Conformational Analysis of DNA-Trinucleotide-Hairpin-Loop Structures Using a Continuum Solvent Model

Martin Zacharias
AG Theoretische Biophysik, Institut für Molekulare Biotechnologie, 07745 Jena, Germany

ABSTRACT A number of trinucleotide sequences in DNA can form compact and stable hairpin loops that may have significance for DNA replication and transcription. The conformational analysis of these motifs is important for an understanding of the function and design of nucleic acid structures. Extensive conformational searches have been performed on three experimentally known trinucleotide hairpin loops (AGC, AAA, and GCA) closed by a four-base-pair stem. An implicit solvation model based on the generalized Born method has been employed during energy minimization and conformational search. In addition, energy-minimized conformers were evaluated using a finite-difference Poisson-Boltzmann approach. For all three loop sequences, conformations close to experiment were found as lowest-energy structures among several thousand alternative energy minima. The inclusion of reaction-field contributions was found to be important for a realistic conformer ranking. Most generated hairpin loop structures within $\sim$5 kcal mol^{-1} of the lowest-energy structure have a similar topology. Structures within $\sim$10 kcal mol^{-1} could be classified into about five structural families representing distinct arrangements of loop nucleotides. Although a large number of backbone torsion angle combinations were compatible with each structural class, some specific patterns could be identified. Harmonic mode analysis was used to account for differences in conformational flexibility of low-energy sub-states. Class-specific differences in the pattern of atomic fluctuations along the sequence were observed; however, inclusion of conformational entropy contributions did not change ranking of structural classes. For an additional loop sequence (AAG) with no available experimental structure, the approach suggests a lowest-energy loop topology overall similar to the other three loop sequences but closed by a different non-canonical base-pairing scheme.

INTRODUCTION

Hairpin loop structures in RNA or DNA consist of a base-paired stem structure and a loop sequence with unpaired or non-Watson-Crick-paired nucleotides. These common structural motifs can be of functional importance as ligand recognition elements or folding initiation sites. In DNA, a number of trinucleotide sequences at the center of palindromic sequences can form stable and compact hairpin structures (Hirao et al., 1992, 1994; Sandusky et al., 1995; Yu et al., 1995; Zhu et al., 1995; Chou et al., 1996, 1999a,b; Aslani et al., 1996; Yoshizawa et al., 1997). Formation of hairpin loops in DNA may have biological significance during DNA replication and transcription (Astell et al., 1985; Glucksmann-Kuis et al., 1992, 1996; Zhu et al., 1996). It has been proposed that hairpin formation of triplet repeat sequences during DNA replication could play a role for the expansion of such repeats associated with several genetic diseases (Gacy et al., 1995; Chen et al., 1995; Mitas et al., 1995). Such DNA structures may also be interesting drug targets because their overall shape and groove geometry significantly differ from regular double-stranded DNA. The most stable trinucleotide hairpins are formed by the GNA motif (G, guanine; A, adenine; N, guanine, adenine, or cytosine) (Hirao et al., 1994; Yoshizawa et al., 1997). However, some other sequences such as AAA, AGC, and AAG can also form hairpin structures albeit with a melting transition at lower temperatures than for the GNA motifs (Yoshizawa et al., 1997). NMR spectroscopic studies have revealed similar compact folding topologies for the three loop sequences GCA, AAA, and AGC (Hirao et al., 1994; Zhu et al., 1995; Chou et al., 1996, 1999a).

Due to their small size and the characteristic and well defined fold, these loop sequences are useful test systems for energy-based structure prediction and conformational analysis approaches. Such computational studies can also be valuable to better understand the energetic origins of the stability of hairpin structures and why a specific loop topology is formed and preferred over other possible conformations.

A prerequisite for energy-based structure prediction is the ability to sort out realistic structures as those of low or lowest energy from a large pool of sterically possible structures. Recent studies have shown that inclusion of solvation effects in the framework of a continuum solvent description based on solutions to the Poisson-Boltzmann (PB) equation can improve conformational search and modeling attempts on nucleic acids (Zakrzewska et al., 1996; Zacharias and Sklenar, 1997, 1999; Maier et al., 1999; Ayadi et al., 1999; Rohs et al., 2000; Zacharias, 2000a). However, the standard finite-difference method (FDPB) to solve the PB equation is only of limited usefulness for energy minimization, and therefore electrostatic solvation contributions can be in-

Received for publication 17 October 2000 and in final form 24 February 2001.

Address reprint requests to Dr. Martin Zacharias, AG Theoretische Biophysik, Institut für Molekulare Biotechnologie, Beutenbergstrasse 11, 07745 Jena, FRG. Tel.: 49-3641-656491; Fax: 49-3641-656495; E-mail: zacharia@imb-jena.de.

0006-3495/01/05/2350/14 $2.00

cluded only after conformer generation and energy minimization (with a simpler, for example, distance-dependent dielectric treatment of electrostatic interactions). The generalized Born (GB) method (Still et al., 1990; Hawkins et al., 1996; Qui et al., 1997; Jayaram et al., 1998; Srinivasan et al., 1998, 1999) allows one to calculate electrostatic interactions and solvation based on effective Born solvation radii for each atom that are calculated from the distribution of all atoms in the molecule. This approach allows a principally less accurate but more rapid calculation of electrostatic solvation than the FDPB approach and in addition rapid calculation of gradients useful for energy minimization and molecular dynamics. Several recent molecular dynamics simulation studies on nucleic acids have shown that the GB approach can lead to trajectories that largely resemble the molecular motions observed in simulations with explicit solvent and ions (Williams and Hall, 1999; Tsui and Case, 2000). In the present study, the GB method has been used during extensive conformational search and energy-minimization studies on DNA hairpin loop structures and compared with energetic evaluation based on the FDPB method and simpler electrostatic treatments. In addition, the conformational flexibility (conformational entropy) around resulting low-energy minima was calculated based on a harmonic mode treatment and used for the evaluation of the generated conformers. The approach has been successful in identifying conformations close to the corresponding known experimental structures for all three loop sequences as lowest-energy structures among more than 3000 alternative energy-minimized loop conformations. The comparison with alternative structures allows us to draw some conclusions about structural requirements for loop closure and the interactions that stabilize the native hairpin loop structure compared with other structural classes. Conformational searches with the same approach have also been used to suggest a low-energy structure for the AAG-loop sequence. Biochemical and thermodynamic studies indicate that this loop sequence may also form a stable loop (Yoshizawa et al., 1997) for which no experimental structure is presently available.

MATERIALS AND METHODS

Energy minimization

Energy minimization calculations have been carried out using a modified version of the Jumna (junction minimization of nucleic acids) program (Lavery et al., 1995) and the Amber4.1 force field (Cornell et al., 1995). In Jumna, each nucleic acid strand is considered as a chain of 3′-monophosphate nucleotides that are placed in space using helicoidal coordinates. These are three translational variables (Xdisp, Ydisp, and Rise) and three rotational variables (Inclination, Tip, and Twist) obeying the Cambridge convention for nucleic acids (Dickerson et al., 1989). In addition, single-bond torsion and valence angles are used to describe the internal nucleotide flexibility. Except for the connection between each nucleotide all bond lengths are assumed to be fixed at their optimum values. The energy function for energy minimization and harmonic mode calculations consists of pairwise additive nonbonded Coulomb (ΔE_{Coul}) and Lennard Jones (ΔE_{LJ}) terms (no cutoff) and valence angle and dihedral angle contributions (ΔE_{AT}).

$$V = \sum_{angles} \tfrac{1}{2} K_\theta(\theta - \theta_o)^2 + \sum_{\mathrm{torsions}} K_\phi(1 + \cos(n\phi - \delta)) + \sum_{i<j} \{C_{12} r_{ij}^{-12} - C_6 r_{ij}^{-6}\} + \sum_{i<j} q_i q_j (r_{ij})^{-1} - \tfrac{1}{2}(1 - \varepsilon_w^{-1}) \sum_{i,j} q_i q_j (r_{ij}^2 + \alpha_{ij}^2 \exp[-r_{ij}^2/4\alpha_{ij}^2])^{-0.5} \quad (1)$$

During energy minimization, electrostatic reaction-field contributions were calculated using a GB model (Still et al., 1990; last term in Eq. 1) added to the Jumna program. In the GB model, electrostatic solvation or reaction-field contributions (ΔE_{reGB}) due to differences in the assigned dielectric constant for the molecule and surrounding solvent are calculated from the charge and atom distribution in the molecule. For calculating the so-called effective Born radii, α_{ij}, the pairwise descreening approximation described by Hawkins et al. (1996) was used with the following atom radii (in Å): $R_H = 1.2$, $R_O = 1.5$, $R_N = 1.6$, $R_C = 1.7$, and $R_P = 2.0$, with a radius offset of -0.12. The screening factors were 0.8 (for H), 0.83 (O), 0.76 (N), 0.76 (C), and 0.86 (P), respectively. These parameters were found to give a very reasonable correlation to reaction-field energies calculated with a finite-difference solution of the Poisson equation using the UHBD program (Madura et al., 1995) and the same set of atomic radii. In the FDPB calculations, a 1.4-Å water probe was used to define the molecular surface of the molecules. A comparison of reaction-field energies using the GB and FDPB (ΔE_{rePB}) approaches for ~60 low-energy hairpin loop structures (all AAA-loop conformers from a conformational search that are within 15 kcal mol^{-1} of the lowest-energy conformer) is shown in Fig. 1. Surface-area-dependent nonpolar solvation contributions (ΔG_{SAS}) were evaluated from the accessible surface area (with $\gamma = 0.0055$ kcal·mol^{-1} Å^{-2}; Sitkoff et al., 1994). This term was found to vary very little even between very

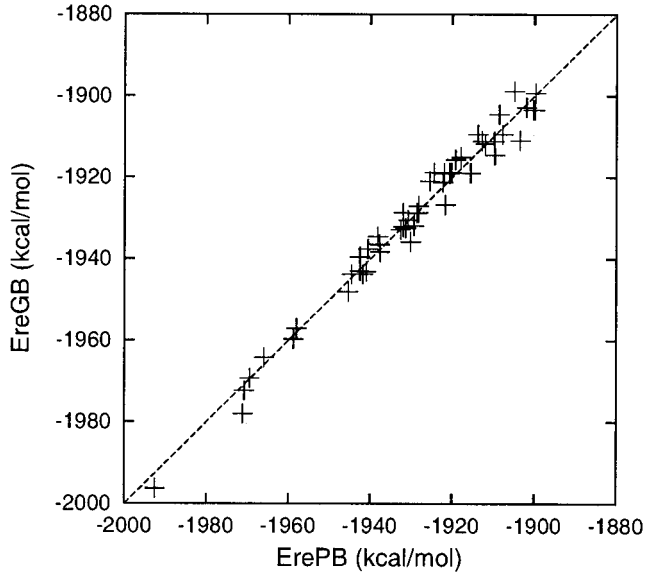

FIGURE 1 Calculated reaction field energies for low-energy DNA hairpin AAA-loops (within ~15 kcal mol^{-1} of the lowest-energy structure) using either the FDPB approach (ErePB, *x* axis) or the GB method (EreGB, *y* axis).

different loop conformers ($<\sim$0.3 kcal mol^{-1}) and was therefore calculated only for the final energy-minimized structures. The total energy of a conformer ($\Delta E_{\text{totPB}}/\Delta E_{\text{totGB}}$) is given as a sum of Coulomb (ΔE_{Coul}), Lennard-Jones (ΔE_{LJ}), valence and torsion angle (ΔE_{TA}), electrostatic solvation ($\Delta E_{\text{rePB}}/\Delta E_{\text{reGB}}$), and nonpolar solvation (ΔE_{SAS}) contributions.

Harmonic mode calculations and estimation of conformational entropies

Harmonic mode calculations were performed essentially as described in Zacharias and Sklenar (2000) and Zacharias (2000b). Briefly, the matrix of second derivatives of the potential energy for a hairpin structure was calculated numerically at an energy minimum using a finite-difference scheme (Levitt et al., 1985). The harmonic modes are the eigenvectors of this matrix. Within the harmonic approximation, the equilibrium conformational distribution around one energy minimum is given by the second moment (variance and covariance) of the conformational variables and associated multi-variant Gaussian distribution. Thermodynamic quantities can be calculated by integration over this Gaussian distribution. Backbone torsion angle variances and covariances, σ_{ij}, were obtained by deforming each structure in eigenvector direction and summing over each harmonic mode contribution (excited by $\frac{1}{2}kT$, where T is room temperature and k is the Boltzmann constant). Within this approach conformational entropies are given by (Karplus and Kushick, 1981)

$$S_{\text{tor}} = \frac{1}{2} k \ln[(2\pi)^{n}\sigma] \tag{2}$$

$$\sigma = \det[\text{Matrix}(\sigma_{ij})] \tag{3}$$

$$\sigma_{ij} = \sum_{k}(\Delta\text{tor}_{i,k} \times \Delta\text{tor}_{j,k}) \tag{4}$$

In Eq. 4, the summation of the covariance of calculated torsion angle deformations ($\Delta\text{tor}_{i,k} \times \Delta\text{tor}_{j,k}$) goes over all harmonic modes (k).

Conformational search

Conformational searches started from two model-built structures (7 nucleotides, 5′-AC**AGC**GT, loop sequence in bold), one with the central G stacked on the 5′-stem and in the other case stacked on the 3′-stem sequence. For the systematic search starting from the two structures, various combinations of backbone torsion angle window constraints were generated for each of the three loop nucleotides. Because the number of required energy minimizations increases very rapidly with the number of backbone torsion angles included explicitly for generating combinations of constraints, only a subset of the most flexible nucleic backbone torsion angles α, γ, and ζ could be considered. For each loop nucleotide, α was constrained to one of two windows, ($-70°$; $-50°$) or (60°; 120°), γ to (50°; $+70°$) or (170°; $-170°$), and ζ to ($-70°$; $-50°$) or (170°; $-170°$), corresponding to the most common ranges of these three backbone torsion angles in DNA (Saenger, 1984), respectively. In addition, for the central loop nucleotide, two window constraints for the δ torsion angle were added corresponding to the C3′-endo (70°; 90°] and C2′-endo regimes (140°; 160°), respectively. With the two start conformations and the resulting 2048 combinations of window constraints applied during energy minimization (200 steps), a total of 4096 different conformations were generated (during this phase the backbone stem sequence was constrained to stay reasonably close to B-form geometry). Subsequently, the structures were relaxed using (unconstrained) energy minimization resulting in $\sim$3100 different energy minima. The B-DNA stems of the 500 lowest-energy structures of the two searches were extended to form the full GTA-CAGCGTAC sequence, and the structures were again energy minimized. In molecular dynamics simulations of double-stranded nucleic acids using the GB continuum model, inclusion of salt effects was necessary to prevent strand dissociation (Tsui and Case, 2000). Although salt effects have not been considered during loop structure generation, no strand dissociation was observed during the final relaxation step, presumably because the stem start structure was close to B-form in a base-paired geometry for all generated conformers.

The protocol took $\sim$2 days of computer time on a SGI R12000 Octane workstation. The resulting conformers served as start structures for the GCA-, AAA-, and AAG-loop sequences (with the same stem sequence). The 100 lowest-energy structures of each loop sequence were evaluated using the FDPB model, sorted according to their energy, and further classified according to the helical placement of the loop nucleotides and backbone torsion angle pattern. Harmonic mode calculations were performed on some of the low-energy structures to get an impression of the conformational flexibility around the calculated energy minimum. It should be emphasized that although $\sim$85% of the 4096 start structures resulted in unique energy minima, the present conformational search still does not cover all possible energy minima for the hairpin loop because not all principally possible torsion angle combination have been explicitly considered. However, additional searches using random variations of the helicoidal coordinates or the MCSYM conformational generator (Major et al., 1991; Gautheret et al., 1993) to obtain additional start structures for energy minimization did not yield new structures of lower energy than the above searches and were not considered further.

RESULTS

Comparison of calculated low-energy and experimental structures

The conformational searches on trinucleotide DNA hairpin loops with a four-base-pair stem resulted in more than 3000 different (energy-minimized) backbone topologies forming a closed loop. The generated structures were ranked according to the FDPB continuum approach because this model allows a principally more accurate definition of the molecule/solvent boundary than the GB model for calculating reaction field energies. The total energy (ΔE_{totPB}) of only a fraction of all conformers from the searches on the three loop sequences, AGC, AAA, and GCA, was within 10 kcal mol^{-1} of the total energy of the lowest-energy conformation ($<$100 conformers). For each loop sequence, the majority of these low-energy structures have a similar helical or topological arrangement of the loop nucleotides in the following, termed class 1. Structures within this family are similar to the corresponding experimentally determined structures for all three loop sequences (Fig. 2). It is characterized by a stacking of the central loop nucleotide mainly on the first nucleotide or the 3′-side of the B-DNA stem with the central loop base facing toward the DNA major groove. Among the many backbone torsion angle combinations compatible with the class 1 hairpin topology, the FDPB model selected for all three loop sequences a lowest-energy backbone torsion angle pattern (structure 2171) that is close to the pattern reported for the experimental loop structures (Zhu et al., 1995; Chou et al., 1996, 1999a; see also Table 3). For example, in case of the AAA-loop for which the experimental stem sequence is identical to the sequence used in the present study the all-atom root mean square deviation between experimental and calculated structure

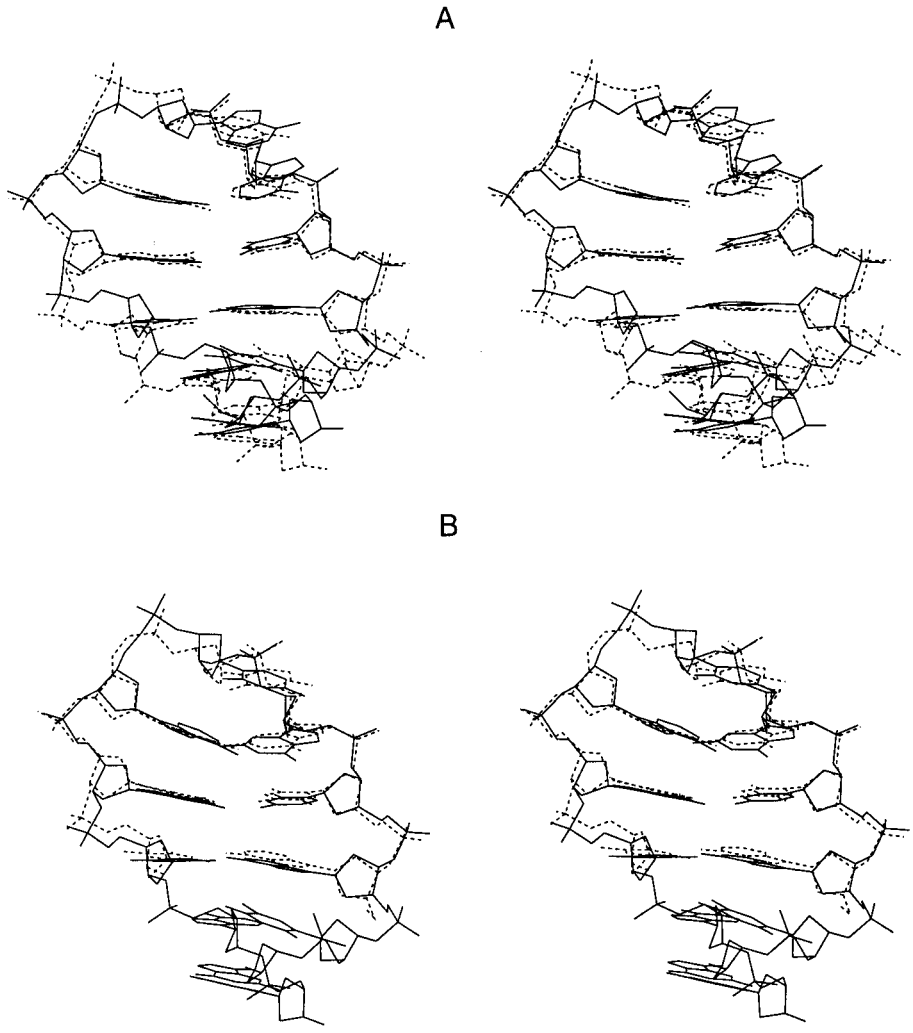

FIGURE 2 Comparison of lowest-energy hairpin loop structures (in stereo) from conformational searches (conformer 2171; *solid line*) and corresponding experimental structures based on NMR spectroscopy (*dashed line*). (*A*) GTAC**AAA**GTAC-loop sequence (loop nucleotides in bold, PDB entry of the experimental structure: first structure of 1bjh). (*B*) GTAC**GCA**GTAC-loop sequence (PDB entry of experimental structure: first structure of 1zhu). For the GCA-loop, only part of the experimental structure with the same sequence as used in the calculations is shown. For clarity, hydrogen atoms have been omitted.

(2171) is <1.4 Å. Energy minimization of the experimental structure (first structure in Protein Data Bank (PDB) entry: 1bjh) results in a structure identical to 2171 (note that all published NMR structures in PDB entry 1bjh have the same basic backbone torsion angle pattern).

The reaction-field energies calculated with the GB model for various hairpin loop conformers show a very reasonable correlation to the FDPB method (Fig. 1), and for both models similar class 1 conformers are predicted to be of lowest energy (Table 1). However, with the GB model for all three loop sequences, conformer 1387 was selected as the lowest-energy structure. The backbone torsion angle pattern for this structure slightly deviates from the experimentally defined pattern (Table 3) based on NMR spectroscopy (mainly due to a B_{II} state at the first loop nucleotide).

The effect of small GB parameter changes (offset and screening factors; see legend of Fig. 3) that still result in reasonable correlation with FDPB reaction-field energies on conformer ranking was tested for the AAA-loop sequence. The alternative set of parameters affected the relative ranking by <1 kcal mol^{-1} for most conformers (Fig. 3), and conformer 1387 was still selected as lowest-energy conformer followed by several other class 1 structures. In the present study, the GB model has been used to approximately reproduce electrostatic solvation energies obtained with the more accurate FDPB method. More drastic changes in GB parameters that result in a loss of correlation between GB and FDPB approaches can affect conformer ranking more strongly (not shown).

Classification of calculated hairpin loop structures

Conformers within ~10 kcal mol^{-1} of the lowest-energy structures have been classified into five classes for the three loop sequences according to the helical arrangement of the loop nucleotides. Although low-energy loop structures are dominated by class 1 for all three sequences (all low-energy loop structures with ~3.5 kcal mol^{-1} of the lowest-energy

TABLE 1 Ranking of hairpin loop conformers

Rank	Class	Number	ΔE_{totPB}	ΔE_{totGB}	ΔE_{Coul}	ΔE_{LJ}	ΔE_{AT}	ΔE_{rePB}	ΔE_{reGB}	ΔE_{SAS}	ΔE_{salt}	$T\Delta S_{vi}$
GTAC**AAA**GTAC*												
1	1a	2171	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	−10.3	0.0
2	1a	1933	0.2	1.2	21.3	2.1	−0.8	−22.5	−21.5	0.2	−10.4	0.1
3	1a	1387	0.5	−2.6	27.6	1.5	−1.2	−27.4	−30.6	0.0	−10.5	−0.2
4	1a	1147	1.1	0.4	2.4	0.2	0.7	−2.1	−2.9	−0.1	−10.7	0.9
5	1a	1131	1.2	0.6	14.5	0.0	0.8	−14.1	−14.7	0.0	−10.5	0.7
6	1a	1419	2.9	−0.5	31.1	−3.8	3.8	−28.0	−31.4	−0.1	−10.6	1.2
7	1a	2604	3.3	−0.8	31.5	0.3	0.2	−28.6	−32.7	−0.1	−10.6	0.2
8	1b	2084	3.5	0.2	37.9	−0.5	−0.9	−32.8	−36.4	−0.1	−10.8	1.0
9	2	3417	3.6	5.8	9.0	8.4	0.1	−14.3	−12.1	0.4	−10.4	−0.9
15	1c	2175	4.9	4.7	12.0	2.4	4.4	−13.9	−14.2	0.1	−10.5	0.9
21	3	2155	5.7	0.4	38.3	−3.3	0.2	−29.1	−34.6	0.1	−10.5	0.8
43	5	3881	8.4	7.2	57.1	1.8	5.6	−56.5	−57.6	0.2	−10.7	−0.7
53	4	1967	10.3	4.9	21.1	−0.4	7.7	−17.4	−23.1	−0.3	−10.6	0.6
GTAC**GCA**GTAC												
1	1a	2171	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	−10.3	0.0
2	1a	1933	0.8	4.0	29.6	4.5	−0.3	−33.1	−29.9	0.1	−10.7	0.1
3	1a	1387	1.2	−1.6	27.1	2.2	0.3	−28.1	−31.0	−0.2	−10.5	−1.1
4	1a	1131	1.7	1.0	14.1	−0.2	0.9	−13.0	−13.7	−0.1	−10.4	−0.1
5	1a	1147	1.9	0.7	3.0	0.2	0.7	−1.9	−3.1	0.0	−10.4	0.2
6	1a	1899	2.7	3.1	17.8	3.7	1.8	−20.7	−20.3	0.0	−10.6	−1.2
7	2	3417	3.2	5.9	6.7	10.0	0.9	−14.5	−11.9	0.1	−10.4	−1.7
8	1a	2172	5.3	2.2	41.0	0.5	1.9	−37.9	−41.0	−0.2	−10.8	−0.4
9	1a	1915	5.5	3.1	40.2	5.0	2.0	−41.7	−44.1	0.0	−10.8	−1.9
12	1b	2084	6.3	4.0	50.6	1.2	1.5	−47.1	−49.4	0.0	−10.9	−0.4
19	3	2155	7.1	2.1	37.0	−1.9	0.8	−28.6	−33.6	−0.2	−10.5	0.4
52	6	1451	10.9	7.7	75.4	5.4	2.4	−72.4	−75.7	0.2	−10.9	−0.9
63	5	3881	11.2	8.8	59.2	4.3	6.3	−59.2	−61.3	0.4	−10.8	−1.0
GTAC**AGC**GTAC												
1	1a	2171	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	−10.5	0.0
2	1a	1147	0.8	0.1	−2.8	1.1	1.1	1.4	0.7	0.0	−10.6	0.6
3	1a	1387	1.5	−1.1	23.1	4.2	−0.6	−25.1	−27.7	−0.1	−10.6	0.9
4	1a	1933	1.7	2.5	4.0	5.2	−1.2	−6.4	−5.5	0.0	−10.7	1.0
5	1a	2604	2.0	−0.4	24.5	2.5	0.7	−25.6	−28.0	−0.1	−10.7	1.4
6	1a	2027	2.3	0.7	17.5	2.8	0.7	−18.6	−20.2	0.0	−10.8	1.3
7	1a	1963	3.5	3.8	3.4	1.8	2.5	−4.2	−4.0	0.0	−10.7	0.1
8	1a	2491	4.5	3.6	16.5	5.8	−0.1	−17.7	−18.6	0.0	−10.8	0.2
9	1a	2316	4.5	4.9	2.2	4.6	4.0	−6.4	−6.0	0.1	−10.7	0.9
10	1a	1003	4.9	4.5	−0.8	3.2	3.4	−0.9	−1.3	0.0	−10.6	0.2
18	3	2155	7.0	1.6	27.7	1.7	−0.6	−21.6	−27.0	−0.1	−10.5	0.5
24	2	3289	8.0	5.8	22.2	0.9	3.3	−18.3	−20.5	−0.1	−10.6	0.9
46	1b	1459	11.2	5.5	37.8	6.1	−0.4	−32.5	−38.2	0.1	−10.9	−0.3
74	5	3881	13.2	9.4	59.2	5.2	5.1	−56.5	−60.2	0.2	−10.8	−0.1
80	4	1967	15.4	5.8	−0.7	2.0	5.1	9.3	−0.3	−0.3	−10.7	−0.3

Column 1 indicates the ranking of DNA hairpin loop structures obtained from conformational searches. Classification and numbering of conformers are given in columns 2 and 3, respectively. The total energy ($\Delta E_{totPB}/\Delta E_{totGB}$) is a sum of Coulomb ($\Delta E_{Coul}$), Lennard Jones ($\Delta E_{LJ}$), electrostatic reaction field ($\Delta E_{rePB}/\Delta E_{reGB}$), valence + torsion angle (ΔE_{TA}), and surface-area-dependent nonpolar solvation (ΔE_{SAS}) contributions and is given with respect to the lowest-energy conformer (first rows). Only the energies of the 9 or 10 lowest-energy conformers and lowest-energy class 2–5 conformers for each loop sequence are given. Salt dependencies are calculated as the change in electrostatic energy upon addition of 150 mM monovalent salt and solving the nonlinear FDPB (Sharp and Honig, 1990). Differences in conformational entropy (last column) are obtained from a harmonic mode analysis (T: room temperature, see Materials and Methods) and are not included in the total energies (columns 4 and 5).
*Loop nucleotides are in bold.

form belong to class 1), the analysis of higher-energy topologies is of interest because it may give insights as to why the class 1 topology is energetically preferred and what other loop arrangements may be energetically feasible upon interaction with other nucleic acids or proteins. Within the above energy range, similar structural families were found for all three loop sequences. The classification for the AAA-hairpin loop is shown in Figs. 4 and 5.

In the case of the AAA-loop, two class 1 sub-forms could be distinguished that showed the same basic stacking pattern (central loop nucleotides stacks on the 3′-end of the stem and points toward the major groove; Fig. 4) but small

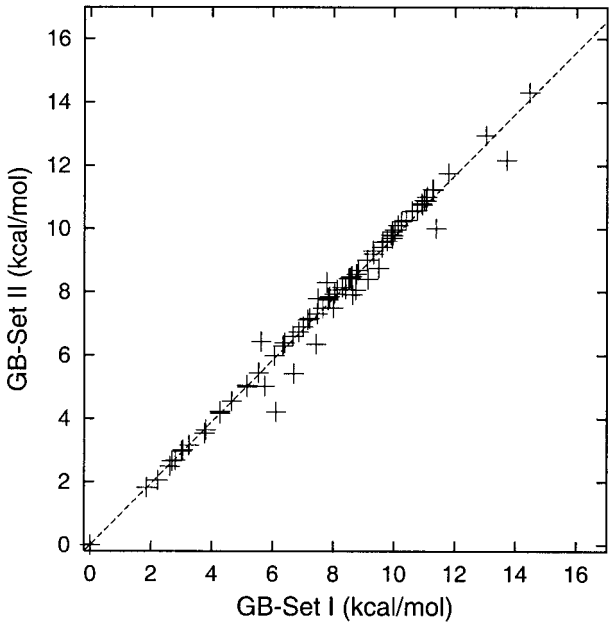

FIGURE 3 Dependence of AAA-hairpin loop conformer ranking on GB parameters. Hairpin loop conformers (~80 low-energy structures) were energy minimized using either GB parameters as described in Materials and Methods (GB set I with a radius offset = −0.12 and screening factors fs(H) = 0.8, fs(O) = 0.83, fs(N) = 0.76, fs(C) = 0.76, and fs(P) = 0.86) or GB set II (radius offset = −0.13 and fs(H) = 0.82, fs(O) = 0.84, fs(N) = 0.77, fs(C) = 0.77, and fs(P) = 0.86). Conformer ranking is given with respect to the lowest-energy conformer (1387) obtained with the GB model (without surface area term).

differences in the arrangement of the non-canonical loop closing base pair formed by the first and third loop nucleotides, respectively. Class 1b differs from 1a by a slightly tilted first loop nucleotide with respect to the stem sequence and an increased stacking interaction of the central loop base with the third loop nucleotide (Fig. 4).

Other classes for which at least one member was found within ~10 kcal mol^{-1} of the lowest-energy structures are, for example, characterized by a central loop base pointing away from the major groove and partially stacked on the third loop nucleotide (and this one stacked on the 5′-end of the helical stem). This class 2 arrangement is reminiscent of the loop topology of GNRA tetraloops in RNA (Heus and Pardi, 1991; Jucker and Pardi, 1995), which is also called type III tetraloop topology (van Dongen et al., 1997). This class also includes some conformers with the central loop base moved toward the DNA minor groove (reminiscent of type II in the tetraloop nomenclature of van Dongen et al., 1997). The present calculations predict the class 2 topology to be energetically less stable than the class 1 form by >3 kcal mol^{-1} (for all three loop sequences). In class 3–5 conformers, the first and second loop bases form a continuous stack with the 3′-end of the helical stem pointing toward the DNA major groove. In class 3 and 4, the third loop nucleotide is placed in the DNA minor groove in two perpendicular orientations with respect to the minor groove floor (Fig. 5). In contrast, in class 5 structures, the third nucleotide points toward the major groove (Fig. 5). Recently, Chou et al. (1999b) observed for pyrimidine-rich trinucleotide loops (sequence TCC) a topology with the first loop nucleotide located in the DNA minor groove and the other loop nucleotide partially stacked. Such an arrangement was found to be of considerably higher energy for the present loop sequences (>12 kcal mol^{-1} with respect to the lowest-energy class 1 structure, not shown), indicating that a purin nucleotide at the first loop position might be incompatible with this experimentally observed pyrimidine-rich loop topology.

Additional class 1 sub-forms have been observed in case of the AAA- and GCA-loop sequence, respectively, that differ from 1a by an alternative pairing scheme of the non-canonical A:A or G:A closing base pair (termed 1c). For the GCA-loop, this alternative pairing scheme is illustrated in Fig. 6. The energy of these 1c GCA- and AAA-loop forms is, however, considerably higher than the energy of the lowest-energy 1a forms (Table 1).

For an understanding of hairpin loop formation in nucleic acids it is interesting to compare the backbone torsion angle pattern observed for the various classes of conformers compatible with a closed hairpin loop (Fig. 7). For example, all low-energy class 1 conformers show a shift of the ϵ, from the standard *trans* state toward $-$*gauche*, and ζ torsion angles, from $-$*gauche* to $+$*gauche*, respectively, at the central loop nucleotides (Fig. 7). This is also seen in the experimental GCA-, AAA-, and AGC-loop structures (Zhu et al., 1995; Chou et al., 1996, 1999a). The conformers are more variable in the pattern for the α, β, and γ torsion angles of the loop and flanking nucleotides. Conformers belonging to class 1b show a characteristic shift of the δ torsion angle toward the C3′-endo regime at the first loop nucleotide not seen for the class 1a forms. Interestingly, for class 2, no characteristic flips in δ, ϵ, and ζ (compared with the standard B-DNA values) can be seen. However, in this case either the α or β torsion angles between the first and second loop nucleotide shift toward the *trans* or *gauche* (α) or $-$*gauche* (β) regimes, respectively, to form the loop.

The calculated salt dependence of the electrostatic energies showed little variation among low-energy conformers. Conformers of the subclass 1b appeared to be slightly more stabilized by the addition of salt than, for example, conformers of the 1a type. This could be related to the finding that 1b conformers appeared to be stabilized by a more favorable reaction-field contribution but destabilized by a less favorable Coulomb contribution compared with 1a conformers (Table 1). The other classes of conformers are less favorable than class 1a due to a combination of electrostatic and packing interactions. Because most low-energy conformers are compact structures, the surface-area-dependent nonpolar solvation term did not vary significantly. It is interesting that differences in Coulomb contributions for each conformer are largely compensated by reaction-field

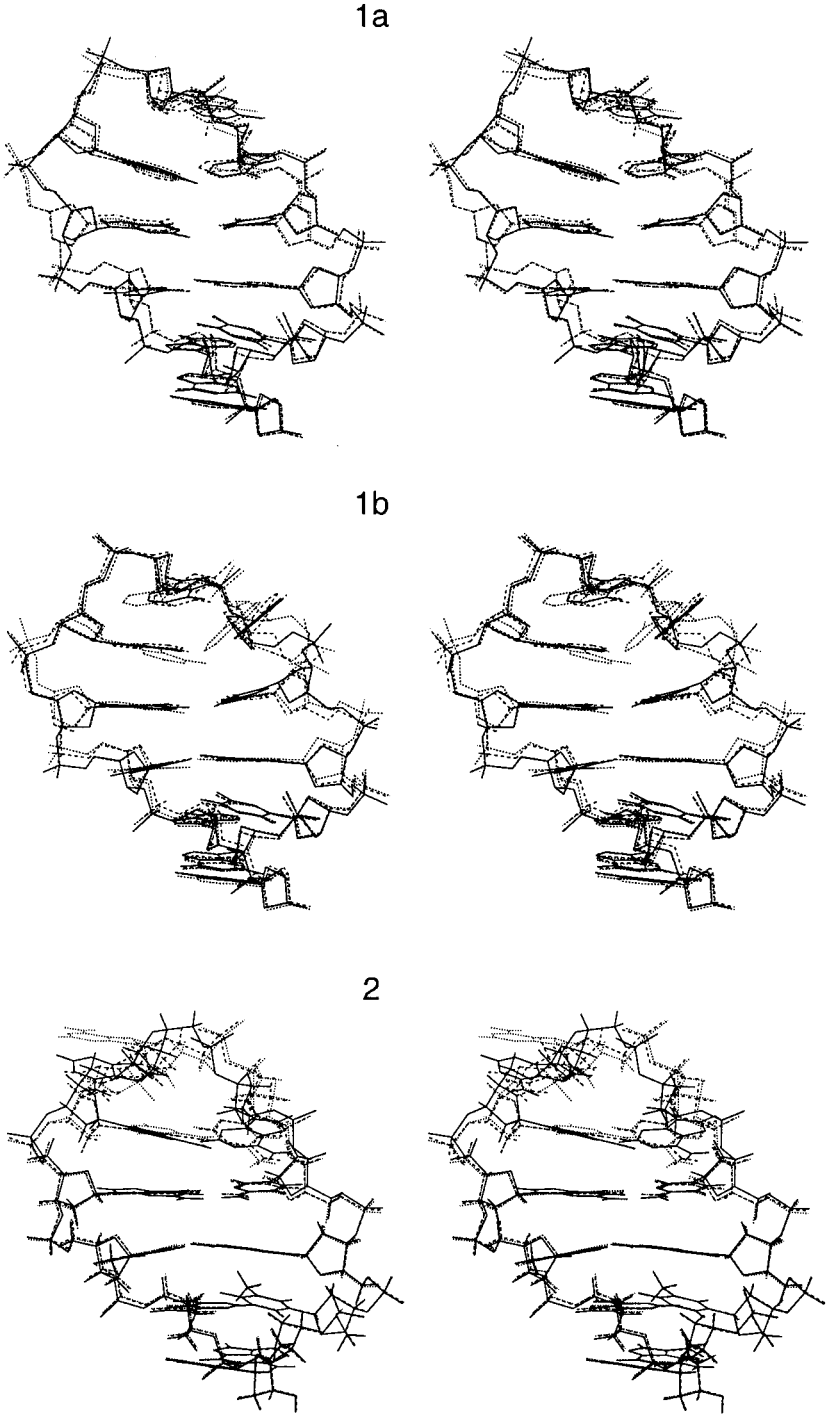

FIGURE 4 Superposition of three to four conformations (represented by different line types, in stereo) of the AAA-hairpin loop belonging to class 1a, 1b, or 2 with representative conformers found within ~5 kcal mol^{-1} of the lowest-energy (class 1a) conformation. For clarity, hydrogen atoms have been omitted.

contributions (Table 1) resulting in relatively small total energy differences. Neglecting electrostatic solvation could therefore result in unrealistically large energy differences between conformers.

As demonstrated for the AAA-loop sequence (Table 2), energetic ranking with a simpler treatment of electrostatic interactions using a distance-dependent or constant effective dielectric constant can result in a substantial reordering of low-energy conformers. In these cases, structures of lowest energy do belong to classes other than class 1 and are in disagreement with experimental structures. Interestingly, the ranking obtained with a constant effective dielectric constant ($\epsilon_{eff} = 4.0$) shows some agreement with the FDPB ranking for a number of conformers. However, there are conformers outside the range of low-energy conformers given in Table 1 that include deformations in the stem

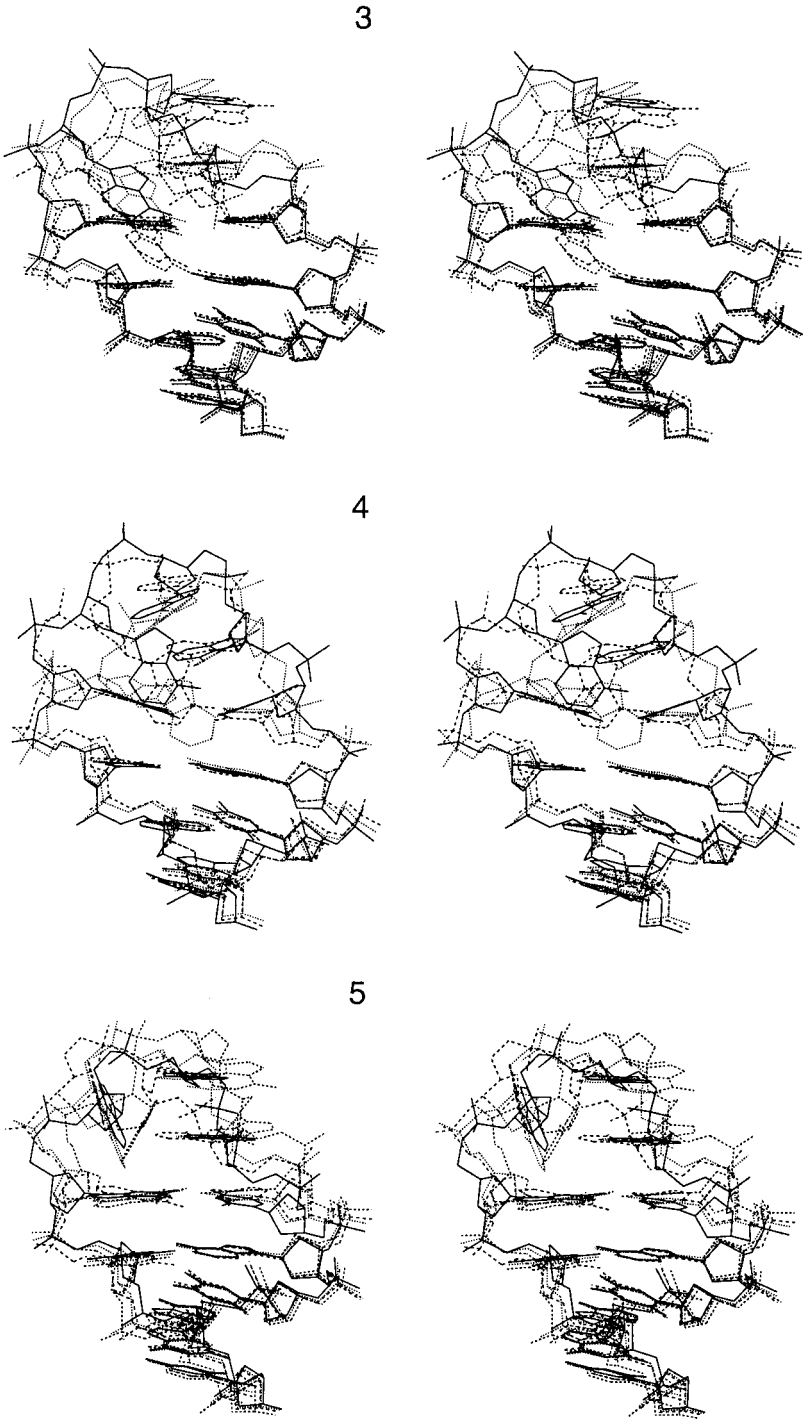

FIGURE 5 Superposition of three to four conformations (in stereo) of class 3–5 (AAA)-hairpin loops found within ~5–10 kcal mol^{-1} of the lowest-energy structure. For clarity, hydrogen atoms have been omitted.

region with contacts between bases and phosphate groups that are ranked better than class 1 conformers when using a uniform constant permittivity (e.g., conformer 2828).

Harmonic mode analysis of loop structures

Although solvent entropy effects are implicitly included in the continuum solvent model, the energetic evaluation of the generated conformers does not include possible differences in conformational flexibilities of the energy-minimized structures. However, the preference for a given conformational sub-state (energy minimum) is determined by both the energy and conformational flexibility. The harmonic mode method was used to get an estimate of the conformational flexibility (or vibrational entropy contribution) of the low-energy structures and its influence on the

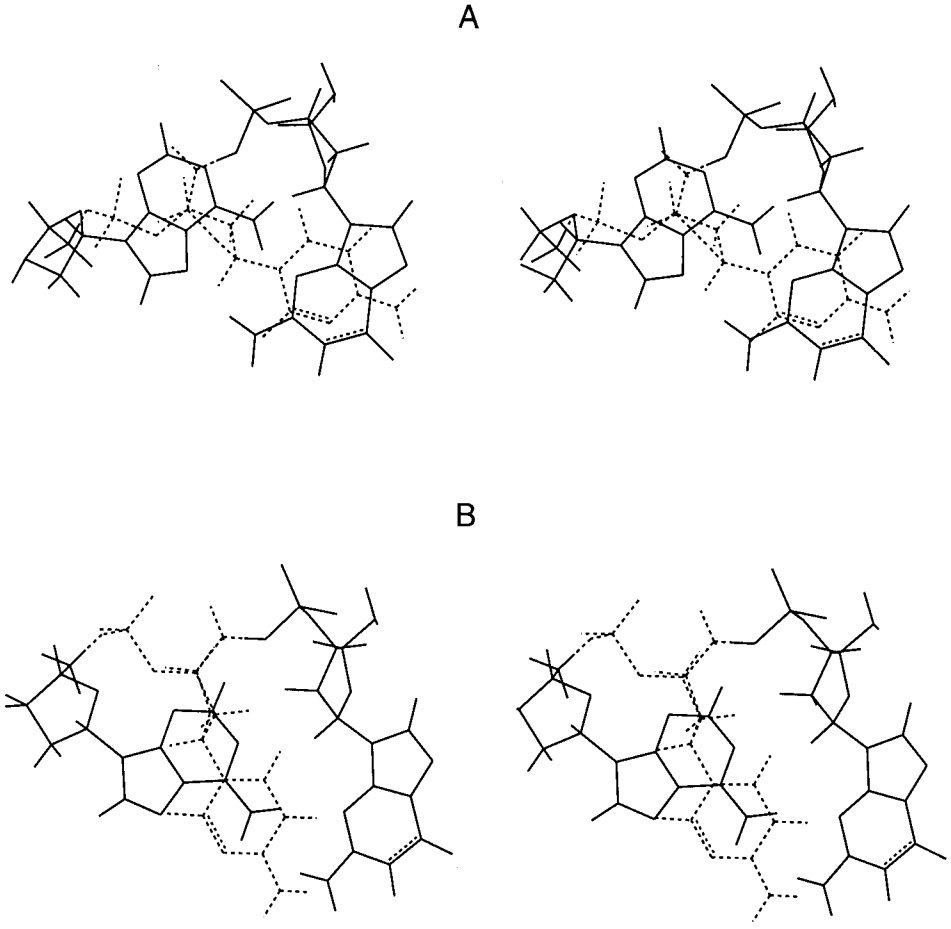

FIGURE 6 Comparison of two pairing schemes found for the G:A loop closing base pair (*bold line*; *dashed line*: central cytosine loop nucleotide) in the case of the conformational search on the GCA-hairpin loops. The (stereo) view is along the helical axis of the stem helix. For clarity, the stem helix is not shown. In the case of the form with the lower calculated energy (*A*) termed class 1a (see Table 1) the closing base pair forms a sheared G:A base pair. The higher-energy class 1c conformers (*B*) are characterized by a hydrogen bond between the guanine $H1N_2$ and the adenine N_4.

conformer ranking. The conformational entropy was calculated based on the backbone torsion angle (co)variances obtained from a harmonic mode analysis. Contributions to the conformational entropy due to fluctuations in bond length or bond angles have not been considered. These hard degrees of freedom may make a substantial contribution to the total conformational entropy within one sub-state; however, it is assumed in the present study that these contributions depend to a much lesser degree on the hairpin conformation than the softer torsion angle fluctuations. For the compact low-energy structures (in Table 1) the free energy differences attributed to conformational entropy differences between the sub-states were found to be on the order of 1–2 kcal mol^{-1}. On average, inclusion of these contributions seems to lower the free energy differences between the various conformers. However, no general rule such as that higher-energy conformers are always more flexible than lower-energy conformers was found (Table 1). Although conformational entropy contributions of 1–2 kcal mol^{-1} can in principle significantly affect conformational equilibrium at room temperature, in the present case it does not change the ranking of the classes of low-energy hairpin loop structures (Table 1). Fig. 8 indicates that the pattern of atomic position fluctuations along the loop sequence can vary between the different low-energy structures depending on the number of contacts made by each nucleotide with other parts of the structure. Conformers of the same class show qualitatively similar atomic fluctuation patterns along the sequence (compare Fig. 8, *A* and *B*) whereas for conformers belonging to different classes there are also qualitative differences in the calculated atomic position variances.

Low-energy conformations for the AAG-hairpin loop

The AAG-loop sequence may also form a hairpin structure, although of lower stability compared with, for example, GNA-loops (Yoshizawa et al., 1997). A conformational search on this sequence for which no experimental structure is available indicates a lowest-energy backbone torsion angle pattern that slightly differs from the one found for the other three loop sequences (Table 3). However, the stacking arrangement for the lowest-energy conformers appears to be similar to the GCA-, AAA-, and AGC-loops. The closing A:G base pair (see Fig. 9) contains a hydrogen bond between the nitrogen, N_3, atom of the adenine (first loop nucleotide) and the hydrogen atom, HN_1, of the third loop

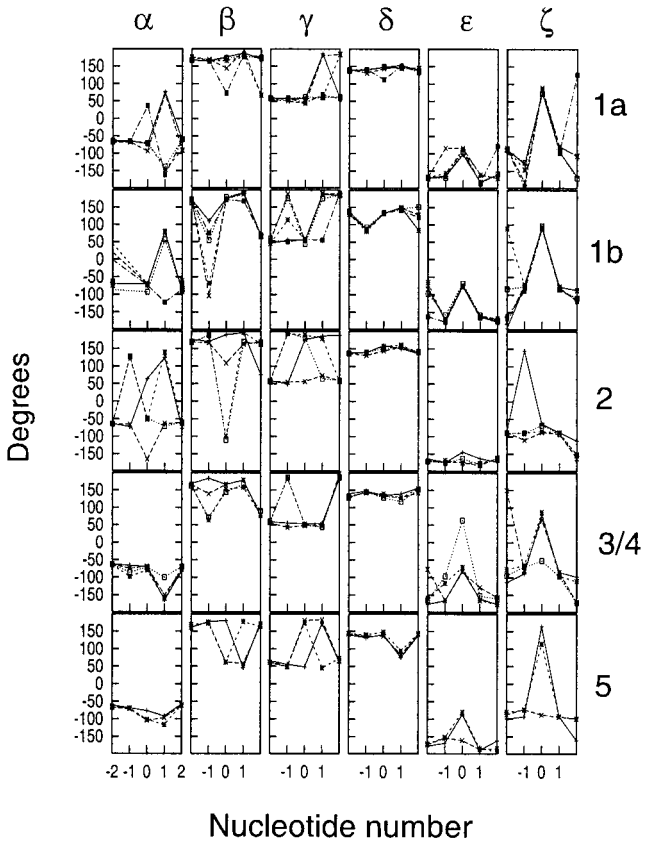

FIGURE 7 Backbone torsion angle distribution around the hairpin loop for three to five representative AAA-hairpin loop conformations (represented by different line types) belonging to class 1a, 1b, 2, 4, and 5. The nucleotide number is given with respect to the central loop nucleotide (0).

nucleotide (G). Additional contacts (distance < 2.3 Å) are made between the adenine HC_2 atom and the guanine O_6 atom and between the sugar O_4' atom of the first loop nucleotide and the amino group of the last loop nucleotide (G).

DISCUSSION

The development and improvement of methods for a realistic conformational analysis and structure prediction of nucleic acids is important for a better understanding of the mechanism of structure formation and energetic origins of conformational preferences. Ultimately, such methods could be very valuable for the design of stable structural motifs and ligands that interfere with the biological function of DNA or RNA molecules. Preferably, computational studies on nucleic acids should include surrounding water molecules and ions explicitly. However, such approaches presently do not allow systematic conformational searches on structural motifs in nucleic acids because the energetic evaluation of each generated conformer requires the equilibration of an explicit solvent and ion atmosphere, which is beyond current computational capabilities. Continuum

TABLE 2 Ranking of AAA-hairpin loops using pairwise electrostatics

Rank	Class	Structure	ΔE_{totPB}	$\Delta E_{\varepsilon=4.0}$	$\Delta E_{\varepsilon=4r}$	$\Delta E_{\varepsilon\text{sigmo}}$
1	1a	2171	0.0	0.0	0.0	0.0
2	1a	1933	0.2	5.1	4.6	4.0
3	1a	1387	0.5	4.2	2.3	0.1
5	1a	1131	1.2	3.7	3.2	2.8
8	1b	2084	3.5	4.7	2.2	1.8
9	2	3417	3.6	5.4	3.8	6.7
21	3	2155	5.7	1.6	−3.5	−1.5
29	6	1451	9.4	10.6	−11.0	8.6
43	5	3881	8.4	13.4	9.5	12.2
53	4	1967	10.3	10.5	9.4	7.8
56	3	2515	11.1	13.4	−1.9	−3.5
>200		2828	26.7	−3.2	−6.8	−5.4

$\Delta E_{\varepsilon=4.0}$, $\Delta E_{\varepsilon=4r}$, $\Delta E_{\varepsilon\text{sigmo}}$ = total energy (see Materials and Methods) without reaction field contribution and using an effective dielectric constant of either $\varepsilon = 4.0$, $\varepsilon = 4r$ or a sigmoidal distance-dependent function (Lavery et al., 1995): $\varepsilon(r) = \varepsilon_{\text{wat}} (\varepsilon_{\text{wat}} - \varepsilon_{\text{o}}) \{(sr)^2 + 2(sr) + 2\} \exp(-sr)$, $\varepsilon_{\text{wat}} = 78.0$, $s = 0.365$, $\varepsilon_{\text{o}} = 2.0$. Phosphate group charges were scaled by 0.5. These parameters were found to yield reasonable energy-minimized structures. Structure ranking is given according to the FDPB model.

(mean-field) solvent models offer an attractive alternative to approximately include solvation effects during energy minimization and conformational search. In contrast to several previous systematic search studies (Maier et al., 1999; Zacharias and Sklenar, 1997, 1999) the present approach includes electrostatic solvation effects during energy minimization and in addition accounts for the conformational flexibility of generated conformers within a harmonic approximation.

By using the GB solvation model it has been possible to select realistic hairpin loop conformations as those of lowest energy out of a large set of sterically possible conformers for three sequences with experimentally known structures. This is an important result because it is a prerequisite to use approaches such as the present force-field-based modeling attempt for structure prediction on small nonhelical motifs in DNA and RNA. Molecular dynamics simulations on nucleic acids have also shown that the GB solvation model results in trajectories that are similar to explicit solvent simulations (Tsui and Case, 2000; Williams and Hall, 1999, 2000a,b). However, MD simulations even in the nanosecond regime usually cover only conformations close to the start structure. In the present study a variety of conformations have been considered that span different helical topologies of loop nucleotides.

Although the reaction-field energies calculated by the GB method showed overall a very reasonable correlation to those calculated with the principally more accurate FDPB approach, residual differences between the two methods lead to some reordering of the conformer ranking (within each class) when applying the latter approach to the final structures. The ranking obtained with the FDPB approach appeared to be in slightly better agreement with the exper-

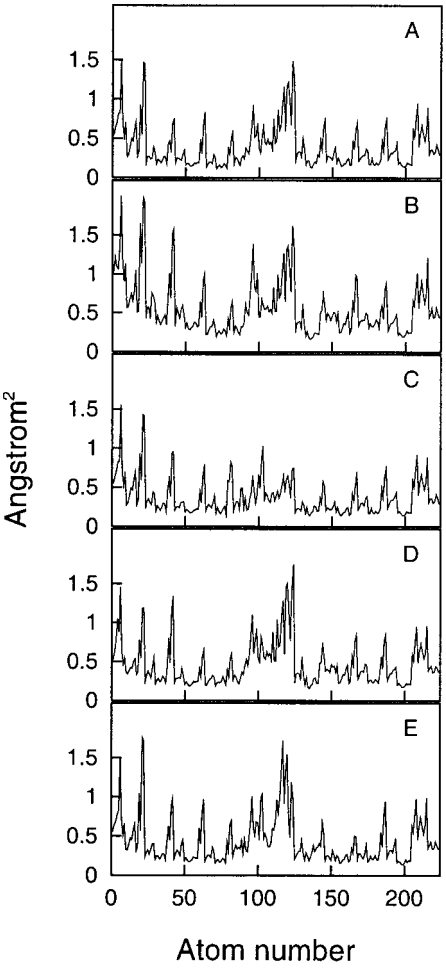

FIGURE 8 Comparison of calculated heavy atom position fluctuations for five AAA-loop sub-states based on harmonic mode analysis: (*A*) class 1a conformer 1147; (*B*) class 1a conformer 1387; (*C*) class 1b conformer 2084; (*D*) class 2 conformer 3417; (*E*) class 3 conformer 2155.

TABLE 3 Backbone torsion angles (in degrees) for low-energy loop structures

N	Structure	α	β	γ	δ	ε	ζ
AAA loop*							
A_{-1}	Exp.	−68	172	62	143	−176	−126
	2171	−66	165	56	138	−166	−127
	1387	−69	168	52	133	−90†	177†
A_0	Exp.	−81	159	80	147	−147	97
	2171	−71	175	57	146	−101	78
	1387	−88	146	43	143	−86†	86
A_{+1}	Exp.	88	−142	−173	146	177	−91
	2171	77	−178	−175	153	180	−95
	1387	72	−173	−177	151	−159†	−78
AAG-loop							
A_{-1}	1131	−63	169	54	142	−161	−141
A_0	1131	−66	170	54	146	−91	74
G_{+1}	1131	−99	168	59	146	−159	−81
G_{+2}	1131	−93	60	179	87	−173	−87

*Numbering is with respect to central nucleotide. Experimental (Exp.) data are from Chou et al. (1996).
†Significant differences between torsion angles in calculated and experimental structures.

imental result than the ranking obtained using the GB method. Similar to observations in previous conformational search studies on nucleic acids, the inclusion of electrostatic solvation contributions significantly improves conformational searches on DNA hairpin structures compared with simpler treatments of electrostatic interactions.

The number of hairpin loop conformations explicitly considered in the present study is considerably larger than in previous studies (Erie et al., 1993; Maier et al., 1999). In these studies, possible loop conformations are generated by combining a set of fixed (rigid) nucleotide conformations. A loop closing condition composed of distance and angle conditions between atoms of the helical stem and loop nucleotides is then used to select a given combination of nucleotide conformations. The number of discrete nucleotide conformations and the loop closing conditions can limit the number of accepted conformers so that important loop conformations might be overlooked. For example, only eight structures out of a large set of conformer combinations were found for the AAA trinucleotide loop to fulfill a set of distance and angle criteria to form a hairpin loop (Erie et al., 1993) compared with several thousand in the present study. The use of torsion angle window constraints during structure generation allows an adjustment of the backbone to form a closed loop during energy minimization. It shows that many backbone torsion angle combinations can result in similar helical arrangements of nucleotides. This can complicate conformational searches and energy-based structure prediction approaches. For example, the lowest-energy structures of the present searches belong to class 1a. However, the generated pool of structures does not contain only low-energy class 1a conformers but also many energetically quite unfavorable class 1a structures. For a correct ranking of the various structural families it is therefore critical not only to generate many different classes or clusters of conformers but also to generate many different conformers within one class to identify low-energy backbone torsion angle combinations.

A possible solution to this problem is to split the conformational search into one for globally distinct helical or topological arrangements of nucleotides and a subsequent efficient search for the best possible backbone combination for each arrangement. The generation of globally distinct structures could, for example, be achieved by using a conformational generator based on rigid substructures such as the MCSYM program (Major et al., 1991; Gautheret et al., 1993).

Most present systematic conformational search studies consider only the energy of generated conformations (Maier et al., 1999; Zacharias and Sklenar, 1999; Ayadi et al., 1999). Depending on the model energy function, this may at

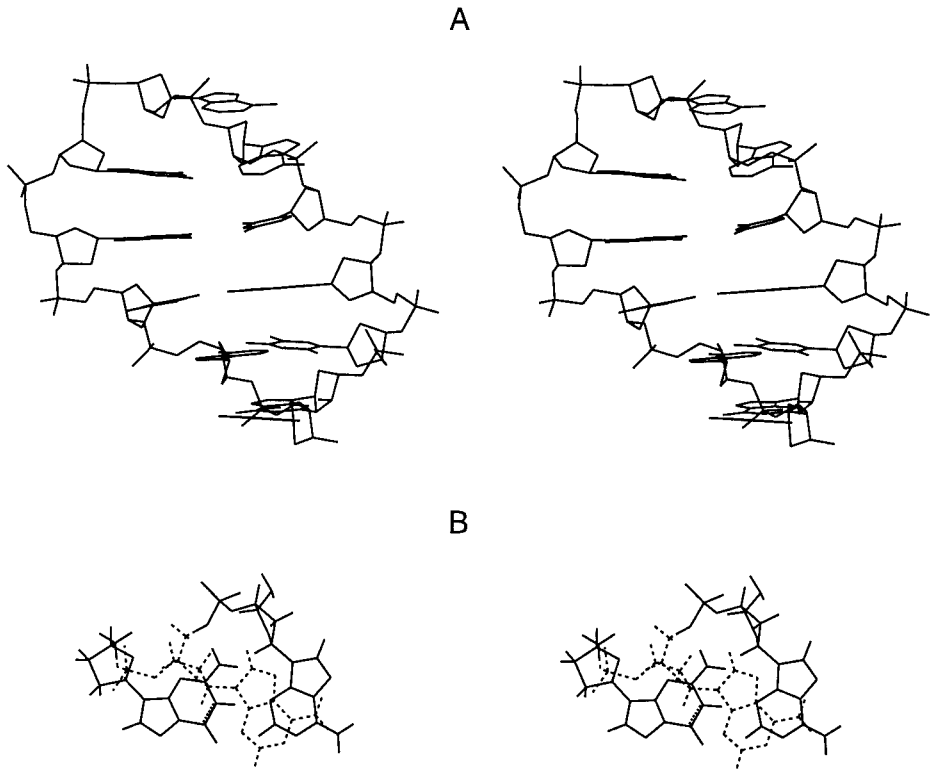

FIGURE 9 Structure of lowest-energy structure obtained from the conformational search on the GTACAAGGTAC-loop sequence. (*A*) Stereo view into the major groove; (*B*) Stereo view along the helical axis of the stem; for clarity, only the loop closing A:G base pair (*solid line*) and the central nucleotide (*dashed line*) are shown.

least in principle include solvent entropy effects; however, possible differences in the conformational flexibility of generated conformers around a stable state are often neglected. In the present study the harmonic mode method was used to estimate conformational flexibilities of the generated sub-states. An advantage of the harmonic mode method is that it is relatively rapid (~5 min computer time) and that it does not depend on an appropriate sampling to achieve convergence such as Monte Carlo or molecular dynamics simulations. However, anharmonic effects of the potential energy surface around a stable state are not included. Nevertheless, a recent comparison of harmonic mode calculations and molecular dynamics simulations on regular RNA indicated that the calculated directions of largest mobility can be quite similar (Zacharias, 2000b). Even if flexibility is accounted for using the present harmonic approximation, the ranking for the three loop sequences does not change, and the free energy gap between the lowest-energy class 1 structure and other structural classes remains at ~2–4 kcal mol^{-1}. This difference is considerably larger than the thermal energy at room temperature and suggests that other than class 1 conformations are represented by less than 5%. The relatively small influence of conformational flexibility on sub-state ranking might be due to the fact that in the present case all low-energy structures have a relatively compact fold that overall limits conformational fluctuations.

Based on the successful conformational search on the three loop sequences with experimentally known structures a search was also performed on the AAG-loop sequence for which no experimental structure is available. In a comparison of all possible DNA trinucleotide loop sequences, Yoshizawa et al. (1997) found for this sequence a higher than average melting temperature, suggesting the possibility of loop formation. The calculated structure of lowest energy is similar to a class 1 type, but the lowest-energy backbone torsion angle pattern differs from the other loop sequences of the present study, and the loop closing A:G base pair differs from the sheared G:A base pair found for the GCA-loop.

It should be emphasized that with the present approach it is possible to calculate an energetic ranking of various hairpin loop conformations. However, it is not possible to get an estimate of the absolute stability of a given hairpin loop (or free energy of formation) because this requires calculation of the free energy of the unfolded (single-stranded) state. It is difficult to calculate an accurate stability of the unfolded state because it requires a summation over a very large number of sterically possible states. Even estimates based on a dominant state approximation are complicated by the fact that it is, for example, not clear if the unfolded state is dominated by stacked or at least partially stacked conformers or completely unstacked nucleotides.

The development of efficient conformational analysis approaches with a realistic energetic ranking may also have important implications for studies of ligand binding to nucleic acids and in turn for the rational design of inhibitors for DNA and RNA function. Several structural studies on RNA-ligand complexes indicate that the RNA conformation

can change significantly upon complex formation (Patel, 1999; Cusack, 1999). A prerequisite for modeling of RNA-ligand complexes is therefore not only a reasonable treatment of intermolecular interactions including solvent effects but also to properly account for RNA conformational adaptation and change in conformational flexibility. It is straightforward to apply the present approach to complexes of ligands and nucleic acid structural motifs by applying conformational searches in the absence and presence of putative ligands.

The calculated low-energy hairpin loop structures can be downloaded at http://www.imb-jena.de/www_tbp (see instructions in http://www.imb-jena.de/www_tbp/structures.html).

I thank Dr. F. Pineda for helpful discussions. This work was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG, ZA 153/3–1).

REFERENCES

Aslani, A-A., O. Mauffret, F. Sourgen, S. Neplaz, R. G. Maroun, E. Lescot, G. Tevanian, and S. Fermandjian. 1996. The hairpin structure of a topoisomerase II site DNA strand analysed by combined NMR and energy minimization methods. *J. Mol. Biol.* 263:776–788.

Astell, C. R., M. B. Chow, and D. C. Ward. 1985. Sequence analysis of the termini of virion and replicative forms of minute virus of mice DNA suggests a modified rolling hairpin model for autonomous parvovirus replication. *J. Virol.* 54:171–177.

Ayadi, L., C. Coulombeau, and R. Lavery. 1999. Abasic sites in duplex DNA: molecular modeling of sequence-dependent effects on conformation. *Biophys. J.* 77:3218–3226.

Cornell, W. D., P. Cieplak, C. I. Bayley, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Cadwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.

Chen, X., S. V. Santhana Mariappan, P. Catasti, R. Ratliff, R. K. Moyzis, A. Laayoun, S. S. Smith, E. M. Bradbury, and G. Gupta. 1995. Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. *Proc. Natl. Acad. Sci. U.S.A.*. 92:5199–5203.

Chou, S.-H., Y.-Y. Tseng, and S.-W. Wang. 1999a. Stable sheared A C pair in DNA hairpins. *J. Mol. Biol.* 287:301–313.

Chou, S.-H., Y.-Y. Tseng, and B.-Y. Chu. 1999b. Stable formation of a pyrimidine-rich loop hairpin in a cruciform promoter. *J. Mol. Biol.* 292:309–320.

Chou, S.-H., L. Zhu, Z. Gao, J.-W. Cheng, and B. R. Reid. 1996. Hairpin loops consisting of single adenine residues closed by sheared A:A or G:G pairs formed by DNA triplets AAA and GAG: solution structures of the d(GTACAAAGTAC) hairpin. *J. Mol. Biol.* 264:981–1001.

Cusack, S. 1999. RNA-protein complexes. *Curr. Opin. Struct. Biol.* 9:66–73.

Dickerson, R. E., M. Bansal, C. R. Calladine, S. Diekmann, W. N. Hunter, O. Kennard, E. von Kitzing, R. Lavery, H. C. Nelson, W. K. Olson, W. Saenger, Z. Shakked, H. Sklenar, D. M. Soumpasis, C. S. Tung, A. H.-J, Wang, and V. B. Zhurkin. 1989. Definitions and nomenclature of nucleic acid structure parameters. *J. Mol. Biol.* 205:787–791.

Erie, D. A., A. K. Suri, K. J. Breslauer, R. A. Jones, and W. K. Olson. 1993. Theoretical predictions of DNA hairpin loop conformations: correlations with thermodynamic and spectroscopic data. *Biochemistry.* 32:436–454.

Gacy, A. M., G. Goellner, N. Juranic, S. Macura, and C. T. McMurray. 1995. Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell.* 81:533–540.

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

Glucksmann-Kuis, M. A., X. Dai, P. Markiewicz, and L. B. Rothman-Denes. 1996. *E. coli* SSB activates N4 virion RNA polymerase promoters by stabilizing a DNA hairpin required for promoter recognition. *Cell.* 84:147–154.

Glucksmann-Kuis, M. A., C. Malone, P. Markiewicz, and L. B. Rothman-Denes. 1992. Specific sequences and a hairpin structure in the template strand are required for N4 virion RNA polymerase promoter recognition. *Cell.* 70:491–500.

Hawkins, G. D., C. J. Cramer, and D. G. Truhlar. 1996. Parameterized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* 100:19824–19839.

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

Hirao, I., G. Kawai, S. Yoshizawa, Y. Nishimura, Y. Ishido, K. Watanabe, and K. Miura. 1994. Most compact hairpin-turn structure exerted by a short DNA fragment, d(GCGAAGC) in solution: an extraordinarily stable structure resistant to nuclease and heat. *Nucleic Acids Res.* 22: 576–582.

Hirao, I., Y. Nishimura, Y. Tagawa, K. Watanabe, and K. Miura. 1992. Extraordinarily stable mini-hairpins: electrophoretical and thermal properties of the various sequence variants of d(GCGAAAGC) and their effects on DNA sequencing. *Nucleic Acids Res.* 20:3891–3896.

Jayaram, B., Sprous, D., and D. L. Beveridge. 1998. Solvation free energies of biomacromolecules: parameters for a modified generalized Born model consistent with the Amber force field. *J. Phys. Chem.* 102: 9571–9576.

Jucker, F. M., and A. Pardi. 1995. GNRA tetraloops make a U-turn. *RNA.* 1:219–222.

Karplus, M., and J. N. Kushick. 1981. Method for estimating the configurational entropy of macromolecules. *Macromolecules.* 14:325–332.

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

Levitt, M. A., C. Sander, and P. S. Stern. 1985. Protein normal-mode dynamics: trypsin inhibitor, crambin, ribonuclease and lysozyme. *J. Mol. Biol.* 181:423–447.

Madura, J. D., M. E. Davis, R. Wade, B. A. Luty, A. Ilin, A. Anosiewicz, M. K. Gilson, B. Bagheri, L. Ridgway-Scot, and J. A. McCammon. 1995. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics program. *Comput. Phys. Commun.* 91:57–95.

Maier, A., H. Sklenar, H. F. Kratky, A. Renner, and P. Schuster. 1999. Force field based conformational analysis of RNA structural motifs: GNRA tetraloops and their pyrimidine relatives. *Eur. Biophys. J.* 28: 564–573.

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

Mitas, M., A. Yu, J. Dill, and I. S. Haworth. 1995. The trinucleotide repeat d(CGG)15 forms a heat-stable hairpin containing G^{syn}:G^{anti} base pairs. *Biochemistry.* 34:12803–12811.

Patel, D. 1999. Adaptive recognition in RNA complexes with peptides and protein modules. *Curr. Opin. Struct. Biol.* 9:74–87.

Qui, D., P. S. Shenkin, F. P. Hollinger, and W. C. Still. 1997. The GB/SA continuum model for solvation: a fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem.* 101:3005–3014.

Rohs, R., and H. Sklenar, R. Lavery, and B. Röder. 2000. Methylene blue binding to DNA with alternating GC base sequence: a modeling study. *J. Am. Chem. Soc.* 122:2860–2866.

Saenger, W. 1984. Principles of Nucleic Acids Structure. Springer Verlag, New York.

Sandusky, P., E. W. Wooten, A. V. Kurochkin, T. Kavanaugh, W. Mandecki, and E. R. P. Zuiderweg. 1995. Occurrence, solution structure and

stability of DNA hairpins stabilized by a GA/CG helix unit. *Nucleic Acids Res.* 23:4717–4725.

Sharp, K. A., and B. Honig. 1990. Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *J. Phys. Chem.* 94: 7684–7692.

Sitkoff, D., K. A. Sharp, and B. Honig. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 98:1978–1988.

Srinivasan, J., J. Miller, P. A. Kollman, and D. A. Case. 1998. Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dyn.* 16:671–682.

Srinivasan, J., M. W. Trevathan, P. Beroza, and D. A. Case. 1999. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor. Chem. Acc.* 101:426–434.

Still, W. C., A. Tempczyk, R. C. Hawley, and T. Hendrikson. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112:6127–6129.

Tsui, V., and D. A. Case. 2000. Molecular dynamics simulations of nucleic acids with a generalized Born solvation model. *J. Am. Chem. Soc.* 122:2489–2498.

van Dongen, M. J. P., M. M. W. Mooren, E. F. A. Willems, G. A. van der Marel, J. H. van Boom, S. S. Wijmenga, and C. W. Hilbers. 1997. Structural features of the DNA hairpin d(ATCCTA-GTTA-TAGGAT): formation of a G-A base pair in the loop. *Nucleic Acid Res.* 25: 1537–1547.

Williams, D. J., and K. B. Hall. 1999. Unrestrained stochastic dynamics simulations of the UUCG tetraloop using an implicit solvation model. *Biophys. J.* 76:3192–3205.

Williams, D. J., and K. B. Hall. 2000a. Experimental and theoretical studies of the effects of desoxyribose substitutions on the stability of the UUCG tetraloop. *J. Mol. Biol.* 297:251–265.

Williams, D. J., and K. B. Hall. 2000b. Experimental and computational studies of the G[UUCG]C RNA tetraloop. *J. Mol. Biol.* 297:1045–1061.

Yoshizawa, S., G. Kawai, K. Watanabe, K. Miura, and I. Hirao. 1997. GNA trinucleotide loop sequences producing extraordinary stable DNA minihairpins. *Biochemistry.* 36:4761–4767.

Yu, A., J. Dill., and M. Mitas. 1995. The purine-rich trinucleotide repeat sequences d(CAG)15 and d(GAC)15 form hairpins. *Nucleic Acids Res.* 23:4055–4057.

Zacharias, M. 2000a. Simulation of the structure and dynamics of nonhelical RNA motifs. *Curr. Opin. Struct. Biol.* 10:307–311.

Zacharias, M. 2000b. Comparison of molecular dynamics and harmonic mode calculations on RNA. *Biopolymers.* 54:547–560.

Zacharias, M., and H. Sklenar. 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.

Zacharias, M., and H. Sklenar. 1999. Conformational analysis of single-base bulges in A-form DNA and RNA using a hierarchical approach and energetic evaluation with a continuum solvent model. *J. Mol. Biol.* 289:261–275.

Zacharias, M., and H. Sklenar. 2000. Conformational deformability of RNA: a harmonic mode analysis. *Biophys. J.* 78:2528–2542.

Zakrzewska, K. A. Madami, and R. Lavery. 1996. Poisson-Boltzmann calculations for nucleic acids and nucleic acid complexes. *Chem. Phys.* 204:263–269.

Zhu, L., S. H. Chou, and B. R. Reid. 1995. Structure of a single-cytidine hairpin loop formed by the DNA triplet GCA. *Nat. Struct. Biol.* 2:1012–1017.

Zhu, L., S. H. Chou, and B. R. Reid. 1996. A single G-to-C change causes human centromere TGGAA repeats to fold back into hairpins. *Proc. Natl. Acad. Sci. U.S.A.* 93:12159–12164.