RNA (1998), *4*:739–749. Cambridge University Press. Printed in the USA.

Modeling active RNA structures using the intersection of conformational space: Application to the lead-activated ribozyme

SÉBASTIEN LEMIEUX,[1]* PASCAL CHARTRAND,[2]* ROBERT CEDERGREN,[2] and FRANÇOIS MAJOR[1]

[1]Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, Montréal, Québec, Canada H3C 3J7
[2]Département de Biochimie, Université de Montréal, Montréal, Québec, Canada H3C 3J7

ABSTRACT

The Pb^{2+} cleavage of a specific phosphodiester bond in yeast $tRNA^{Phe}$ is the classical model of metal-assisted RNA catalysis. In vitro selection experiments have identified a $tRNA^{Phe}$ variant, the leadzyme, that is very active in cleavage by Pb^{2+}. We present here a three-dimensional modeling protocol that was used to propose a structure for this ribozyme, and is based on the computation of the intersection of conformational space of sequence variants and the use of chemical modification data. Sequence and secondary structure data were used in a first round of computer modeling that allowed identification of conformations compatible with all known leadzyme variants. Common conformations were then tested experimentally by evaluating the activity of analogues containing modified nucleotides in the catalytic core. These experiments led to a new structural hypothesis that was tested in a second round of computer modeling. The resulting proposal for the active conformation of the leadzyme is consistent with all known structural data. The final model suggests an in-line SN2 attack mechanism and predicts two Pb^{2+} binding sites. The protocol presented here is generally applicable in modeling RNAs whenever the catalytic or binding activity of structural analogues is known.

Keywords: base triple; catalytic RNA; internal loop; MC-SYM; metalo-nucleotide complex; molecular modeling; noncanonical base pairs

INTRODUCTION

The leadzyme is a catalytic RNA that cleaves a specific ribophosphodiester bond in the presence of Pb^{2+} (Pan & Uhlenbeck, 1992b). It was originally isolated by in vitro selection of molecules undergoing cleavage from partially randomized sequence libraries related to the sequence of yeast $tRNA^{Phe}$ (Pan & Uhlenbeck, 1992a). The secondary structure of the leadzyme, shown in Figure 1, consists of two helical domains sandwiching an asymmetric internal loop of six nucleotides. Cleavage is effected at the phosphodiester bond between C1 and G2 (see Fig. 1A). Rather than producing 2′-3′-cyclic phosphates via a transphosphorylation/cleavage reaction as other catalytic RNA domains (Buzayan et al., 1986; Hutchins et al., 1986; Epstein & Gall, 1987; Forster & Symons, 1987), the leadzyme also performs a second hydrolytic reaction of the cyclic phosphate intermediate reminiscent of protein ribonucleases. Thus, the final products possess either a 5′-hydroxyl group or a 3′-phosphomonoester terminus (Pan & Uhlenbeck, 1992b). Although some NMR structural data are available on the conformation of this RNA domain, its detailed three-dimensional structure, a requisite to the understanding of the details of this reaction, is unknown.

Over the past years, we have been developing a computer-based modeling protocol for RNA consisting of first translating primary, secondary, and tertiary structure data into geometrical constraints, then applying a constraint satisfaction solver, MC-SYM (Major et al., 1991), to generate all-atom three-dimensional models, and finally, refining the preliminary models with molecular mechanics energy minimization (Nilsson & Karplus,

Reprint requests to: François Major, Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, Montréal, Québec, Canada H3C 3J7; e-mail: major@iro.umontreal.ca.
*The first two authors contributed equally to the work.

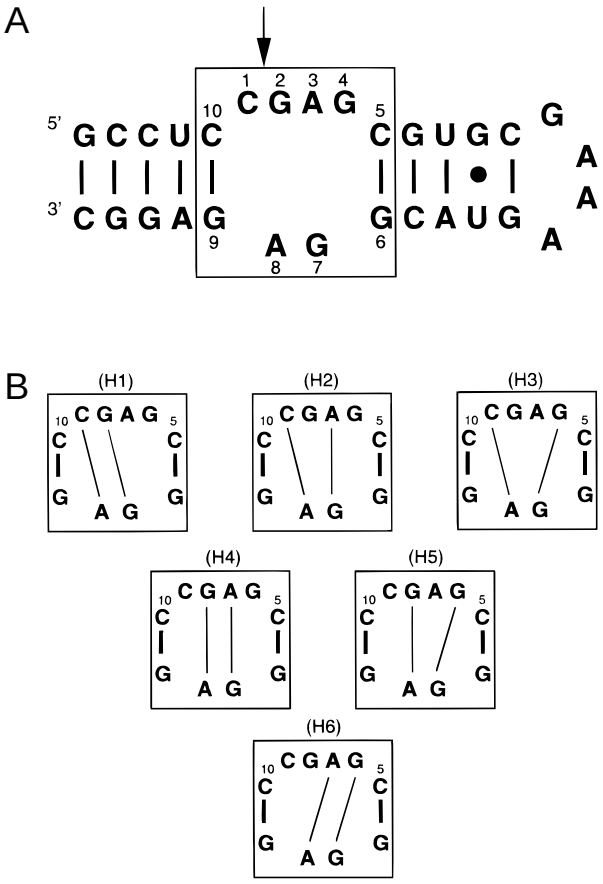

FIGURE 1. A: Primary and secondary structure of the leadzyme. Thick lines indicate Watson–Crick base pairs and bullets indicate non-Watson–Crick base pairs. Arrow indicates the cleavage site between C1 and G2. **B:** Six structural hypotheses that maximize base pairing in the internal hexaloop. Thin lines indicate non-Watson–Crick base pairs.

1986; Weiner et al., 1986). The small size of the leadzyme rendered it an ideal subject for our procedure. Moreover, the abundance of activity data on structural analogues of the leadzyme provided us with the challenge of including these data in an automatic protocol.

During the modeling of the Rev protein binding site of HIV-1, the ability of a collection of RNA molecules (aptamers) to engage in a common bimolecular interaction was taken to mean that the aptamers share many aspects of their three-dimensional structures (Leclerc et al., 1994). In particular, the geometry of base–base interactions at the binding site should be conserved among aptamers in spite of sequence variation. In the case of the leadzyme, many structural analogues have been isolated and their catalytic activity has been determined. Catalytic activity may be a more powerful reflection of the three-dimensional structure of the molecule than the binding ability of above because analogues have to mimic all conformations along the reaction pathway as well as the energy levels. Therefore, the active conformation of the leadzyme should be found among the conformations common to the leadzyme and its active analogues. We present here the principle of the intersection of conformational space (ICS) of catalytically active molecules, which allows the use of activity data. ICS consists of the generation of a conformational library of each active analogue, and the selection of common conformers from each library. A preliminary three-dimensional model of the leadzyme was produced by ICS, and subsequently refined by experimental data and further modeling (see Table 1). The final structure is consistent with all structural and activity data and suggests a plausible reaction mechanism.

TABLE 1. Overview of the modelling strategy.[a]

Data available	Modelling steps
	Initial models
Primary and secondary structure	1. Formation of six structural hypotheses
	2. Generation of conformations for the wild-type sequence using MC-SYM
	3. Classification
	Intersection of Conformational Space
In vitro selection	4. Threading of the variant sequences
	5. Assignment of probabilities using fuzzy set theory
	Final models
Chemical modification	6. Formation of new structural hypotheses
	7. Generation of conformations for the wild-type sequence using MC-SYM
	8. Refinement using molecular mechanics energy minimization
	9. Insertion of the lead ions
	10. Refinement using molecular mechanics energy minimization

[a]Initial models for the wild-type sequence were generated from primary and secondary structure information. Then, the initial models were scanned for common conformations among the wild-type and other active sequences derived from in vitro selection data. Finally, chemical modification experiments were used for verification of the best initial model and generation of a new structural hypothesis. Molecular mechanics energy minimization was then used to obtain the final model.

RESULTS

Initial models

The internal loop of the leadzyme is composed of six nucleotides, four on one strand and two on the other (see Fig. 1A). No firm structural constraint is known with the exception of a pairing between C1 and A8 at pH 6.5 in the absence of lead ions, as suggested by NMR spectroscopy (Legault, 1995; Legault & Pardi, 1994, 1997). To derive the first set of structural hypotheses, a maximum of base pairing and base stacking was assumed because known RNA structures and basic knowledge of RNA thermodynamics suggest that base stacking and base pairing predominate in the stabilization of internal loops (Tinoco et al., 1987; Varani et al., 1989; Wimberly et al., 1993; Cai & Tinoco, 1996). The 4×2 asymmetry of the leadzyme internal loop allows for a maximum of two cross-strand base pairs and six ways in which these two base pairs could be distributed (see Fig. 1B). Each of these combinations was evaluated by defining six structural hypotheses: Hypothesis 1 (H1) proposes the pairs C1·A8 and G2·G7. Hypothesis 2 (H2) to hypothesis 6 (H6) involve, respectively, C1·A8 and A3·G7, C1·A8 and G4·G7, G2·A8 and A3·G7, G2·A8 and G4·G9, and A3·A8 and G4·G9.

The conformational library for each hypothesis was generated by the combinatorial assembly of every hydrogen bonding pattern of each base pair. Here, we evaluated only hydrogen bonding involving at least two hydrogen bonds as described by Saenger (1984). Consideration of one hydrogen bond interaction is possible (see below), but would greatly expand the number of structures that would be generated. Thus, a G·G base pair can be constructed using four different hydrogen bonding patterns: GG(III) [Roman numerals refer to the Saenger system of base pair conformations (Saenger, 1984)]: GG(IV), GG(VI), and GG(VII); the A·G base pair by four hydrogen bonding patterns: AG(VIII), AG(IX), AG(X), and AG(XI); the A·A base by three hydrogen bonding patterns: AA(I), AA(II), and AA(V); and the A·C base pair by two hydrogen bonding patterns: AC(XXV) and AC(XXVI).

In addition, conformational patterns involving protonated bases or less than two hydrogen bonds were also implemented for the case of the A·C interaction (see Fig. 2). Arabic numerals refer to this class of interactions, which include: the pseudo-wobble AC(75) (Rould et al., 1989); the major tautomeric forms of the bases AC(71) and AC(78) (Hutchins et al., 1986); and other patterns generated mathematically, AC(68, 69, 70, 72, 73, 74, 76, 77, 79, 80). These latter patterns were evaluated by MC-SYM, even though they have never been observed experimentally. The protonated A·C base pair of type AC(75) suggested by NMR data (Legault & Pardi, 1994) was considered in hypotheses H2 and H3. Using the wild-type sequence, 60 different base pair patterns for H1, H2, and H3, 16 patterns for H4 and H5, and 12 patterns for H6 could be con-

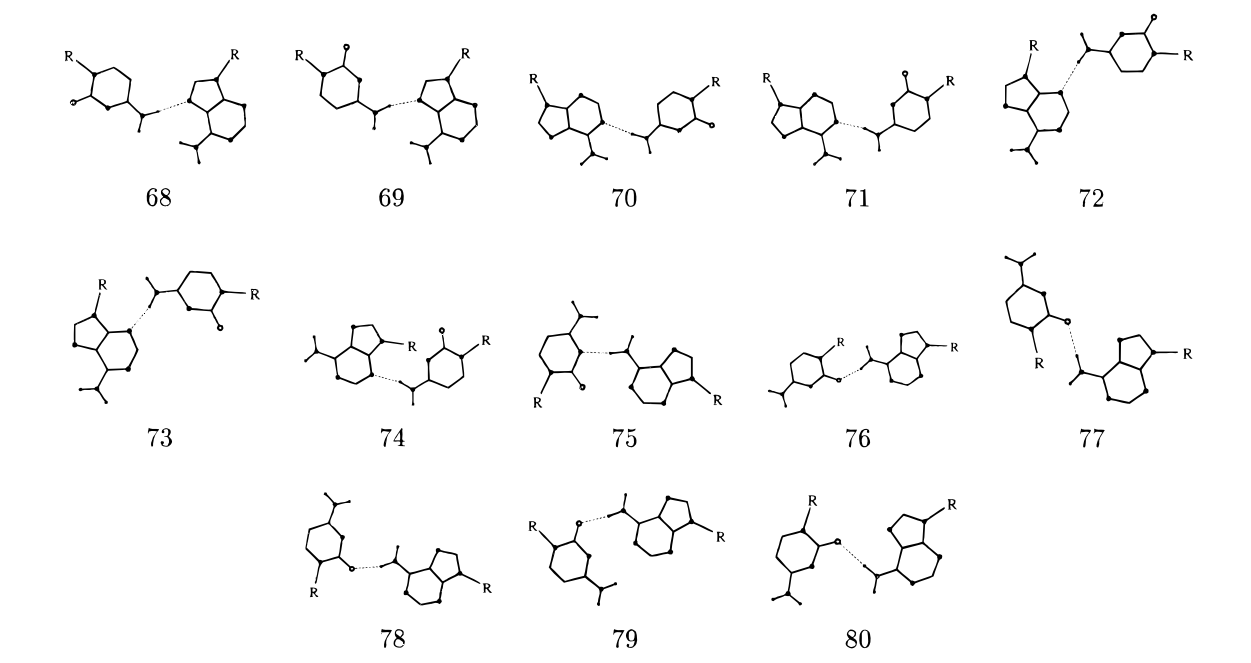

FIGURE 2. A·C base pairing patterns implemented in MC-SYM to test the C1·A8 base pair. All patterns involve a single hydrogen bond, except when protonation occurs, then two hydrogen bonds can be formed.

structed. Additional conformational enrichment was created by consideration of the sugar pucker and glycosyl bond torsion angles. Nucleotides were assigned either the C2′- or C3′-*endo* sugar puckers and either the *anti* or *syn* glycosyl bond torsion angles. Based on existing structures, all nucleotides in double-helical stems including the flanking nucleotides were assigned the A-RNA helix conformation.

The data sets corresponding to the six hypotheses were submitted to MC-SYM. Figure 3 shows the order in which individual nucleotides were added to the nascent models in the step-by-step construction procedure. For each hypothesis of Figure 1B, the base pairing pattern and nucleotide conformation combinatorics defined conformational landscapes of more than 5×10^{11} different conformations. Each of these conformations was evaluated for the wild-type sequence. At each step in the construction, the MC-SYM constraint satisfaction solver pruned conformations that violated basic stereochemical rules (see Major et al., 1993). MC-SYM found nearly 15,000 conformations that were compatible with structural hypotheses H2, H3, H4, and H5, but no structures could be constructed for H1 nor H6. To simplify the subsequent structural analysis, the conformers for each hypothesis were classified into 19 classes based on the base pairing patterns (see Table 2). Classes had up to 15 subclasses based on sugar puckers and glycosidic angles.

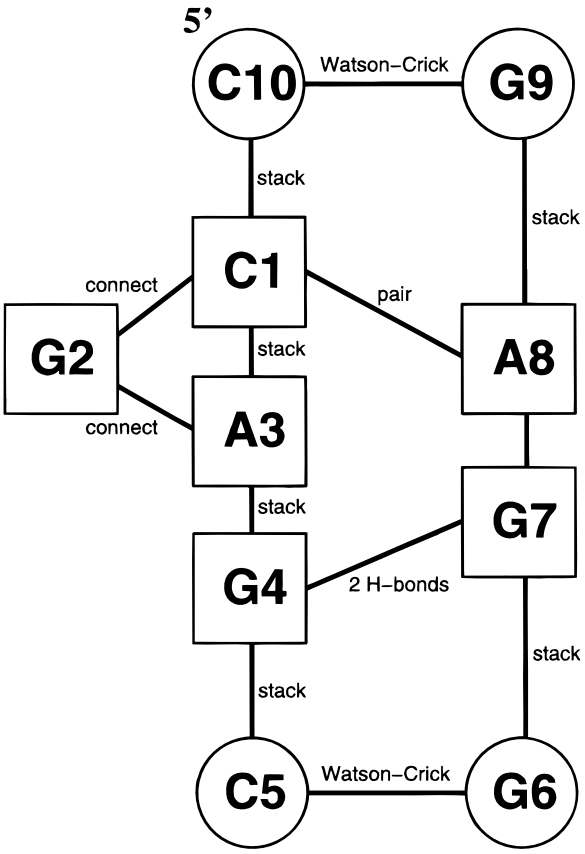

FIGURE 3. Graph of relations to model the leadzyme under H3. Nucleotides are represented by vertices and their spatial relations by edges. Circled nucleotides were assigned C3′-*endo* sugar puckers and *anti* glycosyl torsion (A-RNA helix type). Boxed nucleotides were assigned either the C2′- or C3′-*endo* sugar puckers and either the *anti* or *syn* glycosyl bond torsion angles (sampled). The interpretation of this figure is as follows. The base of C1 stacks on the base of C10. The base of A3 stacks on the base of C1. G2 bulges outside the helix. The base of G4 stacks on the base of A3. The base of C5 stacks on the base of G4. G6 forms a Watson–Crick base pair with C5. G7 forms a base pair involving at least two hydrogen bonds with G4 and stacks on the base of G6. A8 forms a base pair with C1. G9 forms a Watson–Crick base pair with C10 and stacks on the base of A8.

TABLE 2. Summary of MC-SYM simulations for the wild-type leadzyme sequence.[a]

Hypothesis	Class	Subclasses (#)	Structures (%)
H2	AC(69)AG(XI)	8	12.9
	AC(75)AG(XI)	2	6.3
	AC(77)AG(XI)	2	3.8
	AC(78)AG(XI)	4	10.4
	AC(79)AG(XI)	8	4.4
	AC(69)AG(VIII)	9	10.4
	AC(71)AG(VIII)	15	51.7
H3	AC(69)GG(VI)	2	14.0
	AC(71)GG(VI)	2	86.0
H3′	CG(XIX)GG(110)	1	58.3
	CG(XXII)GG(VI)	1	41.7
H4	AG(VIII)AG(VIII)	7	7.2
	AG(VIII)AG(IX)	2	0.4
	AG(IX)AG(VIII)	7	2.4
	AG(IX)AG(XI)	3	3.0
	AG(X)AG(XI)	1	6.8
	AG(XI)AG(XI)	6	80.2
H5	AG(VII)GG(VI)	1	66.7
	AG(IX)GG(VI)	1	33.3

[a]Conformations are defined by hydrogen bonding patterns. Subclasses are based on sugar pucker modes and base orientation at the glycosyl bond. The percentage of models refers to the number of models containing the indicated hydrogen bonds relative to the total number of models in that class.

Computing the ICS of the leadzyme

In theory, sets of conformations should be computed for each active analogue to determine the ICS. In this case, the analogues were identified from in vitro selection (Pan & Uhlenbeck, 1992a), but generally any molecule having the same activity could be considered. A computationally more efficient technique, however, consists of trying to thread the analogue sequences through the conformations found for the wild-type sequence of the leadzyme. The results of this exercise are shown in Figure 4.

The wild-type sequence is the only sequence that generated conformations under H2, H3, H4, and H5 (Fig. 4). The set of hypotheses associated with this molecular modeling experiment is $WT = \{H2, H3, H4, H5\}$ and $m(WT) = 1/9$ (the wild type is one out of nine

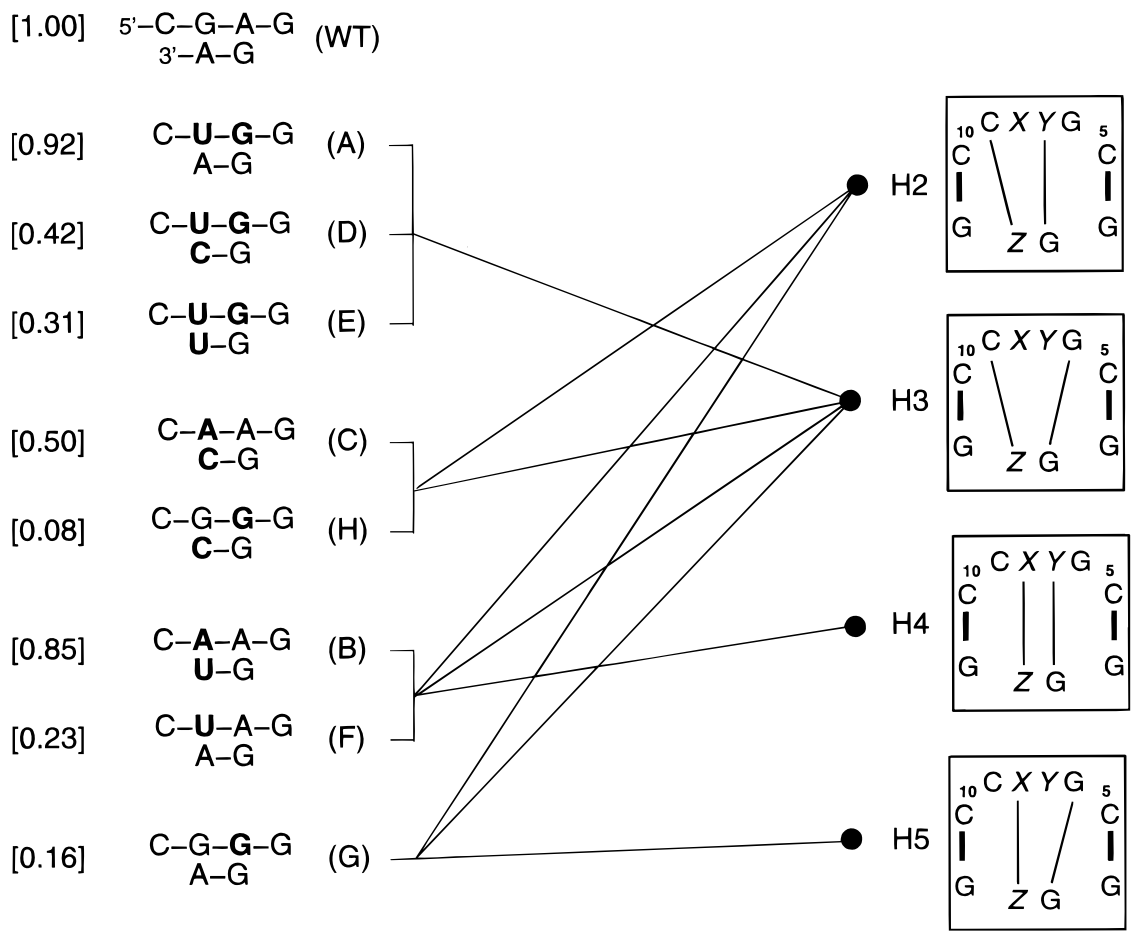

FIGURE 4. Network of folding compatibilities. Numbers in brackets indicate relative activity (Pan et al., 1994). Variant identifiers are in parentheses. Edges connect sequences to the structural hypotheses for which MC-SYM has found consistent conformations. Wild-type sequence can be folded into structural hypotheses H2, H3, H4, and H5. Only H3 accommodates all active analogue sequences.

variants, see Materials and Methods). Variants *A*, *D*, and *E* fold only according to H3, *ADE* = {H3}; variants *B* and *F* fold to H2, H3, and H4, *BF* = {H2, H3, H4}; variants *C* and *H* fold to H2 and H3, *CH* = {H2, H3}; and variant *G* folds to H2, H3, and H5, *G* = {H2, H3, H5}. These folding patterns imply the following basic probabilities: m(*ADE*) = 1/3; m(*BF*) = m(*CH*) = 2/9; and m(*G*) = 1/9, and thus, the likelihood for {H2}, {H3}, {H4}, and {H5} are [0, 2/3], [1/3, 1], [0, 1/3], and [0, 2/9], respectively (see Materials and Methods). These probabilities indicate that H3 has the highest degree of belief, and, thus, is most likely to contain active conformations of the leadzyme. The plausibility of 1 indicates that H3 is the only hypothesis that can accommodate all active variants (see Materials and Methods).

Two major conformational classes are found for H3 (Table 2): 14% feature the AC(69) and GG(VI) pairing patterns and 86% contain the AC(71) and the GG(VI) pairing patterns. The sugar pucker and the glycoside bond rotation further divide these classes into minor subclasses. The RMSD between any pair of models in a given subclass is less than 2.0 Å. Both major classes contain the GG(VI) pairing pattern and allow for the isosteric substitution of C1·A8 by C1·C8 or C1·U8 found among the analogue sequences. Note that the AC(69) and AC(71) base pairing patterns are present among the possible active conformations, but not the AC(75), the protonated pair suggested by the NMR data (Legault & Pardi, 1994). This structure was found only in the conformations compatible with H2.

A new structural hypothesis

During the course of these modeling studies, the activity of different leadzymes having a modification at virtually every functional group in the asymmetric internal loop became available (Chartrand et al., 1997). Functional groups whose modification decreased the catalytic activity of the ribozyme by at least a factor of one order of magnitude when compared to the wild-type leadzyme are shown in Figure 5.

The major inconsistency between the preliminary model and the modified leadzyme data concerned the C1·A8 base pair. Although the functional groups of C1

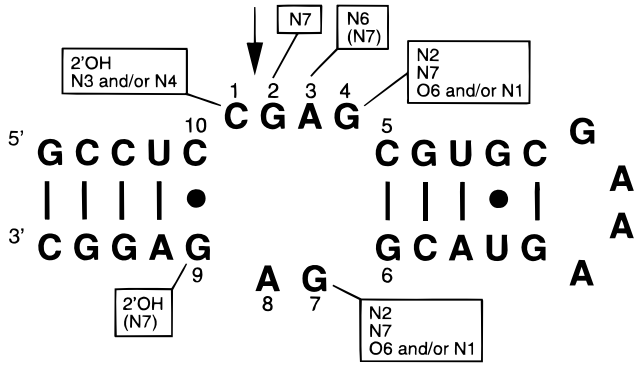

FIGURE 5. Summary of the functional groups found important for the catalytic activity of the leadzyme. Deletion of these functional groups leads to a decrease in activity by more than one order of magnitude compared with the wild-type. Those in parentheses reduce the activity by a factor less than 10×.

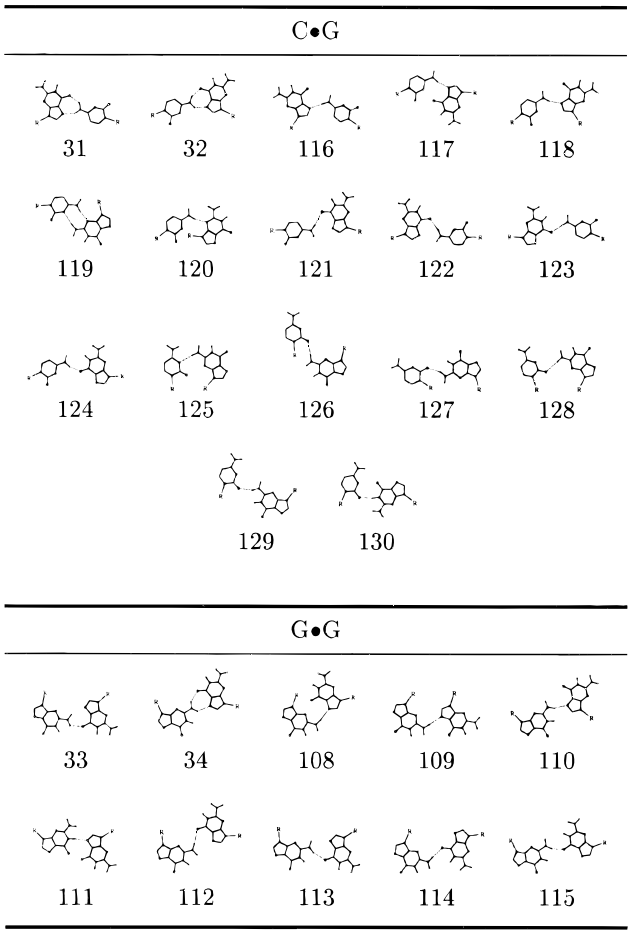

FIGURE 6. C·G and G·G base pairing patterns involving one hydrogen bond, as implemented in MC-SYM to test the C1·G7 and G4·G7 base pairs. The C·G base pairs were assigned canonical Watson–Crick patterns. The G·G base pair was assigned two hydrogen bond pairing patterns only.

were found to be crucial in the modified leadzyme data, A8 was clearly not involved in any interaction (Chartrand et al., 1997). This inconsistency raised the possibility that C1 could be involved in an interaction with G7 because type VI base pairing of G4·G7 does not implicate the functional groups of G7 involved in a Watson–Crick base pair. This possibility gave rise to H3′, a new structural hypothesis not considered in the initial modeling process. H3′ is derived from H3, but contains the base triple C1·G7·G4. Given the invariability of C1, G7, and G4, the C1·G7·G4 combination is the only possible triple under H3.

The conformational space of this triple is defined by 19 base pairing patterns for the C1·G7 base pair (Watson–Crick, reverse Watson–Crick, and 17 patterns involving one hydrogen bond; see Fig. 6), and 14 base pairing patterns for the G7·G4 base pair (the three G·G base pairing patterns described in the initial model and 11 patterns involving one hydrogen bond; see Fig. 6). There are 266 theoretical conformations of this base triple, however, only 61 C·G·G triples are sterically sound and, among these, only 11 are possible in the context of the leadzyme hexaloop.

Because stereochemical defects are inherent in models derived from the systematic use of discreet nucleotide conformation sets, molecular mechanics refinement (see Materials and Methods) was necessary to optimize the geometry of the models. Models that were not able to maintain their base pairing patterns during optimization were eliminated from further consideration, leaving only two models that were successfully optimized (see Table 2). One of these two models featured the CG(XIX) and GG(110) base pairing patterns, and was therefore discarded because many of the functional groups found to be important from the modified nucleotide data, i.e., the significant A3:N^6, G7:N^7, G4:N^1, G4:O^6, G4:N^2, and G2:N^7, were not involved in stabilizing interactions and A3:N^6 was positioned far from the active site. On the other hand, the model featuring the triple composed of the CG(XXII) (reverse Watson–Crick) and GG(VI) base pairs satisfied all available structural data.

DISCUSSION

A conformation consistent with catalytic activity

The conformation proposed by this procedure, as shown in Figure 7, satisfies many criteria that could be expected of an active conformation. In particular, the atom C1:$O^{2'}$, the presumed nucleophile, is aligned with the G2:P and G2:$O^{5'}$, as required of an in-line cleavage mechanism. A potential Pb^{2+} binding site is found adjacent to the cleavage site and is bounded by the base triple on one side and the backbone atoms of C1 and G2 on the other (see Fig. 7A). The Pb^{2+} cofactor is presumed to be bound in proximity to the catalytic site

to promote the nucleophilic character of the C1:$O^{2'}$. In this position, the electronegative binding pocket is defined by A3:N^6, A3:N^7, G2:P, C1:$O^{2'}$, G4:O^6, and G4:N^7, all of which have been shown to be important for catalytic activity and could contribute to the affinity of the cation at this site.

The in-line mechanism suggested by this structure, although likely, differs from the adjacent mechanism suggested for the yeast $tRNA^{Phe}$, which involves pseudorotation at the scissile phosphate (Brown et al., 1985). Current experimental data do not permit a distinction between these two mechanisms. The model also suggests that the catalytic lead cation could be involved in the stabilization of the transition-state trigonal bipyramid conformation of the phosphate group. The distance between the Pb-O^1P is less than 4 Å compared with the much longer distance observed in the lead binding pocket of $tRNA^{Phe}$ (Brown et al., 1985). Our model correctly predicts the specificity of cleavage because the C1:$O^{2'}$H is the only 2′-OH within the catalytic pocket (see Fig. 7B).

A second lead binding site?

The loss in free energy of transition-state stabilization, ΔG, due to the deletion of G7:N^2, is 3.2 kcal/mol, a value similar to that observed at G12:N^2 in the magnesium ion binding pocket of the hammerhead ribozyme (Scott et al., 1995). In the leadzyme model, G7:N^2 is involved in only one hydrogen bond with C1:N^3 and could therefore be available for ion binding. Also, modifications in this region reduced the cooperative binding of lead cations, whereas modifications in the catalytic site had no effect (Chartrand et al., 1997).

We thus propose a second cationic binding pocket involving G7:N^2, G9:$O^{2'}$, and C1:N^4. This binding could explain the finding that conversion of the G9–C10 base pair into a C9–G10 base pair or removing the G9:$O^{2'}$H led to an important loss of activity. A standard Watson–Crick base pair positions the G9:$O^{2'}$H toward the adenine base in position A3, whereas a reverse Watson–Crick base pair, as shown in Figure 7B, positions it below the plane of the G9 base, increasing its exposure to the Pb^{2+}, in the presumed binding pocket. This region contains several phosphodiester bonds that would be good potential coordination sites for the cation as well (Fig. 7).

To permit energy minimization of the metalo-nucleotide complex, Pb^{2+} ions were parameterized as indicated in Materials and Methods. The energy-minimized model is shown in Figure 7C. During minimization, the guanine base at position 2 moved slightly back toward the outside of the catalytic core, creating a better pocket for Pb^{2+} than in the initial model. However, the most striking aspect of this optimized structure is how the phosphodiester chain folds back on itself in the G9, A8, and G7 region. This folded-back conformation is due to a reverse Watson–Crick at the G9–C10 pair and the bulged-out nucleotide A8. Undoubtedly, this chain reversal would normally be unlikely due to the close juxtaposition of the charged phosphates in the looped back region; however, in the leadzyme structure, these charges could be shielded very effectively by the Pb^{2+} in the second binding pocket. The finding that an equimolar amount of neodymium, Nd^{3+}, and Pb^{2+} increases the catalytic activity of the leadzyme (Ohmichi & Sugimoto, 1997) could be due to the fact that only large ions provide sufficient shielding of the phosphate ions at this site. In the context of our two metal-ion binding site model, Nd^{3+} would bind in the structural site and Pb^{2+} in the catalytic site, because Nd^{3+} is inactive alone. Higher Nd^{3+}/Pb^{2+} ratios reduce the activity because Nd^{3+} ions may compete for the catalytic binding site.

No data are currently available on the functional groups involved in the formation of the catalytic site of the second step of the leadzyme reaction, the cyclic phosphate hydrolysis. However, according to our model, the propensity of the leadzyme to catalyze the hydrolysis of the cyclic phosphate could be due to: (1) the fact that the 5′ leaving group is an unpaired nucleotide and thus less likely to be in position for a reversal of the reaction; (2) the availability of water due to the state of hydration of the duodecavalent Pb^{2+} is much greater than Mg^{2+}, the usual metallic catalyst of ribozyme reaction; and, (3) the active conformation could exist long enough for a second chemical step.

Applicability of the ICS method

Structural modeling using the ICS approach, like any modeling project, is an attempt to reconcile all data into a three-dimensional form. This technique should not be confused with structural determination, because modeling necessarily depends on the quantity and, above all, quality of the data used during the process. In this sense, computer modeling with qualitative data can be thought of as a low-resolution data transformation process. Nevertheless, these models, even those of low resolution, offer the experimentalist a composite view of a variety of experimental results in an easy to understand form, thereby facilitating the design of more definitive experiments (Cedergren & Major, 1998).

In the present case, additional assumptions such as that concerning the conformation of the helical regions (the A-RNA helix conformation) and maximizing base pairing in the internal loop region of the molecule have been used. Basic knowledge of RNA thermodynamics suggests that base stacking and base pairing predominate in the stabilization of internal loops (Tinoco et al., 1987; Varani et al., 1989; Wimberly et al., 1993; Cai & Tinoco, 1996). Also, it is assumed that the nucleotide conformation at each position is present in the conformational set used in the preliminary model. However, energy minimizations used to construct refined models

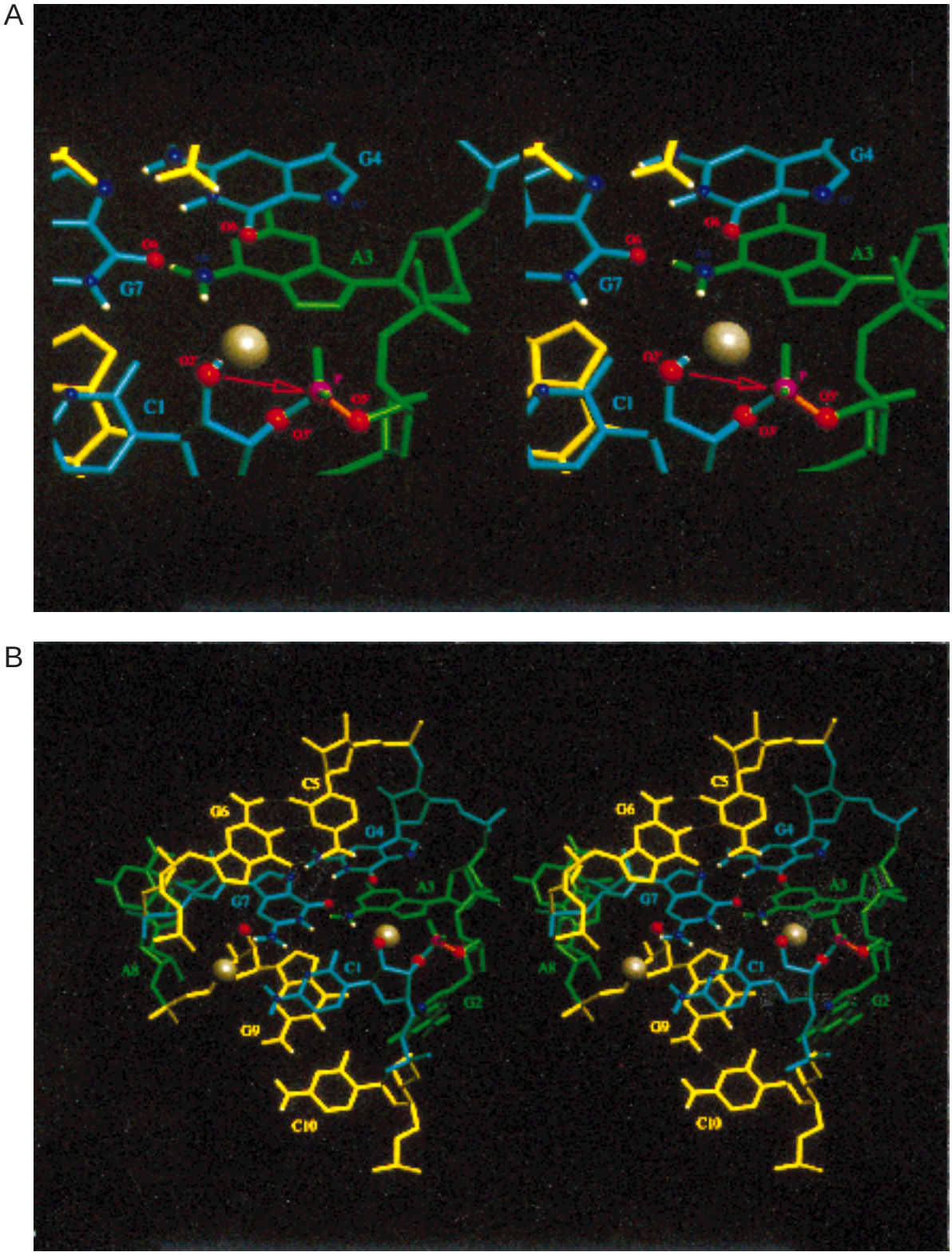

FIGURE 7. (*Figure continues on facing page.*)

limit the effect of this assumption. These types of assumptions are more practical than fundamental because they are used primarily to speed up the algorithm, and in some cases could be done away with entirely. It is important to understand that the ICS technique that we developed is in itself not dependent on these assumptions and could be applied to cases where no such assumptions were made.

C

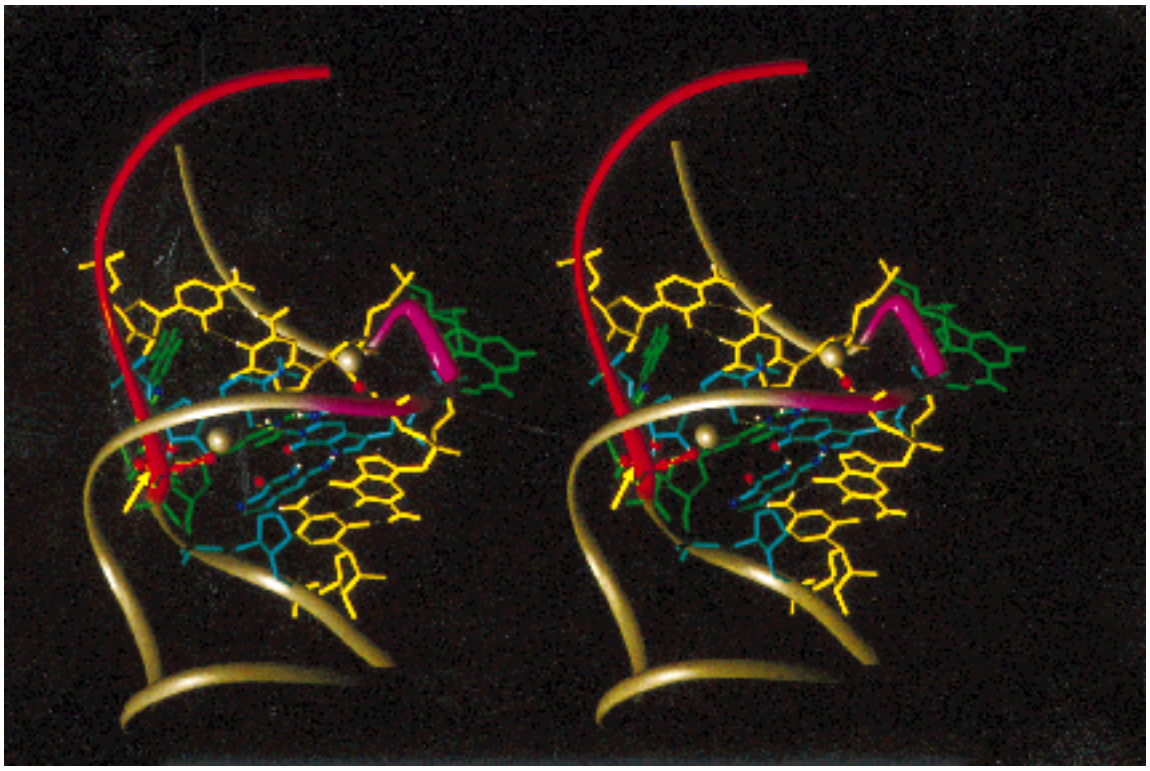

FIGURE 7. The proposed model. Nucleotides of the C·G·G base triple are shown in cyan. The C1:O$^{2'}$H, the P between C1 and G2 and the G2:O$^{5'}$ are aligned, suggesting the in-line attack indicated by the red arrow. The P-O$^{5'}$ scissile bond is shown in orange. The lead ions are shown in gray. The flanking G9·C10 and C5-G6 base pairs are shown in yellow. G2, A3, and A8 are shown in green. Atoms determined to be important for the catalytic activity are drawn with CPK surfaces using blue for nitrogen, red for oxygen, magenta for phosphate, and white for hydrogen atoms. Atoms with smaller CPK surfaces were not tested individually with modification experiments (see Fig. 4). **A:** In-line attack. A metal binding pocket can be observed whose internal surface is made up of many functional groups that were found important in modification experiments. **B:** Stereo view of the active site. **C:** Stereo view of the leadzyme complex with lead. The backbone ribbon of the nucleotides from the leadzyme 5′ end to C1 is shown in red. Near the second lead ion binding pocket, the phosphodiester chain folds back on itself as indicated by the magenta ribbon of nt 6–9. The backbone ribbon of all other nucleotides is shown in gray. Hydrogen bonds are shown as white dashed lines.

A model being dependent on available data is subject to change, as more structural data become available. A key element of the present work is that we were able to propose a single conformation that responds to a certain number of criteria expected of a catalytic molecule and is consistent with all available data on the leadzyme except for the proposed AC base pair, as is the modified nucleoside data as well. We believe that the ICS method should produce a truer picture of the active conformation of the leadzyme as the number of analogues increases. Even more precise structural data would be necessary to establish the nature of the structural interrelationships among the functional groups found important in the chemical modification data that are not involved directly in interactions.

MATERIALS AND METHODS

Formalism for the ICS approach

If we assume that each active sequence variant has an equal probability of adopting any given conformation, then probabilities can be assigned to structural hypotheses, X, in the following way: $m(X) = 0$, if none of the variants are compatible with the models in X, and $m(X) = 1$, if all variants are compatible with at least one model of the hypothesis in X. The likelihood of a single structural hypothesis, $\{x\}$, is given by an interval of probabilities, [Bel($\{x\}$), Pl($\{x\}$)], where Bel($\{x\}$) $= \Sigma_{Y\subseteq\{x\}} m(Y)$ and Pl($\{x\}$) $= \Sigma_{Y\cap\{x\}} m(Y)$ (Zadeh, 1983). The belief of x, Bel($\{x\}$), corresponds to the sum of the basic probabilities for the sets that contain exactly x, equal to $\{x\}$, and the plausibility of x, Pl($\{x\}$), corresponds to the sum of the basic probabilities of the sets that include x (Major et al., 1998).

Molecular mechanics energy minimization

MC-SYM structures were refined using molecular mechanics calculations performed by the molecular simulation program sander, from the Amber 4.1 suite of programs (Pearlman et al., 1995) using the Amber 94 forcefield. All 1–4 electrostatic interactions were reduced by a 1.2 factor as suggested for the 94 Amber forcefield. A distance-dependent dielectric model, $\epsilon = 4R_{ij}$, for the Coulombic representation of electrostatic interactions was used, as suggested by Weiner et al.

(1984; 1986). As a first step, energy minimization has been performed using the steepest descent for 50 steps, then the conjugate gradient method was applied until the maximum derivative was less than 0.01 kcal/mol/Å. The alignment of the C1:$O^{2'}$ with G2:P and G2:$O^{5'}$, and the base pairs assigned by MC-SYM were restrained during the first 5,000 steps. No cutoff was used during the minimization.

Parameterization of the lead ion

The charge parameter was fixed at 2.0. The van der Waals parameters were initially assigned a basic radius of 1.4 Å, corresponding to the Pauling radii. The potential well depth parameter, *e*, was fixed at 0.15 kcal/mol, corresponding to a slightly higher potential than that of the phosphate atom.

Molecular mechanics energy minimization with lead ions

Two Pb^{2+} were added to the leadzyme structure. The first was placed between atoms G4:O^6 and C1:$C^{1'}$. The second was inserted in the region bounded by the positions of atoms G6:P, G7:N^2, C1:N^4, and G9:$O^{2'}$. Both sites correspond to a high electronegativity surface (not shown). Molecular mechanics calculations were performed as described above except that the Amber 94 forcefield was modified to include the Pb^{2+} parameters. For the first 5,000 steps, the cations were restrained to their starting positions. The conjugate gradient method was applied until the maximum derivative was less than 0.1 kcal/mol/Å. Then, the restraints were removed and the minimization was continued until the maximum derivative was less than 0.01 kcal/mol/Å. The calculated empirical potential energy of the minimized lead structure was −168.6 kcal/mol. The atomic coordinates of the minimized model can be found at: http://www-lbit.iro.umontreal.ca/structures/leadzyme.pdb.

GAAA loop

The GAAA tetranucleotide loop of the leadzyme was modeled from the structural features of the solution structure of the GNRA class of tetranucleotide loops (Heus & Pardi, 1991). The MC-SYM script encoding the structural constraints defined for the GAAA loop describes a search tree of 28,946 nodes from which only eight models were consistent. The maximum RMSD (excluding H atoms) among the eight models was approximately 0.9 Å. One of the GAAA models found by MC-SYM was appended to complete the leadzyme structure.

ACKNOWLEDGMENTS

We thank Dr. Marielle Foucrault for preliminary work on the initial model, Majid Ftouhi for discussions on the mathematical formalism of the ICS technique, Stanislaw Oldziej for his help with the parameterization of the lead ion, Dr. Daniel Gautheret for providing a program to compute the non-Watson–Crick base pairing geometries, and the referees for useful and constructive comments. P.C. was supported by a predoctoral fellowship from the Medical Research Council (MRC) of Canada. R.C. is Richard Ivey Fellow of the Canadian Institute for Advanced Research programme in evolutionary biology. F.M. is a fellow of the MRC of Canada. This work was supported by the MRC of Canada.

Received August 12, 1997; returned for revision October 2, 1997; revised manuscript received April 10, 1998

REFERENCES

Brown R, Dewan J, Klug A. 1985. Crystallographic and biochemical investigation of the lead(II)-catalyzed hydrolysis of yeast phenylalanine tRNA. *Biochemistry 24*:4785–4801.

Buzayan J, Gerlach W, Bruening G. 1986. Satellite tobacco ringspot virus RNA: A subset of the RNA sequence is sufficient for autolytic processing. *Proc Natl Acad Sci USA 83*:8859–8862.

Cai Z, Tinoco IJ. 1996. Solution structure of loop a from the hairpin ribozyme from tobacco ringspot virus satellite. *Biochemistry 35*: 6026–6036.

Cedergren R, Major F. 1998. Modeling the tertiary structure of RNA. In: Simons RW, Grunberg-Manago M, eds. *RNA structure and function*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press. pp 37–75.

Chartrand P, Usman N, Cedergren R. 1997. The effect of structural modifications on the activity of the leadzyme. *Biochemistry 36*: 3145–3150.

Epstein L, Gall J. 1987. Self-cleaving transcripts of satellite DNA from the newt. *Cell 48*:535–543.

Forster AC, Symons RH. 1987. Self-cleavage of plus and minus RNAs of a virusoid and a structural model for the active sites. *Cell 49*:211–220.

Heus H, Pardi A. 1991. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science 253*:191–194.

Hutchins C, Rathjen P, Forster A, Symons R. 1986. Self-cleavage of plus and minus RNA transcripts of avocado sunblotch viroid. *Nucleic Acids Res 14*:3627–3640.

Leclerc F, Cedergren R, Ellington A. 1994. A three-dimensional model of the Rev-binding element of HIV-1 derived from analyses of aptamers. *Nature Struct Biol 1*:293–300.

Legault P. 1995. Structural studies of ribozymes by heteronuclear NMR spectroscopy [thesis]. Boulder, Colorado: University of Colorado.

Legault P, Pardi A. 1994. In situ probing of adenine protonation in RNA by 13C NMR. *J Am Chem Soc 116*:8390–8391.

Legault P, Pardi A. 1997. Unusual dynamics and pk_a shift at the active site of a lead-dependent ribozyme. *J Am Chem Soc 119*: 6621–6628.

Major F, Gautheret D, Cedergren R. 1993. Reproducing the three-dimensional structure of a transfer RNA molecule from structural constraints. *Proc Natl Acad Sci USA 90*:9408–9412.

Major F, Lemieux S, Ftouhi A. 1998. Computer RNA three-dimensional modeling from low-resolution data and multiple-sequence information. In: Leontis NB, SantaLucia J Jr, eds. *Molecular modeling of nucleic acids*. Washington DC: American Chemical Society. ACS Symposium Series 682. pp 394–404.

Major F, Turcotte M, Gautheret D, Lapalme G, Fillion E, Cedergren R. 1991. The combination of symbolic and numerical computation for three-dimensional modeling of RNA. *Science 253*:1255–1260.

Nilsson L, Karplus M. 1986. Empirical energy functions for energy minimization and dynamics of nucleic acids. *J Comp Chem 7*:591–616.

Ohmichi T, Sugimoto N. 1997. Role of Nd^{3+} and Pb^{2+} on the {RNA} cleavage reaction by a small ribozyme. *Biochemistry 36*:3514–3521.

Pan T, Dichtl B, Uhlenbeck O. 1994. Properties of an in vitro selected Pb^{2+} cleavage motif. *Biochemistry 33*:9561–9565.

Pan T, Uhlenbeck O. 1992a. In vitro selection of RNAs that undergo autolytic cleavage with Pb^{2+}. *Biochemistry 31*:3887–3895.

Pan T, Uhlenbeck O. 1992b. A small metalloribozyme with a two-step mechanism. *Nature 358*:560–563.

Pearlman D, Case D, Caldwell J, Ross W, Cheatham T, Ferguson D, Seibel G, Singh U, Weiner P, Kollman P. 1995. *Amber 4.1*. San Franciso, California: University of California.

Rould M, Perona J, Soll D, Steitz T. 1989. Structure of *E. coli* glutaminyl-tRNA synthetase complexed with $tRNA^{Gln}$ and ATP at 2.8 Å resolution. *Science 246*:1135–1142.

Saenger W. 1984. *Principles of nucleic acid structure*. New York: Springer-Verlag.

Scott W, Finch J, Klug A. 1995. The crystal structure of an all-RNA hammerhead ribozyme: A proposed mechanism for RNA catalytic cleavage. *Cell 81*:991–1002.

Tinoco IJ, Davis P, Hardin C, Puglisi J, Walker G, Wyatt J. 1987. RNA structure from A to Z. *Cold Spring Harbor Symp Quant Biol 52*:135–146.

Varani G, Wimberly B, Tinoco IJ. 1989. Conformation and dynamics of an RNA internal loop. *Biochemistry 28*:7760–7772.

Weiner S, Kollman P, Nguyen D, Case D. 1986. An all atom force field for simulations of proteins and nucleic acids. *J Comput Chem 7*:230–252.

Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S, Weiner P. 1984. A new forcefield for molecular mechanical simulation of nucleic acids and proteins. *J Am Chem Soc 106*:765–784.

Wimberly B, Varani G, Tinoco IJ. 1993. The conformation of loop E of eukaryotic ribosomal RNA. *Biochemistry 32*:1078–1087.

Zadeh LA. 1983. Commonsense knowledge representation based on fuzzy logic. *Computer 16*:61–65.