Research Article

The Combination of Symbolic and Numerical Computation for Three-Dimensional Modeling of RNA

FRANÇOIS MAJOR,* MARCEL TURCOTTE, DANIEL GAUTHERET, GUY LAPALME, ERIC FILLION, ROBERT CEDERGREN†

Three-dimensional (3-D) structural models of RNA are essential for understanding of the cellular roles played by RNA. Such models have been obtained by a technique based on a constraint satisfaction algorithm that allows for the facile incorporation of secondary and other structural information. The program generates 3-D structures of RNA with atomic-level resolution that can be refined by numerical techniques such as energy minimization. The precision of this technique was evaluated by comparing predicted transfer RNA loop and RNA pseudoknot structures with known or consensus structures. The root-mean-square deviation (2.0 to 3.0 angstroms before minimization) between predicted and control structures reveal this system to be an effective method in modeling RNA.

THE REALIZATION, MANY YEARS AGO, THAT POLYPEPTIDES may encode sufficient structural information in their amino acid sequences to self-assemble into functional conformations is largely responsible for the fascination surrounding attempts to predict 3-D structures from protein sequences (*1*). In addition, detailed knowledge of protein conformation is considered a crucial prerequisite to the comprehension and eventual manipulation of protein function (*2*). However, even though very powerful methods have been used, protein structure prediction has proved elusive (*3*).

By comparison, the prediction of single-stranded nucleic acid conformations, in particular those of RNA, might be considered even more challenging, since with the exception of short oligomers only one type of RNA (tRNA) has ever been crystallized and subjected to x-ray analysis (*4, 5*). However, predictions of the secondary structure of RNA, that is, its base-pairing pattern, whether based on free energy calculations or inferred from compensatory mutations, are more reliable than those predicted for proteins from amino acid sequence (*6*). Furthermore, these patterns make it possible to juxtapose nucleotides that are distant in the primary structure. This advantage, coupled with the richness of RNA cellular functions, makes RNA structure prediction and modeling particularly attractive.

In the search for an acceptable modeling procedure, many schemes have been considered and their weaknesses have been analyzed. Those based on energy minimization or distance geometry (*7–9*) are limited because solutions are likely to represent local minima, depending on the input structure, rather than the global minimum. Nevertheless, these methods could be helpful to refine low-resolution models (*10*). Other methods developed for proteins sample the conformational space and then screen potential solutions on the basis of energy or hydrophobicity criteria (*11, 12*). Algorithms of this type require computer time proportional to n^m, where n is the number of variables and m the number of permitted values. Therefore, even though satisfactory solutions are found, modeling of only relatively small molecules at atomic resolution can be realistically attempted (*11, 13*).

The limitations of the above methods and the desire to make use of accumulated structural data led us to consider RNA modeling as a constraint satisfaction problem (CSP), that is, all potential solutions must be consistent with a given ensemble of structural information. The facility of data incorporation and the ability to implement CSP algorithms suggested the use of symbolic programming, where symbols or simplified models rather than mathematical formulas are used to represent complex phenomena (*14*). Solutions from these algorithms are less precise but could serve as good starting structures for refinement by computational methods such as energy minimization (*15*). We therefore have implemented the following scheme for RNA structure modeling and prediction and include: (i) the definition of double-helical and single-stranded regions of the RNA by existing secondary structure algorithms; (ii) the use of secondary and other structural information in order to define a CSP; (iii) the generation of structures that are consistent with these constraints with the use of the symbolic program; and (iv) the refinement of solutions produced by the program with an energy minimization routine. Here we show that this combination of symbolic and numerical techniques can be successfully applied to a variety of RNA structural problems.

The CSP algorithm. A CSP algorithm appropriate for macromolecular modeling is defined by the variables $X = \{x_1, x_2 \ldots x_n\}$, the values of which are taken from the domains of permitted values $D = \{d_1, d_2 \ldots d_n\}$, and a set of constraints, $C = \{c_{p,q,} \ldots | p \in \{1 \ldots n\}, q \in \{1 \ldots p - 1\}\}$. Although in this case the constraints are defined as binary, that is, between any two structural features, the case of constraints applied to more than two features has been implemented. Solving the CSP means finding the value or values of X (from its domain of values D) that satisfy all constraints in C. This

F. Major, M. Turcotte, and G. Lapalme are in the Département d'Informatique et de Recherche Opérationnelle and D. Gautheret, E. Fillion, and R. Cedergren are in the Département de Biochimie, Université de Montréal, Montréal, Québec, Canada H3C 3J7.

*Present address: National Center for Biotechnology Information, National Institutes of Health, Bethesda, MD 20894.
†To whom correspondence should be addressed.

algorithm generates a search tree where each node corresponds to the assignment of a value to a variable. At each assignment, the consistency of that value with the constraints is evaluated. If consistent, the next variable is assigned and the process is continued. If inconsistent, this node and attached branches are pruned from the search tree and the algorithm "backtracks" to the previous node (Fig. 1A) (*16*). The backtracking feature can increase efficiency of the algorithm by reducing the number of assignments even though the complexity of the problem remains the same.

For the purposes of RNA modeling, the variable (X) is the set of nucleotides corresponding to an RNA sequence; the domain of values attributed to the nucleotides (D) is the set of cartesian products of various permitted nucleotide conformations and 3-D transformational matrices. An RNA molecule is specified by its nucleotide sequence, in which each nucleotide is annotated by a conformational indicator, and by any other known constraints. The conformational indicator describes the spatial relation of this nucleotide to others and provides the means to introduce problem-specific information into the algorithm. Each indicator uses a value domain that is generated by a specific computational "help" function. These functions reduce the number of variable assignments (see below) and the execution time of the algorithm (*15*).

The program called MC-SYM (Macromolecular Conformation by SYMbolic generation) has been written in Miranda, a "fully lazy" functional programming language (*17, 18*). Since a Miranda program is both easy to write and modify, it constitutes an ideal prototyping environment. Programs are 10 to 20 times shorter than those written in C or Pascal, although execution times are greater. The informational flow chart for this system and its relation to the numerical units are shown in Fig. 1B.

Representation of nucleotides. Since the conformational flexibility of a single nucleotide is great, its domain of permitted values would be virtually infinite, if some simplifications were not made. Therefore, an RNA structural database was assembled that included all internucleotide and internal nucleotide torsion angles found in the crystal structures of tRNA (*4, 5, 19*). From this base, a set of ten typical conformations was constructed by taking representative samplings of each internal torsion angle: α, β, δ, and χ (Fig. 2 and Table 1). The set includes C2′-*endo*, C3′-*endo* sugar puckers, nitrogen base *anti* and high *anti* conformations (*20*). In addition to these conformations, the value domains contain transformational matrices, which in biochemical terms represent the different ways in which two nucleotides can be joined.

Examination of the database revealed that most internucleotide bonds, ζ, are encompassed in one of three P–O5′ torsions (−90°, +90°, and 180°). The multiplication of one of the ten nucleotide conformations by one of the three transformational matrices yields

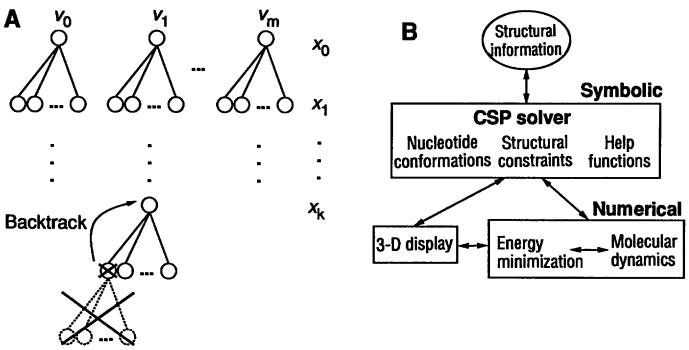
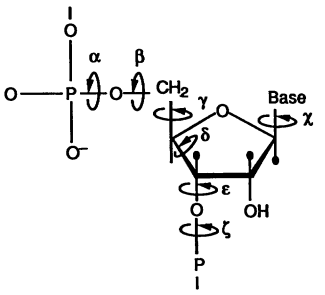


Fig. 1. (**A**) Search tree and backtracking. The values $\{v_0, v_1, \ldots v_m\}$ for variable x_0 are taken from the value domain d_0. The algorithm backtracks when an assignment is not consistent with a constraint. (**B**) Informational flow chart for the modeling system.

Fig. 2. Definition of nucleotide torsion angles.



the atomic coordinates of the nucleotide and is the assignment of the variable x. The use of ten torsion angle sets reduces somewhat the conformational freedom of the modeling procedure and consequently the execution time of the algorithm. However, this limitation does not hinder the use of any atom of the nucleotide in constraints evaluated during CSP solving, since the exact location of every atom is known, nor does it interfere with the generation of structures consistent with constraints, since variability in intranucleotide torsion angles can be simulated by declaring more approximate constraints. The problem of further refining intranucleotide torsion angles is relegated to energy minimization in the last step of our scheme.

Constructing polynucleotides. The generation of the polynucleotide chain is accomplished stepwise in accordance with available structural information. Consider the case of the dinucleotide pApG where no structural information is available. First the procedure assigns one by one the ten different conformations described above for pA. In the next step pG is added and the three transformational matrices in its domain of values are used to calculate the position of pG. The ten possible conformations of pG raise the total number of potential solutions (the conformational search tree) to $10 \times 3 \times 10 = 300$.

Now consider the above example if pApG were in an A-helical region of an RNA molecule. This information, which would reduce dramatically the number of solutions, is incorporated into the program by the conformational indicator of the two nucleotides. Here, the indicator would be Helix5′ (generates helices in the 5′ to 3′ direction). The help function of the same name would generate the domain of values permitted for each nucleotide. Since only one conformation and one transformational matrix for each of the two nucleotides is generated by this help function, only one 3-D structure exists for the dinucleotide. Other help functions that have been implemented are given in Table 2.

Finally, other structural information is incorporated by a constraint. This information can be expressed as distance equalities or inequalities between specific atoms of any nucleotide or by other more complex formulations. The generation of an RNA loop provides an example of the use of constraints. In order to generate a three-nucleotide loop at the end of a double-helical stem region, the problem can be stated as follows: A Wc indicator relates the two stem nucleotides at the base of the loop and the Connect indicator relates the three nucleotides in the loop. The program would normally produce 30^3 structures (from 30 structures of each of the three nucleotides of the loop). However, not all of these possibilities could be a solution, since the loop must be closed. In order to ensure closure, a distance constraint is imposed between the last nucleotide of the loop and the first nucleotide of the stem region; the constraint is that the two nucleotides must be close enough so that a phosphodiester bond can link them.

Although one might think intuitively that constraints of high precision would lead to better solutions, this modeling system, making use of a limited sampling of intra- and internucleotide

Table 1. Intranucleotide angles. The ten sets of angles derived from those most frequently observed in $tRNA^{Phe}$. For the pseudorotation angle, *P*, C3′ corresponds to C3′-*endo* and C2′ corresponds to C2′-*endo*. For the χ angle, −160° corresponds to an *anti* orientation and −90 to a high-*anti* orientation.

Angle	1	2	3	4	5	6	7	8	9	10
P	C3′	C3′	C3′	C3′	C3′	C3′	C2′	C2′	C2′	C2′
α	−60	180	60	−60	180	60	−60	180	−60	180
β	180	180	180	180	180	180	−150	−150	−150	−150
γ	50	50	50	50	50	50	50	50	180	180
χ	−160	−160	−160	−90	−90	−90	−90	−90	−90	−90

torsion angles, requires that the precision of constraints be relaxed somewhat so that good solutions are not eliminated. An ideal data set contains a large number of constraints, even if they are approximate, since each would help to prune the search tree.

Energy minimization. The role of minimization in our scheme is threefold. First, even though atomic collisions are minimized by the judicious use of torsion angles for the intra- and internucleotide bonds, the structure-generation feature in MC-SYM does not interdict spatial overlapping of atoms unless appropriate constraints are specified. Consequently, energy minimization provides a way of removing possible atomic collisions. However, this procedure cannot correct or refine structures resulting from intertwining of helical regions. Such problems must be treated with constraints in the polynucleotide generation step.

Second, the rigid nucleotides in the modeling procedure require that constraints be approximated. In a subsequent example, a 3 Å constraint is used for a P–O5′ bond that is normally 1.5 Å. Other torsion angles in the molecule can be adjusted so that the energy minimization procedure can rapidly rectify the P–O bond length. Our experience with MC-SYM indicates that about ten cycles of minimization are sufficient. Also, energy minimization has the effect of increasing flexibility in the model, since previously fixed torsion angles can be modified during minimization. We do not expect major changes in torsion angles, however, unless thousands of cycles are executed, which, based on the approximate nature of simple potential energy functions for nucleic acids, would likely introduce other problems.

Finally, because the interplay between MC-SYM and energy minimization is mutually beneficial, MC-SYM can play an important role in providing good starting structures for extensive energy minimization or molecular dynamics, since a major problem of these computational techniques is the starting structure dependence of solutions. The multiple solutions of MC-SYM help to legitimize the minimization techniques because computations are performed on a more complete sampling of possible conformations. In order to assist in communication between these modules, MC-SYM generates conventional coordinate files.

RNA loops. Energy minimization, structure homology, and interactive computer graphics have been used to resolve the complexities inherent to the 3-D modeling of looped regions in macromolecules (*21–23*). Because of the importance of loops in RNA structure, we have generalized the above example of loop generation. The search for nucleotide values that satisfy a bond-distance constraint between two selected nucleotides of a loop has been applied to predict the conformation of the anticodon and T-loops found in tRNA.

In the definition of the anticodon loop, the five nucleotides 34 to 38 are assumed to be stacked (Fig. 3A). This structural information had been predicted for the anticodon loop of tRNA before the crystal structure was available (*24*), but would not necessarily be available for an unknown loop. In this example, we have allowed a bond distance of 3.0 Å for the single-loop–closure constraint between the O3′ of U33 and the P of G34.

The generation of the stem region with the Wc and Helix5′ indicators takes only 0.05 s (*25*). The stacked region of the loop was generated by Stacked3′ and gave $2^5 = 32$ possible solutions in 0.43 s. The least-restrained area of the loop consists of the two nucleotides, C32 and U33, for which no conformational information is available other than the nucleotides must be positioned so that the loop can be closed. The Connect5′ indicator here generated 27 different values for each nucleotide such that the search tree for the entire stem loop was $27 \times 27 \times 2^5 = 23{,}328$, and of these, 171 structures satisfied the closure constraint (0.73 percent of the conformational space). The simplicity of the coded data for this loop is illustrated in Fig. 3B.

In order to evaluate the precision of the system, we compared the 171 solutions with the known crystal structure of the anticodon loop, and the congruence was expressed as the root-mean-square (rms) deviation in angstroms. The energy level of each solution was determined after 1000 cycles of energy minimization by the adapted Newton-Raphson method available in CHARMm (*26*). The solution with the lowest energy level also had the lowest rms value. The correlation between the energy level and the rms deviation was determined by plotting the values for each solution; the correlation coefficient ρ was 0.385 (*27*). The superposition of the crystal structure and the predicted structure with the lowest deviation (2.00 Å for 730 atoms) is shown in Fig. 3C. Further analysis revealed that this value is surprisingly low, since the rms deviation for the 426 atoms in the base-paired region alone is 0.92 Å. Therefore 26 percent of the rms deviation of the total structure derives from the region that we assumed was known with precision. Obviously, full congruency was not expected because the duplex was built with idealized torsion angles that do not take into account sequence or other local effects. Prediction of the phosphate position for every nucleotide in the structure is better (1.5 Å) than that for the entire ribophosphate backbone (1.65 Å). Therefore, the imprecision in placing the nitrogen base may represent a significant contribution to the rms deviation.

In a similar manner, we modeled the T-loop and stem of $tRNA^{Phe}$. The structural information used for the model consisted of the reverse Hoogsteen pairing between T54 and A58 and the stacking of bases G53, T54, and ψ55, all of which were proposed prior to the crystal structure (*28*). The bridging of the loop resulting from the T54-A58 pair divides the problem into two smaller loops, each with a ring-closing distance constraint of 3.5 Å. The search tree size for this problem is 26,244,000; MC-SYM found 168 structures (less than 0.001 percent) that satisfied both closure constraints in 9.1 hours of CPU time. The T-loop and stem solutions were compared with the crystal structure, and the solution with the

Table 2. Help functions called by conformational indicators. The normal number of conformations used are indicated; however, these can be modified according to the particular case. The total number of values for a given nucleotide is the number of matrices multiplied by the number of conformations.

Indicator	Matrices (no.)	Conformations (no.)	Description
Wc	1	1	Watson-Crick base pair in an A helix
Rev_hoog	1	1	Reverse Hoogsteen base pair
Helix3′	1	1	A-helix form appended to 3′ end
Helix5′	1	1	A-helix form appended to 5′ end
Connect3′	3	10	Free connection to 3′ end
Connect5′	3	9	Free connection to 5′ end
Stacked3′	2	2	Stacked connection to 3′ end
Stacked5′	2	2	Stacked connection to 5′ end

lowest rms deviation and incidently the lowest energy value after 1000 cycles of minimization was selected for further study. The rms deviation was 1.76 Å for the phosphates, 2.04 Å for the backbone, and 2.35 Å for all atoms (with the exception of the nonpolar hydrogens).

In comparison with the results of the anticodon loop, the predicted helical region and the stacked nucleotides of the T-loop are much closer to the observed structure. However, the predicted conformation of the nucleotides where no structural information was available is slightly worse. Overall the juxtaposition of the backbone of the calculated and crystal structure (Fig. 3D) is quite good, particularly because the several known tertiary interactions of this loop were not used in the prediction.

Prediction of a pseudoknot. A pseudoknot is characterized by base pairings between a loop and another region of RNA (Fig. 4A). This structure, first proposed in the tRNA-like 3′-terminus of turnip yellow mosaic virus RNA, has since been suggested in many RNA's including 16*S* RNA and autocatalytic group I introns (*29*). Dumas *et al.* produced a theoretical model for the pseudoknot structure by computer graphics (*30*). A model derived from nuclear magnetic resonance (NMR) studies was proposed and essentially confirmed the earlier theoretical model (*31*). The information used in the theoretical model was introduced into MC-SYM to reconstruct the model automatically. Reconstruction was performed with three types of conformational indicators, (i) Wc and (ii) Helix3′ for the base-pairing pattern in Fig. 4A and (iii) Stacked5′ for the stacking of bases 4, 5, and 6. The Connect5′ was used for the conformationally unknown nucleotides in loops L1 and L2. As an additional constraint, the two helices were made coaxial.

The modeling of the helical regions took 0.13 s. Modeling of loop L1 implies a search tree length of 108 (2 × 2 × 27), which in 34 s generated only one solution that satisfied a loop-closure constraint of 4.0 Å between the O3′ of C6 and the P of U7. The L2 loop generated a search tree of size 19,683 and required 37 min to produce 62 solutions that satisfied a loop-closure constraint of 4.0 Å between the O3′ of A18 and the P of U19. Therefore, the entire problem implies a search tree size of 2,125,764 (108 × 19,683) and required ~1 hour to be explored by the system. The 62 solutions, representing only 0.003 percent of the conformational space, were evaluated as to their congruence with the theoretical model (*30*). The solution with the lowest energy level also had the smallest rms deviation from the consensus model. This model predicts an L1 region in which the phosphates are more toward the inside of the molecule than in the theoretical model (Fig. 4B). The overall rms deviation is 2.80 Å.

A recent study of the constraints on closing loops L1 and L2 of the pseudoknot by NMR offered an opportunity to test the reliability of MC-SYM (*31*). We used the same constraints as above to

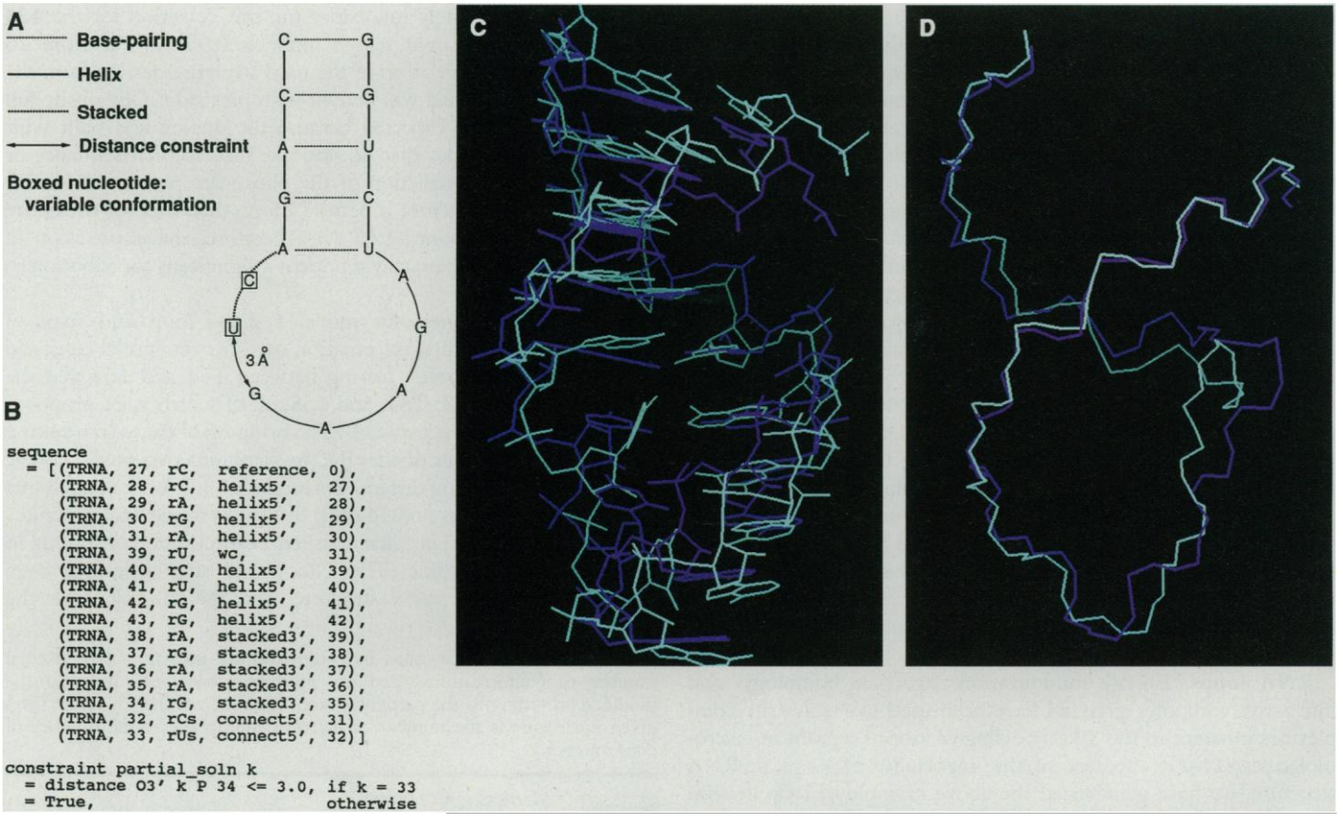


Fig. 3. (**A**) Secondary structure of the anticodon loop used by MC-SYM. (**B**) The Miranda script of the structural information used to model the anticodon loop of tRNA. The function "sequence" describes the conformational information for each nucleotide of the molecule and the function "constraints," the constraints on the 3-D structure. Each data line in the sequence list contains the name of the molecule, the position number, the nucleotide at that position, the name of the help function which applies to the nucleotide, and finally the nucleotide to which the help function is referred; rA, rC, rG, and rU are the four ribonucleotides; the "s" following the name of the nucleotide indicates the use of ten different intranucleotide conformations. "Reference" is used to denote the starting nucleotide in the procedure. The constraint is read as follows: the constraint on the partial solution up to the present nucleotide, *k*. When *k* = 33, the constraint is satisfied, if the distance between the O3′ of nucleotide 33 and the phosphorus of nucleotide 34 is less than or equal to 3 Å. There is no verification in any other case, that is when *k* is greater or less than 33. (**C**) The superposition of the predicted (white) and crystal (violet) structures of the anticodon stem loop. (**D**) The superposition of the backbone atoms of the predicted (white) and the crystal (violet) structures of the T stem loop.

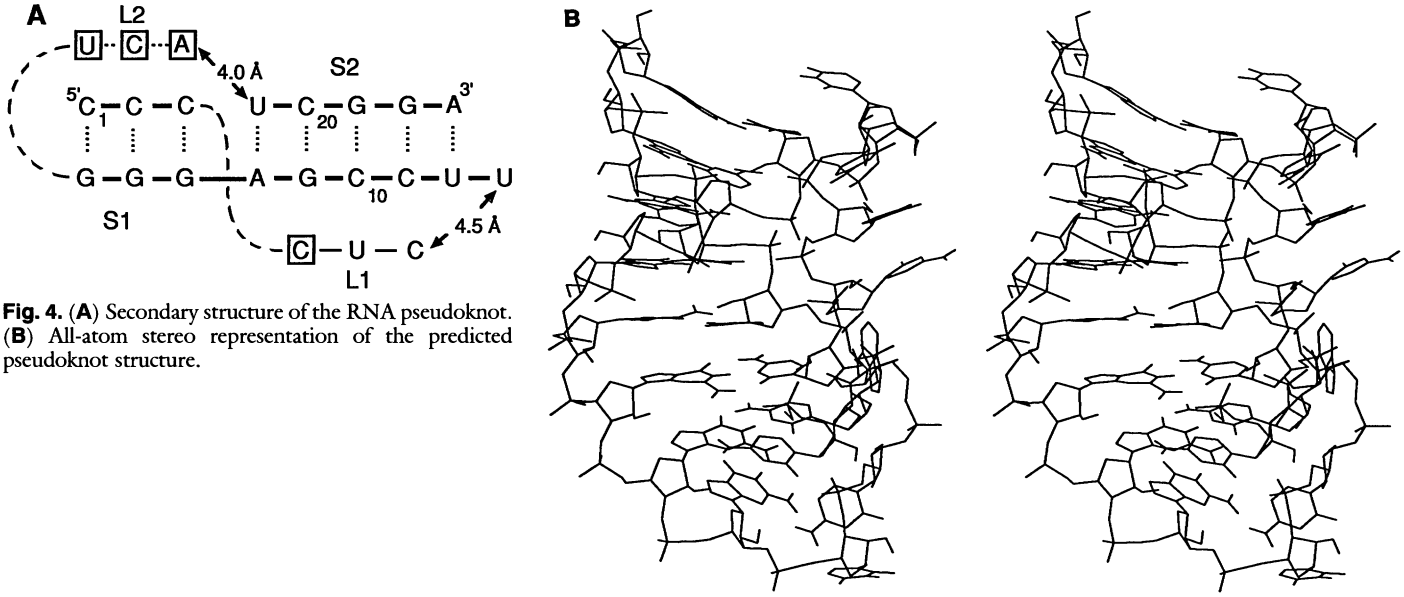


Fig. 4. (**A**) Secondary structure of the RNA pseudoknot. (**B**) All-atom stereo representation of the predicted pseudoknot structure.

determine the number of consistent solutions when the length of the stems and the loops were varied. The minimum number of nucleotides in the loop L1 necessary to bridge the S2 region is three for zero to three base pairs and two for four to eight base pairs. For region S1, three nucleotides are necessary in loop L2 when there are zero to two base pairs and four are necessary when there are three base pairs. In cases where NMR indicates that the pseudoknot structure is unlikely, such as for S2 = 5 and L1 = 1 or S1 = 3 and L2 = 2, MC-SYM found no solution. In addition, other restrictions on the structure are apparent; there is a complex relation between the length of L1 and the stem S2 that it spans. In most cases at least two loop nucleotides are needed; however, helix lengths of three nucleotide pairs or less require three loop nucleotides. Loop L2 must contain three or four nucleotides for helices up to three base pairs. Judging from this data, we estimate that at least four nucleotides are needed to span helices of four or more nucleotide pairs. The high concordance between the NMR and modeled structures gives some confidence that MC-SYM could be used to formulate structural criteria for pseudoknots.

Precision and perspective. We have implemented a CSP-based computational system that can serve as an "intelligent" tool for generating RNA tertiary structures when topological constraints are available. Results show that all-atom representations of macromolecules are generated relatively easily when conformational space is restricted. Moreover, the system is flexible enough to evaluate hypotheses concerning RNA structure (pseudoknots) and to use a wide variety of RNA structural knowledge. We have presented a set of conformations and help functions that have been extensively tested in RNA modeling and prediction. These parameters can easily be changed and adapted as new information on RNA structure becomes available.

The question as to the precision of the models is of primary importance. For this reason solutions generated by MC-SYM summarized in Table 3 were compared to the known crystal or theoretical structure. Although it is very difficult to compare the precision of our models with crystal structures, the rms deviation between the predicted and the control structures suggests that the resolution of the modeling system is somewhat less than that of a good x-ray model. On the other hand, both x-ray and MC-SYM structures represent a conformational space, and the question is to what extent the two spaces overlap (*32*).

Particularly remarkable is the observation that as the number of constraints and the size of the potential search tree increases, the solution ratio decreases dramatically. Also, the CPU time is not related to the search tree space, which is to be expected if the constraints play a major role in pruning the tree. This result is very encouraging because it raises hopes that much larger molecules can be treated by MC-SYM provided that sufficient constraints are known. The apparent correlation between the structures with the smallest rms deviation and their energy level suggests that the selection among solutions of unknown structures could be done by choosing the solutions with the lowest energies.

The combination of symbolic and numerical techniques such as CSP solving and energy minimization constitutes a novel computational tool (*14*). These two methods are mutually beneficial: the precision lost due to simplification of the model in the symbolic programming environment is recovered by the numerical module at a later step in the process, and the numerical module is used only for the minimization of reasonable structures provided by the CSP module. In the same way, MC-SYM structures could be very useful for molecular dynamic calculations. NMR studies are likely to be the base of many RNA structure determinations, and although this kind

Table 3. Comparison of the three models.

Structure	Search tree size	Solutions (no.)	Solution ratio	CPU time* (s)	Energy† (kcal/mol)	P-P‡ (Å)	Bk§ (Å)	H‖ (Å)
Anticodon	2×10^4	171	7×10^{-3}	2,679	−460.7	1.50	1.65	2.00
T-loop	2×10^6	168	1×10^{-4}	33,100	−423.6	2.04	1.76	2.35
Pseudoknot	2×10^6	62	3×10^{-5}	3,617	−618.1	3.29	2.95	2.80

*On a Sun SPARCstation 1+ with the use of a Miranda interpreter, version 2.014. †Determined by CHARMm (*26*). ‡P-P refers to the rms deviation of phosphorus atoms only. §Bk to the atoms of the backbone (P, O5′, C5′, C4′, C3′, and O3′). ‖H includes all atoms except nonpolar hydrogens (polar hydrogens are those involved in base pairing and the 2′-OH).

of information can be treated by minimization routines (*23, 31, 33, 34*), these data could be used more elegantly in the system we describe in this article.

REFERENCES AND NOTES

1. C. B. Anfinsen, *Science* **181**, 223 (1973).
2. M. Mutter, *Trends. Biol. Sci.* **13**, 260 (1988).
3. G. D. Fasman, *ibid.* **14**, 295 (1989).
4. S. H. Kim *et al.*, *Science* **185**, 435 (1974).
5. J. Robertus *et al.*, *Nature* **250**, 546 (1974).
6. D. Turner, N. Sugimoto, S. Freier, *Annu. Rev. Biophys. Chem.* **17**, 167 (1988).
7. H.-Y. Mei, T. W. Kaaret, T. C. Bruice, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9727 (1989).
8. M. Summers *et al.*, *Biochemistry* **29**, 329 (1990).
9. L. Nilsson *et al.*, *ibid.*, p. 10317.
10. M. Karplus and G. A. Petsko, *Nature* **347**, 631 (1990).
11. R. E. Bruccoleri and M. Karplus, *Biopolymers* **26**, 137 (1987).
12. J. Skolnick and A. Kolinski, *Science* **250**, 1121 (1990).
13. A. Malhotra, R. K. Tan, S. C. Harvey, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 1950 (1990).
14. H. Abelson *et al.*, *Commun. A.C.M.* **32**, 546 (1989).
15. F. Major, G. Lapalme, R. Cedergren, *J. Funct. Program.* **1**, 213 (1991).
16. R. Haralick, *Artif. Intell.* **14**, 263 (1980).
17. D. A. Turner, in *Proceedings on Functional Programming and Computer Architecture no. 201*, P. Jouannaud, Ed. (Springer-Verlag, New York, 1985).
18. More information about MC-SYM can be obtained by sending an Email message to mcsym-request@ncbi.nlm.nihgov.
19. For the crystal structure of $tRNA^{Asp}$, see D. Moras *et al.*, *Nature* **228**, 669 (1980). Data from the Brookhaven Protein Data Bank [F. C. Bernstein *et al.*, *J. Mol. Biol.* **112**, 535 (1977)].
20. W. Saenger, *Principles of Nucleic Acid Structure* (Springer-Verlag, New York, 1984).
21. H. Scheraga, *Adv. Phys. Org. Chem.* **6**, 103 (1968).
22. I. Haneef, S. J. Talbot, P. G. Stockley, *J. Mol. Graphics* **7**, 186 (1989).
23. J. Puglisi, J. R. Wyatt, I. Tinoco, *Biochemistry* **29**, 4215 (1990).
24. W. Fuller and A. Hodgson, *Nature* **215**, 817 (1967).
25. Evaluated on a Sun SPARCstation 1+.
26. B. Brooks *et al.*, *J. Comput. Chem.* **4**, 187 (1983).
27. This value was obtained with $N = 148$, since 23 structures with major steric conflicts have been eliminated from the calculation; ρ has values between -1 and $+1$.
28. M. Levitt, *Nature* **224**, 759 (1969).
29. C. Pleij, *Trends Biol. Sci.* **15**, 143 (1990).
30. P. Dumas *et al.*, *J. Biomol. Struct. Dyn.* **4**, 707 (1987).
31. J. R. Wyatt, J. Puglisi, I. Tinoco, *J. Mol. Biol.* **214**, 455 (1990).
32. Although the rms deviation is an unbiased criterion of fit, it does not measure the relative position among atoms of the predicted structure.
33. J. Puglisi, J. R. Wyatt, I. Tinoco, *J. Mol. Biol.* **214**, 437 (1990).
34. H. A. Heus, O. C. Uhlenbeck, A. Pardi, *Nucleic Acids Res.* **18**, 1103 (1990).

35. We thank L. Jolicoeur for assistance in the early stages of this project. Supported by the Medical Research Council of Canada and the Fonds pour la formation de Chercheurs et l'Aide à la Recherche of Québec (FCAR); by Natural Sciences and Engineering Research Council of Canada predoctoral fellowships (F.M. and M.T.); by a FCAR postdoctoral fellowship (F.M.) and a fellowship of the Canadian Institute for Advanced Research (R.C.).

11 February 1991; accepted 29 July 1991

"But we spend all our money creating *toxic waste. We were hoping someone* else *would figure out how to* detoxify *it."*