the catalytic core have been modeled based upon biochemical

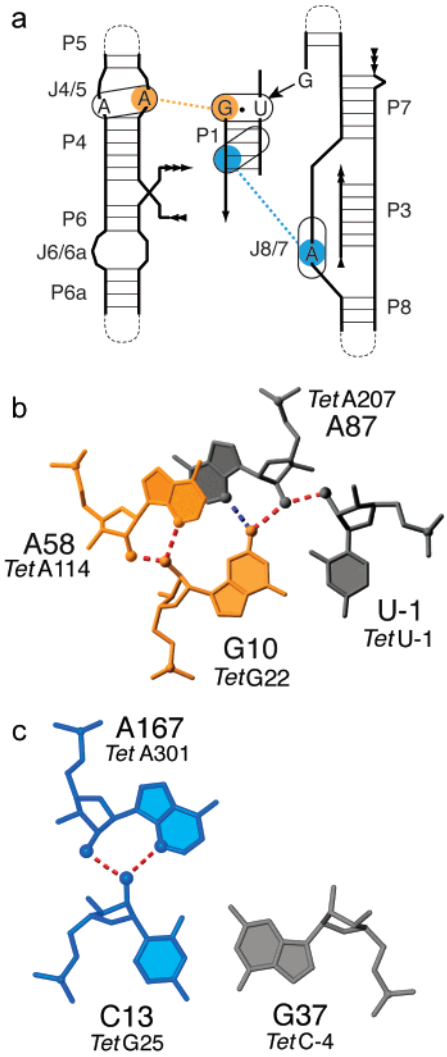

FIGURE 2: (a) Generic diagram of group I intron secondary structure showing the P1 to J4/5 (orange) and P1 to J8/7 (blue) tertiary interactions (straight dashed lines). Connectivity throughout the RNA is depicted with one, two, or three arrowheads. The variable loop segments are shown as curved dashed lines. (b) Model of the G·U wobble receptor between P1 and J4/5 emphasizing the type II A-minor motif (orange) between *Azoarcus* A58 and G10 (*Tet*A114 and G22) that comprises part of the interaction. In addition to the type II A-minor motif, a hydrogen-bonding network between A87 (*Tet*207) and the G10·U-1 (*Tet*G22·U-1) wobble pair defines the group I intron 5′-splice site. A58 and A87 are involved in sheared A·A pairs with A86 and A57 (*Tet*A113 and A206), respectively (not shown). The hydrogen bonds that mediate the J4/5−P1 tertiary interaction are indicated as dashed lines. Hydrogen bonds supported by an interference suppression (NAIS) experiment in this or a previous study (*13*, *47*) are shown in red. The interaction between the 2′-OHs of U-1 and A87 (*Tet*207) is specific to the transition state (*47*). The remaining hydrogen bond that is consistent with primary interference (NAIM) results and inferred from modeling is shown in blue. For technical reasons, this contact could not be tested in either the *Tetrahymena* (c^3AαS not incorporated) or the *Azoarcus* intron (insufficient signal with 2′-deoxy U-1 substrate). (c) Model of an A-minor motif between the P2 helix (*Tetrahymena* P1 helix) and the J8/7 region (blue). *Azoarcus* A167 (*Tet*301) utilizes its 2′-OH and N3 to contact the 2′-OH of C13 (*Tet*G25). The formation of both hydrogen bonds (red) is experimentally supported by this study. However, it is possible that *Azoarcus* A167 forms these interactions within a type 0 motif, rather than the type II motif depicted in this figure.

rather than crystallographic data. The 5′-exon base-pairs with a complementary sequence in the intron, termed the internal

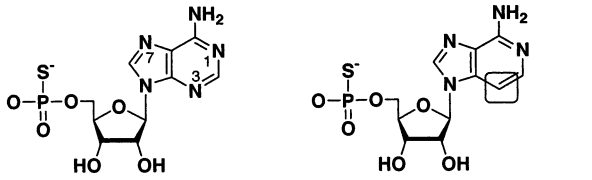

FIGURE 3: Adenosine and 3-deazaadenosine (c^3A). The N3 to C substitution in c^3A is highlighted by an oval. Both nucleotides are shown as the monophosphorothioate (AαS and $c^3A\alpha S$, respectively), the form in which they are incorporated into the RNA. Numbering of the heterocyclic bases is indicated on the adenosine.

guide sequence (IGS), to form the P1 or substrate helix (*16*). This helix is proposed to make an extensive set of interactions with adenosines that line the catalytic core of the ribozyme (*13*, *14*). The G·U wobble pair immediately adjacent to the cleavage site is recognized by two consecutively stacked sheared A·A pairs in the J4/5 region (*13*) (Figure 2a,b), while residues in P1 more distal from the cleavage site are recognized by adenosines in the single-stranded J8/7 segment (Figure 2a,c) (*14*).

Interacting pairs of 2′-OH groups involved in P1 helix docking within the *Tetrahymena* intron have been defined by Nucleotide Analogue Interference Mapping (NAIM)[1] and Nucleotide Analogue Interference Suppression (NAIS) analysis (*13*, *14*). The 2′-OH of the wobble pair G (*Tet*G22) at the cleavage site hydrogen-bonds to the 2′-OH of *Tet*A114 in J4/5 (Figure 2b); and a 2′-OH in the IGS distal from the cleavage site (*Tet*G25) hydrogen-bonds to the 2′-OH *Tet*A301 in J8/7 (Figure 2c). The nucleotides involved in these interactions are highly conserved among all introns in the group I family (*17*, *18*). Interference data and molecular modeling suggested that these interactions include what has now been termed a type II A-minor motif (*1*), wherein both the 2′-OH and the N3 imino groups of the adenosines form hydrogen bonds to the P1 hydroxyl groups (*13*, *14*). However, participation by the N3 groups of the putative A-minor adenosines in this motif has not been investigated.

The lack of information about the proposed A-minor contacts in the group I active site reflects a larger biochemical deficiency. The functional role of the adenosine N3 imino group is not amenable to investigation by any biochemical approach reported to date, in part because the N3 is not a site of efficient chemical modification in RNA (*19*). Given the prevalence of the A-minor motif in many RNA tertiary structures, the development of an efficient biochemical assay to identify the hydrogen bonds in these minor groove helix packing interactions is warranted. A nucleotide analogue that would provide such information is 3-deazaadenosine (c^3A), which has a carbon in place of the minor groove N3 imino group (Figure 3). This substitution would disturb the hydrogen bonding surface of the donor adenosine, a feature that appears to be of particular importance for the type II A-minor motif.

[1] Abbreviations: c^3A, 3-deazaadenosine[c3]; ATPαS, 5′-*O*-(1-thio)-3-deazaadenosine triphosphate[c3]; AαS, 3-deazaadenosine 5′-monophosphorothioate nucleotide as incorporated into RNA by T7 RNA polymerase; AαS, adenosine 5′-monophosphorothioated; AαS, 2′-deoxyadenosine 5′-monophosphorothioate; DAPαS, diaminopurine riboside 5′-monophosphorothioate; PAGE, polyacrylamide gel electrophoresis; NAIM, Nucleotide Analogue Interference Mapping; NAIS, Nucleotide Analogue Interference Suppression.

As a rapid means to screen all the positions in an RNA for minor groove effects, we postulated that the phosphorothioate-tagged version of c^3A could be utilized in a NAIM strategy (*20*, *21*). This application requires the synthesis of the α-phosphorothioate-tagged triphosphate of c^3A (c^3-ATPαS), and the ability of the T7 RNA polymerase to utilize the analogue as a substrate for random incorporation at adenosines within the RNA transcript. After selecting the active RNA variants in the population, the nucleotide positions where the N3 group is important for activity can be determined readily by cleavage of the phosphorothioate linkage with iodoethanol (*22*) and resolution of the cleavage products by polyacrylamide gel electrophoresis (PAGE). A gap in the sequencing ladder defines those positions intolerant of $c^3A\alpha S$ substitution (*20*). In this way, it would be possible to simultaneously, yet individually, assess the functional importance of every N3 imino group of adenosine within an RNA.

To implement this scheme, we utilized the self-splicing RNAs from *Tetrahymena* and the purple bacteria *Azoarcus* (*23*, *24*). The *Tetrahymena* intron has become the archetype for the group I intron family and has been characterized by a wide variety of methods (*16*). The *Azoarcus* intron is located within the anticodon loop of pre-$tRNA^{Ile}$ and is the smallest self-splicing intron identified to date (*24*, *25*). It is less than half the size of the *Tetrahymena* intron primarily due to its lack of peripheral domains. The small size of the *Azoarcus* intron has made it ideal for NAIM and NAIS studies, where individual nucleotide resolution of the interference data is critical (*26*). Based upon analogy to the *Tetrahymena* intron and consistent with 2′-deoxyadenosine (dAαS) NAIM analysis, *Azo*A58 in J4/5 and *Azo*A167 in J8/7 are predicted to form type II A-minor contacts with G10 and C13 in the substrate helix (*26*) (Figure 2).

Here we report the synthesis of the phosphorothioate-tagged triphosphate of c^3A, conditions for its incorporation into in vitro RNA transcripts, and its use in NAIM and NAIS analysis of the *Azoarcus* and *Tetrahymena* group I intron active sites. These studies define a biochemical signature that can be used to identify A-minor interactions in other RNA tertiary structures.

MATERIALS AND METHODS

Synthesis of c^3ATPαS. General Methods. Physical data were measured as follows: ^1H and ^{13}C NMR spectra were recorded on 125, 270, or 500 MHz instruments in $CDCl_3$ or D_2O. Tetramethylsilane or 3-(trimethylsilyl)propionic acid sodium salt was used as an internal standard. ^{31}P NMR spectra were recorded on a 109 MHz instrument in D_2O with H_3PO_4 as an internal standard. Chemical shifts are reported in parts per million (δ), and signals are expressed as s (singlet), d (doublet), m (multiplet), or br (broad). Thin-layer chromatography was performed on Merck Kieselgel F254 precoated plates. Silica gel YMC 60A (70−230 mesh) was used for column chromatography.

*4-[(Dibutylamino)methylene]amino-1-(2,3-di-O-acetyl-5-O-dimethoxytrityl-β-*D*-ribofuranosyl)imidazo[4,5-c]pyridine (**3**)*. To a suspension of 3-deazaadenosine (**1**) (*27*) (200 mg, 0.75 mmol) in dry DMF (3 mL) was added *N*,*N*-dibutylformamide dimethylacetal (0.54 mL, 2.26 mmol), and the mixture was stirred at room temperature for 3 days (*28*,

29). The solvent was removed in vacuo, and the residue was coevaporated with toluene to give crude **2**. A mixture of the resulting **2** and dimethoxytrityl chloride (762 mg, 2.25 mmol) in dry pyridine (4 mL) was stirred at room temperature for 2.5 h, and then acetic anhydride (0.7 mL, 7.5 mmol) and 4-(dimethylamino)pyridine (10 mg) were added to the mixture, which was stirred at the same temperature for 24 h. The reaction was quenched by addition of EtOH, and the solvent was removed in vacuo. The residue was partitioned between $CHCl_3$ and H_2O, and the organic layer was washed with saturated aqueous $NaHCO_3$, followed by brine. The organic layer was dried (Na_2SO_4) and concentrated in vacuo. The residue was purified through a silica gel column, eluted with 0−6% EtOH in $CHCl_3$, to give **3** (400 mg, 67% as a brown oil): FAB-LRMS *m*/*z* 792 (MH^+); FAB-HRMS calcd for $C_{45}H_{54}N_5O_8$ (MH^+) 792.3972, found 792.3949; ^{1}H NMR ($CDCl_3$) δ: 8.66 (s, 1 H), 7.99 (s, 1 H), 7.86 (d, 1 H, J = 5.9 Hz), 7.44 (m, 2 H), 7.28 (m, 7 H), 7.10 (d, 1 H, J = 5.9 Hz), 6.82 (m, 4 H), 6.02 (d, 1 H, J = 7.9 Hz), 5.87 (dd, 1 H, J = 5.2, 7.9 Hz), 5.66 (dd, 1 H, J = 2.7, 5.2 Hz), 4.31 (m, 1 H), 3.78 and 3.77 (each s, each 3 H), 3.67 (m, 2 H), 3.49 (br s, 2 H), 3.32 (m, 2 H), 2.14 and 2.03 (each s, each 3 H), 1.61 (m, 4 H), 1.26 (m, 4 H), 0.93 (m, 6 H); ^{13}C NMR ($CDCl_3$) δ: 169.5, 169.0, 158.6, 156.1, 155.8, 143.9, 141.5, 139.8, 138.9, 135.2, 135.0, 133.6, 130.2, 130.1, 128.2, 127.9, 127.0, 113.2, 100.9, 87.1, 86.0, 82.5, 72.7, 71.2, 62.8, 55.1, 51.2, 44.8, 31.1, 29.2, 20.5, 20.2, 19.7, 13.9, 13.6.

*4-[(Dibutylamino)methylene]amino-1-(2,3-di-O-acetyl-β-*D*-ribofuranosyl)imidazo[4,5-c]pyridine (**4**).* To a solution of **3** (94 mg, 0.12 mmol) in dry CH_2Cl_2 (3 mL) was added trichloroacetic acid (30 mg, 0.18 mmol), and the mixture was stirred at room temperature for 1 h. The reaction was quenched by addition of saturated aqueous $NaHCO_3$, and the reaction mixture was diluted with $CHCl_3$. The separated organic layer was washed with H_2O and brine. The organic layer was dried (Na_2SO_4) and concentrated in vacuo. The residue was purified over a silica gel column, eluted with 0−8% EtOH in $CHCl_3$, to give **4** (34 mg, 59% as a pale yellow oil): FAB-LRMS *m*/*z* 490 (MH^+); FAB-HRMS calcd for $C_{24}H_{36}N_5O_6$ (MH^+) 490.2665, found 490.2651; ^{1}H NMR ($CDCl_3$) δ: 8.61 (s, 1 H), 8.45 (s, 1 H), 8.02 (d, 1 H, J = 5.3 Hz), 7.12 (d, 1 H, J = 5.3 Hz), 6.01 (d, 1 H, J = 6.0 Hz), 5.63 (dd, 1 H, J = 5.3, 6.0 Hz), 5.48 (dd, 1 H, J = 4.0, 5.3 Hz), 4.26 (m, 2 H), 3.94 (m, 2 H), 3.65 (m, 2 H), 3.30 (m, 2 H), 2.05 and 2.00 (each s, each 3 H), 1.59 (m, 4 H), 1.31 (m, 4 H), 0.90 (m, 6 H); ^{13}C NMR ($CDCl_3$) δ: 169.7, 169.2, 156.1, 155.5, 141.4, 140.4, 139.0, 133.1, 100.6, 86.9, 83.9, 74.3, 71.1, 61.0, 51.4, 44.8, 31.1, 29.2, 20.5, 20.3, 20.2, 19.8, 13.9, 13.7.

*3-Deazaadenosine 5′-O-(1-Thiotriphosphate) (**7**).* Compound **4** (72 mg, 0.15 mmol) was dissolved in a mixture of dry pyridine (150 μL) and dry dioxane (450 μL). A freshly prepared 1 M solution of 2-chloro-4*H*-1,3,2-benzodioxaphosphorin-4-one in dry dioxane (33 mg in 165 μL) was added to the mixture. After being stirred at room temperature for 10 min, a 0.5 M solution of bis(tri-*n*-butylammonium)pyrophosphate in dry DMF (600 μL) and tri-*n*-butylamine (200 μL) were added to the reaction mixture. After 10 min, a suspension of sulfur (9.6 mg, 0.3 mmol) in dry DMF (300 μL) was added, and the reaction was stirred for an additional 30 min. Water (5 mL) was added, and the reaction mixture was stirred for 1 h. The solvent was removed in vacuo, and the residue was treated with ammonium hydroxide (28%, 25 mL) at 55 °C for 48 h in a steel container. The reaction mixture was concentrated in vacuo, the residue was dissolved in H_2O (200 mL), and the solution was applied to a DEAE-Sephadex A-25 column (2.3 × 30 cm). Chromatography was performed with a linear gradient of 1 L each of 0.05 and 1 M TEAB buffer. Fractions containing the desired product were combined and concentrated to dryness in vacuo to give **7** as a triethylammonium salt, which was converted to the sodium salt with DIAINO WK20 (Na^+ form, 39 mg, 51%) as a mixture of diastereomers: FAB-LRMS (negative) *m*/*z* 521 [$(M-H)^-$]; FAB-HRMS (negative) calcd for $C_{11}H_{16}N_4O_{12}P_3S$ [$(M-H)^-$] 520.9698, found 520.9678; ^{1}H NMR (D_2O, sodium salts) δ: 8.57 (s, 0.5 H), 8.54 (s, 0.5 H), 7.63 (br s, 1 H), 7.21 (br s, 1 H), 5.96 (m, 1 H), 4.85 (m, 2 H), 4.46 (br s, 1 H), 4.43 (m, 2 H); ^{31}P NMR (D_2O, triethylammonium salts) δ: 44.0 (br d, J = 26.1), −7.0 (br d, J = 17.5), −22.1 (m).

Ribozyme Preparation. Azoarcus L-9 G206 and *Tetrahymena* L-21 G414 group I ribozymes were prepared by in vitro transcription using *Ear*I-cut plasmid pUCL-9 G206 and pUCL-21G414, respectively. c^3AαS or AαS was randomly incorporated into the RNAs by inclusion of the triphosphate in the transcription reaction. c^3AαS containing RNAs were transcribed in 40 mM Tris-HCl, pH 7.5, 4 mM spermidine, 10 mM DTT, 15 mM $MgCl_2$, 1 mM $Mn(OAc)_2$, 0.05% Triton X-100, 0.05 μg/μL DNA template, the Y639F mutant form of RNA polymerase, and final triphosphate concentrations of 0.25 mM ATP and 1 mM each of GTP, CTP, UTP, and c^3ATPαS. AαS was incorporated as previously described except the final concentration of ATPαS was 0.01 mM in the presence of 1 mM GTP, CTP, UTP, and ATP. This produced a level of approximately 1% analogue incorporation (1 substitution per 100 adenosine residues incorporated) to match that achieved for c^3AαS. RNAs were purified by PAGE (6% denaturing), eluted into 10 mM Tris-HCl, pH 7.5, 0.1 mM EDTA (TE), precipitated with ethanol, resuspended in TE, and stored at −20 °C.

Interference suppression analysis of the bacterial intron was performed using an L-6 G206 ribozyme form with specific functional group substitutions introduced at positions G10, C13, or C14. These RNAs were prepared by enzymatic ligation of a synthetic oligonucleotide onto the 5′-end of an L-22 G206 transcript prepared using the *Ear*I-cut plasmid pUC L-22 G206 as a template (*26*). AαS, dAαS, or c^3AαS substitutions were randomly incorporated during transcription as described above or in Ryder et al. (*30*) except 10 mM GMP was included in the transcription reaction. This produced L-22 G206 molecules beginning with a 5′-monophosphate for use in ligation with T4 DNA ligase (*31*). The L-22 G206 RNAs were purified by PAGE, eluted into TE, ethanol-precipitated, and resuspended in TE. Full-length (L-6 G206) RNAs were prepared by ligating onto the 5′-end of the transcript a 16-mer synthetic oligonucleotide containing a single 2′-deoxy or inosine substitution at G10 (UUUd-GUGCCUUGCGCCG or UUUIUGCCUUGCGCCG), or a 2′-deoxy substitution at C13 (5′-UUUGUGdCCUUGCGCCG) or C14 (5′-UUUGUGCdCUUGCGCCG). Ligation reactions were performed and RNAs purified as described (*26*).

Interference Reactions and Quantitation. Azoarcus L-9 G206 RNAs (250 nM final concentration) were preincubated in 25 mM MES, pH 6.0, 3 mM $MgCl_2$, and 1 mM $Mn(OAc)_2$

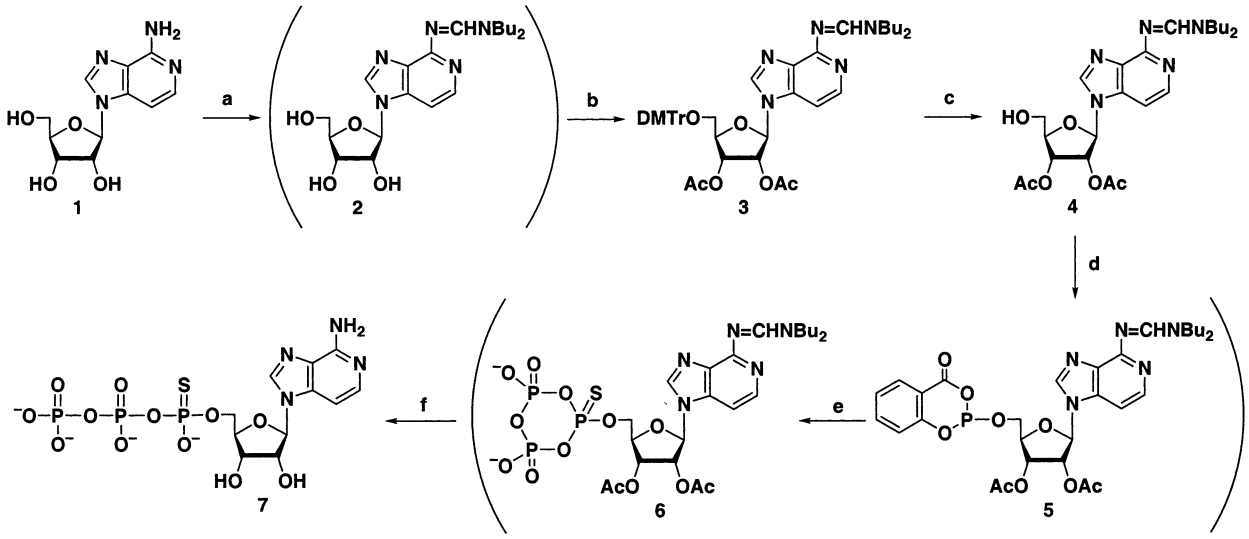

FIGURE 4: Scheme for the synthesis of c^3ATPαS. Reagents: (a) *N*,*N*-dibutylformamide dimethylacetal, DMF, rt, 3 days; (b) DMTrCl, pyridine, rt, 2.5 h, then Ac_2O, DMAP, 24 h; (c) Cl_3CCOOH, CH_2Cl_2, rt, 1 h; (d) 2-chloro-4*H*-1,3,2-benzodioxaphosphorin-4-one, dioxane-pyridine, rt, 10 min; (e) (i) bis(tri-*n*-butylammonium)pyrophosphate, tri-*n*-butylamine, DMF, rt, 10 min; (ii) sublimed sulfur, rt, 30 min; (f) (i) H_2O, rt, 1 h; (ii) NH_4OH, 55 °C, 48 h.

at 50 °C for 5 min and cooled to 30 °C for 2 min. The substrate oligonucleotide, CAUAAA, was radiolabeled at its 3′-terminus with [α-^{32}P]cordycepin using poly-A polymerase and gel-purified (*32*). It was dissolved in the same reaction buffer as the ribozyme, warmed to 30 °C for 2 min, combined with the ribozyme solution, and allowed to react at 30 °C for 4 min. The reactions were quenched by addition of 2 volumes of stop buffer (95% formamide, 50 mM EDTA, 0.01% bromphenol blue, 0.01% xylene cyanol) and split into two tubes. One-tenth volume of 100 mM iodine in ethanol was added to one tube to cleave the phosphorothioate linkages (*22*). The samples were heated to 90 °C for 2 min and the cleavage products resolved by 6% denaturing PAGE. Reactions containing no iodine were run in parallel to confirm that the cleavage pattern was specific to the iodine treatment and not due to nonspecific degradation.

Tetrahymena L-21 G414 RNAs (50 nM final concentration) were preincubated in 50 mM HEPES, pH 7.5, 3 mM $MgCl_2$, and 1 mM $Mn(OAc)_2$ at 50 °C for 10 min. The 3′-end-labeled substrate oligonucleotide, CCCUCdTAAAAAA, was dissolved in the same reaction buffer as the ribozyme, warmed to 50 °C for 2 min, combined with the ribozyme solution, and allowed to react at 50 °C for 10 min. The reactions were stopped and prepared for separation by gel electrophoresis as described above.

To control for sequence-dependent variability in analogue incorporation at various positions along the RNA, each ribozyme was radiolabeled at its 5′-end and analyzed as previously described (*26*). Individual band intensities for both the parental nucleotides and the nucleotide analogues were quantitated by PhosphorImager analysis at each nucleotide position in both the 3′-exon ligation experiments and the unselected 5′-labeled control RNAs (*18*). These values were used to calculate the extent of analogue interference (κ) at each position as previously described (*18*). In this calculation, a κ value of 1 indicates no analogue interference at the position, while κ values of 2 or greater indicate sites of interference.

Interference Suppression Experiments. Interference suppression analysis was performed using chimeric RNAs with site-specific substitutions. These ribozymes were impaired in their reactivity, so the reaction time was extended to obtain a reactivity level comparable to that of the wild-type unmodified ribozyme. All reactions included 250 nM L-6 G206 ribozyme, the 3′-end-labeled substrate (CAUAAAp*A_H) in 25 mM HEPES, pH 7.0, and divalent cations [28 mM $MgCl_2$ and 2 mM $Mn(OAc)_2$ for c^3AαS suppression or 3 mM $MgCl_2$ and1 mM $Mn(OAc)_2$ for dAαS suppression]. Ribozymes were preincubated at 50 °C for 5 min and allowed to slow-cool to room temperature. For the c^3AαS interference suppression, the chimeric ribozymes and substrates were cooled for 10 min, mixed, and incubated for 60 min on ice. The ribozymes without site-specific substitutions were incubated with substrate for 20 min on ice. For the dAαS interference suppression, the chimeric ribozymes were incubated with substrate for 20 min at 30 °C, and the unsubstituted RNAs were incubated for 4 min. In each case, the reaction workup and quantitation were performed as described above.

RESULTS

Synthesis of c^3ATPαS. The synthesis of c^3ATPαS could not be achieved using the standard protocol, which typically involves modification of the unprotected nucleoside with $PSCl_3$ followed by addition of pyrophosphate (*33*). This was due apparently to low nucleophilicity of the 5′-OH relative to other positions on the nucleoside. Thus, it was essential to utilize a more reactive phosphitylating reagent and to protect not only the 2′- and 3′-OH groups, but also the exocyclic amino group. For this reason, we utilized the approach reported by Ludwig and Eckstein using 2-chloro-4*H*-1,3,2-benzodioxaphosphorin-4-one (Figure 4) (*34*). Protection of the amino group with a dimethylaminomethylene group appeared to be ideal because of its selectivity for the exocyclic amino group and the ease of its removal under basic conditions. However, when **1** was treated with *N*,*N*-dimethylformamide dimethylacetal in DMF, the desired dimethylaminomethylene protected nucleoside was contaminated with side products that could not be separated. Instead, the selective protection of the amino group was achieved

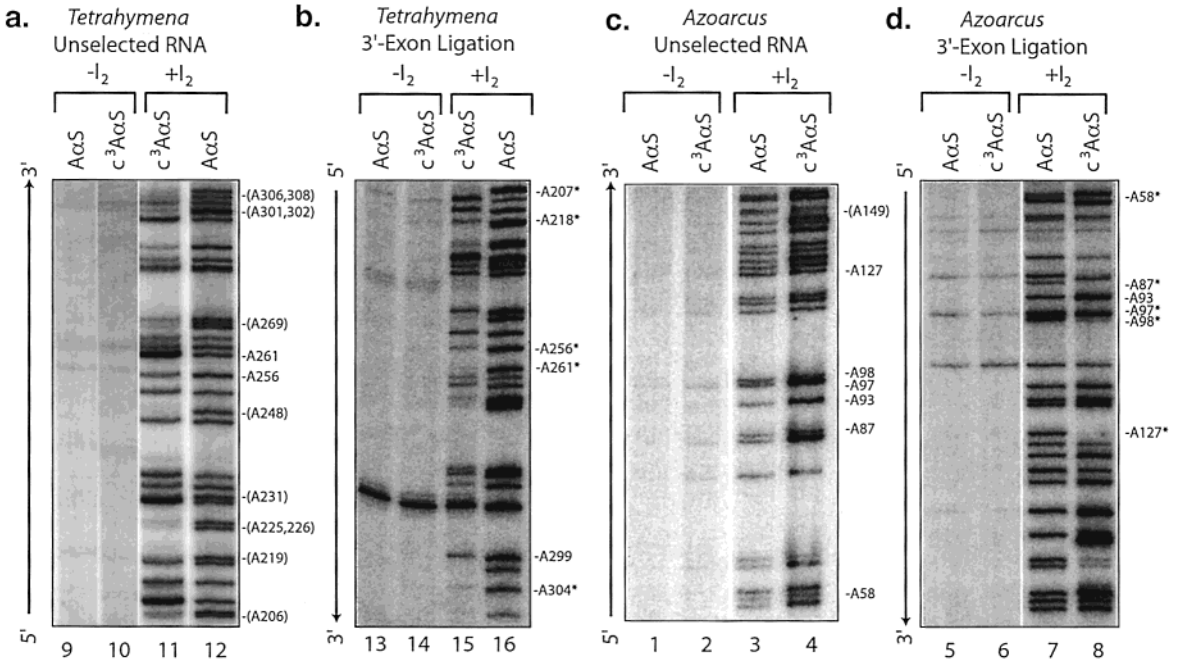

FIGURE 5: Interference analysis with $c^3A\alpha S$. (a) Incorporation of $c^3A\alpha S$ into the *Tetrahymena* L-21 G414 ribozyme. Autoradiogram of the unselected 5′-end-labeled *Tetrahymena* L-21 G414 RNA shows the distribution and efficiency of $c^3A\alpha S$ incorporation throughout the intron. Sites that were insufficiently incorporated to facilitate analysis are indicated by parenthesis around the nucleotide number. (b) Interference analysis of $c^3A\alpha S$ within the *Tetrahymena* intron. Active RNAs were selected using the 3′-exon ligation reaction. Sites of interference are highlighted with an asterisk. (c) Same as panel a except for the *Azoarcus* L-9 G206 ribozyme. (d) Same as panel b for the *Azoarcus* ribozyme. In each panel, arrows to the left of each gel denote the direction of the sequence from 5′ to 3′. Note that because the unselected control RNA is 5′-end-labeled and the 3′-exon ligation reaction produces a 3′-end-labeled product, the direction of the sequence between the two panels (a and b, c and d) is reversed. Weak interference at *Azoarcus* A97 is somewhat difficult to visualize from direct inspection of these panels. $c^3A\alpha S$ is hyper-incorporated at A97 in the unselected control, yet the relative signal intensity is reduced in the 3′-exon ligation reaction (for example, compare A97 to A93 in both panels).

using *N*,*N*-dibutylformamide dimethylacetal (*28*, *29*) to give **2**, which could be deprotected with ammonium hydroxide at 55 °C. Unpurified **2** was successively treated with dimethoxytrityl chloride, followed by acetic anhydride in pyridine to give fully protected **3** in 67% yield after 3 steps. The dimethoxytrityl group was removed selectively by treating with trichloroacetic acid in dichloromethane to give **4** in 59% yield. The synthesis of **7** was achieved by treatment of **4** with 2-chloro-4*H*-1,3,2-benzodioxaphosphorin-4-one following the Ludwig and Eckstein protocol (*34*) via the intermediates **5** and **6**. Compound **7** was isolated in 51% yield as a mixture of diastereomers after chromatography on DEAE-Sephadex.

Transcriptional Incorporation of $c^3A\alpha S$. To be used in NAIM, $c^3A\alpha S$ must be incorporated into the RNA polymer by in vitro transcription. Simple inspection of the analogue suggested that it might present a challenge to the polymerase. During standard transcriptional elongation, the N3 imino group of A may be recognized by the RNA polymerase at the nucleotide insertion and/or the elongation step (*35*). As a result, eliminating the N3 functional group may severely affect the ability of the polymerase to recognize the nucleotide as a substrate for nucleotide addition or as a primer for transcriptional elongation. This might cause the nucleotide to act as an inhibitor of transcription. Furthermore, elimination of the N3 imino group in c^3A raises the pK_a of the remaining N1 imino group to 7.0 compared to 3.5 for adenosine (*36*). Thus, at the near-neutral pHs used for transcription, a larger fraction of the nucleotide may be in its protonated form, which might adversely affect the ability of c^3A to compete effectively with A during the doped transcription required for NAIM experiments.

Despite these challenges we set out to determine if $c^3A\alpha S$ could be incorporated into RNA as an adenosine by T7 RNA polymerase. *Azoarcus* or *Tetrahymena* group I intron RNAs were transcribed at various concentrations of $c^3ATP\alpha S$ and ATP with either the wild-type or the Y639F mutant form of T7 RNA polymerase (*37*). The Y639F polymerase mutant efficiently incorporates 2′-deoxynucleotides into RNA due to reduced chemical discrimination in the minor groove (*38*). Given that c^3A is also a minor groove substitution, we reasoned that this polymerase mutation might facilitate $c^3A\alpha S$ incorporation as well. RNAs were transcribed, purified, and 5′-end-labeled. The phosphorothioate linkages were cleaved with iodine and the cleavage products resolved by PAGE to reveal the sites of $c^3A\alpha S$ incorporation (*Tetrahymena* intron data Figure 5a, *Azoarcus* intron data Figure 5c).

The $c^3A\alpha S$ incorporation was rather inefficient. The Y639F mutant polymerase appeared to incorporate the analogue slightly better than the wild-type polymerase (unpublished data, E. McMillan and S.A.S.), so all RNAs were synthesized with the mutant enzyme. Inclusion of 1 mM Mn^{2+} in the transcription buffer also promoted inclusion of the analogue into the RNA polymer (*39*). As expected, $c^3A\alpha S$ was incorporated exclusively at adenosines within the sequence (Figure 5a), but a high ratio of $c^3ATP\alpha S$ to ATP (1 mM $c^3ATP\alpha S$ and 0.25 mM ATP) was required to achieve even 1−2% analogue incorporation (based upon comparison of band intensity from an RNA transcribed with a 1:50 ratio of ATPαS (S_P diastereomer) and ATP).

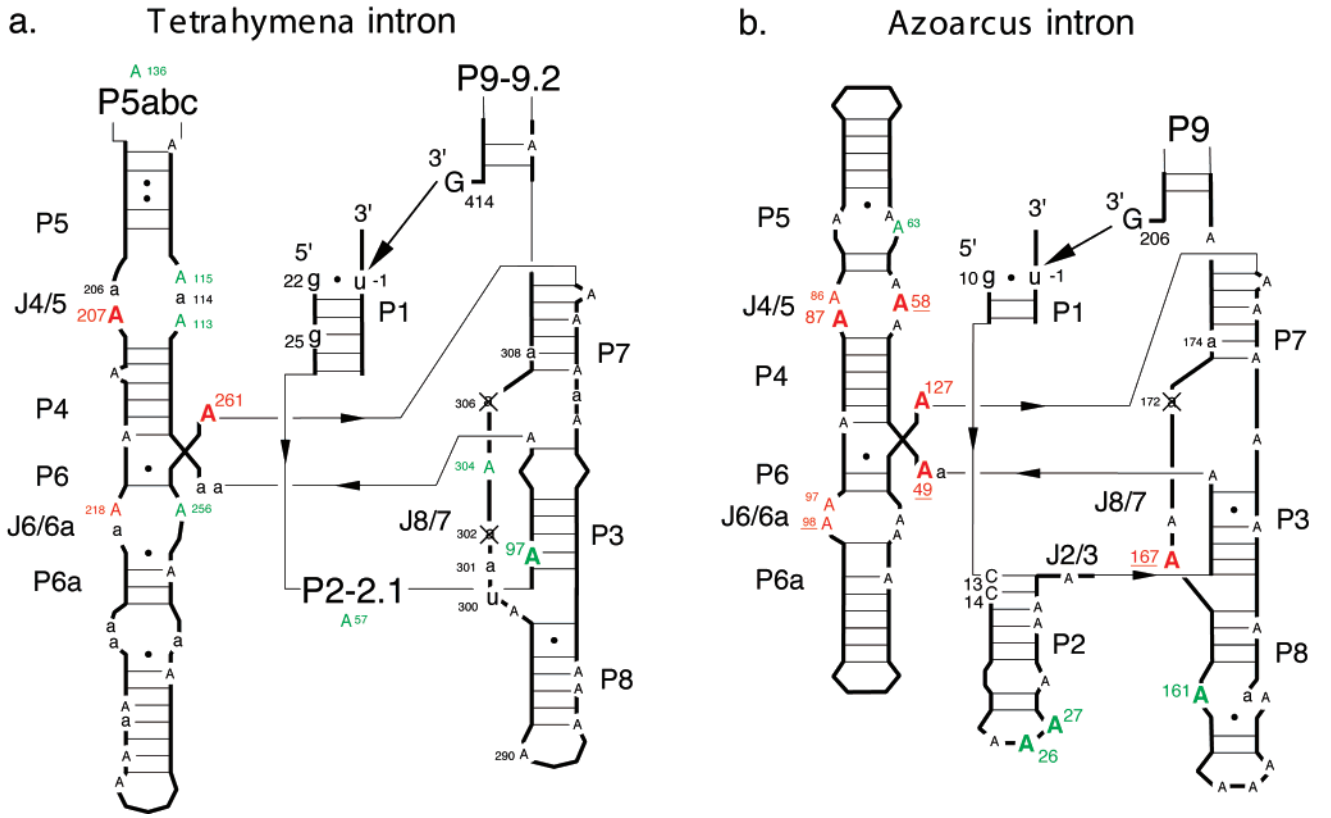

FIGURE 6: Summary of interference data for two group I introns. (a) *Tetrahymena* L-21 G414 ribozyme. (b) *Azoarcus* L-9 G206 ribozyme. Sites of $c^3A\alpha S$ interference observed in both introns are colored red. In cases where $c^3A\alpha S$ was not incorporated into the equivalent site in the other intron, the interference is also indicated in red, but the nucleotide number is underlined. Nucleotides are colored green if interference occurred at an A in one intron for which there is no equivalent A in the other intron (*Azo*26, 27, 63, 161 and *Tet*57, 97, 136, 256, 304). Nucleotides are also colored green if interference was observed in one intron, but not the other (*Tet*113 and 115). Positions of moderate to strong $c^3A\alpha S$ interference ($k \geq 4$) are depicted in large capital red or green letters. Positions of weak to moderate interference ($2 \leq k \geq 4$) are depicted in small capital red or green letters. Adenosine positions where $c^3A\alpha S$ was incorporated but where no interference was observed ($0.5 \leq k \geq 2$) are shown as small black capital letters. Positions where $c^3A\alpha S$ was not incorporated sufficiently to facilitate analysis are depicted in small case black letters. An X is placed over those residues where the phosphorothioate tag caused sufficient interference to make the site uninformative for $c^3A\alpha S$. The G nucleophile, the G·U wobble pair at the reaction center, and a few other nucleotides that play an important role in these experiments are also shown. Heavy lines represent the schematic secondary structure of the intron; thin lines illustrate connectivity between sequence elements. The peripheral extensions P5abc and P9−P9.2 (*Tetrahymena*) or P9 (Azoarcus) are present in the constructs used in these experiments, but they are not shown. Although not depicted in the figure, there were several sites in the *Tetrahymena* peripheral extensions where $c^3A\alpha S$ failed to incorporate. P2−2.1: incorporated at 21 A's, but not at *Tet*70, 87, and 95. P5abc: incorporated at 18 A's, but not at *Tet*123, 140, 172, 173, 183, and 186. P9−9.2: incorporated at 18 A's, but not at *Tet*324, 373, 374, 380, 392, and 399. $c^3A\alpha S$ was incorporated efficiently into all nine of the A's in the *Azoarcus* P9 extension. Paired (P) and joiner (J) regions of the introns are labeled. The arrow denotes attack by the 3′-terminal G on the oligonucleotide substrate with subsequent transfer of the 3′-end of the substrate onto the 3′-end of the ribozyme.

Transcription reactions were also performed at elevated pH (8.5 or 9.0) in an effort to improve incorporation by favoring the deprotonated form of the nucleoside, but pH had no effect on $c^3A\alpha S$ incorporation efficiency, and the overall yield of RNA transcript was reduced at elevated pHs with or without the addition of analogue.

Although the analogue was incorporated, the incorporation efficiency at individual positions throughout the RNA was uneven. There were a few sites of hyper-incorporation, usually at the 5′-most adenosine in a sequence of two or more consecutive adenosines. In contrast, $c^3A\alpha S$ was not incorporated at a level sufficient for NAIM analysis at a few sites, though the sequence specificity of nonincorporation followed no obvious pattern. For unknown reasons, nonincorporation of $c^3A\alpha S$ was more prevalent in the *Tetrahymena* intron sequence (28 of 115 A's, 24% of sites) than in *Azoarcus* (3 of 51 total A's, 6% of sites). The improved incorporation efficiency in *Azoarcus* may be related to the lower A/U content of the bacterial intron. Despite these limitations, $c^3A\alpha S$ was incorporated at the majority of sites within both introns at an efficiency sufficient to be informative in a NAIM assay.

Interference Mapping of the Tetrahymena Intron with $c^3A\alpha S$. Sites of $c^3A\alpha S$ interference within the L-21 G414 *Tetrahymena* intron were determined using a 3′-end-labeled substrate with a 2′-deoxy substitution at the ligation site in a reaction analogous to the reverse of the second step of splicing (*40*). Three sites of strong interference were observed in the *Tetrahymena* intron at A97, A207, and A261 (Figures 5a,b and 6a). All three of these positions have shown interference with other adenosine analogues, including $dA\alpha S$. Each has been implicated as playing an important role in intron function (see Discussion) (*18*). Seven other positions (A57, A113, A115, A136, A218, A256, and A304) showed weak to moderate interference (Figures 5b and 6a). Given the increased pK_a of the nucleotide analogue relative to adenosine, we tested the pH dependence of $c^3A\alpha S$ interference. None of the interferences were affected by the reaction

pH between pH 6 and 8 (data not shown), which argues that the interferences are not due to an altered protonation state of the nucleotides. Unfortunately, $c^3A\alpha S$ was not incorporated at either A114 or A301 (Figures 5a,b and 6a), the two adenosines expected to form A-minor tertiary interactions with the P1 helix (Figure 2). As a result, the hypothesis that substrate helix docking is mediated by interactions with the N3 imino group could not be tested with $c^3A\alpha S$ interference in this intron.

Interference Mapping of a Bacterial Intron with $c^3A\alpha S$. NAIM data have also been collected for the bacterial intron from *Azoarcus* in an effort to generate a chemical phylogeny of intron function (*26*). Its small size (<200 nt), unusual thermal stability (*25*), and lack of peripheral domains makes it an ideal molecule for interference studies where single-nucleotide resolution is increasingly compromised as the sequence length exceeds a few hundred nucleotides. The $c^3A\alpha S$ was efficiently incorporated at *Azoarcus* A58 and A167, the A's analogous to *Tetrahymena* A114 and A301 (Figures 5c and 6b). Thus, it was possible in the *Azoarcus* intron to test for interference at the two sites predicted to form A-minor interactions with P1.

The *Azoarcus* L-9 G206 ribozyme was tested for sites of $c^3A\alpha S$ interference using a 3′-end-labeled all-ribose substrate in a reaction analogous to the reverse of the second step of splicing (*26*, *40*). There were several sites of strong $c^3A\alpha S$ interference throughout the bacterial intron including: *Azo*A26, A27, A49, A58, **A87**, **A127**, A161, and A167 (Figures 5d and 6b). Two of these positions (in boldface type) also displayed strong interference in the *Tetrahymena* intron (*Tet*A207 and *Tet*A261) (Figure 6a). The third site of strong interference in *Tetrahymena* (*Tet*A97) is not an adenosine in *Azoarcus* (Figure 6). Again, none of these interferences were dependent upon the reaction pH (data not shown). Of particular interest, both *Azo*A58 and *Azo*A167 show strong $c^3A\alpha S$ interference (Figure 6b), consistent with A-minor interactions in the intron core.

An A-Minor Interaction between P1 and J4/5. While $c^3A\alpha S$ interferences at *Azo*A58 and *Azo*A167 are in agreement with the model for P1 helix docking (*13*, *14*), it is not necessarily the case that these groups interact with the specific bases predicted. NAIM with $c^3A\alpha S$ can define which N3 imino groups are important for function, but it does not define how these groups participate in the tertiary structure of the RNA. Nucleotide analogue interference suppression (NAIS) is a derivative method that combines site-specific analogue substitution and interference modification approaches, and thus can identify specific tertiary interaction partners within an RNA structure (*13*). The concept is that if an interaction is disrupted by deleting one functional group in an interacting pair, such as the critical 2′-OH in the P1 helix, then no additional energetic penalty will result from alteration of the second functional group, namely, the N3 imino or 2′-OH group. Interference suppression is scored by the specific reappearance of a band on the sequencing gel. Such experiments were used to generate constraints for the original *Tetrahymena* P1 helix docking model using 2′-OH effects (*13*, *14*), and here we report their use to test for P1 tertiary contacts in the bacterial intron using 2′-OH and N3 effects.

We first tested the interaction between the G·U wobble pair at the ribozyme cleavage site and adenosines in the J4/5 region of the active site. By analogy to the *Tetrahymena* intron, *Azo*A58 and *Azo*A87 are the minor groove bases in each of two consecutively stacked sheared A·A pairs (Figure 2b). They create a wobble receptor by forming hydrogen bonds to the 2′-OH and the N2 exocyclic amine of G10, respectively (*13*). Consistent with this structural prediction, $dA\alpha S$ causes interference in the bacterial intron at *Azo*A58 and *Azo*A87 (*26*). Thus, a 2′-deoxy substitution at *Azo*G10 (dG10) should suppress $dA\alpha S$ interference at *Azo*A58, while a *Azo*G10 to inosine substitution (I10), which deletes the N2 exocyclic amine of the base, should suppress $dA\alpha S$ interference at *Azo*A87.

Chimeric L-9 G206 introns were prepared with a single site-specific 2′-deoxy or inosine substitution at *Azo*G10 and random $dA\alpha S$ substitutions throughout the rest of the bacterial RNA. The dG10-substituted ribozyme suppressed $dA\alpha S$ interference at *Azo*A58, as predicted from the model (Figures 2b and 7a,b). As controls, $dA\alpha S$ interference persisted at *Azo*A26, A27, A87, and A167 (Figure 7a,b). Furthermore, the I10 substitution did not suppress *Azo*A58 $dA\alpha S$ interference. Thus, the effect is specific both to the nature of the substitution and to the site of the suppression. These data suggest that the 2′-OH of *Azo*G10 interacts with the 2′-OH of *Azo*A58, which is comparable to the interaction observed in the *Tetrahymena* intron (*13*) and is analogous to a crystal contact in the P4−P6 crystal structure (*10*).

The tertiary contact to the *Azo*G10 exocyclic amine was also tested by NAIS. The prediction is that $dA\alpha S$ interference at *Azo*A87 should be suppressed upon I10 substitution; however, no such suppression was observed (Figure 7a,b). *Azo*A87 $dA\alpha S$ interference persisted in the I10-substituted ribozyme. Efforts were made to repeat the experiment using a substrate with a 2′-deoxy substitution at the cleavage site, as per the experiment originally performed in *Tetrahymena*, but the low reactivity of the substrate made it impossible to obtain sufficient ligation signal for the experiment. This is likely due to the extremely short internal guide sequence in the *Azoarcus* intron (3 base pairs). Thus, while the primary *Azo*A87 $dA\alpha S$ interference data are consistent with a contact to the G10 amine, this could not be explicitly confirmed by a NAIS experiment (see Discussion).

In addition to accepting a hydrogen bond from the *Azo*A58 2′-OH, the G10 2′-OH is predicted to donate a hydrogen bond to the A58 N3 (Figure 2b). To test the latter contact, we screened for $c^3A\alpha S$ interference suppression upon dG10 substitution using chimeric ribozymes similar to those described above. The $c^3A\alpha S$ interference at *Azo*A58 was largely suppressed ($\kappa \approx 2$ compared to a $\kappa \geq 6$) upon dG10 substitution (Figure 7c,d). Interference at *Azo*A27, A87, and A127 persisted (Figure 7c,d). In contrast, $c^3A\alpha S$ interference at A58 was not affected by the I10 substitution, which argues against a global change in ribozyme structure. Thus, like the $dA\alpha S$ interference suppression at A58 described above, the $c^3A\alpha S$ interference suppression was site-specific and unique to the nature of the substitution. This argues that the 2′-OH of G10 is energetically coupled to both the N3 and the 2′-OH of A58, which is consistent with an A-minor interaction for P1 docking into J4/5.

A Second A-Minor Interaction between J8/7 and the P1/P2 Stack. We next explored the predicted contact between J8/7 and nucleotides on the substrate helix distal from the cleavage site (Figure 2c). Interference from $dA\alpha S$ and $c^3A\alpha S$

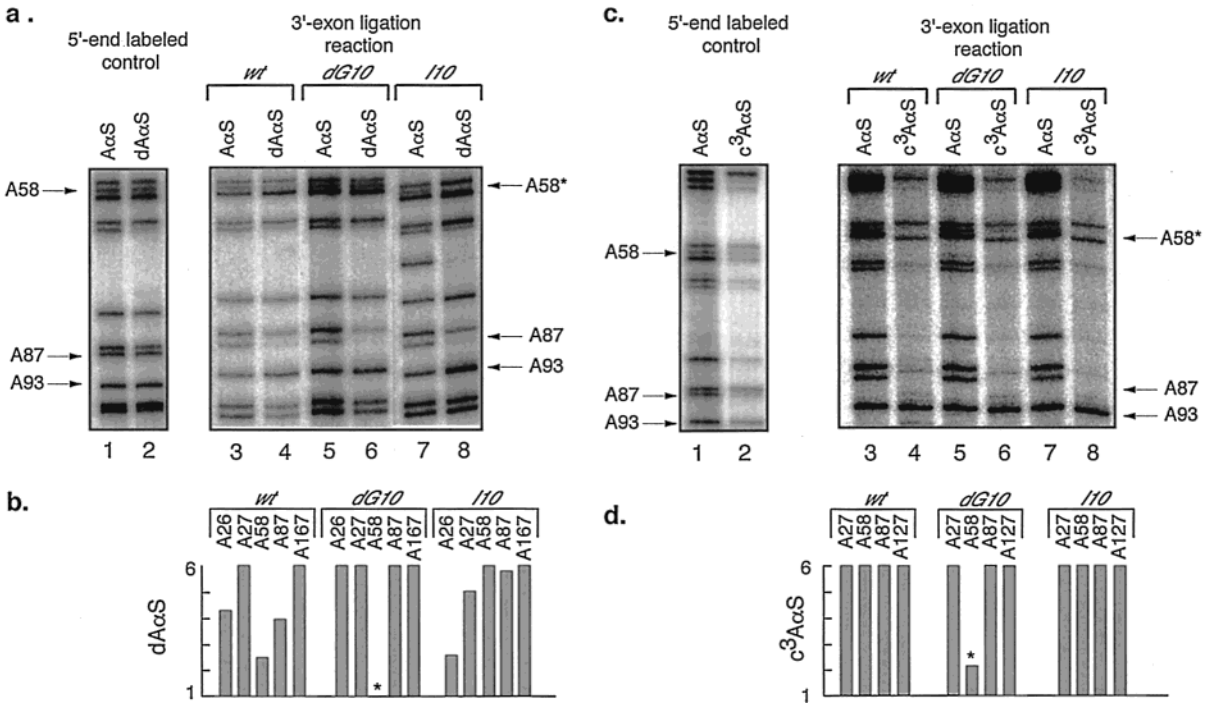

FIGURE 7: Interference suppression with $c^3A\alpha S$ and dAαS demonstrates an A-minor interaction between *Azo*G10 and A58. (a) dG10 suppression of dAαS interference at *Azo*A58. 5′-End-labeled control (lanes 1−2) and 3′-exon ligation (lanes 3−8) lanes are shown for ribozymes with a single-site G, dG, or I substitution at *Azo*G10 (ribozyme substitution indicated above bracket) and AαS or dAαS incorporated throughout the RNA (analogue identity shown above each lane). The nucleotide numbers that correspond to bands in the sequence are shown to the left and right of the gels. The site of interference suppression is indicated with an asterisk to the right of the respective lane. The orientation of the 5′-end-labeled control has been inverted for ease of comparison. (b) Quantitation of dAαS interference data from panel a. The nature of the ribozyme substitution is indicated above the bracket. The position of the dAαS interference is indicated within the bracket. Bar height indicates the magnitude of interference. A value of 1 indicates no interference, a value between 2 and 4 indicates moderate interference, and a value of 6 indicates maximum quantifiable interference at that site. An asterisk highlights the position of dAαS interference that is suppressed upon introduction of the dG10 substitution. Values represent the average of at least two independent experiments. (c) dG10 suppression of $c^3A\alpha S$ interference at *Azo*A58. Panels are labeled as in (a). (d) Quantitation of $c^3A\alpha S$ interference data from panel c. Figure is labeled as in (b).

substitution at *Azo*A167 in J8/7 is potentially consistent with an A-minor interaction with this nucleotide. The two most likely acceptors are *Azo*C13 or C14 within the P2 helix (Figure 6b). The 2′-OH groups of both residues contribute substantially to ribozyme activity (*26*). They are located at positions comparable to those important for P1 helix docking in the *Tetrahymena* intron, specifically *Tet*G25 and G26. However, the comparison is uncertain, because *Azo*C13 and C14 are base-paired within the P2 rather than P1 helix (Figure 6). Assuming that P2 stacks under P1 in the *Azoarcus* intron, then *Azo*C13 and C14 would be equivalently distant from the cleavage site as *Tet*G25 and G26, respectively (Figure 6). The comparison is further complicated by the observation that *Tet*U300, the base which contacts *Tetrahymena* G26 (*Azo*C14), is deleted from the *Azoarcus* intron. Deletion of the first nucleotide in the J8/7 region is the primary feature that distinguishes the *Tetrahymena* IC1 from the *Azoarcus* IC3 class of introns (*17*, *41*). Thus, it is possible that alternative helix packing interactions are responsible for the J8/7−P1 contacts in these two introns.

NAIS was performed to test for A-minor motifs between *Azo*A167 and candidate residues in P2. Chimeric L-6 G206 ribozymes were constructed with 2′-deoxy-C at position 13 or 14 (designated dC13 and dC14) and dAαS or $c^3A\alpha S$ randomly incorporated throughout the RNA. Both dAαS and $c^3A\alpha S$ interference at *Azo*A167 was suppressed upon dC13 substitution (Figure 8). No dAαS or $c^3A\alpha S$ suppression was observed at *Azo*A167 in the context of the dC14 or the G10 substitutions tested in the previous section (Figure 8c,d and data not shown). Interference persisted with dAαS at *Azo*A58, A87, and A127 or with $c^3A\alpha S$ at *Azo*A27, A58, A87, A127, and A161 (Figure 8c,d). Again, the suppression was particular to the specific substitution and the nucleotide position. This argues that the 2′-OH of C13 is energetically coupled to both the 2′-OH and the N3 of A167, which is the same signature observed in the J4/5 region. These data are consistent with an A-minor interaction between A167 and C13.

DISCUSSION

Structural evidence suggests that the A-minor motif is the most prevalent tertiary element for helix packing in RNA structure (*1*). It occurs at functionally critical and conserved positions throughout the ribosome and several smaller catalytic RNAs (*1*, *2*, *4*−*12*). This work extends these observations by providing evidence that the A-minor motif is also used for helix packing within the catalytic core of the group I intron. We have demonstrated by NAIM and NAIS that A-minor motifs provide the basis for substrate helix recognition within a bacterial group I intron active site. Specifically, A58 in the J4/5 region contacts the G·U wobble pair at the cleavage site in the P1 helix, and A167 in the J8/7 region contacts the C13-G37 base pair in the P2 helix. This study utilized the phosphorothioate-tagged nucleotide

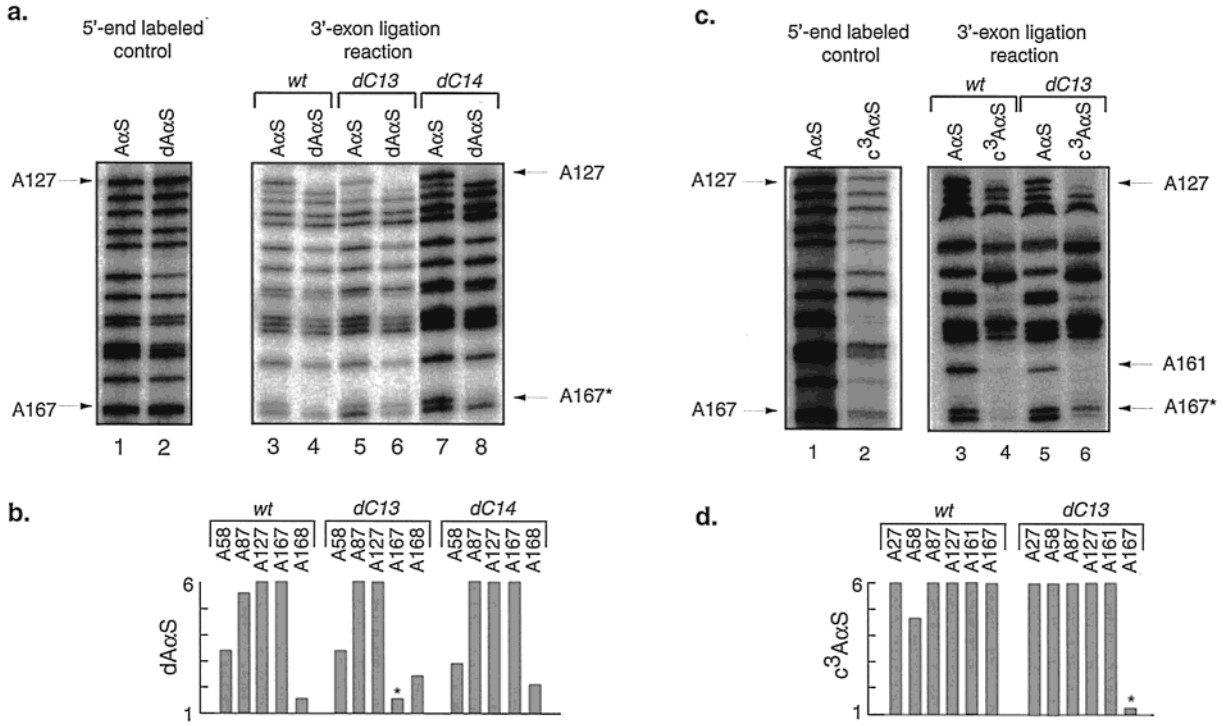

FIGURE 8: Interference suppression with $c^3A\alpha S$ and dAαS to demonstrate an A-minor interaction between *Azo*C13 and A167. Each panel is labeled as described in Figure 7. The site-specific ribozyme substitutions tested are 2′-deoxy-C at position *Azo*C13 or C14. The analogues incorporated into the RNA are dAαS or $c^3A\alpha S$. (a) dC13 suppression of dAαS interference at A167. (b) Quantitation of interferences in panel a. (c) dC13 suppression of $c^3A\alpha S$ interference at A167. (d) Quantitation of interference in panel c.

3-deazaadenosine, which makes it possible for the first time to explore the functional role of the N3 imino group of the adenine base.

It is formally possible that the energetic coupling identified by interference analysis does not reflect direct hydrogen bonding between the functional groups. For example, it is conceivable that the interactions are mediated by water or metal ions. While the data do not exclude such models, there are no data to favor them over the more simplistic interpretation that the interferences are indicative of A-minor motifs. It is important to note that these motifs are stabilized by both hydrogen bonding and van der Waals contacts, contacts which are maximized by the relatively smooth surface of the adenosine minor groove edge. Inclusion of a water or metal ion between the interacting nucleotides would both disrupt the van der Waals surface and substantially alter the hydrogen bonding donor−acceptor relationship between the residues' functional groups. Given that all of the biochemical data are consistent with A-minor motifs, we favor the more straightforward interpretation.

Distinguishing between A-Minor Motif Variations by Interference. While these data identify likely A-minor motifs, can these interference results discriminate which variation of the A-minor motif forms at each position? There are several classes of A-minor motifs, each distinguished by the position of the adenosine relative to the base pair and by the set of hydrogen bonds that mediate the interaction (Figure 1) (*1*). In both the *Azo*A58 and the *Azo*A167 A-minor interactions characterized in this study, the data are most consistent with a type II motif. In this variation, the N3 of the adenosine hydrogen bonds to the 2′-OH in the base pair. In contrast, within the type I motif, either the adenosine N3 is not contacted (as for A-U, U-A, and possibly G-C acceptor base pairs) or it hydrogen bonds to the N2 amine in the C−G base pair (*1*). In both of the A-minor candidates explored in this study, interference suppression supported an energetic coupling between the N3 of the adenosine (*Azo*A58 or A167) and the 2′-OH of the base-paired nucleotide, consistent with the type II motif. Other evidence also argues against the type I A-minor motif in both cases. One prediction of type I is that deletion of the N2 amine of the G would suppress the $c^3A\alpha S$ interference since the guanosine amine is hydrogen-bonded to the adenosine N3 in this variation. This was tested in the case of A58 A-minor, and no suppression was observed. Instead, as shown in the *Tetrahymena* intron, the N2 amine of the wobble pair is coupled to the neighboring sheared A·A pair (*13*).

This experiment could not be performed for the J8/7 A-minor contact without great technical difficulty, but other data provide compelling evidence against a type I motif between A167 and the *Azoarcus* P2 helix. The type I A-minor motif results in intimate packing of the adenosine against the minor groove edge of the C-G pair. Introduction of steric bulk at the minor groove edge of the base pair is predicted to substantially destabilize the type I motif, while the more limited packing interface of the type II motif would be substantially more tolerant of such a perturbation. We previously reported interference analysis of the *Azoarcus* intron with *N*2-methylguanosine ($m^2G\alpha S$), an analogue that introduces a methyl group on the minor groove edge of a C-G base pair (*42*). This analogue does not affect the stability of duplex formation (*43*), but it strongly interferes with any tertiary contact made to the minor groove of a G within a base-paired region (*42*). No $m^2G\alpha S$ interference was observed at *Azo*G37, the base that pairs with C13 (*26*). This argues against a close contact of A167 with the N2 amine in the C13-G37 base pair. This observation is inconsistent with the type I and consistent with the type II A-minor motif.

Another prediction of the type I motif is that both 2′-OH groups in the acceptor base pair would be important for the interaction. Within the C13-G37 pair, only C13 shows a 2′-deoxy effect, as would be consistent with a type II motif where only one of the 2′-OH groups is contacted. Interestingly, the adjacent C14-G36 base pair shows strong 2′-deoxy effects at both nucleotides and is intolerant of methylation of the G36 exocyclic amine (*26*). This suggests there is close packing in the C14-G36 minor groove in a manner consistent with an A-minor type I motif. The interaction partner(s) is (are) unknown, but it is possible that a residue(s) in J2/3 may play this role.

To be complete, the other two possible A-minor motifs must also be considered. In the type III motif, neither the N3 nor the 2′-OH of the adenosine contact the base pair (*1*), so the data are wholly inconsistent with this structural variation. The type 0 motif is a substantially less common A-minor variant (10 examples of type 0 in the 50S ribosome compared to 83 examples of type I and 54 of type II) (*1*), but it involves the same set of hydrogen bonds as type II. The difference is that the ribose sugar, rather than the adenine base, packs into the base pair minor groove (Figure 1). As a result, the type II interaction is A-specific, but the type 0 interaction is observed for other bases.

Given the hydrogen-bonding ambiguity between the type II and the type 0 motifs, is there an interference signature that can distinguish between them? The answer is somewhat equivocal. Closer packing of the adenine base in type II than in type 0 suggests that type II should be more sensitive to perturbation of the minor groove edge of the adenosine. The best analogue for such a perturbation is 2-aminoadenosine (diaminopurine riboside, DAPαS), which introduces an extra amino group onto the adenosine's smooth van der Waals surface. As expected for a type II interaction, this analogue causes interference at A58 in the *Azoarcus* intron and at A114 and A301 in the *Tetrahymena* intron (*26*). A somewhat surprising inconsistency is that DAPαS does not cause interference at *Azoarcus* A167 (whereas it is a very strong effect at the equivalent A301 in *Tetrahymena*). Although this is a negative result that could derive from angle differences between how the J8/7 strand approaches the substrate helix in these two introns, it cannot be completely excluded that the C13-A167 A-minor interaction in *Azoarcus* is type 0 rather than type II. The c^3AαS interference experiment could not be performed in *Tetrahymena* due to lack of analogue incorporation, but other interference data and modeling results favor a type II interaction in the *Tetrahymena* J8/7 (*14*). Thus, it is possible that *Azoarcus* and *Tetrahymena* make tertiary interactions with equivalent residues, but utilize different A-minor motifs. This could be indicative of larger differences in substrate helix packing between the two introns that result from a shortened J8/7 segment (6 instead of 7 nt), a lengthened J2/3 segment (3 instead of 2 nt), and P1−P2 helix stacking in the bacterial intron (*26*).

Other Possible A-Minor Motifs within the Bacterial Intron. The primary focus of this work is the A-minor motifs responsible for substrate helix recognition, but interference data were collected on most of the other adenosines throughout the bacterial intron. This raises the following question: Is the interference pattern at any other positions consistent with A-minor motifs and of what variety?

Extensive biochemical evidence indicates that the P2 tetraloop contacts a tetraloop receptor in P8 using structural motifs similar to those observed by X-ray crystallography within the P4−P6 domain (*10*, *25*, *26*, *44*−*46*). Assuming structurally equivalency, *Azo*A26 and A27 (Figure 6b) should form type II and type I contacts, respectively. Both positions display c^3AαS interference. Interference at *Azo*A26 is consistent with a type II hydrogen bond from the N3 to the *Azo*U148 2′-OH. Interference at *Azo*A27 is consistent with a type I motif as it is predicted to form a *Azo*A27 N3 to G163 N2 hydrogen bond with the *Azo*C147-G163 base pair (based upon homology to the tetraloop receptor interaction in *Tetrahymena* P4−P6) (*10*). This underscores the observation that c^3AαS interference alone does not distinguish which variety of A-minor motif each residue forms. But the clustering of such interferences can be highly informative. A-minor motifs within the 50S ribosome were observed to occur in clusters where the variety of the motif decreased going from the 5′ to the 3′ direction (*1*). Thus, where two consecutive As show c^3AαS and dAαS interference, a reasonable working hypothesis would be that the 5′ A is type II, while the 3′ A is type I.

Interferences with several other analogues in P2 and P6 agree with this structural assignment [summarized in Table 3 of (*26*)]. Most notable are the intolerance of *Azo*A26 and A27 to introduction of an N2 amine (DAPαS interference), the intolerance of *Azo*A26, A27, U148, C147, or G163 to 2′-OH deletion, and the intolerance of the *Azo*G163 N2 amine to methylation or deletion. These interferences support the expectation that the symmetry of 2′-OH effects in the acceptor pair is a distinguishing feature between the type I and type II motifs. In the *Azo*A27·C147-G163 type I motif, the 2′-OH groups of both C147 and G163 show 2′-deoxy interference, whereas only *Azo*U148 shows 2′-deoxy interference in the type II motif (*10*, *26*).

Another region in which c^3AαS interferences are clustered is J4/5. Interference at *Azo*A58 is explained by the type II A-minor interaction with G10 (see above), but what is the explanation for the interferences at A86 and A87 on the opposing strand? Based upon the G·U wobble receptor model established for the *Tetrahymena* intron (Figure 2b), the two sheared A·A pairs in J4/5 contact the 2′-OH and N2 amine of the wobble G. The modest interference at *Azo*A86 could derive from disruption of the sheared A86·A58 pair, which involves a A56 N6 to A86 N3 hydrogen bond. Assuming equivalency between *Tetrahymena* and *Azoarcus*, the amine of *Azo*G10 should contact the 2′-OH and N3 of A87, which would explain all the interferences observed on both residues. We attempted a suppression experiment to establish this contact in the bacterial intron, but the *Azo*A87 dAαS and c^3AαS interference persisted when an inosine was substituted at G10. At first glance, this might suggest that the *Azoarcus* intron defines the 5′-splice site by a molecular mechanism distinct from that in *Tetrahymena*; however, persistence of *Azo*A87 interference could have a much simpler explanation. Based upon the wobble receptor model, the *Azo*A87 2′-OH would make a ground-state contact to the G10 amine, and a transition-state contact to the 2′-OH of U-1 (*47*) (Figure 2b). Because of this second important contribution, *Azo*A87 dAαS interference would not be suppressed so long as there is a ribose at the U-1 cleavage site. The analogous experiment in *Tetrahymena* did result in suppression, but it was done in

the context of a U-1 2′-deoxy-substituted substrate (*13*). Extensive effort was made to perform the suppression experiment in *Azoarcus* using a substrate with a 2′-deoxy substitution, but it proved impossible to obtain sufficient signal. The extremely short *Azoarcus* internal guide sequence when combined with the substantial reaction rate reduction due to the deoxy substitution proved intractable. Thus, while the primary *Azo*A87 dAαS interference data are consistent with a contact to the G10 amine, this could not be directly confirmed.

A third region where c^3AαS interferences are clustered is the P6 internal loop (*Azo*A97 and A98). These interference are unlikely to result from disruption of A-minor motifs because neither of the residues show 2′-OH effects (*26*). Lack of 2′-OH effects also argues against A-minor motifs at *Azoarcus* A49, A63, and A161. The N3 groups of these residues are likely to make their contributions by other types of interactions.

The only other *Azoarcus* residue that shows interference upon N3 and 2′-OH substitution, consistent with a type 0, I, or II A-minor motif, is *Azo*A127 in the J6/7 region (*26*). Equivalently strong interferences were reported for the analogous residue in *Tetrahymena* (A261) (*18*). In fact, *Azo*A127 is the only adenosine in the *Azoarcus* intron that shows dAαS and c^3AαS interference but whose tertiary contact has not been identified. A127 is very highly conserved among almost all 700 group I introns sequenced and aligned (conserved in 663 of 671 introns, 98.8%) (*48*). *Azo*A127 shows little or no effect upon introduction of an N2 amine (i.e., no DAPαS interference) (*26*), which argues strongly against A-minor type I, modestly disfavors type II, and hints at type 0. The A127 interference results suggest that in addition to the A-minor-like contacts, the Hoogsteen face (N7 and N6) of the residue also makes tertiary interactions (*18*, *26*). This implicates *Azo*A127 as a bridging element that brings multiple secondary structural elements together in the active site. If *Azo*A127 forms an A-minor motif that is conserved between these two introns, then it would be expected to contact a 2′-OH whose functional importance is also conserved. Despite the low-resolution nature of the docked model and the undocked crystal structure (*14*, *15*), the overall spatial positioning of these residues leaves only a few possible candidates, including *Azo*A172 at the 3′-end of J8/7 and *Azo*C46 at the 3′-end of P3 (*26*). Interference suppression experiments are underway to explore these possibilities.

Characteristic Interference Patterns of the A-Minor Motifs. The results presented here suggest that nucleotide analogue interference patterns can identify and partially distinguish A-minor interactions in RNA tertiary structure, particularly the type I and type II varieties. The diagnostic patterns are as follows.

Type I motif: The adenosine donor is intolerant of 2′-OH deletion, intolerant of N3 substitution (there is a strong effect if the adenosine contacts a C-G pair, but interference in the context of other pairs has not been tested), and strongly impaired by functional substitution at the C2 position. The base pair acceptor shows strong effects upon 2′-OH deletion at each residue and is intolerant of steric bulk introduced into the minor groove.

Type II motif: The adenosine donor is intolerant of 2′-OH and N3 substitution, and adversely affected by amine substitution at C2. The base pair acceptor only shows interference upon deletion of one of its 2′-OH groups, and there is greater tolerance for steric bulk on the minor groove edge of the base pair (i.e., methylation of the exocyclic amine of G) relative to the type I motif. Consecutive adenosines that are affected by 2′-OH and N3 substitution are good candidates for A patches that involve both type I and type II motifs.

Type III motif: Interference on the adenosine donor in this motif is uninformative because neither the 2′-OH nor the N3 is involved in contacts to the base pair. The N1 position is the only site of interaction, but no nucleotides have been reported that make it possible to analyze this position in a NAIM assay. Thus, the type III is distinguished only by interference at one of the 2′-OHs on the acceptor base pair. No examples of this motif were explicitly investigated in this study, but one such contact was identified in the *Tetrahymena* J8/7 region using dimethyl sulfate footprinting analysis (*49*).

Type 0 motif: This motif's interference pattern is predicted to be almost identical to that of the type II motif except there may be a greater tolerance for steric substitution on the adenosine C2.

CONCLUSION

Like the ribosome, the group I intron utilizes A-minor motifs at functionally critical positions. Additionally, the nucleotide analogue c^3AαS described in this report can be used to investigate A-minor motifs in any RNA that can be transcribed in vitro and biochemically selected. In combination with the 2′-deoxynucleotide analogue series previously reported (*21*), it is now possible to identify experimentally the donor adenosines and the acceptor base pairs and even provide a reasonable prediction about which A-minor motif forms between them. Identification of these interactions will provide valuable constraints for RNA structure prediction and lead to a molecular explanation for RNA structure and function relationships.

ACKNOWLEDGMENT

We thank E. McMillan for optimizing c^3ATPαS transcription conditions, L. Weinstein, D. Klein, and R. Anderson for comments on the manuscript, and M. Schmeing for assistance with figures.

REFERENCES

1. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B., and Steitz, T. A. (2001) *Proc. Natl. Acad. Sci. U.S.A. 98*, 4899−4903.
2. Doherty, E. A., Batey, R. T., Masquida, B., and Doudna, J. A. (2001) *Nat. Struct. Biol. 8*, 339−343.
3. Gutell, R. R., Weiser, B., Woese, C. R., and Noller, H. F. (1985) *Prog. Nucleic Acid Res. Mol. Biol. 32*, 155−216.
4. Gutell, R. R., Cannone, J. J., Shang, Z., Du, Y., and Serra, M. J. (2000) *J. Mol. Biol. 304*, 335−354.
5. Ban, N., Nissen, P., Hansen, J., Moore, P., and Steitz, T. (2000) *Science 289*, 905−920.
6. Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Jr., Morgan-Warren, R. J., Carter, A. P., Vonrhein, C., Hartsch, T., and Ramakrishnan, V. (2000) *Nature 407*, 327−339.
7. Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H., and Noller, H. F. (2001) *Science 29*, 29.
8. Nissen, P., Hansen, J., Ban, N., Moore, P., and Steitz, T. (2000) *Science 289*, 920−930.

9. Ogle, J. M., Brodersen, D. E., Clemons, W. M., Jr., Tarry, M. J., Carter, A. P., and Ramakrishnan, V. (2001) *Science 292*, 897−902.
10. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) *Science 273*, 1678−1685.
11. Ferre-D'Amare, A. R., Zhou, K., and Doudna, J. A. (1998) *Nature 395*, 567−574.
12. Rupert, P. B., and Ferre-D'Amare, A. R. (2001) *Nature 410*, 780−786.
13. Strobel, S. A., Ortoleva-Donnelly, L., Ryder, S. P., Cate, J. H., and Moncoeur, E. (1998) *Nat. Struct. Biol. 5*, 60−66.
14. Szewczak, A. A., Ortoleva-Donnelly, L., Ryder, S. P., Moncouer, E., and Strobel, S. A. (1998) *Nat. Struct. Biol. 5*, 1037−1042.
15. Golden, B. L., Gooding, A. R., Podell, E. R., and Cech, T. R. (1998) *Science 282*, 259−264.
16. Cech, T. R., and Herschlag, D. (1996) in *Catalytic RNA* (Eckstein, F., and Lilley, D. M. J., Eds.) pp 1−17, Springer, New York.
17. Damberger, S. H., and Gutell, R. R. (1994) *Nucleic Acids Res. 22*, 3508−3510.
18. Ortoleva-Donnelly, L., Szewczak, A. A., Gutell, R. R., and Strobel, S. A. (1998) *RNA 4*, 498−519.
19. Lawley, P., and Brookes, P. (1963) *Biochem. J. 89*, 127−138.
20. Strobel, S. A., and Shetty, K. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 2903−2908.
21. Conrad, F., Hanne, A., Gaur, R. K., and Krupp, G. (1995) *Nucleic Acids Res. 23*, 1845−1853.
22. Gish, G., and Eckstein, F. (1988) *Science 240*, 1520−1522.
23. Cech, T. R., Zaug, A. J., and Grabowski, P. J. (1981) *Cell 27*, 487−496.
24. Reinhold-Hurek, B., and Shub, D. A. (1992) *Nature 357*, 173−176.
25. Tanner, M. A., and Cech, T. R. (1996) *RNA 2*, 74−83.
26. Strauss-Soukup, J., and Strobel, S. A. (2000) *J. Mol. Biol. 302*, 339−358.
27. Minakawa, N., and Matsuda, A. (1993) *Tetrahedron 49*, 557−570.
28. Froehler, B. C., and Matteucci, M. D. (1983) *Nucleic Acids Res. 11*, 8031−8036.
29. Caruthers, M. H., McBride, L. J., Bracco, L. P., and Dubendorff, J. W. (1985) *Nucleosides Nucleotides 4*, 95−105.
30. Ryder, S. P., Ortoleva-Donnelly, L., Kosek, A. B., and Strobel, S. A. (2000) *Methods Enzymol. 317*, 92−109.
31. Moore, M. J., and Query, C. C. (1998) in *RNA−Protein Interactions: A Practical Approach* (Smith, C., Ed.) pp 75−108, Oxford University Press, Oxford, U.K.
32. Lingner, J., and Keller, W. (1993) *Nucleic Acids Res. 21*, 2917−2920.
33. Arabshahi, A., and Frey, P. A. (1994) *Biochem. Biophys. Res. Commun. 204*, 150−155.
34. Ludwig, J., and Eckstein, F. (1989) *J. Org. Chem. 54*, 631−635.
35. Steitz, T. A., Beese, L., Freemont, P. S., Friedman, J. M., and Sanderson, M. R. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 465−471.
36. Minakawa, N., Kojima, N., and Matsuda, A. (1999) *J. Org. Chem 64*, 7158−7172.
37. Sousa, R., and Padilla, R. (1995) *EMBO J. 14*, 4609−4621.
38. Sousa, R. (2000) *Methods Enzymol. 317*, 65−74.
39. Huang, Y., Beaudry, A., McSwiggen, J., and Sousa, R. (1997) *Biochemistry 36*, 13718−13728.
40. Beaudry, A. A., and Joyce, G. F. (1992) *Science 257*, 635−641.
41. Michel, F., and Westhof, E. (1990) *J. Mol. Biol. 216*, 585−610.
42. Ortoleva-Donnelly, L., Kronman, M., and Strobel, S. A. (1998) *Biochemistry 37*, 12933−12942.
43. Rife, J., Cheng, C., Moore, P. B., and Strobel, S. A. (1998) *Nucleic Acids Res. 26*, 3640−3644.
44. Costa, M., and Michel, F. (1995) *EMBO J. 14*, 1276−1285.
45. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Szewczak, A. A., Kundrot, C. E., Cech, T. R., and Doudna, J. A. (1996) *Science 273*, 1696−1699.
46. Basu, S., Rambo, R. P., Strauss-Soukup, J., Cate, J. H., Ferre-D Amare, A. R., Strobel, S. A., and Doudna, J. A. (1998) *Nat. Struct. Biol. 5*, 986−992.
47. Strobel, S. A., and Ortoleva-Donnelly, L. (1999) *Chem. Biol. 6*, 153−156.
48. Cannone, J. J., Subramanian, S., Schnare, M. N., Collett, J. R., D'Souza, L. M., Du, Y., Feng, B., Lin, N., Madabusi, L. V., Muller, K. M., Pande, N., Shang, Z., Yu, N., and Gutell, R. R. (2002) *BioMed Central Bioinformatics 3*, URL: http://www.rna-.icmb.utexas.edu/.
49. Pyle, A. M., Murphy, F. L., and Cech, T. R. (1992) *Nature 358*, 123−128.

BI020265L