Eur Biophys J (1999) 28: 564–573

ARTICLE

Alexandra Maier · Heinz Sklenar · Herbert F. Kratky
Alexander Renner · Peter Schuster

Force field based conformational analysis of RNA structural motifs: GNRA tetraloops and their pyrimidine relatives

Received: 20 January 1999 / Revised version: 4 June 1999 / Accepted: 10 June 1999

Abstract The protocol of conformational analysis applied here to ribonucleotide oligomers combines conformational search in the space of torsion angles and energy minimization using the AMBER4.1 force field with a continuum treatment of electrostatic solute-solvent interactions. RNA fragments with 5′-GGGCG-NNAGCCU-3′ sequences commonly fold into hairpins with four-membered loops. The combinatorial search for acceptable conformations using the MC-SYM program was restricted to loop nucleotides and yielded roughly 1500 structures being compatible with a double-stranded stem. After energy minimization by the JUMNA program (without applying any experimental constraints), these structures converged into an ensemble of 74 different conformers including 26 structures which contained the sheared G-A base pair observed in experimental studies of GNRA tetraloops. Energetic analysis shows that inclusion of solvent electrostatic effects is critically important for the selection of conformers that agree with experimentally determined structures. The continuum model accounts for solvent polarization by means of the electrostatic reaction field. In the case of GNRA loop sequences, the contributions of the reaction field shift relative stabilities towards conformations showing most of the structural features derived from NMR studies. The agreement of computed conformations with the experimental structures of GAAA, GCAA, and GAGA tetraloops suggests that the continuum treatment of the solvent represents a definitive improvement over methods using simple damping models in electrostatic energy calculations. Application of the procedure described here to the evaluation of the relative stabilities of conformers resulting from searching the conformational space of RNA structural motifs provides some progress in (non-homology based) RNA 3D-structure prediction.

Key words RNA structural motifs · GNRA tetraloops · Conformational search · Solvent effects · Electrostatic continuum model

Abbreviations *MC-SYM* Macromolecular conformations by symbolic programming · *JUMNA* Junction minimization of nucleic acids · *AMBER* Assisted model building with energy refinement · *RF* Reaction field · *FIESTA* Field integrated electrostatic approach · *FD* Finite difference · *sc* Synclinal · *ac* Anticlinal · *ap* Antiperiplanar · *ED* Electrostatic damping

Introduction

Understanding RNA structure is important owing to the variety of biological functions fulfilled by different classes of these molecules. The currently available information on RNA 3D structures is, however, quite limited (Uhlenbeck et al. 1997). Crystallization and X-ray crystallography have been successful only for tRNAs, small oligonucleotides (e.g. Perbandt et al. 1998) as well as recently for the hammerhead ribozyme (Pley et al. 1994; Scott et al. 1995), the catalytic domain of a group I intron (Cate et al. 1996), and a hepatitis delta virus ribozyme (Ferré-D'Amaré et al. 1998). In the last few years, substantial progress was made by combining NMR spectroscopic methods and distance constraint calculations or restrained molecular dynamics in the determination of the structure of small RNA fragments. The structures obtained for a variety of hairpins, bulges, internal loops, and pseudoknots (Moore 1993; Shen et al. 1995; Feigon et al. 1996) revealed many ways in which nucleotides can be involved in hydrogen bonding interactions (for an updated review, see also Schuster et al.

A. Maier · H. Sklenar (✉)
Max-Delbrück-Centrum für Molekulare Medizin,
Robert-Rössle-Strasse 10, D-13122 Berlin, Germany
e-mail: sklenar@mdc-berlin.de

H.F. Kratky · A. Renner · P. Schuster
Institut für Theoretische Chemie
und Molekulare Strukturbiologie der Universität Wien,
Währingerstrasse 17, A-1090 Vienna, Austria

1997). The remarkable intrinsic stability of certain classes of structural motifs and their reoccurrance in many RNA structures indicate that they play an important role in tertiary folding and in the biological functions of RNA molecules (Michel and Westhof 1990; Murphy and Cech 1994). The emerging "tool-kit" of RNA structural motifs will help to understand better the relationships between sequences, structures, and functions of RNA molecules. It can be also expected to substantially aid model building of RNA 3D structures. For this purpose, however, the database of RNA structural motifs needs to be expanded and should also include data on the relative stabilities of alternative structures as functions of their nucleotide sequences. Since investigations by crystallography or NMR are often costly and/or very time consuming, a computational approach to this task would be helpful in the sense that it extends reliably the experimental data set. In addition, information on suboptimal conformations is highly desirable for the discussion of molecular interactions. It can be easily obtained only by constrained computational optimization. In the present study we have developed and explored a force field based method for the conformational analysis of oligoribonucleotide structures.

The use of conformational search methods for predicting molecular structures of biological interest is presently hampered by several problems and, hence, the results obtained by such techniques are met with some scepticism. The difficulties in treating large molecular systems are twofold. Firstly, the numbers of states which must be considered in complete conformational analysis are enormous. This problem is closely related to global minimization of a cost function on a high-dimensional multi-minima landscape. The progress achieved in this field is documented, in particular, by successful applications to searching the conformational space of oligopeptides (Scheraga 1993). The problem to find optimal solutions in large search spaces can be considerably simplified by using the empirical knowledge on conformational preferences in similar systems. Nevertheless, one commonly ends up with large sets of different structures, among which only a few are relevant for the biologically relevant conformation or the thermodynamic ground state. Secondly, the apparently even more serious difficulty is the evaluation of the relative stability of the conformers found by the search procedure. Thus, the choice of the force field is of critical importance for application of conformational searches as predictive tools. Predicted stabilities of different nucleic acid conformers in solution depend, in particular, on the description of solvent-induced electrostatic interactions. In their modeling study on NUUN tetraloops, Kajava and Rüterjans (1993) have stated that "it is impossible to choose properly the most probable 3D structure on the basis of the value of calculated energy". The authors, however, have adopted in their calculations the common practice of describing dielectric damping by a simple linear distance dependence of the dielectric constant, which is inappropriate for a realistic treatment of electrostatic solute-solvent interactions.

In the present study, conformational search and energy minimization are combined with calculations of solvation energies which are based on a continuum treatment of the solvent. The solvent is characterized by its dielectric permittivity constant. The energy of an unsolvated conformer is defined by the sum of deformation energies of bond and torsion angles and the pairwise additive Lennard-Jones and Coulomb electrostatic contributions. The transfer of the solute from vacuum to water results in a polarization of the dielectric medium. Interaction with the charges of the solute can be described by the so-called reaction field (RF), which contributes to the electrostatic energy of the system (Elcock et al. 1997). Salt-dependent solvent contributions, entropic effects related to reordering of water molecules at the solute-solvent interface, and van der Waals interactions between solute and water are neglected in this approach. In a recent study of DNA bulge conformers (Zacharias and Sklenar 1997) these contributions were found to have only small effects on the relative stability of different conformations of the DNA bulge molecule. One may argue, however, that such effects are generally not negligible. Studies on the stacking interactions and the stability of short single-stranded RNA fragments in explicit solvent simulations have indeed shown that the strict enthalpic ranking of the conformers is not sufficient to account for the observed stacking probabilities (Norberg and Nilsson 1998). In the case of the especially stable tetraloops the magnitude of entropic contributions can, nevertheless, be expected to lie within the uncertainty range of the force field parameters, for example of the charges used in this study. Thus, as a first step at least, it seems to be justified to focus on the major contributions to the electrostatic energy which are due to interactions of solute charges with the aqueous solvent.

Our protocol for conformational analysis starts with a search of the conformational space of the four nucleotides forming the loop by using the program MC-SYM (macromolecular conformations by symbolic programming) (Major et al. 1991; Gautheret et al. 1993). All conformers with distances between the 3′- and 5′-ends that roughly fit into the backbone of the closing base pair of the double-stranded stem were selected. Subsequently they are passed over to the JUMNA (junction minimization of nucleic acids) program (Lavery et al. 1995) for chain closure and energy minimization with the AMBER4.1 (assisted model building with energy refinement) force field (Pearlman et al. 1991; Weiner et al. 1984, 1986). The results are finally corrected by adding the RF contributions of the solvent to the intramolecular energy in vacuum. The RF energy was calculated with the field integrated electrostatic approach (FIESTA) (Sklenar et al. 1990), which uses an analytical virtual source technique for solving the Poisson equation and is less demanding than numerical finite difference (FD) methods (Gilson et al. 1988; Davis et al. 1991). A similar approach (using anti-dipoles for solving the boundary problem) has been proposed by Davis (1994).

Oligoribonucleotides of chain length 12 and 5′-GGGCGNRAGCCU-3′ sequences (Fig. 1; N stands for any base and R for purine, either G or A) were chosen as test cases for the method. RNA fragments with such sequences are known to fold into hairpins with four-membered loops. Loops containing the tetranucleotide sequences GNRA, UNCG, and CUUG belong to the most frequently occurring hairpins in native RNAs and are referred to as especially stable tetraloops (Wolters 1992). Conservation during evolution and thermodynamic stability of these tetraloops indicate that they may play an important role in RNA folding and as recognition sites in intra- and intermolecular interactions. The GAAA, GCAA, and GAGA loops are the most abundant members of the GNRA family and they have been extensively studied by a variety of biochemical and biophysical methods (Groebe and Uhlenbeck 1989; Woese et al. 1990; Antao et al. 1991). The solution structures of these tetraloops have been determined using NMR spectroscopy (Heus and Pardi 1991; Orita et al. 1993; Jucker et al. 1996). They are represented by an ensemble of refined conformers which fulfill the constraints derived from the NMR experiments and share, as the most prominent structural feature, a sheared G-A base pair closing the loop.

In the present study, these structures served as a reference for comparison with the predictions of conformational search and energy evaluation. The results obtained for GNRA tetraloops show that the conformations with lowest energies, which were selected from 74 alternative structures by using an energy criterion in the final step of the search procedure, are in good agreement with the NMR data. In an energy range of 2 kcal/mol above the ground state, only one conformer was found which does not form the sheared G-A base pair closing the loop. There is also a strong resemblance between the backbone torsion angles of the selected conformers and the conformational fluctuations observed in a 200 ps molecular dynamics simulations of the 5′-GGGCGCAAGCCU-3′ hairpin which started from one of the NMR-derived structures for this RNA molecules (Zichi 1995).

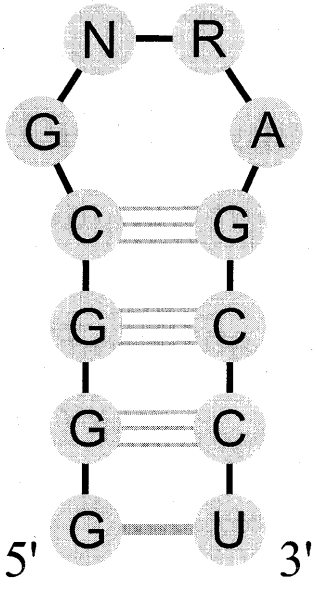

Fig. 1 Secondary structure of the GNRA tetraloops

Methods

Conformational search

In our protocol of conformational search we used the program MC-SYM to generate starting structures for the three tetraloop sequences GAAA, GCAA, and GCUA. The combinatorial search of the conformational space was restricted to the four nucleotides in the loop, while the stem region was assumed to adopt a structure close to the standard A-form double helix. This limitation was necessary because of the amount of computation time needed for the combinatorial search procedure. All the tetranucleotide conformers, which roughly fit to the closing base pair of the stem, were joined to a double-stranded helix structure and subsequently passed to the JUMNA program for chain closure and energy minimization.

Energy minimization

The JUMNA algorithm was specifically designed to build, manipulate, and optimize nucleic acid structures. Special features of JUMNA arise from breaking down each nucleic acid strand into a series of 3′-monophosphate nucleotides. This enables use of helical coordinates (three translations and three rotations) for positioning the bases of individual nucleotides in space, in conjunction with internal coordinates (dihedral and valence angles) for describing the internal flexibility of each nucleotide (Sklenar et al. 1986). Bond lengths are assumed to be fixed at their optimum value, and the junction between successive nucleotides are closed using harmonic distance constraints during energy minimization. In this way, not only a considerable reduction in the number of variables, compared to the use of Cartesian coordinates, but also versatile control over local and global characteristics of the structure are achieved. In the present study, such features of the JUMNA concept turned out to be very useful in order to relax the starting structures provided by MC-SYM to structures with a well-defined local minimum of the conformational energy. Both the standard force field of JUMNA termed FLEX (Lavery et al. 1986) and the AMBER4.1 force field were used to minimize the energy of the full set of starting structures for the three sequences studied. From the structures obtained, a subset of 74 conformers having the lowest conformational energies was selected and used for the final round of conformational analysis.

This subset contains 26 conformers with a G-A base pair closing the loop. The conformational energies calculated at this stage of the procedure did not take into account the electrostatic RF contributions. They do not show any clear preference for conformers that are close to the structures derived from the NMR experiments. We finally carried out one further step of conformational searching. Since representation of nucleic acid structures in JUMNA is independent of the associated sequences, it was simple to use the whole subset of the low-energy conformers as starting structures for all the 16 GNNA sequences. In this step, the AMBER4.1 force field was used for energy minimization and the RF contributions were calculated and added to the energies for all the final structures obtained for each of the sequences.

Evaluation of energies

The force field model of molecular mechanics describes the internal energy of a molecule by pairwise additive Coulomb-type electrostatic and 6–12 dependent Lennard-Jones potential terms complemented by harmonic bond length stretching and valence angle deformation terms, and cosine dihedral barriers. Within JUMNA, fixed bond lengths are assumed so that the bond length stretching term is not relevant. Both the FLEX and the AMBER force fields calculate internal energies under pseudo-vacuum conditions. In order to account for the dielectric damping induced by the solvent, a distance-dependent dielectric function $\varepsilon(R)$ can be introduced in the electrostatic term. In contrast to the often used linear dependence, we have preferred a sigmoidal function based on the Hingerty model (Hingerty et al. 1985). The formulation of this function allows variation of both the plateau value of the dielectric reached at long distances (D) and the slope of the sigmoidal segment of the function (S):

$$\varepsilon(R) = D - (D-1)/2[(RS)^2 + 2RS + 2]\exp(-RS)$$

The values $D = 78$ and $S = 0.356$ were found to be most appropriate for modeling RNA molecules in solution. This damping is combined with a reduction of all phosphate net charges to $-0.25e$ in order to mimic the effect of counter-ions. The use of electrostatic damping (ED) is, however, only a crude model. In the case of highly charged nucleic acids it is therefore not surprising that such a simple approach cannot be used for estimating the relative stability of different conformers. For this purpose, we have replaced the electrostatic term of the AMBER force field with a continuum treatment of solvent-induced electrostatic interactions. In this approach the most important term is the RF potential, which describes the interactions of the solute charges with the polarized aqueous medium and adds to the Coulomb potential. Compared to this term, salt-dependent solvent contributions calculated by a complete Poisson–Boltzmann treatment turned out to have only minor effects on the relative stability of the conformers. The same is true for the entropic contributions related to reordering of solvent molecules at the solvent-solute interface and van der Waals interactions between solute and water. This has been shown in a study on the stability of bulge conformers in DNA (Zacharias and Sklenar 1997). For the sake of simplicity, the current study therefore has been confined to the calculation of the RF contribution based on the Poisson equation

$$\nabla\varepsilon(\mathbf{r})\nabla\Phi(\mathbf{r}) = -4\pi\sum_{i=1}^{N} q_i\delta(\mathbf{r}-\mathbf{r}_i)$$

in which q_i are the atomic charges of the solute and $\Phi(\mathbf{r})$ is the potential. The function $\varepsilon(\mathbf{r})$ describes the change of the dielectric permittivity from the solute ($\varepsilon_{\text{solute}}$) to the solvent ($\varepsilon$) at the solute/solvent interface. With $\varepsilon_{\text{solute}} = 1$, this leads to a RF potential defined as follows:

$$\begin{aligned}\Phi_R(\mathbf{r}) &= \Phi(\mathbf{r}) - \sum_{i=1}^{N} q_i/|\mathbf{r}-\mathbf{r}_i| \\ &= -\int_{\text{solvent}} \mathrm{d}\tau'\mathbf{P}(\mathbf{r}')\nabla'(1/|\mathbf{r}-\mathbf{r}'|)\end{aligned}$$

where $\mathbf{P}(\mathbf{r})$ describes the polarization of the solvent by the solute charges. This potential can be redefined in terms of virtual sources, which are located exclusively within the volume defined by the solute surface envelope:

$$\begin{aligned}\Phi_R(\mathbf{r}) = \frac{1}{\varepsilon}\Bigg[&\int_{\text{solute}} \mathrm{d}\tau'\mathbf{P}(\mathbf{r}')\nabla'(1/|\mathbf{r}-\mathbf{r}'|) \\ &-(\varepsilon-1)\sum_{i=1}^{N} q_i/|\mathbf{r}-\mathbf{r}_i|\Bigg]\end{aligned}$$

where the virtual polarization is defined by $\mathbf{P}(\mathbf{r}) = (\varepsilon-1)\nabla\Phi(\mathbf{r})/4\pi$. This representation is the starting point for FIESTA which introduces a series expansion in terms of spherical harmonics and other limited approximations in order to calculate the polarization $\mathbf{P}(\mathbf{r})$ analytically. The accuracy of this technique is comparable to the numerical FD method, but it is computationally less demanding and can be applied to a large number of structures. High computational speed of the procedure results from two technical improvements: (1) the virtual polarization of chemically bound atoms (with fixed geometry) is taken from a pre-calculated library and (2) the solvent-induced interactions between more distant atoms are described by point dipoles which are derived from a fast converging iteration. The computational time of FIESTA increases with the square of the number of atoms. For a 12-nucleotide long RNA structure the CPU time on a MIPS R10000 processor is 1.4 s, which is several hundred times faster than a conventional FD calculation.

The final expression that has been used to calculate the molecular energy in solution is:

$$E = \frac{1}{2}\sum q_i \Phi_R(\mathbf{r}_i) + \sum \left(q_i q_j / R_{ij} - A_{ij}/R_{ij}^6 + B_{ij}/R_{ij}^{12} \right) + \frac{1}{2}\sum V_s(1 \pm \cos(N_s \tau_s)) + \sum K_a \left(\sigma_a - \sigma_a^0\right)^2$$

The partial charges q_i, the Lennard-Jones parameters A_{ij} and B_{ij}, and the parameters V_s, N_s, and K_a defining the distortion energy associated with torsion angle τ_s and valence angle σ_a, respectively, were taken from the AMBER4.1 parametrization; R_{ij} is the distance between atoms i and j.

Ideally, the solvent treatment by FIESTA should be integrated into the energy minimization procedure. This, however, requires an analytical calculation of the corresponding force components, a problem that has not been completely solved yet. Accordingly, we had to confine this study to "single point" corrections.

Results

Conformational search

The program MC-SYM has been used for searching the conformational space of three tetraloop sequences, GAAA, GCAA, and GCUA, consisting of 30 conformations per nucleotide with distinct sugar puckers and different backbone and glycosidic torsion angles. In each case, $30^4 = 810\,000$ conformations were tested by the MC-SYM program. A simple distance criterion for the 5′- and 3′-ends of the loop yielded 1435 GAAA, 1459 GCAA, and 1457 GCUA loop structures which roughly fit to the backbone of the closing base pair in a double-stranded stem. The fact that the MC-SYM search led to almost the same number of acceptable conformers for each of the three sequences implies that in this case the results are largely sequence independent. It justifies the usage of these structures also as starting points for the other 13 tetraloop structures of the GNRA and GNYA families by simple replacement of the corresponding bases. The structures obtained were joined to the stem and subsequently passed to the JUMNA program (Lavery et al. 1995) for chain closure and energy minimization, using both the FLEX (Lavery et al. 1986) and the AMBER4.1 force fields in conjunction with a sigmoidal electrostatic damping model. The results, given by altogether 74 different loop conformations with acceptable stereochemistry, were used as starting structures for the final round of conformational and energetic analysis.

GNRA tetraloops

Table 1 shows relative energies for the 10 conformers of the lowest energies, which were found by energy minimization and addition of the solvent-induced RF energy for each of the GNRA molecules. The conformations were selected by the criterion that their energy, relative to the conformer with lowest energy, is less than 2 kcal/mol for at least one of the GNRA sequences. In order to show the effect of the RF, the energy values calculated using simple electrostatic damping are given in parentheses. Nine out of the 10 structures have stacking patterns corresponding to the NMR data described by Jucker and Pardi (1995), but show differences in backbone angles. The first and the last residues in the loops of these structures (in the following termed "canonical tetraloops") form a sheared G5-A8 base pair stacked on top of the four Watson-Crick base pairs in the stem. All the conformers that share this and many other structural features with the NMR derived structures are marked by an asterisk in Table 1. The other structure (number 576) is distinct at the position A8. In this conformer, A8 lies in the major groove of the stem and forms a hydrogen bond with residue G1. Comparing the calculated energies shows, as the most important result of our study, that the electrostatic RF contributions stabilize the canonical tetraloop structures to a significantly larger amount than the non-canonical conformers, which would be ranked at the top if only AMBER energies (calculated with electrostatic damping) were used.

GNYA tetraloops

The results obtained for GNRA tetraloop sequences encouraged us to extend our study to the GNYA

Table 1 Lowest energy conformers of the GNRA tetraloops[a]

ID	GAAA	GGAA	GCAA	GUAA	GAGA	GGGA	GCGA	GUGA
404*	**0.0** (7.5)	0.6 (16.1)	0.6 (6.7)	0.4 (7.8)	1.9 (5.1)	1.7 (25.5)	1.6 (12.1)	1.1 (17.6)
418*	0.2 (6.3)	1.1 (14.6)	1.0 (6.2)	1.2 (7.2)	2.2 (4.2)	2.2 (24.7)	2.5 (11.5)	2.3 (17.6)
419*	0.3 (7.0)	**0.0** (14.9)	2.2 (7.6)	1.8 (8.1)	2.1 (5.2)	1.3 (25.0)	2.9 (13.3)	2.9 (18.5)
445*	0.9 (8.9)	2.0 (17.6)	1.7 (8.8)	2.0 (9.9)	2.9 (6.8)	3.0 (27.8)	3.1 (14.3)	2.9 (20.4)
508*	0.7 (5.2)	0.9 (13.0)	1.5 (5.1)	1.8 (5.6)	**0.0** (3.5)	**0.0** (23.8)	**0.0** (10.7)	**0.0** (16.8)
576	1.8 (1.5)	1.9 (11.7)	1.0 (0.9)	0.6 (1.6)	5.2 (2.7)	4.3 (22.4)	4.1 (8.6)	0.9 (12.6)
799*	2.9 (5.1)	2.0 (13.0)	1.9 (4.8)	1.8 (5.4)	1.6 (5.9)	1.1 (25.7)	2.5 (12.7)	1.6 (18.2)
838*	1.4 (9.0)	1.5 (17.4)	1.6 (8.5)	1.5 (9.2)	3.4 (6.9)	2.7 (26.9)	2.8 (13.9)	2.2 (19.0)
851*	0.7 (5.3)	0.3 (13.4)	0.6 (4.9)	**0.0** (5.3)	4.2 (4.1)	3.0 (24.1)	3.5 (10.3)	2.9 (15.3)
857*	0.5 (5.2)	0.2 (13.4)	**0.0** (4.6)	0.3 (5.0)	3.1 (3.3)	1.9 (23.1)	2.3 (9.7)	1.7 (14.6)

[a] The energies (in kcal/mol) are relative to the respective conformers with lowest energies. In addition to the energies including the RF contributions, the minimized energies calculated by means of the simple electrostatic damping model are given in parentheses. Since the structures obtained after energy minimization stay close to the respective starting structures, the same identifier (ID) can be used for all sequences. The conformers which are close to the NMR derived structures are marked by an asterisk

tetraloops (having a pyrimidine at position 7). The data given in Table 2 show that, in the ranking of AMBER energies, GNYA tetraloops again prefer one of the non-canonical conformers, similar to that found for GNRA tetraloops. In pyrimidine containing loops, the GNCA and GNUA sequences, respectively, behave differently with respect to the inclusion of the RF potential. Most of the canonical tetraloop conformers with GNUA sequences become more stable than non-canonical alternative structures. Surprisingly, such a clear effect is not seen in GNCA loops: the non-canonical conformer 576 is found to have the lowest energy. Since GNYA loop sequences occur very rarely in nature or in RNAs with stable structures and thus no experimental data are available on structural level, it is a matter of speculation whether or not our computational results reflect the behavior of real systems.

Structural analysis

The most prominent structural features of GNRA tetraloops derived from NMR studies, and also seen in the majority of the most stable calculated structures, are: (1) the sharp bend, a so-called U-turn (Jucker and Pardi 1995), between the first and second nucleotide in the loop which is associated with a change of the torsion angle α at G5 from the common *–sc* range to $+ac/+ap$, (2) the formation of an asymmetric G5-A8 base pair with a strong hydrogen bond between one of the N2 amino protons of G5 and N7 of A8, (3) an additional hydrogen bond between the other N2 amino proton of G5 and a R7pA8 phosphate oxygen, and (4) roughly coaxial stacking of the N, R, and A bases with the double-helical part which continues the stack at the 3′-end of the stem. These structural elements are shared by all the conformers marked by an asterisk in Table 1. Other hydrogen bonds located in the NMR derived structures (Jucker et al. 1996) are found in some of the calculated conformers. This concerns the hydrogen bond between the 2′-OH of G5 and N7 of the purines R7, which is formed in the conformers 508 and 851 for GNRA and in conformer 857 for GNGA sequences, and likewise the hydrogen bond between the 2′-OH oxygen and an amino proton of A7 formed in the conformers

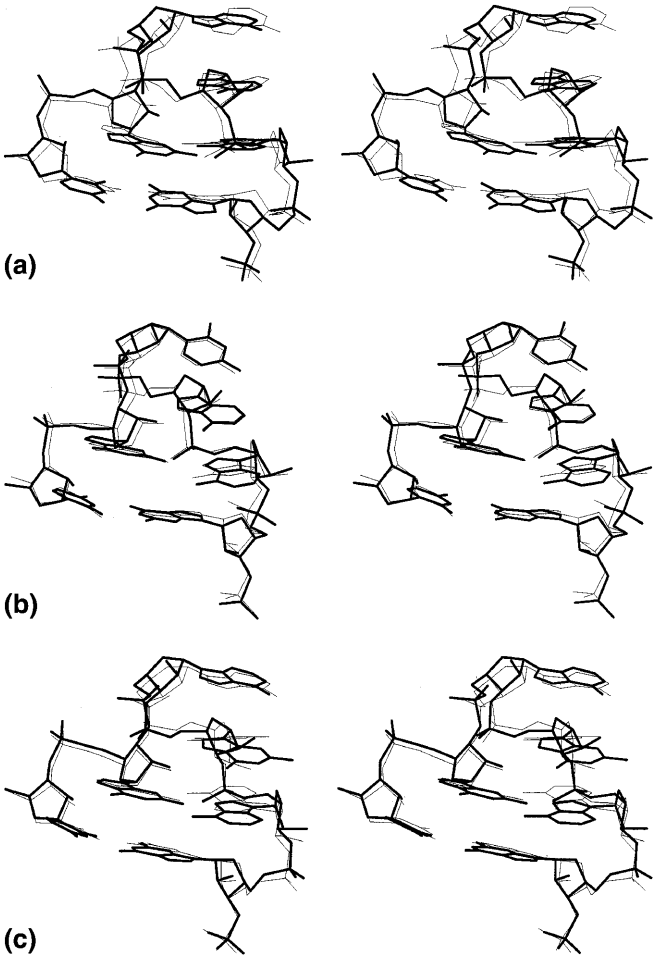

Fig. 2a–c Stereo view of the lowest energy conformers (*thick lines*) superimposed on the NMR structures having the lowest RMSD to the respective conformers. For clarity, only the four loop nucleotides and the first stem base pair of conformer 404 for GAAA (**a**), conformer 857 for GCAA (**b**), and conformer 508 for GAGA (**c**) are shown

Table 2 Lowest energy conformers of the GNYA tetraloops[a]

ID	GACA	GGCA	GCCA	GUCA	GAUA	GGUA	GCUA	GUUA
404*	2.4 (6.6)	3.1 (19.3)	3.9 (5.0)	3.8 (9.0)	3.3 (9.6)	3.7 (11.3)	2.0 (8.2)	1.4 (8.6)
418*	1.6 (4.3)	2.3 (16.3)	3.6 (3.4)	3.7 (7.1)	2.6 (7.1)	3.1 (8.6)	1.9 (6.4)	1.7 (6.8)
419*	1.1 (5.0)	0.9 (16.3)	3.5 (4.6)	4.0 (6.5)	1.7 (7.6)	1.8 (8.7)	1.5 (7.2)	2.3 (6.8)
508*	1.5 (4.3)	1.6 (15.3)	3.1 (3.9)	3.2 (4.5)	**0.0** (7.2)	**0.0** (8.5)	**0.0** (6.8)	1.6 (6.3)
576	**0.0** (**0.0**)	**0.0** (10.8)	**0.0** (**0.0**)	**0.0** (**0.0**)	5.9 (1.7)	5.4 (4.3)	3.9 (0.5)	3.5 (0.9)
610	2.6 (1.9)	2.7 (12.5)	3.6 (3.8)	1.1 (2.0)	7.7 (4.1)	7.3 (6.4)	5.2 (5.8)	4.4 (4.5)
797	3.7 (7.1)	1.9 (18.5)	4.5 (5.0)	4.6 (8.5)	12.3 (10.6)	12.5 (13.0)	10.3 (8.8)	9.7 (9.9)
799*	2.9 (2.5)	2.7 (13.5)	3.5 (2.2)	3.3 (2.8)	2.8 (6.1)	2.8 (8.8)	0.7 (5.1)	**0.0** (5.4)
833*	4.9 (5.1)	4.4 (16.0)	5.7 (4.9)	5.6 (5.4)	4.3 (8.0)	4.3 (10.8)	2.3 (6.9)	1.8 (7.4)
838*	2.7 (8.0)	2.8 (20.2)	4.4 (6.7)	4.2 (9.9)	3.3 (10.8)	3.6 (12.3)	1.9 (9.5)	1.5 (9.7)
850*	5.5 (10.6)	5.4 (22.4)	7.8 (10.8)	7.6 (13.4)	5.8 (9.6)	5.6 (11.3)	5.2 (9.7)	1.7 (9.3)
853*	3.3 (4.2)	2.8 (15.1)	3.9 (3.9)	3.7 (4.4)	2.8 (7.3)	2.9 (9.9)	0.7 (6.2)	0.2 (6.7)

[a] See Table 1 for explanations

508, 799, 851, and 857 for GNAA. For the loops with sequences GAAA, GCAA, and GAGA, the lowest energy structures are close to the experimental solution structures determined by NMR. This good agreement is visualized in Fig. 2 that shows, for the sequences GAAA, GCAA, and GAGA, the superposition of the lowest energy conformers (404, 857, and 508, respectively) and the NMR structures having the lowest RMSD to the respective conformers.

The RMSD ranges, calculated for each of the selected conformers with respect to the ensemble of NMR derived structures, are given in Table 3. For the purpose of comparison, Table 4 shows, in addition, the same data for a set of most stable structures selected using the simple ED model instead of the RF potential. In this case, the number of canonical tetraloop conformers is much smaller and the conformer with lowest energy is generally a structure that does not agree with the experimental data. The comparison of both sets of results indicates that solvent electrostatic effects, described by the RF, have a significant impact on the stability of nucleic acid structures in solution.

The differences between individual conformations are most clearly seen by looking at the torsion angle space for the backbone structures. Figure 3 shows the deviations of the individual backbone angles from their common values in A-form helices (shown by open rectangles) for the nucleotides G5 to G9. Apart from the U-turn specific change in α for G5 (Quigley and Rich 1976), different combinations of backbone angles are apparently possible in order to close the loop. The most abundant change is clearly the α/γ flip which occurs at junctions C4/G5, N6/R7, R7/A8, and A8/G9 in various combinations. The change of β on G9 from the normal *trans* conformation to the unusual *gauche* range, which is evident from the NMR studies (Heus and Pardi 1991), is only seen in conformer 508. Interestingly, this conformer, which is compatible with virtually all the structural features derived from the NMR data, is the most stable conformer found by our conformational search procedure (see Table 1). It should also be noted that the so-called crankshaft transitions of residues A8 and G9, which has been observed in the 200 ps molecular dynamics simulation of the GCAA hairpin (Zichi 1995), correspond to the α/γ flips seen in conformers (799, 851) and (404, 418, 445), respectively (see Fig. 3).

The only exception to the overall good agreement concerns the sugar puckers which are exclusively of the N-type (C3′-*endo* or C2′-*exo*) in the calculated structures, whereas an equilibrium of N- and S-type sugars for some of the loop nucleotides has been detected by NMR. Attempts to generate low-energy loop structures,

Table 3 Comparison of low-energy conformers (selected by energies including RF) with NMR derived structures[a]

GAAA			GCAA			GAGA		
ID	E_{RF}	RMSD range	ID	E_{RF}	RMSD range	ID	E_{RF}	RMSD range
404*	**0.0**	1.2–1.6	857*	**0.0**	0.6–1.6	508*	**0.0**	0.9–1.4
418*	0.2	1.2–1.6	404*	0.6	1.0–1.7	799*	1.6	1.1–1.5
419*	0.3	1.0–1.5	851*	0.6	0.8–1.7	404*	1.9	1.1–1.5
857*	0.5	0.8–1.6	576	1.0	3.5–3.9	419*	2.1	1.1–1.5
851*	0.7	1.1–1.8	418*	1.0	1.0–1.6	418*	2.2	1.1–1.4
508*	0.7	1.3–1.8	508*	1.5	1.0–1.6	445*	2.9	1.1–1.4
445*	0.9	1.2–1.6	838*	1.6	0.9–1.6	857*	3.1	1.3–1.6
838*	1.4	1.0–1.5	445*	1.7	1.0–1.6	838*	3.4	1.1–1.5
576	1.8	3.4–3.8	799*	1.9	0.7–1.6	851*	4.2	1.3–1.6
799*	2.9	1.0–1.8	419*	2.2	0.8–1.6	576	5.2	3.4–3.7

[a] The atoms of nucleotides 4–9 (loop nucleotides and the stem-closing pair) were used in the calculation of RMSD ranges (Å). The calculated energies (E_{RF}) include the electrostatic reaction field (RF) contributions

Table 4 Comparison of low-energy conformers (using electrostatic damping in energy calculations) with NMR derived structures[a]

GAAA			GCAA			GAGA		
ID	E_{ED}	RMSD range	ID	E_{ED}	RMSD range	ID	E_{ED}	RMSD range
915	**0.0**	4.1–4.4	188	**0.0**	5.4–5.9	188	**0.0**	6.0–6.2
188	0.6	5.6–5.9	576	0.9	3.5–3.9	600	0.1	4.2–4.3
175	1.3	5.4–5.6	175	1.6	5.4–5.9	175	2.0	5.6–5.7
576	1.5	3.4–3.8	600	2.6	3.8–4.5	576	2.7	3.4–3.7
600	3.0	3.8–4.2	164	2.8	4.0–4.5	613	2.7	2.1–2.6
164	3.1	4.6–4.9	915	3.0	3.7–3.8	531	2.8	3.1–3.4
171	3.3	4.6–4.8	171	3.3	3.9–4.4	857*	3.3	1.3–1.6
799*	5.1	1.0–1.8	857*	4.6	0.6–1.6	508*	3.5	0.9–1.4
508*	5.2	1.3–1.8	799*	4.8	0.7–1.6	851*	4.1	1.3–1.6
857*	5.2	0.8–1.6	851*	4.9	0.8–1.7	418*	4.2	1.1–1.4

[a] In contrast to Table 3, the energies (E_{ED}) were calculated using the simple electrostatic damping (ED) instead of RF

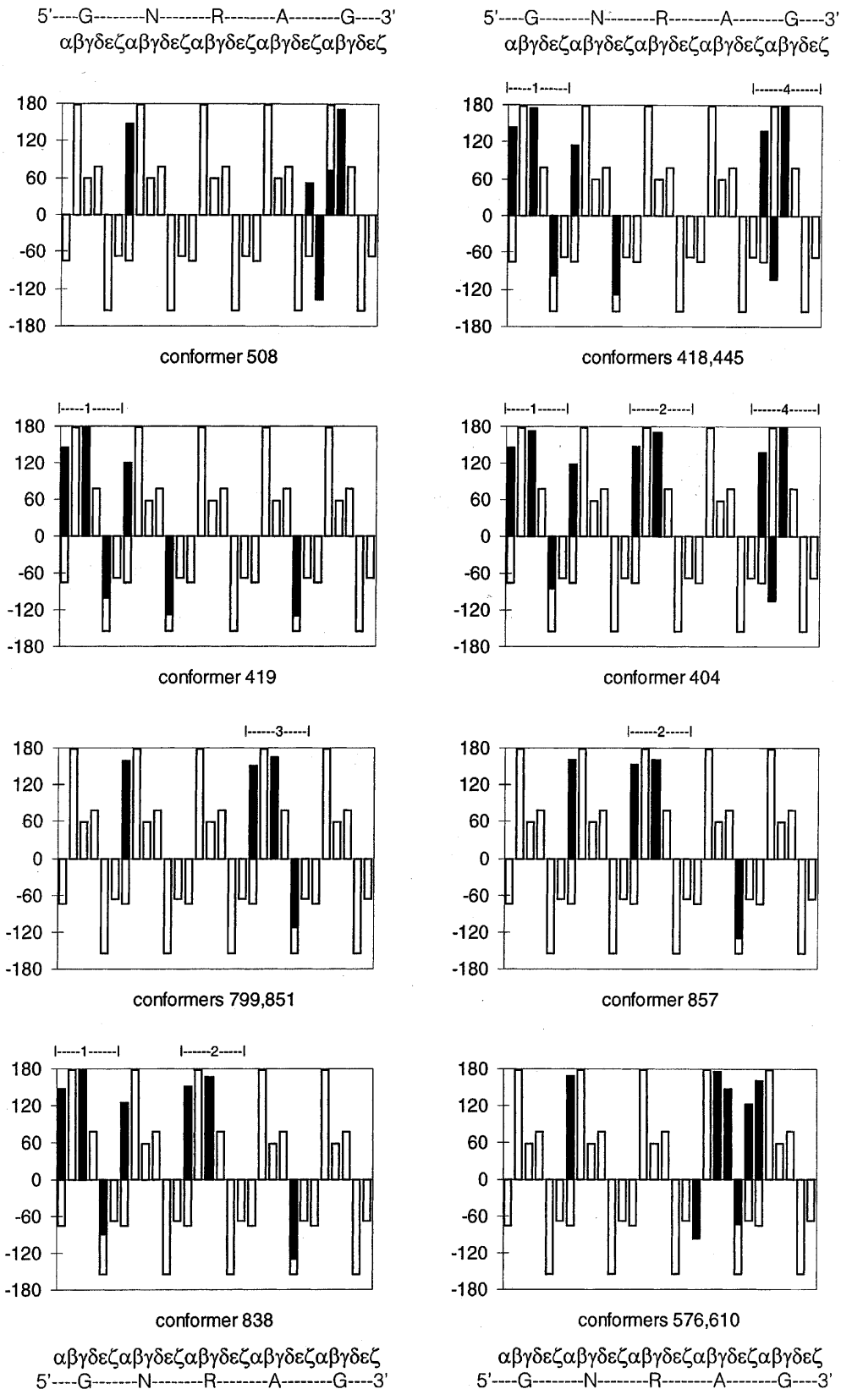

Fig. 3 Diagrams of the backbone torsion angles (degrees) for the lowest energy conformers of GNRA tetraloops. The deviations from common values in A-form helices (*unfilled columns*), larger than 30 degrees, are shown by *filled columns*. The α/γ flips occurring at junctions C4/G5, N6/R7, R7/A8, and A8/G9 are marked by *numbers 1–4*

starting from S-type (C2′-*endo*) sugar puckers and using energy minimization, have always reestablished the N-type sugar pucker (data not shown). Interestingly, in the molecular dynamics simulation of Zichi, which was started with the A7 ribose in C3′-*endo*, also no transition to C2′-*endo* could be observed. This indicates that this

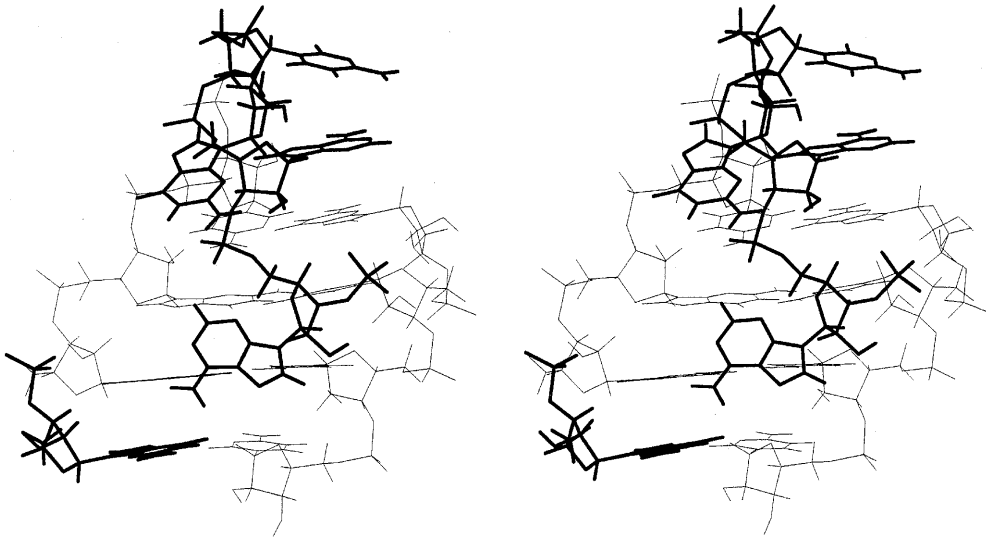

Fig. 4 Stereo view of the non-canonical conformer 576 showing the position of residue A8 in the major groove of the stem and its interaction with base G1

discrepancy to the NMR data could be due to a general problem of the AMBER force field which was used in both calculations.

One of the alternatives to the canonical tetraloop structures (conformer 576) is shown in Fig. 4. In this case, the nucleotide A8 is turned down into the major groove of the stem and the adenine base interacts with the G1 base pair via a hydrogen bond. Such a structure has not yet been detected experimentally, and thus we can only conjecture that it is presumably formed in hairpins with specific loop sequences, such as GNCA, where this conformer was indeed obtained as the structure with lowest energy.

Discussion

A force field based conformational analysis without using experimental constraints has been applied to RNA hairpins with GNRA and GNYA tetraloop sequences. The structures are represented by an ensemble of conformers that were selected purely on the basis of calculated energies from a large set of conformations generated by combinatorial loop search. The objective of the study was to investigate the capability of a continuum treatment of solvent electrostatic effects to describe relative stabilities of different conformers in agreement with experimental observations in solution. In contrast to the well-known failure of such predictions done on the basis of quasi-vacuum force fields (FLEX and AMBER4.1 with electrostatic damping by a sigmoidal dielectric function), the inclusion of reaction field contributions of the solvent result in a selection of low-energy conformers in accordance with experimental data.

The most encouraging result is that nine out of 10 lowest energy conformers, being selected from 74 starting structures in the final round of the conformational search, fit well the experimental data obtained for GAAA, GCAA, and GAGA hairpins. First, an asymmetric G5-A8 base pair is formed which stacks on the closing base pair of the stem. Second, the change of the α torsion angle for G5 from the common *–sc* to the unusual *+ac*/*+ap* range results in a sharp bend, a so-called U-turn, between the first and second nucleotide in the loop. Third, the bases of nucleotides N6, R7, and A8 continue the stacking from the 3′-end of the stem. Structural differences between individual low-energy conformers indicate that different combinations of backbone angles are apparently compatible with loop closure. In particular, an α/γ flip at junctions C4/G5, N6/R7, R7/A8, and A8/G9 is seen in various combinations. One of the conformers (number 508), having the lowest energy for GNGA and only slightly higher energies in the case of GRAA sequences, is consistent with almost all structural features derived from the NMR data.

RNA hairpins closed by GNYA tetranucleotide sequences show remarkable differences compared to the GNRA tetraloops. For GNCA, the canonical tetraloop structures are less stabilized by the electrostatic reaction field of the solvent and hence alternative structures, for example those locating nucleotide A8 in the major groove of the stem, are found with higher ranking in the conformational ensemble.

The results of this study indicate a definitive improvement of the molecular mechanics force field, used in searching the conformational space of a small RNA structural motif, by including a continuum treatment of electrostatic solute-solvent interactions.

Acknowledgements We thank Dr. M. Zacharias for helpful discussions. Support of this work by the priority program "RNA Biochemistry" (DFG Grant Sk 35/2-1) of the Deutsche Forschungsgemeinschaft is gratefully acknowledged. The work in Vienna has been partially supported financially by the Austrian Fonds zur Förderung der wissenschaftlichen Forschung (Project P-11065-CHE) and by the European Commission in the frame of contract study (PSS*0884) and the biotechnology program (BIO4-98-0189).

References

Antao VP, Lai SY, Tinoco I Jr (1991) A thermodynamic study of unusually stable RNA and DNA hairpins. Nucleic Acids Res 19: 5901–5905

Cate JH, Gooding AR, Podell E, Zhou K, Golden BL, Kundrot CE, Cech TR, Doudna JA (1996) Crystal structure of a group I ribozyme domain: principles of RNA packing. Science 273: 1696–1699

Davis ME (1994) The inducible multipole solvation model: a new model for solvation effects on solute electrostatics. J Chem Phys 100: 5149–5159

Davis ME, Madura JD, Luty BA, McCammon JA (1991) Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian dynamics program. Comput Phys Commun 62: 187–198

Elcock AH, Potter MJ, McCammon JA (1997) Application of Poisson-Boltzmann solvation forces to macromolecular simulations. In: Gunsteren WF van, Weiner PK, Wilkinson AJ (eds) Computer simulation of biomolecular systems: theoretical and experimental applications, vol 3. Kluwer, Dordrecht, p 244

Feigon J, Diekmann T, Smith FW (1996) Aptamer structures from A to zeta. Chem Biol 3: 611–617

Ferré-D'Amaré AR, Zhou K, Doudna JA (1998) Crystal structure of a hepatitis delta virus ribozyme. Nature 395: 567–574

Gautheret D, Major F, Cedergren R (1993) Modeling the three-dimensional structure of RNA using discrete nucleotide conformational sets. J Mol Biol 229: 1049–1064

Gilson MK, Sharp KA, Honig BH (1988) Calculating the electrostatic potentials of molecules in solution: method and error assessment. J Comput Chem 9: 327–335

Groebe DR, Uhlenbeck OC (1989) Thermal stability of RNA hairpins containing a four-membered loop and a bulge nucleotide. Biochemistry 28: 742–747

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

Hingerty B, Richie RH, Ferrel TL, Turner JE (1985) Dielectric effects in biopolymers: the theory of ionic saturation revisited. Biopolymers 24: 427–439

Jucker FM, Pardi A (1995) GNRA tetraloops make a U-turn. RNA 1: 219–222

Jucker FM, Heus HA, Yip PF, Moors EH, Pardi A (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. J Mol Biol 264: 968–980

Kajava A, Rüterjans H (1993) Molecular modelling of the 3D structure of RNA tetraloops with different nucleotide sequences. Nucleic Acids Res 21: 4556–4562

Lavery R, Sklenar H, Zakrzewska K, Pullman B (1986) The flexibility of the nucleic acids: (II) the calculation of internal energy and applications to mononucleotide repeat DNA. J Biomol Struct Dyn 3: 989–1014

Lavery R, Zakrzewska K, Sklenar H (1995) JUMNA (junction minimisation of nucleic acids). Comput Phys Commun 91: 135–158

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

Michel F, Westhof E (1990) Modeling of the three-dimensional architecture of group-I catalytic introns based on comparative sequence analysis. J Mol Biol 216: 585–610

Moore PB (1993) Recent RNA structures. Curr Opin Struct Biol 3: 340–344

Murphy FL, Cech TR (1994) GAAA tetraloop and conserved bulge stabilize tertiary structure of a group I intron domain. J Mol Biol 236: 49–63

Norberg J, Nilsson L (1998) Solvent influence on base stacking. Biophys J 74: 394–402

Orita M, Nishikawa F, Shimayama T, Taira K, Endo Y, Nishikawa S (1993) High-resolution NMR study of a synthetic oligoribonucleotide with a tetranucleotide GAGA loop that is a substrate for the cytotoxic protein, ricin. Nucleic Acids Res 21: 5670–5678

Pearlman DA, Case DA, Caldwell JC, Seibel GL, Singh C, Weiner P, Kollman PA (1991) AMBER4.0. University of California, San Francisco

Perbandt M, Nolte A, Lorenz S, Bald R, Betzel C, Erdmann VA (1998) Crystal structure of domain E of thermus flavus 5S rRNA: a helical RNA structure including a hairpin loop. FEBS Lett 429: 211–215

Pley HW, Flaherty KM, McKay DB (1994) Three-dimensional structure of a hammerhead ribozyme. Nature 372: 68–74

Quigley GJ, Rich A (1976) Structural domains of transfer RNA molecules. Science 194: 796–806

Scheraga HA (1993) Searching conformational space. In: Gunsteren WF van, Weiner PK, Wilkinson AJ (eds) Computer simulation of biomolecular systems: theoretical and experimental applications, vol 2. ESCOM, Leiden, pp 231–248

Schuster P, Stadler PF, Renner A (1997) RNA structures and folding. From conventional to new issues in structure predictions. Curr Opin Struct Biol 7: 229–235

Scott WG, Finch JT, Klug A (1995) The crystal structure of an all-RNA hammerhead ribozyme: a proposed mechanism for RNA catalytic cleavage. Cell 81: 991–1002

Shen LX, Cai Z, Tinoco I Jr (1995) RNA structure at high resolution. FASEB J 9: 1023–1033

Sklenar H, Lavery R, Pullman B (1986) The flexibility of nucleic acids: (I) "SIR", a novel approach to the variation of polymer geometry in constraint systems. J Biomol Struct Dyn 3: 967–986

Sklenar H, Eisenhaber F, Poncin M, Lavery R (1990) Including solvent and counterion effects in the force fields of macromolecular mechanics: the field integrated electrostatic approach (FIESTA). In: Beveridge DL, Lavery R (eds) Theoretical biochemistry & molecular biophysics, vol 2. Adenine Press, Schenectady, NY, pp 317–335

Uhlenbeck OC, Pardi A, Feigon J (1997) RNA structure comes of age. Cell 90: 833–840

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

Weiner SJ, Kollman PA, Nguyen DT, Case DA (1986) An all atomic force field for simulations of proteins and nucleic acids. J Comput Chem 7: 230–252

Woese CR, Winker S, Gutell RR (1990) Architecture of ribosomal RNA: constraints on the sequence of 'tetra-loops'. Proc Natl Acad Sci USA 87: 8467–8471

Wolters J (1992) The nature of preferred hairpin structures in 16S-like rRNA variable regions. Nucleic Acids Res 20: 1843–1850

Zacharias M, Sklenar H (1997) Analysis of the stability of looped-out and stacked-in conformations of an adenine bulge in DNA using a continuum model for solvent and ions. Biophys J 73: 2990–3003

Zichi DA (1995) Molecular dynamics of RNA with the OPLS force field. Aqueous simulation of a hairpin containing a tetranucleotide loop. J Am Chem Soc 117: 2957–2969