

# Use of Circular Permutation and End Modification to Position Photoaffinity Probes for Analysis of RNA Structure

Michael E. Harris<sup>1</sup> and Eric L. Christian

Center for RNA Molecular Biology, Department of Molecular Biology and Microbiology,  
Case Western Reserve University School of Medicine, Cleveland, Ohio 44106

Photocrosslinking allows first-order structural analysis with relatively small amounts of biological material and can be applied in complex *in vitro* systems. In this article we describe methods for positioning both arylazide and thionucleotide photoagents within an RNA of interest by end modification of circularly permuted RNAs. Application of this technique provided a library of constraints that, together with biochemical and phylogenetic comparative data, were used to develop a structure model of the bacterial ribonuclease P ribozyme-substrate complex. Circularly permuted genes for *in vitro* transcription are generated by PCR from tandem genes. Circularly permuted RNA transcripts can be modified with high efficiency at both the 5' and 3' termini with arylazide crosslinking reagents, or transcription can be primed with photoactive nucleotide analog monophosphates such as 6-thioguanosine. These crosslinking agents can be used over a wide range of experimental conditions but remain inert until they are activated by UV light. Crosslinked sites are subsequently mapped by reverse transcriptase primer extension of gel-purified crosslinked species. In addition to providing basic protocols for these methods, we discuss approaches for establishing the relevance of crosslinking data to native RNA structure. © 1999

Academic Press

Detailed information on RNA structure is becoming increasingly available from X-ray crystallography and NMR; however, many experimental systems are either too complex or provide insufficient amounts of material to allow application of these

very powerful methods of structure determination. In these cases, chemical crosslinking using precisely positioned photoaffinity probes provides a practical alternative for examining RNA structure. Crosslinking can be used as a probe for known RNA structure motifs and can provide distance constraints between specific elements of RNA structure. This information can be used, together with data from phylogenetic and biochemical analysis, to develop models of ribonucleoprotein structure (e.g., 1, 2). While general crosslinking approaches using UV irradiation and reactive species such as psoralen have been applied in a wide variety of experimental systems involving large RNAs, site-specific positioning of crosslinking agents in RNA allows a more focused exploration of RNA structure and function.

Initially, site-specific incorporation of photocrosslinking agents took advantage of photoactive or chemically reactive nucleotides in naturally occurring RNAs (3, 4). The development of oligonucleotide-directed ligation of RNA fragments has allowed a broader application of site-specific modification with nucleotide analogs, including photoreactive reagents (5, 6; see also articles by Yu, Konarska, and Reed and Chiara, this issue). An alternative method that is simple and efficient involves end modification of circularly permuted RNAs (cpRNAs) for site-specific positioning of crosslinking agents for structural analysis of RNA. The observation that RNA structures are often stabilized by a complex network of hydrogen-bonding interactions and are quite tolerant of individual discontinuities, or nicks, in the polynucleotide backbone (e.g., 7–11) makes possible the use of cpRNAs to systematically reposition the 5' and 3' termini in an RNA without significant per-

<sup>1</sup> To whom correspondence should be addressed. Fax: (216) 368-3055. E-mail: meh2@pop.cwru.edu.

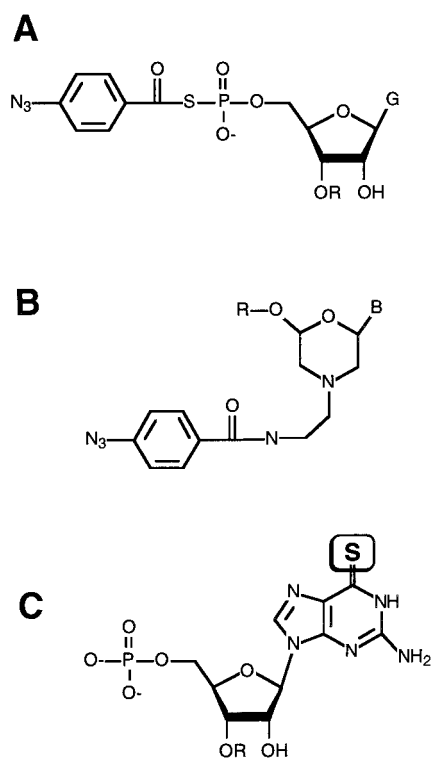
turbation of the native three-dimensional structure. Because the 5' and 3' ends of an RNA can be efficiently and uniquely modified, end modification of cpRNA can offer advantages of speed, simplicity, and yield. This approach was used to analyze the structure of bacterial ribonuclease P RNA; however, it should be generally applicable to the analysis of other large RNAs and ribonucleoprotein complexes.

Azidophenacyl (APA) derivatives and thionucleotide photoagents such as 6-thioguanosine ( $s^6G$ ) can be readily incorporated at RNA termini (Fig. 1). APA is a long-range crosslinking agent with an azido moiety ca. 9 Å from its point of attachment to the RNA (12). It can be used to identify general features of an RNA structure such as position and orientation of RNA helices (13, 14). Thionucleotides like  $s^6G$  are short-range structural probes whose crosslinking thio moiety is attached directly to the nucleotide base and can thus be used to refine the distance

constraints between positions that appear to be proximal by long-range crosslinking or other phylogenetic or biochemical data (15). Both agents share the important chemical property that they are reactive over a wide range of experimental conditions but remain inert until they are activated with UV light. Moreover, photocrosslinking with these reagents is in general very sensitive since most applications use radiochemical detection, which requires as little as femtomole amounts of biological material.

Though very useful, there are definite limitations to what can be accomplished experimentally with crosslinking. The degree of structural resolution obtained by crosslinking is considerably less than that of classical methods of structural analysis such as X-ray crystallography and NMR. Moreover, RNA conformational heterogeneity can complicate application of crosslinking and the interpretation of results. Additionally, modification of conserved nucleotides can potentially disrupt function, and the photocrosslinking reaction can, in principle, be affected by local chemical environments within an RNA. Nevertheless, the relative ease of obtaining intermediate-resolution structural data by crosslinking, together with its sensitivity and the ability to apply this technique under almost any experimental conditions, make end modification of cpRNAs a useful method for obtaining structural information bearing on complex biological systems.

The photocrosslinking approach described here was coupled with molecular modeling to develop a first-order model of the structure of bacterial ribonuclease P RNA, hereafter referred to as P RNA. Initial application of this approach was reported by Nolan *et al.* (16) who defined areas of P RNA adjacent to pre-tRNA by using circularly permuted tRNA substrates 5' end modified with APA. Subsequently, further application of cpRNA end modification to both P RNA and tRNA was used to develop a library of positional constraints for molecular modeling (2, 13, 14). The RNA subunit of bacterial RNase P provides an excellent system for understanding RNA structure and function. P RNA, which catalyzes specific cleavage of pre-tRNA, represents an essential, widespread, and conserved class of catalytic RNAs (ribozymes). Thus, a detailed investigation provides the opportunity to explore important aspects of RNA folding and function. This article focuses on experimental methods for RNA circular permutation and end modification for placing arylazide and thioguanosine crosslinking reagents within



**FIG. 1.** Arylazide and thioguanosine photoagents. (A) Azidophenacyl guanosine monophosphorothioate. G, the guanine base of the GMPS moiety; R, the remainder of the RNA chain. (B) *N*-(*N*-azidophenacyl)aminoethyl-modified NMP. For attachment of the photoagent, the 3'-furanose has been converted to a morpholine ring. B, the 3' terminal base; R, the remainder of the upstream RNA chain. (C) 6-Thioguanosine monophosphate. The sulfur-modified 6 position is highlighted.

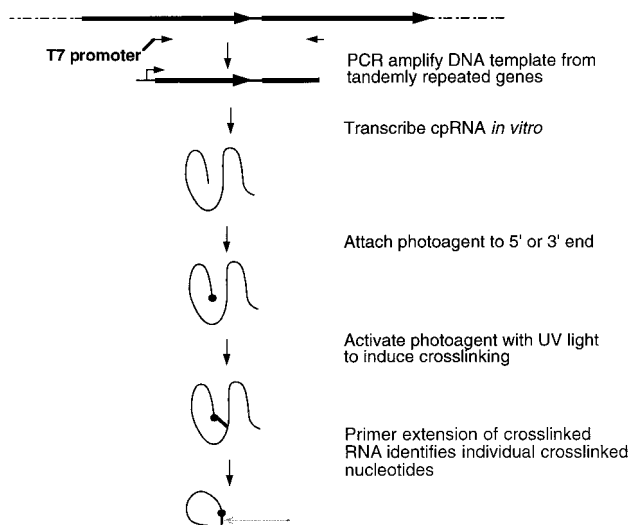
a large RNA molecule for structural analysis. In addition, we describe conditions and experimental considerations for crosslinking and the purification of crosslinked material. Finally, we conclude with a discussion of ways to test the degree to which crosslinking results reflect native RNA structure.

## METHODS

### Generation of Circularly Permuted RNAs

The strategy for photocrosslinking analysis using end-modified cpRNAs is outlined in Fig. 2. Briefly, the method involves first generating cpRNA by *in vitro* transcription from a circularly permuted gene, followed by end modification of the RNA with a photoactivatable crosslinking reagent. The RNA is then placed under a desired set of reaction conditions and the photoreactive species is activated by brief UV irradiation. Subsequently, individual cross-linked species are purified and the individual cross-link sites are mapped, usually by primer-extension assays.

Circularly permuted gene sequences for *in vitro* transcription are generated by PCR from tandemly



**FIG. 2.** Preparation and crosslinking of photoagent-modified cpRNA. Tandem RNA-encoding genes (top) are shown as horizontal arrows. Location of primer binding sites for generation of a single cpRNA construct are shown below. Transcribed RNA sequences are shown as a thin line and the attached photoagent is shown as a sphere. Primer extension of the crosslinked RNA is shown in gray. Descriptions of each step are shown at right.

repeated genes. Forward and reverse primers oriented in the appropriate direction in the upstream and downstream genes define the 5' and 3' ends of the cpRNA that is subsequently generated by *in vitro* transcription (Fig. 2). Tandem genes are joined by a short oligonucleotide leader that can be of any sequence so long as it does not alter the function of the RNA being studied. PCR reactions for studies involving bacterial RNase P RNA genes typically contained 30 mM Tris-HCl, pH 8.3, 1 mM dNTPs, 50 mM potassium chloride, 1.5 mM magnesium chloride, 0.05% NP-40, 1–10 pg tandem P RNA or tRNA genes, and 100 pmol of the forward and reverse primers. For most experimental situations, it is important to design PCR primers containing promoter sequences and restriction enzyme sites that can be used for cloning and *in vitro* transcription.

The locations of the new endpoints in the circularly permuted RNAs are selected primarily based on their strategic positions. Biochemical, structural, and phylogenetic considerations can be used to prioritize the exploration of useful photoagent incorporation sites. However, there are two practical considerations for choice of circular permutation sites. First, the use of phage T7 RNA polymerase for *in vitro* transcription dictates that RNAs initiate with a guanosine residue (17). T7 RNA polymerase optimally requires two consecutive G residues at the 5' end; however, sufficient amounts of RNA can also be obtained from transcripts beginning with GU, GC, and GA (17). The restriction of nucleotide identity at the 5' end can be circumvented in part by using SP6 polymerase which has less stringent requirements than T7; however, one generally finds numerous appropriate circular permutation sites in the sequence or structure of interest. The second important consideration involves generation of the 3' end of the RNA. In order to generate a template for run-off transcription which ends precisely at the desired P RNA nucleotide, restriction sites that cut at a distance from their recognition sequences are placed just downstream of the RNase P sequences. *FokI* and *BbsI* are quite useful in this regard for P RNA template design.

Transcripts of cpRNAs are obtained by standard *in vitro* transcription protocols. In bacterial RNase P studies, nonradiolabeled RNAs are generated by transcription of 2–4  $\mu$ g of linearized template DNA with 40 units of phage T7 RNA polymerase in a 100- $\mu$ l reaction containing 40 mM Tris-HCl, pH 8.0, 1 mM spermidine, 5 mM dithiothreitol, 0.1% Triton X-100, 20 mM magnesium chloride, and 1 mM NTPs

for 5–12 h. RNAs are gel purified by electrophoresis through 4% polyacrylamide/7 M urea gels and visualized by UV shadowing. RNA is passively eluted from gel slices into 0.3 M sodium acetate, 20 mM Tris-HCl, pH 8.0, 1 mM ethylenediaminetetraacetate, 0.1% sodium dodecyl sulfate.

For all cpRNAs generated, it is important to establish that the native RNA structures are maintained. For the RNase P ribozyme, this is demonstrated by their ability to catalyze pre-tRNA cleavage. In general, the kinetic parameters  $k_{\text{cat}}$  and  $K_m$  for the cpRNAs used are approximately the same as those of their native counterparts (2, 13, 14). For structural RNAs that are not ribozymes, cpRNAs can be compared to their native RNA counterparts by diagnostic metal ion cleavage, chemical modification, or the kinetics of their interactions with other molecules. Instances where circular permutation alters catalytic function or some other diagnostic feature is indicative of structural distortion and must be excluded from subsequent structural studies.

#### *End Modification of cpRNA with Arylazide APA and s<sup>6</sup>G*

APA belongs to a larger group of arylazide photoagents (18) which can be covalently attached to the 5' end of RNA by a simple two-step approach (19). The first step involves the transcription of cpRNAs in the presence of guanosine monophosphorothioate (GMPS). Because T7 polymerase will initiate transcripts with nucleotide monophosphates but requires triphosphosphates for elongation, inclusion of GMPS in transcription reactions results in its exclusive incorporation at the RNA 5' end. Transcription reactions with GMPS:GTP ratios between 10:1 and 40:1 result in efficiencies of 5'-GMPS incorporation of 70–90%. GMPS is available from Amersham, but the reagent can also be generated by chemical phosphorylation. A general description of this procedure based on Murry and Atkinson (20) can be found in Burgin and Pace (19). The second step of 5'-APA modification involves reaction of 5'-GMPS RNA with azidophenacyl bromide (Fluka/Pierce); modification is generally quantitative (19). This reaction is performed by resuspending the 5'-GMPS cpRNA in 100  $\mu\text{l}$  of 40% methanol, 20 mM sodium bicarbonate, pH 9.0, 0.1% sodium dodecyl sulfate, and 5 mM azidophenacyl bromide. Following an incubation at room temperature for 1 h, the reaction volume is increased to 200  $\mu\text{l}$  by adding 100  $\mu\text{l}$  of a solution containing 10 mM Tris-HCl, pH 8.0, and 1 mM ethylenediaminetetraacetate. The residual uncou-

pled photoagent is removed by phenol extraction and the RNA recovered by ethanol precipitation.

A convenient 3' end modification approach that was initially applied by Oh and Pace (21) involved chemically modifying the RNA 3' end to contain a primary amine, followed by modification of this moiety with a bifunctional reagent containing both an azido group, for subsequent crosslinking, and a hydroxysuccinimidyl group, for modification of the amine. Briefly, 3' modification is accomplished by oxidation of the 3'-*cis*-diol with 3 mM sodium periodate in 100 mM sodium acetate, pH 5.4, for 1 h at room temperature to create the 3'-dialdehyde in 50- to 100- $\mu\text{l}$  reactions that contain 5–10  $\mu\text{g}$  RNA. The RNA is recovered by ethanol precipitation and the oxidized RNA is alkylated under reducing conditions with ethylenediamine. This involves resuspending the RNA in 100  $\mu\text{l}$  of 20 mM imidazole, pH 8.0, 5 mM sodium cyanoborohydride, 1 mM ethylenediamine, and incubating at 37°C for 1 h. Sodium borohydride is then added to 5 mM and the incubation is continued for an additional 10 min. The RNAs are precipitated twice with ethanol and then reacted with 10 mM photoagent [*N*-hydroxysuccinimidyl 4-azidobenzoate (Pierce)] in 50 mM Hepes/sodium hydroxide, pH 9.0 at room temperature in a volume of 100  $\mu\text{l}$  for 1 h in the dark. This step and subsequent steps are performed in amber tubes and care is taken to minimize exposure to ambient light. Modified RNAs are again recovered by ethanol precipitation.

The general approach to adding s<sup>6</sup>G to the 5' end of an RNA is straightforward in that it involves a single step of priming transcription with the cross-linking agent. s<sup>6</sup>G, however, is only available commercially as the nucleoside and is not sufficiently soluble to function in transcription reactions. The 5'-nucleoside monophosphate (s<sup>6</sup>GMP) is sufficiently soluble to be used in transcription and can be generated by chemical phosphorylation as in Sergiev *et al.* (22) using the approach indicated above for GMPS. The approach involves reaction of the nucleoside with phosphorylchloride in triethylphosphate followed by hydrolysis and purification by anion-exchange chromatography. Specifically, 6-mercaptoguanosine [611 mg (2 mmol), Sigma] is dissolved in 5.0 ml triethyl phosphate and stirred on ice for 10 min. Phosphoryl chloride [0.6 ml (5.8 mmol), Aldrich] is then added and the mixture is stirred on ice in the dark for 4 h to yield a clear yellow solution. After adjusting the volume to 500 ml with water, the pH is adjusted to ~7.5 by the addition of triethylamine (~4.1 ml) over a period of 2 h. The neutralized

reaction mixture is then loaded onto a  $2.5 \times 18$ -cm Supelco TSK-gel Toyopearl DEAE-650M column previously equilibrated with 0.05 M triethylammonium bicarbonate (pH 7.5) at a flow rate of 60 ml/h and the column is washed with 4 vol of 0.05 M triethylammonium bicarbonate (pH 7.5). The 6-thioguanosine 5'-monophosphate is eluted using a 400-ml linear gradient of 0.05–0.5 M triethylammonium bicarbonate (pH 7.5). Fractions containing 6-thioguanosine 5'-monophosphate are located by measuring the absorption at 252 nm after  $200\times$  dilution into 1 mM sodium hydroxide. Peak fractions are subsequently pooled and the buffer is removed by rotary evaporation with the addition of 50 ml methanol. The solid 6-thioguanosine 5'-monophosphate (0.2 mmol) is dissolved in water at a final concentration of  $\sim 50$  mM, the pH is adjusted to  $\sim 7$  with sodium hydroxide, and it is stored in the dark at  $-70^\circ\text{C}$ . This method yields quantities of material that are more than sufficient ( $\mu\text{mol}$ ). For priming transcription with  $s^6\text{GMP}$ , reactions contain 10 mM analog and 1 mM GTP.

It is important to note that  $s^6\text{G}$  is a particularly reactive and efficient crosslinking agent and requires protection from UV light. Indeed, it is important to minimize the exposure of both photoagents to ambient or fluorescent light throughout the purification of RNA transcripts, before and after the crosslinking reaction itself, and prior to the mapping of crosslinked sites by reverse transcription described below. Exposure to light can lead to significant loss of the crosslinking signal by premature activation of the crosslinking moiety. In our experience, performing transcription and subsequent reactions in dark amber tubes has helped to minimize premature activation of these crosslinking agents.

#### *Crosslinking and Mapping of Crosslinked Nucleotides*

The quality of distance constraints obtained from photoaffinity crosslinking is a direct reflection of the extent to which RNAs can be folded into their native structure. Care must be taken to maintain uniformity in the manner in which RNAs are renatured and kept folded throughout the crosslinking reaction. In our studies, renaturing bacterial RNase P RNA involves resuspending transcripts in reaction buffers containing 1 M monovalent salts such as sodium chloride and heating to  $95^\circ\text{C}$  for 3 min. Transcripts are then cooled to  $37^\circ\text{C}$  and incubated 15–30 min in the presence of 25 mM divalent metal ion, optimally magnesium, to allow RNAs to fold. Cross-

linking of P RNA in the presence of pre-tRNA substrate, however, requires that cleavage be inhibited during renaturation and the crosslinking reaction. This is achieved by renaturing the two RNAs separately and in the presence of calcium, which has been shown to maintain the native RNA structure and allow tight binding of substrate, but suppresses the rate of pre-tRNA cleavage (23). Following renaturation, P RNA and pre-tRNA are incubated together at  $37^\circ\text{C}$  to allow for binding prior to crosslinking.

Typical crosslinking reactions containing 100–400 nM modified ribozyme in 100–200  $\mu\text{l}$  of reaction buffer are irradiated for 5–15 min using a common hand-held UV lamp and the crosslinked RNAs subsequently recovered by ethanol precipitation. Note, however, that APA-derived crosslinking agents are activated by a 302-nm light source (UVP, Model UVM-57 lamp) while  $s^6\text{G}$  is activated by a 366-nm light source (UVP, Model UVGL-58 lamp). Also, note that while most crosslinking can be done at room temperature, it is often desirable to keep crosslinking samples on ice to slow catalysis and to reduce thermal motion. Irradiation can be performed by placing the UV lamp approximately 2 cm over open microcentrifuge tubes; however, we have found that performing the reactions in 10- to 20- $\mu\text{l}$  droplets on Parafilm or in 50- $\mu\text{l}$  microtiter dishes results in higher crosslinking efficiency and is the most convenient method when there are many samples that require simultaneous processing. It is also important that the UV lamp be filtered with either a standard 1/8-in. glass plate or a polystyrene petri dish to absorb short wavelength ( $<300$  nm) light which can result in the formation of photoagent-independent crosslinks. Plastic screens, however, can become opaque from UV exposure and should be replaced after several uses.

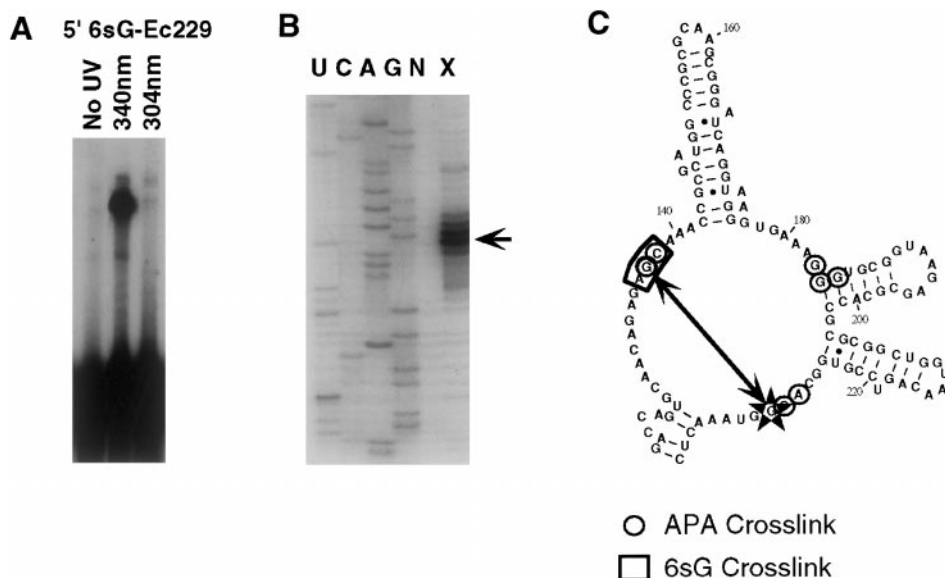
Individual crosslinked species are resolved by electrophoresis on a denaturing polyacrylamide gel. The crosslinked RNAs of interest are passively eluted from individual gel slices by incubating with 0.3 M sodium acetate, 10 mM Tris-HCl, pH 8.0, 1 mM ethylenediaminetetraacetate, and 0.1% sodium dodecyl sulfate. An example of the crosslinking reaction for 5'- $s^6\text{G}$ -modified P RNA circularly permuted at nucleotide 229 is shown in Fig. 3. Intramolecular crosslinking results in the formation of lariat species while intermolecularly crosslinked RNAs, such as that between P RNA and its pre-tRNA substrate, form Y-branched structures. Both lariats and branched species generated by crosslinking migrate

slower than linear forms in denaturing polyacrylamide gels and thus can be easily separated from unreacted RNA. RNAs are generally radiolabeled to facilitate detection; however, visualizing RNAs by ethidium bromide staining of thicker (5 mm) gels also works well.

Individual crosslinks are mapped using a standard reverse transcriptase primer-extension protocol. A method that works well in our hands involves the following. Equal amounts of gel-purified crosslinked material and 5'-<sup>32</sup>P-end-labeled sequencing primers (usually 0.1–0.2 pmol each in 5  $\mu$ l) are incubated at 65°C for 3 min in 50 mM Tris–HCl (pH 8.3), 15 mM sodium chloride, 10 mM dithiothreitol, and then frozen immediately on dry ice. Samples are then thawed on wet ice and MgCl<sub>2</sub> (1  $\mu$ l) is added to a final concentration of 6 mM. Subsequently, all four deoxynucleotides (dATP, dCTP, dGTP, and dTTP) and 2 units of AMV reverse transcriptase (Boehringer Mannheim) are added to the samples (to a final volume of 10  $\mu$ l) followed by incubation at 47°C for 5 min. Reactions are terminated by the addition of 10  $\mu$ l of a solution containing 0.5 M sodium chloride, 20 mM ethylenediaminetetraacetate, 1  $\mu$ g/ $\mu$ l glycogen (Boehringer Mannheim), and precipitated

in 2.5 vol (50  $\mu$ l) ethanol. Reactions are resuspended in 2  $\mu$ l distilled water and 2  $\mu$ l formamide containing 150 mM Tris–HCl (pH 8.0), 15 mM ethylenediaminetetraacetate, and 2.5 mg/ml bromophenol blue and xylene cyanole. Samples are heated to 95°C for 3 min, held on ice for 2 min, and then separated on a 6% denaturing polyacrylamide gel alongside standard dideoxy sequencing reactions of uncrosslinked P RNA generated using 0.5 mM dideoxynucleotides and the conditions above.

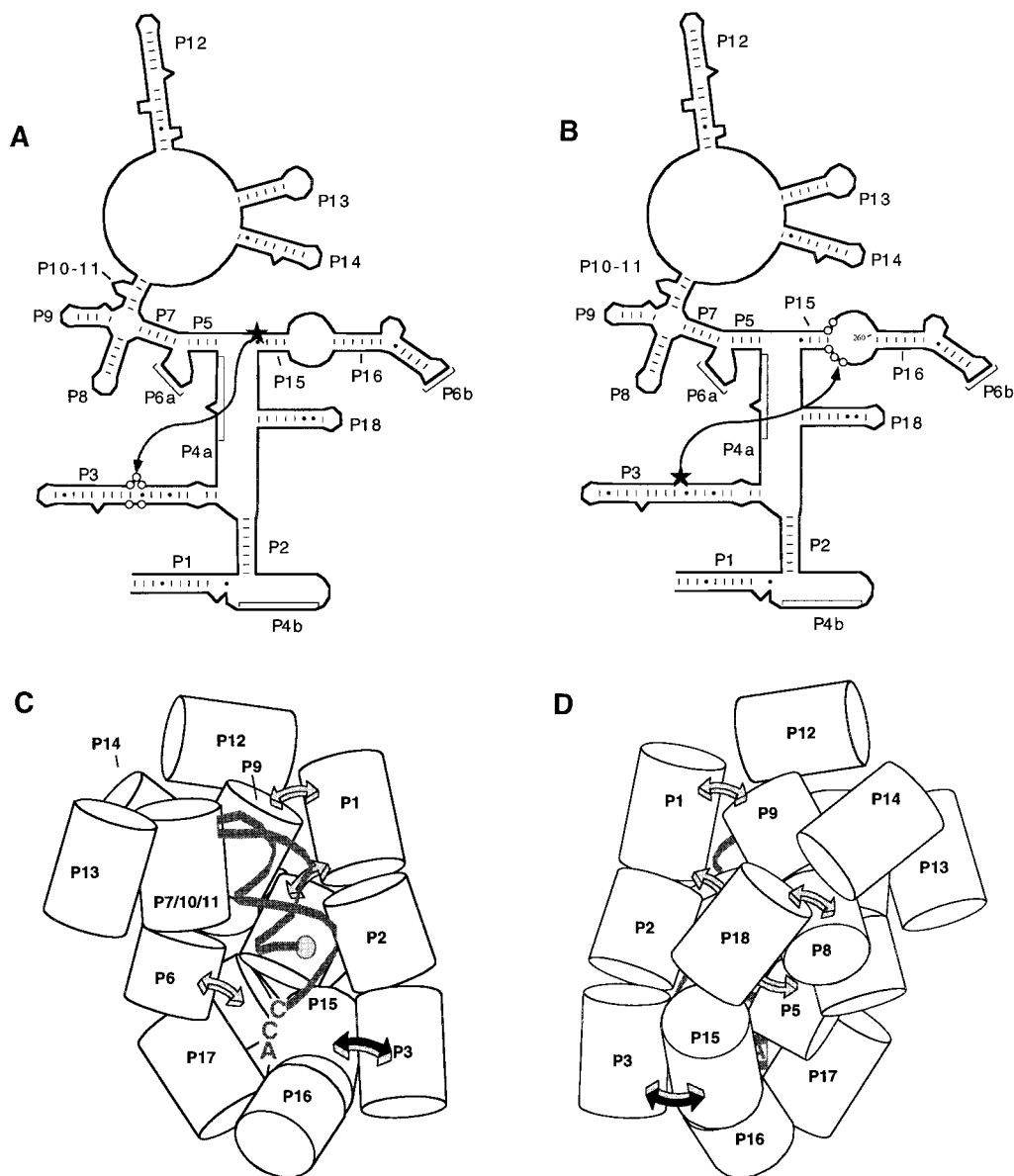
An example of primer-extension mapping data is shown in Fig. 3B. This figure also illustrates a key difference in crosslinking results obtained using the arylazide crosslinking reagents versus the shorter-range s<sup>6</sup>G reagent. Crosslinking using arylazide-modified Ec229 results in several crosslinked species, reflecting crosslinking to three different regions of the ribozyme which are illustrated by circled nucleotides in Fig. 3. This identifies nucleotides which are in proximity to, but may not necessarily interact with, nucleotide 229. Such constraints are useful for global structure modeling where multiple long-range distance constraints can give an overall picture of the position and orientation of helices in an RNA structure. The crosslinking pattern of 5'-s<sup>6</sup>G-



**FIG. 3.** Analytical crosslinking and mapping of crosslinked nucleotides. (A) Example of crosslinking of radiolabeled 5'-s<sup>6</sup>G-cpP RNA with 5' end at nucleotide G229. Photoagent-modified RNA was irradiated as indicated in standard P RNA reaction buffer (see text) and the crosslinked products resolved on a polyacrylamide gel. (B) Primer extension mapping of crosslinked nucleotides. Control reactions containing sequencing reactions and primer extensions using no ddNTPs are marked U, C, A, G, and N, respectively. Primer-extension products from the crosslinked species shown in part A is marked X. The position of crosslinked nucleotides are indicated by an arrow. (C) Summary of mapping results for both 5'-APA and 5'-s<sup>6</sup>G Ec229 RNA. Nucleotides detected with 5'-APA are circled. Nucleotides detected with s<sup>6</sup>G are boxed and are connected to the photoagent attachment site by an arrow. Nucleotide G229 is indicated by a star.

*Ec229* RNA is much more restricted; essentially only a single region of the ribozyme crosslinks with high efficiency. These results suggest that the nucleotides detected by intramolecular crosslinking with  $s^6G$  are in closer proximity to nucleotide 229 and therefore are better candidates for formation of specific

interactions with this nucleotide. In addition to the intramolecular reaction, intermolecular crosslinking with  $s^6G$ -modified tRNA has also helped refine positional information gained using APA reagents. Photocrosslinking between  $s^6G$  at the tRNA 5' end has identified specific nucleotides adjacent to the



**FIG. 4.** Incorporation of long-range crosslinking constraints into structure models. (A and B) Examples of photocrosslinking results between nucleotides in helices P3 and P15 from (13) and (14) demonstrating reciprocity. Sites of APA photoagent attachment are shown as stars; crosslinked nucleotides are shown as circles and are connected to sites of photoagent attachment by an arrow. (C and D) Front and back views of the model of the RNase P ribozyme-substrate complex. The positions and orientations of individual P RNA helices are shown as cylinders. tRNA is shown as a gray ribbon. For clarity, only the acceptor and T-stems of tRNA are shown. The position of the 5' end of tRNA is shown as a sphere. Key crosslinking constraints are shown by arrows. The constraint illustrated in A and B and described in the text is shown as a darker arrow.

substrate phosphate in the ribozyme-tRNA complex (28). Clearly, these results can be used to guide the design of experiments targeting these regions for mutation or modification.

#### *Assessing the Validity of Crosslinking Data*

Any structure determination method requires an assessment of the relationship between the derived structure and the native (active) RNA conformation, particularly since RNAs can potentially adopt multiple conformations (24–26). Structurally heterogeneous populations of RNA might be manifested as an expanded array of crosslinked positions. As shown in Fig. 3, crosslinking usually involves several adjacent nucleotides in one to four distinct regions of the target RNA with long-range structure probes such as APA, while the number of nucleotides and crosslinked regions of RNA is significantly reduced with short-range structural probes such as s<sup>6</sup>G. The cpRNA constructs and renaturing conditions should, therefore, be reexamined if they produce crosslinking to many regions of the target RNA or if crosslinks conflict with other well-established crosslinking or biochemical data. It is also important to note that the absence of crosslinking does not imply the lack of structural proximity. It is likely that crosslinking can be dependent upon the local three-dimensional structure of the target site, which may present an unfavorable geometry for nuclear attack or juxtapose the probe with unreactive functional groups.

The ability to meaningfully interpret structural data from crosslinking experiments depends critically on a few criteria. First, the efficiency of the chemical formation of the crosslinking species is relevant. It can be argued that crosslinked species obtained with high efficiency most likely, but not conclusively, represent the most stable (i.e., native) structure in the population. However, the possibility of kinetically trapping a minor, nonnative conformation that is in rapid equilibrium with the native structure means that efficiency alone does not prove the relevance of any individual crosslink. Second, the ability to reproduce the crosslink in the context of a second homologous RNA can be used to test the generality of the crosslinking results. Identical crosslinking results for two homologous, but structurally distinct, RNAs indicates that the results are not idiosyncratic to any one individual RNA species. For example, many of the constraints developed for P RNA structure were derived from comparative crosslinking experiments in both *Bacillus subtilis*

and *Escherichia coli* RNAs (2, 13, 14). Comparative analysis is particularly useful when gathering constraints on the conserved, and therefore most functionally important, portions of the macromolecule. Third, the ability to link two positions within the RNA chain together with two reciprocal crosslinks lends additional support to individual crosslinking constraints on structure. That is, when a photoagent attached to site A detects a second site, B, it is often useful to position the photoagent in a second experiment at site B in order to assess whether site A is detected. Finally, and perhaps most importantly, it is essential to demonstrate that individual crosslinked species retain biological activity. Although it is possible that crosslinked species reflecting nonnative conformations might retain the flexibility to refold into an active conformation, taken together with the additional criteria described above biological activity gives sufficient validity to the crosslinking results that they may be taken as meaningful constraints on structure.

An example of attempts to fulfill these criteria for a constraint used in modeling the RNase P ribozyme is shown in Fig. 4. Positioning an APA reagent in P15 resulted in efficient crosslinking to nucleotides in P3 (13). The proximity of these two helices is further supported by reciprocal crosslinking detected when the photoagent was positioned in P3 at the nucleotides which crosslinked when P15 was modified (14). Furthermore, experiments with *B. subtilis* P RNA, a species which has the same conserved core secondary structure as the *E. coli* ribozyme but differs significantly in phylogenetically variable sequences, showed that an almost identical group of nucleotides were crosslinked when the photoagents were placed at homologous positions (14). Thus, the proximity in the folded RNA between P15 and P3 is supported by both reciprocal and comparative crosslinking results. In addition, the crosslinked species from *E. coli* P RNA were found to retain catalytic activity. Taken together, these data support the validity of the P15–P3 constraint for the native ribozyme structure.

This constraint, as well as several additional constraints from photocrosslinking, were used along with information from comparative sequence analysis to generate a model of the structure of the RNase P ribozyme–substrate complex (14; Fig. 4). An approach combining molecular mechanics computer modeling and manual structure manipulation was used to derive a single working model of the structure. A detailed description of this structure and,



most importantly, estimates of its resolution have been published recently (13, 14, 27). The available set of data provide sufficient information to specify the relative helix positions and orientations relative to each other and to tRNA. The relatively low resolution of the crosslinking constraints can be dramatically refined when they lend support to elements of previously defined structure identified by phylogenetic sequence comparisons. For example, in P RNA three GNRA-tetraloop interactions, two of which were corroborated by photocrosslinking data, provide crucial constraints on the overall structure. However, in many regions, especially complex helix junctions, the resolution of the model is quite low (10–15 Å). More recent experiments with the shorter range s<sup>6</sup>G probe promise to significantly refine the positions of conserved nonhelical elements of P RNA in the ribozyme-substrate complex (28, 29). The short-range crosslinking approach described above, combined with comparative and biochemical data, promises to allow further refinement of these elements and, in principle, can be generally applied to identify specific elements of RNA structure and to probe RNA structure.

---

## ACKNOWLEDGMENTS

Special recognition is due to Dr. James Nolan, Dr. Jiunn-Liang Chen, Alexei Kazantsev, Dr. Arun Malhotra, Dr. Stephen Harvey, and Dr. Norman Pace, all of whom contributed extensively to the crosslinking and structure development described here. In particular, we acknowledge the enthusiastic support and intellectual contributions of Dr. Norman Pace in whose laboratory these approaches were developed. We gratefully acknowledge Dr. David McPheeters for refinements of the chemical phosphorylation procedure. Also, we thank Dr. Jo Ann Wise and the members of the Harris laboratory for critical reading of the manuscript.

---

## REFERENCES

1. Malhotra, A., and Harvey, S. C. (1994) *J. Mol. Biol.* **240**, 308–340.
2. Harris, M. E., Nolan, J. M., Malhotra, A., Brown, J. W., Harvey, S. C., and Pace, N. R. (1994) *EMBO J.* **13**, 3953–3963.
3. Wetzel, R., and Soll, D. (1977) *Nucleic Acids Res.* **4**, 1681–1694.
4. Hsu, L. M., Lin, F. L., Nurse, K., and Ofengand, J. (1984) *J. Mol. Biol.* **172**, 57–76.
5. Moore, M. J., and Sharp, P. A. (1992) *Science* **256**, 992–997.
6. Moore, M. J., and Query, C. C. (1997) in *RNA-Protein Interactions: A Practical Approach* (Smith, C., Ed.), Oxford Univ. Press, Oxford.
7. Pan, T., Gutell, R. R., and Uhlenbeck, O. C. (1991) *Science* **254**, 1361–1364.
8. Van Der Horst, G., Christian, A., and Inoue, T. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 184–188.
9. Reich, C., Gardiner, K. J., Olsen, G. J., Pace, B., Marsh, T. L., and Pace, N. R. (1986) *J. Biol. Chem.* **261**, 7888–7893.
10. Guerrier-Takada, C., and Altman, S. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1266–1270.
11. Waugh, D. S., and Pace, N. R. (1993) *FASEB J.* **7**, 188–195.
12. Hixson, S. H., Brownie, T. F., Chua, C. C., Crapster, B. B., Satlin, L. M., Hixson, S. S., Boyce, C. O., Ehrich, M., and Novak, E. K. (1980) *Ann. N. Y. Acad. Sci.* **346**, 104–114.
13. Harris, M. E., Kavansav, A., Chen, J.-L., and Pace, N. R. (1997) *RNA* **3**, 561–576.
14. Chen, J.-L., Nolan, J. M., Harris, M. E. and Pace, N. R. (1998) *EMBO J.* **17**, 1515–1525.
15. Favre A., Saintome C., Fourrey J. L., Clivio P., and Laugaa P. (1998) *J. Photochem. Photobiol.* **42**, 109–124.
16. Nolan, J. M., Burke, D. H., and Pace, N. R. (1993) *Science* **261**, 762–765.
17. Milligan, J. F., and Uhlenbeck, O. C. (1989) *Methods Enzymol.* **164**, 51–62.
18. Hixson, S. H., and Hixson, S. S. (1975) *Biochemistry* **14**, 4251–4254.
19. Burgin, A. B., and Pace, N. R. (1990) *EMBO J.* **9**, 4111–4118.
20. Murry, A. W., and Atkinson, M. R. (1968) *Biochemistry* **7**, 4023–4029.
21. Oh, B.-K., and Pace, N. R. (1994) *Nucleic Acids Res.* **22**, 4087–4094.
22. Sergiev, P. V., Lavrik, I. N., Wlasoff, V. A., Dokudovskaya, S. S., Dontsova, O. A., Bogdanov, A. A., and Brimacombe, R. (1997) *RNA* **3**, 464–475.
23. Smith, D., Burgin, A. B., Haas, E. S., and Pace, N. R. (1992) *J. Biol. Chem.* **267**, 2429–2436.
24. Brion, P., and Westhof, E. (1997) *Annu. Rev. Biophys. Biomol. Struct.* **26**, 113–137.
25. Pyle, A. M., and Green, J. B. (1995) *Curr. Opin. Struct. Biol.* **5**, 303–310.
26. Uhlenbeck, O. C. (1995) *RNA* **1**, 4–6
27. Harris, M. E., Frank, D., and Pace, N. R. (1997) in *RNA Structure and Function* (Simons, R. W., and Grunberg-Manago, M., Eds.), Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
28. Christian, E. L., McPheeters, D. S., and Harris, M. E. (1998) *Biochemistry* **37**, 17618–17628.
29. Christian, E. L. and Harris, M. E., submitted.