*Proc. Natl. Acad. Sci. USA*
Vol. 94, pp. 2903–2908, April 1997
Biochemistry

# Defining the chemical groups essential for *Tetrahymena* group I intron function by nucleotide analog interference mapping

SCOTT A. STROBEL† AND KANAKA SHETTY

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

*Communicated by Thomas A. Steitz, Yale University, New Haven, CT, January 27, 1997 (received for review November 18, 1996)*

**ABSTRACT** **Improved atomic resolution biochemical methods are needed to identify the chemical groups within an RNA that are essential to its activity. As a step toward this goal, we report the use of 5′-*O*-(1-thio)inosine monophosphate (IMPαS) in a nucleotide analog interference mapping (NAIM) assay that makes it possible to simultaneously, yet individually, determine the contribution of almost every N2 exocyclic amine of G within a large RNA. Using IMPαS, we identified the exocyclic amines that are essential for 5′ or 3′ exon ligation by the *Tetrahymena* group I intron. We report that the amino groups of three phylogenetically conserved guanosines (G111, G112, and G303) are important for 3′ exon ligation. The amine of G22, as well as the amines of the other four guanosines within the P1 helix, are essential for ligation of the 5′ exon. Previous work has shown that point mutation of either G22 or G303 to an adenosine (A) substantially reduces activity. Like inosine, adenosine lacks an N2 amino group. Interference rescue of the G22A and G303A point mutations was detected at the site of mutation by NAIM using 5′-*O*-(1-thio)diaminopurine riboside monophosphate (DMPαS), an adenosine analog that has an N2 exocyclic amine. The G22A point mutant could also be rescued by incorporation of DMPαS at A24. By analogy to genetics, there are interference phenotypes comparable to loss of function, reversion, and suppression. This method can be readily extended to other nucleotide analogs for the analysis of chemical groups essential to a variety of RNA and DNA activities.**

The *Tetrahymena* group I intron catalyzes two consecutive transesterification reactions in the course of RNA self-splicing (1). Splicing is initiated by nucleophilic attack at the 5′ splice site by the 3′-hydroxyl of an exogenous guanosine (G) cofactor. This results in covalent attachment of the G onto the intron and release of the 5′ exon. Previous to the second step of splicing, the RNA undergoes a conformational transition in which G414, the 3′ terminal nucleotide in the intron, enters the G binding site. This conformational rearrangement brings the exon boundaries into close proximity. In the second step of splicing, the 5′ exon attacks the 3′ splice site to produce ligated exons. The intron can also catalyze the reverse of these two reactions resulting in ligation of the exons onto the intron (2–4).

During the first step of splicing, the location of the 5′ splice site is defined by a phylogenetically conserved G·U wobble pair in the P1 helix (Fig. 1) (5–11). This noncanonical pair forms between a U at the end of 5′ exon and a G located in the internal guide sequence. The contributions made by specific chemical groups within the wobble pair were defined by systematically substituting the G·U with a series of nucleotide analogs, including inosine and diaminopurine riboside (10–12). Inosine is a G analog that lacks the N2 exocyclic amine, while diaminopurine riboside is an A analog with the amine. These experiments demonstrated that the exocyclic amine of G, when presented on the minor groove surface by the wobble configuration, contributes to tertiary interaction with the ribozyme active site, reaction fidelity, and stabilization of the chemical transition state (10, 11).

There are 106 other guanosines within the intron, including several that are phylogenetically conserved (13). It is reasonable to expect that the N2 amino groups of a subset of these are important for proper splicing activity, though there is currently no simple method available to identify them. With substantial effort, specific positions could be tested using a semisynthetic method to prepare full-length RNAs with a single inosine substitution (14). However, this approach is sufficiently labor intensive that only a few nucleotide positions could be tested. A more efficient method is required to identify the essential amines or other important chemical groups within an RNA.

Interference mapping by phosphorothioate substitution is an effective method to identify the nucleotide positions where replacing the nonbridging $R_P$ oxygen with a sulfur is detrimental to activity (15–22). The sites of phosphorothioate incorporation are readily detected by cleavage of the phosphate linkage with iodine (15). This makes it possible to simultaneously, yet individually, quantitate every position within the RNA by an assay that is as simple as RNA sequencing. Krupp and coworkers (23, 24) extended the phosphorothioate interference approach to nucleotide analogs by chemically tagging 2′-deoxy and 2′-methoxy derivatives with a phosphorothioate marker. They used these derivatives to identify the 2′-hydroxyl groups in tRNA that are essential for recognition by RNase P.

In principle, the $R_P$ phosphorothioate could be used as a general chemical tag for numerous nucleotide analogs. Development of a complete library of phosphorothioate-tagged analogs would provide a powerful atomic resolution tool for RNA structure/function studies. As a step toward this goal, we report the use inosine and diaminopurine riboside phosphorothioates to identify the N2 exocyclic amines of G that are essential for 5′ or 3′ exon ligation by the *Tetrahymena* group I intron. In combination with point mutagenesis, this pair of nucleotide derivatives makes it possible to detect essential amino groups by both inhibition and restoration of intron function.

## METHODS

**Synthesis of Phosphorothioate Analogs.** 5′-*O*-(1-thio)-inosine triphosphate (ITPαS) was synthesized from inosine in a procedure analogous to that reported for the synthesis of

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.


0027-8424/97/942903-6$2.00/0
PNAS is available online at **http://www.pnas.org**.

Abbreviations: NAIM, nucleotide analog interference mapping; IMPαS or ITPαS, 5′-*O*-(1-thio)inosine monophosphate or triphosphate; DMPαS or DTPαS, 5′-*O*-(1-thio)diaminopurine riboside monophosphate or triphosphate; GMPαS or AMPαS, 5′-*O*-(1-thio)guanosine or adenosine monophosphate; rT, 5-methyl uridine; dC, 2′-deoxy-cytidine; dT, thymidine; rT(-1)S, GGCCCUC(rT)AAAAA; dT(-1)S, CCCUC(dT)AAAAA; rT(-1)P, GGCCCUC(rT); dT(-1)P, CCCUC(dT); dC(-1)P, CCCUC(dC).

†To whom reprint requests should be addressed. e-mail: strobel@csb.yale.edu.

5′-*O*-(1-thio)-2′-deoxyadenosine triphosphate (25). Inosine (100 mg, 0.37 mmol) was dried under vacuum at 110 °C for 12 h and dissolved in triethylphosphate (5 ml). The nucleoside was reacted with thiophosphorylchloride (42 μl, 0.41 mmol, 1.1 equivalents) in the presence of trioctylamine (180 μl, 0.41 mmol, 1.1 equivalents) at 0°C for 30 min. It was converted directly to the triphosphate by addition of tributylammonium pyrophosphate (0.5 g, 1.0 mmol) in triethylphosphate (3 ml). The product was precipitated by the addition of excess triethylamine (2.5 ml), centrifuged, and decanted. The residue was dissolved in aqueous triethylammonium bicarbonate (10 ml, 1.0 M, pH 7.5). Purification by DEAE-A25 Sephadex chromatography using a triethylammonium bicarbonate buffer gradient (0.05–0.8 M) afforded ITPαS as a diastereomeric mixture in 23% yield. $^{31}$P NMR ($H_2O$): 43.96 (*m*), −10.21 (*d*), −24.13 (*t*). UV: 250 nm $\lambda_{max}$. A similar procedure afforded 5′-*O*-(1-thio)diaminopurine ribosyl triphosphate (DTPαS) in 26% yield. $^{31}$P NMR ($H_2O$): 42.61 (*m*), −6.24 (*d*), −23.18 (*t*). UV: bimodal with a $\lambda_{max}$ at 256 and 280 nm.

**Plasmids.** Plasmids pUCL-21G414R and pUCL+1 were generated from plasmid pT7L-21 by PCR amplification with oligonucleotides containing the appropriate sequence changes. The pUCL-21G414R plasmid contains a new *Ear*I site for use in run-off transcription to produce an RNA with G414 as the terminal nucleotide. pUCL+1 replaces the first 21 nt of the wild-type intron with the sequence 5′-GAAUUACA-CAAUUAUACCAUU. The sequence contains mutations (underlined) that eliminate the circularization sites present within the wild-type L1 loop (26). This sequence includes a G as the first base of the intron which is analogous to the exogenous G added after the first step of splicing (27). pUCL-21G414R G303A and pUCL+1 G22A are single point mutations of the parent plasmids.

**Transcription Reactions.** L-21 $G^{414}$ RNA (Fig. 1) was transcribed from *Ear*I digested pUCL-21G414R by T7 RNA polymerase using 1 mM of each NTP. L+1 *Sca*I RNA was transcribed from *Sca*I-digested pUCL+1 under the same conditions. 5′-*O*-(1-thio)inosine monophosphate IMPαS or DMPαS were randomly incorporated into the RNA transcripts at a level of ≈5 per molecule (5%) using 0.4 mM ITPαS or 25 μM DTPαS in the transcription mixture, respectively. RNAs were also prepared with ≈5% GMPαS or AMPαS using 50 μM

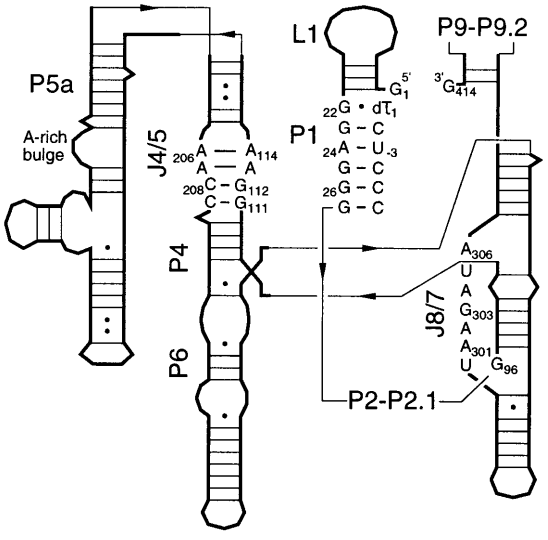


FIG. 1. Composite diagram of the *Tetrahymena* group I intron. The L-21 $G^{414}$ RNA begins at G22 and ends at G414, but does not include the L1 loop. The L+1 *Sca*I RNA begins at G1, includes a mutated version of the L1 loop, and terminates 5 nt before G414. The composite RNA is shown bound to the oligonucleotide dT(-1)P which is the substrate for the reaction with L+1 *Sca*I. The diagram also identifies the nucleotide positions, and the duplex and single-stranded regions discussed in the text.

of the $S_P$ form of the triphosphate (18). Levels of IMPαS and DMPαS incorporation were determined by normalizing to transcripts containing 5% GMPαS and AMPαS, respectively. There was no detectable difference in the interference pattern at any of the levels of phosphorothioate incorporation tested (from 0.5% to 10%). A level of 5% incorporation provided an easily detected signal, but remained low enough to avoid interference between multiple substitutions within a single RNA. Following transcription, the RNAs were purified by PAGE and eluted into 10 mM Tris·HCl, pH 7.5/1 mM EDTA (TE buffer). The RNAs were ethanol precipitated and resuspended in TE buffer.

**A Ribozyme with Inosine in Place of Guanosine.** An L-21 *Sca*I form of the *Tetrahymena* ribozyme was prepared with inosine in place of G at all positions except the 5′ terminal nucleotide, G22. *Sca*I-digested pT7L-21 plasmid was used as a template for T7 RNA polymerase transcription with 4 mM inosine triphosphate and 10 mM GMP in the transcription mixture.

**Radiolabeling.** Oligonucleotides rT(-1)S [GGCCCUC(rT)A-AAAA] and dT(-1)S [CCCUC(dT)AAAAA] were 3′ end labeled with [α-$^{32}$P]cordycepin using poly(A) polymerase. Oligonucleotides rT(-1)P [GGCCCUC(rT)], dT(-1)P [CCCUC(dT)], and dC-(-1)P [CCCUC(dC)] were 5′ end labeled with [γ-$^{32}$P]ATP using polynucleotide kinase. All oligonucleotides were purified on a native polyacrylamide gel prior to use and eluted into TE buffer.

The L-21 $G^{414}$ RNA was treated with alkaline phosphatase and 5′ end-labeled with [γ-$^{32}$P]ATP using polynucleotide kinase. The radiolabeled 5′ control RNAs were gel purified and eluted into TE buffer.

**Reaction Conditions.** The L+1 *Sca*I and L-21$G^{414}$ RNAs perform RNA ligation reactions that are analogous to the reverse of the first and second steps of RNA splicing, respectively. Permissive and nonpermissive reaction conditions were chosen for nucleotide analog interference mapping (NAIM). Permissive conditions are defined as 50 nM enzyme in 20 mM $MgCl_2$, 2 mM $Mn(OAc)_2$, and 50 mM Hepes (pH 7.0) for 10 min at 50°C with 5 nM of an all ribose oligonucleotide [rT(-1)S for L-21$G^{414}$ and rT(-1)P for L+1 *Sca*I RNA]. Nonpermissive conditions are defined as 50 nM enzyme in 4 mM $MgCl_2$, and 50 mM Hepes (pH 7.0) for 10 min at 50°C with an oligonucleotide (5 nM) containing a single 2′-deoxy substitution at the reaction site [dT(-1)S for L-21$G^{414}$ and dT(-1)P for L+1 *Sca*I RNA]. The single deoxy substitution does not affect oligonucleotide binding, but it does reduce the rate of the chemical reaction by at least a 1000-fold (28). RNAs were incubated at 50°C in the appropriate reaction buffer for 10 min prior to the addition of the oligonucleotide. 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). The phosphorothioates were cleaved by the addition of 1/10th volume of 100 mM iodine in ethanol (15). The solution was heated to 90°C for 2 min and the cleavage products resolved by electrophoresis on a 5% polyacrylamide gel. The intensities of individual bands were quantitated by PhosphorImager (Molecular Dynamics) analysis.

## RESULTS

**A Group I Intron Containing Inosine in Place of Guanosine Is Inactive.** Previous work has shown that the N2 exocyclic amine of G22 is important for the first step of group I intron activity (10, 11). We set out to determine if any other N2 amino groups of G play an important role in the activity. To address this question, a variation of the L-21 *Sca*I form of the intron was prepared in which every G except G22 was replaced with inosine. The inosine ribozyme demonstrated no activity under any reaction conditions when incubated for several hours with the oligonucleotide substrate rT(-1)S. This included incubation of the substrate with 1 μM ribozyme, 10 mM GMP, and 100 mM $Mg^{2+}$ ion. Failure to detect activity under these conditions demonstrates that N2 amino groups at positions

other than G22 are important. The locations of the essential groups could not be determined from this assay.

**NAIM.** We employed NAIM to identify the location of the essential N2 amino groups. The method involves four steps: (*i*) the phosphorothioate derivative of the nucleotide analog is randomly incorporated into the RNA by *in vitro* transcription, (*ii*) the active variants in the RNA population are selected from the inactive transcripts, (*iii*) the phosphorothioate linkage is cleaved by treatment with iodine, and (*iv*) the sites of analog substitution that are detrimental to RNA activity are identified by gel electrophoresis and autoradiography (Fig. 2). Gaps in the sequencing ladder define the sites where analog substitution is detrimental to activity. By this approach it is possible to simultaneously, yet individually, test the importance of an inosine substitution at every G within the group I intron.

**Defining the N2 Exocyclic Amines Important for 3′ Exon Ligation.** Variants of the L-21 $G^{414}$ RNA containing ≈5 IMPαS substitutions per molecule were reacted with the 3′ end-labeled oligonucleotide substrate dT(-1)S under nonpermissive conditions (Fig. 2) (4). In this reaction, the 3′ end of the oligonucleotide (AAAAAA*), which is analogous to the 3′ exon, is transferred from the substrate onto the 3′ end of the active variants in the intron population (3). Inactive RNAs are not radiolabeled.

The intensity of the iodine sequencing ladder at three nucleotide positions is substantially reduced among the active RNA variants containing IMPαS compared with the GMPαS control (Fig. 3). This includes G111, G112, and G303. There is a 3-fold phosphorothioate effect at G303, but IMPαS is reduced an additional 8-fold relative to GMPαS at this position (Fig. 3*A* lanes 1 and 2). The 3- and 5-fold effects were observed upon substituting inosine at G111 and G112, respectively. There was no phosphorothioate effect at these positions (Fig. 3*B*). Normal cleavage intensities at G111, G112, and G303 were observed among the total population of RNAs containing IMPαS, as assayed by direct 5′ end labeling (Fig. 3*A*, lanes 7 and 8; Fig. 3*B*). This confirms that the loss of signal does not result from a failure of the polymerase to incorporate IMPαS into the RNA at these sites. This is further supported by the observation that the cleavage intensities at G111, G112, and G303 are partially recovered when the RNAs were reacted with the oligonucleotide rT(-1)S under permissive conditions (Fig. 3*A*, lane 5 and 6). There was no position where IMPαS substitution enhanced activity.

A total of 93 of the 106 guanines in the group I intron could be characterized in this assay. A strong phosphorothioate effect at G96 resulted in a band intensity that was too low in the GMPαS RNA for the IMPαS RNA to be informative. Other positions could not be accurately measured because the cleavage products were not be sufficiently resolved from the full-length ligated intron.

**Defining the N2 Exocyclic Amines Important for 5′ Exon Ligation.** To determine if 5′ exon ligation has the same N2 amino group requirements as ligation of the 3′ exon, variants of the L+1 *Sca*I RNA containing IMPαS substitutions were reacted with the radioactively labeled oligonucleotide dT(-1)P. In this reaction, the 5′ exon is covalently transferred onto the 5′ end of the active introns with displacement of the terminal guanosine (G1) (2).

For the 5′ splicing assay, the iodine sequencing ladder was substantially reduced at all five Gs in the internal guide sequence (Fig. 3*B*). No phosphorothioate effect was measured at any of these positions, so the reported effects are due entirely to inosine substitution. The strongest effect was at G22 where the magnitude of the interference was greater than 10-fold. This is the expected result based upon previous studies of the G·U wobble pair, and reconfirms that the exocyclic amine of G22 is essential for the activity of the intron (10, 11). Smaller effects (between 2- and 5-fold) were observed at other Gs within the P1 helix, including G23, G25, G26, and G27. If the L+1 *Sca*I IMPαS RNA was reacted with rT(-1)P under permissive conditions, cleavage bands were detected for all of the internal guide sequence

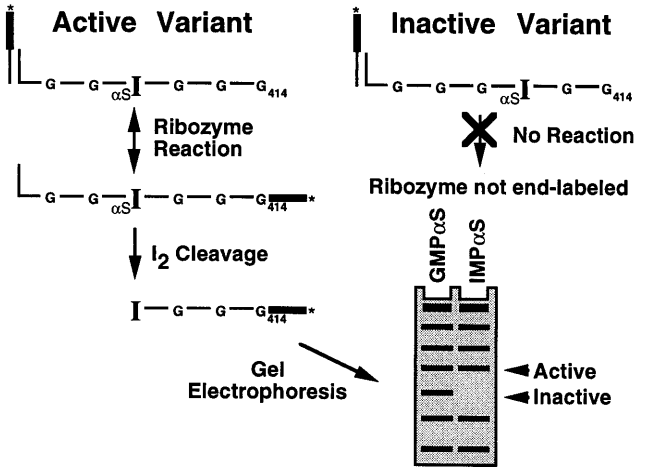


FIG. 2. Scheme for the identification of the chemical groups important for RNA activity by NAIM. In this example, IMPαS is randomly incorporated in place of G at a low ratio into the L-21 $G^{414}$ RNA. Active variants covalently transfer the 3′ end of the exon onto the 3′ terminus of the intron. Hydrolysis of the phosphorothioate linkage with iodine breaks the RNA at the sites of IMPαS incorporation. Separation of the cleavage fragments by PAGE produces a sequencing ladder that represents the sites tolerant of IMPαS substitution. In contrast, inactive variants fail to react with the oligonucleotide which leaves a gap in the sequencing ladder. RNAs containing GMPαS are tested to insure that loss of activity is not due to the phosphorothioate group used to chemically tag the inosine.

residues, confirming that the loss of signal is due to loss of activity, rather than a failure of the RNA polymerase to incorporate IMPαS into the RNA at these sites.

The three positions identified as important for the 3′ exon ligation reaction were tolerated in the 5′ ligation reaction. This included G303, which in the 3′ ligation reaction displayed neither a phosphorothioate effect nor an effect from inosine substitution. The reaction also tolerated a phosphorothioate substitution at G96. This does not appear to be due to a general lack of selectivity in the L+1 *Sca*I reaction. Several of the phosphorothioate effects at adjacent positions, including A302, A304, U305, and A306, remained or were enhanced for this reaction (data not shown).

**Interference Rescue of the G303A Mutation.** The strongest IMPαS interference signal for the L-21 $G^{414}$ reaction was at G303. Mutation of G303 to an A reduces the activity of the L-21 *Sca*I ribozyme 30-fold (29). Based upon the interference results above, this can be at least partially explained by the loss of the exocyclic amine upon mutation of G303 to an A. Diaminopurine riboside (2-amino adenosine; abbreviated as D) is an analog of adenine that contains an exocyclic amine at the N2 position of the purine ring comparable to G. We postulated that if the amine at G303 is essential for catalytic function, then activity could be rescued by repeating the interference experiment in the G303A mutant background using the nucleotide analog DMPαS (Fig. 4*A*).

There are two ways that DMPαS could rescue activity. The first, and most likely outcome, is that DMPαS could provide an exocyclic amine at the position originally mutated to an adenosine. This would appear on the interference sequencing gel as a single band of strong intensity at the same nucleotide position that is missing in the IMPαS interference experiment. Second, DMPαS could rescue activity at a position other than the original site of mutation. Both of these possibilities could be detected in a single interference mapping experiment.

Variants of the L-21 $G^{414}$ G303A RNA containing approximately five random AMPαS or DMPαS substitutions per molecule were reacted with the 3′ end-labeled oligonucleotide rT(-1)S. The resulting interference pattern is shown in Fig. 4*B*. A very weak signal was detected at position A303 in the RNA

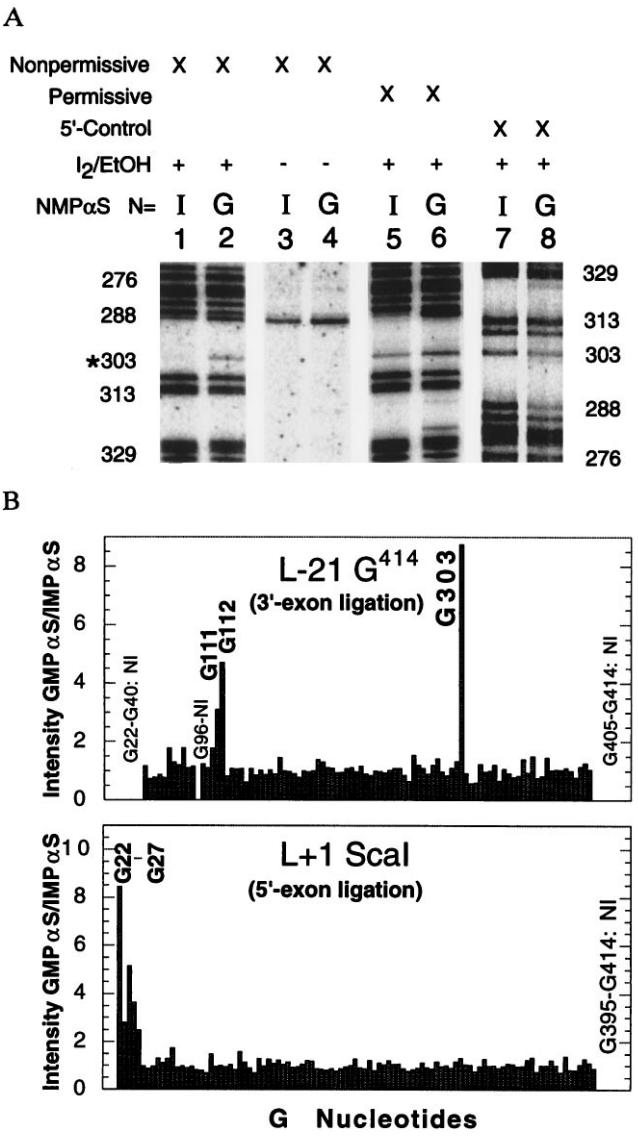


FIG. 3. (*A*) A portion of the sequencing autoradiogram showing the reaction of L-21 $G^{414}$ IMPαS or GMPαS RNA with dT(-1)S under nonpermissive conditions (lanes 1–4) or rT(-1)S under permissive conditions (lanes 5 and 6). The nucleotide positions correlating to the cleavage products are shown to the left for lanes 1–6 and to the right for lanes 7 and 8. For the 5′ control (lanes 7 and 8), the L-21 $G^{414}$ is enzymatically labeled at the 5′ end. This results in a reversal of the sequence orientation on the gel, but shows the actual level of analog incorporation at all nucleotide positions. RNA degradation independent of iodine treatment is consistently observed at A290. This is clearly seen in lanes 3 and 4. (*B*) Histogram showing the ratio of GMPαS intensity divided by IMPαS intensity for each position in both the L-21 $G^{414}$ and L+1 *Sca*I RNAs when reacted under nonpermissive conditions and normalized to the actual levels of incorporation as defined by the 5′ control. Intensities were measured by PhosphorImager analysis. A value of 1 indicates that an inosine substitution at that site has no effect. A value greater than 2 indicates that inosine is detrimental, while a value less than 0.5 (not detected) would indicate that inosine substitution is advantageous at that site. Positions G22–G40 could not be accurately measured in the 5′ control, so these values were normalized using the intensity of each band under permissive reaction conditions. Thus, the magnitude of the effect at G22–G27 is a minimum estimate. All values were measured at least three times in separate experiments and are accurate to within 25% error. Sites that were not informative in the assay are indicated on the graph with the initials NI.

containing AMPαS (lane 2). The A303 signal is restored more than 5-fold in the RNA containing DMPαS (lane 1). This indicates that the activity of the G303A mutant can be at least partially restored by substituting an N2 amino group in the context of an adenosine nucleotide derivative, and confirms the importance of this chemical group to intron activity.

**Interference Rescue of a G22A Mutation.** The strongest site of IMPαS interference for the L+1 *Sca*I RNA was at G22. G22 forms a wobble pair with the U at position −1 in the oligonucleotide. Mutation of G22 to an A reduces the second order rate constant for cleavage by the L-21 *Sca*I form of the ribozyme by almost 500-fold (11). This is due to the loss of the exocyclic amine at G22, and the conversion of the G·U wobble pair to an A-U Watson–Crick pair. We set out to determine if the G22A mutation could also be rescued by interference with DMPαS.

Diaminopurine forms a stable Watson–Crick pair with U and a relatively unstable wobble pair with C (30). We measured the extent of DMPαS rescue using oligonucleotide substrates with either a U or a C at position −1 (Fig. 5). Activity was rescued more than 10-fold (compared with an RNA containing AMPαS and normalized for the extent of reaction and the level of analog incorporation) at position A22 when L+1 *Sca*I G22A RNA containing DMPαS was reacted with dC(-1)P. Only a 3-fold rescue was observed when the RNA was reacted with dT(-1)P. The DMPαS interference rescue experiment confirms the IMPαS result, thus showing by both positive and negative interference that the amine of G22 is important for 5′ exon ligation. The requirement for a mutation at a second site (D·C gave better rescue than D-U) confirms that the exocyclic amine makes its greatest contribution to splicing in the context of a wobble pair, where the amino group is free to make a tertiary contact with the intron active site (10, 11).

The intensity of iodine cleavage at positions other than A22 was much lower in the G22A experiment than for the G303A experiment because there was not phosphorothioate effect at position 22. This made it possible to detect interference rescue at positions other than the site of mutation. In addition to the strong cleavage band at position 22, a second band of moderately strong intensity was present at A24, the only other A nucleotide within the P1 helix (Fig. 5*B*). This band is 5-fold more intense (after correcting for the extent of reaction and analog incorporation) in the DMPαS lane than in the AMPαS lane when the RNAs were reacted with dC(-1)P (Fig. 5*B*, lanes 1 and 2). There was a 3-fold rescue at A24 when the RNA is reacted with dT(-1)P (lanes 5 and 6). Close inspection of cleavage intensities at other positions within the RNA failed to detect other examples of interference rescue.

## DISCUSSION

Nucleotide analog interference mapping is a general method to define the chemical basis of RNA function. In this example, we have shown that inosine can be incorporated into the RNA transcript as a 5′-phosphorothioate in place of guanosine. This simultaneously replaces the N2 amino group of G with a proton and the 5′ nonbridging $R_P$ phosphate oxygen with a sulfur. The ability to cleave with iodine exclusively at the phosphorothioate linkage makes it possible to map the sites of inosine incorporation that are detrimental to RNA function.

In many ways the 5′-α-phosphorothioate is an ideal chemical tag. Sulfur has only a slightly larger van der Waals radius than oxygen, and the length of the P—S bond is only slightly longer than the P—O bond (31). For these reasons, the sulfur substitution is not expected to substantially distort the RNA secondary structure. The most noticeable effect upon RNA structure is a change in the ability of the phosphate group to coordinate metal ions (32); however, this can be compensated by the inclusion of thiophilic metals such as $Mn^{2+}$ or $Cd^{2+}$ in the reaction mixture (33, 34). Thus, for a large majority of nucleotides within an RNA, the phosphorothioate substitution is functionally silent and can serve as an effective tag for analog incorporation. For example, in this experiment, only 1 of 94 positions was not informative because of phosphorothioate interference, and that site was only uninformative for the 3′ exon ligation reaction.

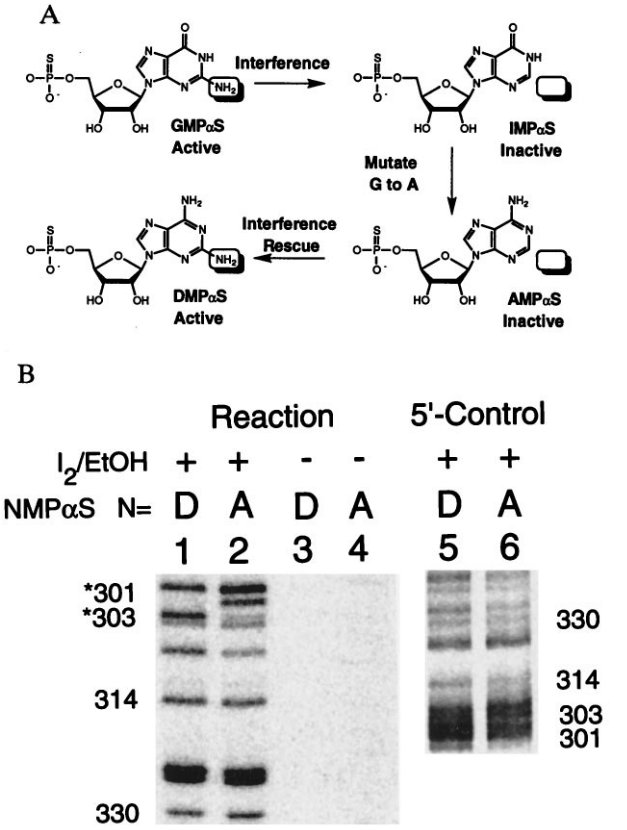


FIG. 4. (*A*) Scheme for interference rescue. Ribozyme activity is reduced by mutating a G that is intolerant of IMPαS substitution to an A. Activity is rescued by substituting the A with DMPαS. (*B*) A portion of the sequencing autoradiogram showing the reaction of L-21 $G^{414}$ G303A DMPαS or AMPαS with rT(-1)S for 10 min at 50°C in 4 mM $MgCl_2$. The nucleotide positions correlating to the cleavage products are shown to the left for lanes 1–4 and to the right for lanes 5 and 6. For the 5′ control (lanes 5 and 6), the L-21 $G^{414}$ RNAs are enzymatically labeled at the 5′ end. This results in a reversal of the sequence orientation on the gel, but shows the actual level of analog incorporation at all nucleotide positions. A visual comparison between the reaction and the 5′ control is complicated by strong phosphorothioate interference at A302 and A306. These bands are absent in both the DMPαS and AMPαS reaction lanes, but they are present in both 5′ controls.

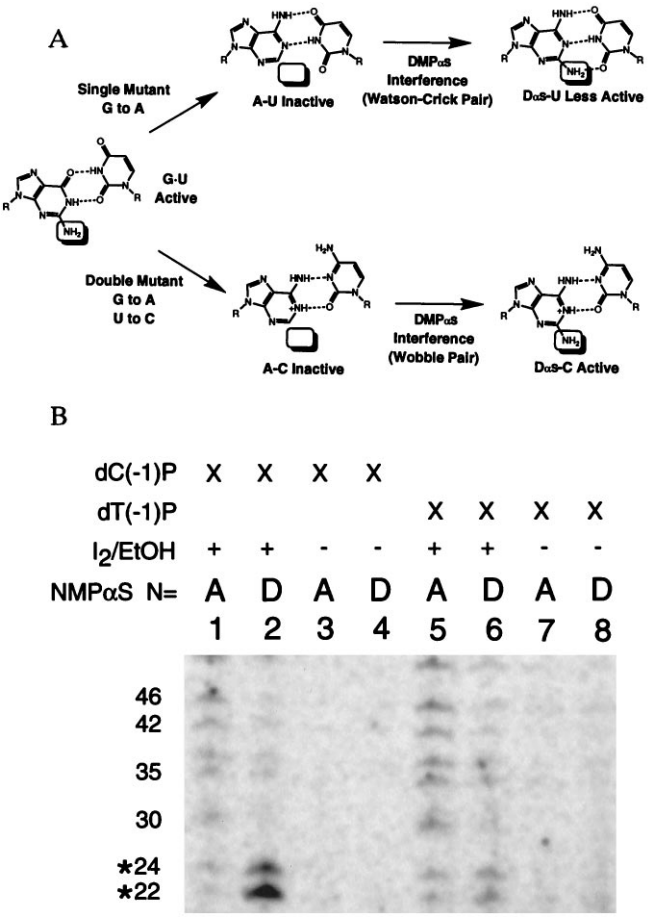


FIG. 5. (*A*) Scheme for interference rescue that requires a second mutation for maximum activity. Unlike the scheme in Fig. 5, activity of the G to A point mutant is not significantly rescued by DMPαS substitution unless a second nucleotide is also changed. In this case, U −1 must be mutated to a C. This restores the wobble configuration of the base pair. (*B*) A portion of the sequencing autoradiogram showing the reaction of L+1 *Sca*I DMPαS or AMPαS at 50°C in 4 mM $MgCl_2$. Reactions with dT(-1)P and dC(-1)P were for 2 and 10 min, respectively. The magnitude of the interference rescue was estimated by comparing the intensity of each band in the DMPαS lane to the intensity in the AMPαS lane and normalizing to the intensity at all other positions in the RNA.

**Genetic Analogies.** It might be useful to consider this approach using the language of genetics. Each nucleotide position within the RNA has a phenotype that is expressed as a band intensity on a sequencing gel. Reduction of signal intensity among the active fraction of RNA molecules under nonpermissive conditions (low metal and a 2′-deoxy substitution at the cleavage site) is analogous to a loss of function phenotype, and defines the positions where analog substitution is detrimental to RNA function. Corollaries of genetic revertants and suppressors can be generated by making a deleterious mutation in the RNA and screening for interference rescue. Restoration of activity by analog substitution at the site of mutation is analogous to a genetic revertant, while rescue by substitution at a distant position is a comparable to a genetic suppressor.

We observed at least one example of each of these phenotypes during our efforts to map the *Tetrahymena* group I intron. The N2 exocyclic amines of G111, G112, and G303 are important for 3′ exon ligation, but are not essential for ligation of the 5′ exon. All three of these nucleotides are located within the catalytic core of the intron. G22, G23, G25, G26, and G27 are important for 5′ exon ligation, but could not be measured for the 3′ exon ligation reaction. These five nucleotides are all in the P1 helix of the intron. Thus, there are eight examples of a loss of function phenotype. Mutation of either G22 or G303 to an A is deleterious, but activity could be restored by incorporating DMPαS at the sites of mutation. This provides two examples of a revertant. The activity of the G22A mutation could also be rescued by incorporating DMPαS at A24. This is an example of a suppressor. Given the parallels between nucleotide analog interference mapping and genetics, we feel that the term "chemogenetics," originally coined by Sigler, provides a convenient and descriptive one-word summary of the approach (35).

**Possible Roles for the Essential N2 Amino Groups.** All of the amino groups identified as important in the exon ligation assays either stabilize the P1 duplex or are in close proximity to nucleotides known to interact with it. The site showing the highest level of interference in the second step of splicing is G303. It is located in the J8/7 region of the active site and is 94% conserved among all group I introns (439 total examples) (13). It is 100% conserved among the IC subgroup of introns which includes *Tetrahymena* (243 examples). This high level of conservation suggests that G303 plays an important role in intron function. The neighboring base, A302, has been shown to make a long-range hydrogen bonding contact with an important 2′-hydroxyl in the P1 helix (29). This tertiary contact is thought to position the 5′ exon into the intron active site for splicing. It is reasonable to expect that the exocyclic amine of G303 is also proximal to the P1 helix. It could contribute to

activity by making a direct hydrogen bond to P1, as proposed in the original model of Michel and Westhof (36), or the contribution could be indirect. For example, the amine might be essential for orienting A302 so that it can contact P1.

There are several possible explanations for the contribution of the G111 and G112 amines. G111 and G112 are both in the P4 region of the intron active site, and both form canonical Watson–Crick pairs with C209 and C208, respectively (37). The amino groups are located in the minor groove of the A-form RNA duplex (37). G112 is 89% conserved among all group I introns and 98% conserved among introns in subgroup IC. G111 is only 58% conserved overall, but it is 92% conserved among the subgroup IC1 and IC2 introns (13). This suggests that the exocyclic amines of G111 and G112 are essential for the *Tetrahymena* intron, but their contribution might not be universal.

The P4 helix is located within a densely packed region of the active site. The major groove face of G111 interacts with U305 in the J8/7 region (38). Photo-crosslinking demonstrates that the G·U of the P1 helix is positioned adjacent to the sheared A-A pairs that are coaxially stacked immediately above the G112-C208 pair (37, 39). The A-rich bulge in P5a docks into the minor groove of the P4 helix (37). Thus, G111 and G112 are in a central core of the intron, close to the J8/7 single-stranded region, the P1 helix, and the A-rich bulge in P5a.

The most trivial explanation for the detrimental effect of inosine substitution at G111 and G112 is that the exocyclic amines stabilize base pairing within the P4 helix. Although an I-C pair is less stable than a G-C pair (30), it seems unlikely that this is the explanation, given the high level of phylogenetic conservation at these two positions. It is more likely that these two amines are important for long-range tertiary interactions between other regions of the active site. The P1 helix is the most likely candidate due to the close proximity of the P1 helix to the minor groove of G111 and G112.

All of the N2 amino groups in the P1 helix are important for 5′ exon ligation including those at positions G22, G23, G25, G26, and G27. Except for G22, the sequence at these positions is not phylogenetically conserved, so the exocyclic amines are unlikely to be involved in essential tertiary interactions. The effects at G23–G27 most likely result from a reduction in P1 duplex stability. Improved P1 stability is also the most probable explanation for suppression of the G22A mutation by DMP$\alpha$S at A24. This creates a D-U pair which is more stable than the wild-type A-U pair (30).

Unlike the other positions in the P1 helix, G22 is 100% conserved among all group I introns and forms a wobble pair with −1 U which is 98.7% conserved (13). The effect at G22 is consistent with previous experiments that characterized an RNA containing a single inosine substitution at position 22 (10, 11).

The crystal structure of the P4–P6 domain suggests that other N2 amino groups could be important for activity (37). These were not detected. It is possible that the P4–P6 domain and other regions within the intron are sufficiently stable and structurally redundant that folding is not impaired by the loss of a single hydrogen bonding group (40).

**Conformational Changes in the Active Site Between the First and Second Steps of Splicing?** It is uncertain why the G111, G112, and G303 amino groups are only essential for 3′ exon ligation. One intriguing possibility is that the difference in the interference pattern for G111, G112, and G303 reflects structural changes in the active site between the two steps of splicing. While this cannot be proven from the data presented here, there is a precedent for this interpretation. The phosphorothioate interference patterns for 5′ exon cleavage and 3′ exon hydrolysis are similar, but not equal (16, 18, 19). Many of the differences are in regions close to G303 and G112. Prior reports have suggested that the identity of the G22·U-1 pair at the 5′ exon cleavage site is only essential for the first step of splicing (7); while covariation between U305 and G111 is essential for 3′ splice site hydrolysis, but not for 5′ exon cleavage (38). Furthermore, methylation interference experiments on the group I intron sunY showed a different interference pattern for 5′ compared with 3′ exon hydrolysis (41). The difference in the inosine interference pattern, provides additional evidence that the conformation of the intron active site (specifically in areas proximal to the P1 helix) is not equivalent for the first and second steps of splicing.

**Other Nucleotide Analogs and Other Nucleic Acid Polymers.** In principle, this approach can be extended to several other nucleotide analogs and to any RNA or DNA polymer that can be enzymatically synthesized and functionally selected. "Functions" might be defined as catalysis, protein binding, secondary or tertiary structure formation, ligand binding, conformational heterogeneity, etc. NAIM will make it possible to identify chemical groups essential for a wide variety of RNA and DNA activities.

We thank Leo Beigelman and Jasenka Adamic for advice on phosphorothioate synthesis, Robin Gutell for performing phylogenetic comparisons, and Lara Weinstein, Jamie Cate, Sean Ryder, and Soumitra Basu for critical reading of the manuscript. This work was supported by National Institutes of Health Grant GM54839 and a Junior Faculty Research Award from the American Cancer Society to S.A.S.

1. Cech, T. R. (1990) *Annu. Rev. Biochem.* **59,** 543–568.
2. Green, R., Ellington, A. D. & Szostak, J. W. (1990) *Nature (London)* **347,** 406–408.
3. Beaudry, A. A. & Joyce, G. F. (1992) *Science* **257,** 635–641.
4. Mei, R. & Herschlag, D. (1996) *Biochemistry* **35,** 5796–5809.
5. Waring, R. B., Towner, P., Minter, S. J. & Davies, R. W. (1986) *Nature (London)* **321,** 133–139.
6. Doudna, J. A., Cormack, B. P. & Szostak, J. W. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 7402–7406.
7. Barfod, E. T. & Cech, T. R. (1989) *Mol. Cell. Biol.* **9,** 3657–3666.
8. Pyle, A. M., Moran, S., Strobel, S. A., Chapman, T., Turner, D. H. & Cech, T. R. (1994) *Biochemistry* **33,** 13856–13863.
9. Knitt, D. S., Narlikar, G. J. & Herschlag, D. (1994) *Biochemistry* **33,** 13864–13879.
10. Strobel, S. A. & Cech, T. R. (1995) *Science* **267,** 675–679.
11. Strobel, S. A. & Cech, T. R. (1996) *Biochemistry* **35,** 1201–1211.
12. Green, R., Szostak, J. W., Benner, S. A., Rich, A. & Usman, N. (1991) *Nucleic Acids Res.* **19,** 4161–4166.
13. Damberger, S. H. & Gutell, R. R. (1994) *Nucleic Acids Res.* **22,** 3508–3510.
14. Moore, M. J. & Sharp, P. A. (1992) *Science* **256,** 992–997.
15. Gish, G. & Eckstein, F. (1988) *Science* **240,** 1520–1522.
16. Waring, R. B. (1989) *Nucleic Acids Res.* **17,** 10281–10293.
17. Ruffner, D. E. & Uhlenbeck, O. C. (1990) *Nucleic Acids Res.* **18,** 6025–6029.
18. Christian, E. L. & Yarus, M. (1992) *J. Mol. Biol.* **228,** 743–758.
19. Christian, E. L. & Yarus, M. (1993) *Biochemistry* **32,** 4475–4480.
20. Jeong, Y.-H., Kumar, P. K. R., Suh, Y.-A., Taira, K. & Nishikawa, S. (1994) *Nucleic Acids Res.* **22,** 3722–3727.
21. Hardt, W.-D., Warnecke, J. M., Erdmann, V. A. & Hartmann, R. K. (1995) *EMBO J.* **14,** 2935–2944.
22. Yu, Y.-T., Maroney, P. A., Darzynkiewicz, E. & Nilsen, T. W. (1995) *RNA* **1,** 46–54.
23. Gaur, R. K. & Krupp, G. (1993) *Nucleic Acids Res.* **21,** 21–26.
24. Conrad, F., Hanne, A., Gaur, R. K. & Krupp, G. (1995) *Nucleic Acids Res.* **23,** 1845–1853.
25. Arabshahi, A. & Frey, P. A. (1994) *Biochem. Biophys. Res. Commun.* **204,** 150–155.
26. Been, M. D. & Cech, T. R. (1987) *Cell* **50,** 951–961.
27. Cech, T. R., Zaug, A. J. & Grabowski, P. J. (1981) *Cell* **27,** 487–496.
28. Herschlag, D., Eckstein, F. & Cech, T. R. (1993) *Biochemistry* **32,** 8312–8321.
29. Pyle, A. M., Murphy, F. L. & Cech, T. R. (1992) *Nature (London)* **358,** 123–128.
30. Strobel, S. A., Cech, T. R., Usman, N. & Beigelman, L. (1994) *Biochemistry* **33,** 13824–13853.
31. Frey, P. A. & Sammons, P. (1985) *Science* **228,** 541–545.
32. Pecoraro, V. L., Hermes, J. D. & Cleland, W. W. (1984) *Biochemistry* **23,** 5262–5271.
33. Dahm, S. C. & Ohlenbeck, O. C. (1991) *Biochemistry* **30,** 9464–9469.
34. Piccirilli, J. A., Vyle, J. S., Caruthers, M. H. & Cech, T. R. (1993) *Nature (London)* **362,** 85–88.
35. Sigler, P. B. (1994) *Nat. Struct. Biol.* **1,** 3–4.
36. Michel, F. & Westhof, E. (1990) *J. Mol. Biol.* **216,** 585–610.
37. Cate, J. H., Gooding, A. R., Podell, E., Zhou, K., Golden, B. L., Kundrot, C. E., Cech, T. R. & Doudna, J. A. (1996) *Science* **273,** 1678–1685.
38. Tanner, M. A. & Cech, T. R. (1997) *Science,* **275,** 847–849.
39. Wang, J.-F., Downs, W. D. & Cech, T. R. (1993) *Science* **260,** 504–508.
40. Celander, D. W. & Cech, T. R. (1991) *Science* **251,** 401–407.
41. von Ahsen, U. & Noller, H. F. (1993) *EMBO J.* **12,** 4747–4754.