Theoretical Predictions of DNA Hairpin Loop Conformations: Correlations with Thermodynamic and Spectroscopic Data†

Dorothy A. Erie,‡ Asif K. Suri, Kenneth J. Breslauer, Roger A. Jones, and Wilma K. Olson*

Department of Chemistry, Rutgers, The State University of New Jersey, New Brunswick, New Jersey 08903

Received August 11, 1992

ABSTRACT: A computational procedure for generating conformations of DNA hairpin loop structures from a broad range of low-energy starting states is described. The starting point of the modeling is the distribution of oligonucleotide chain conformations obtained from Monte Carlo simulations of feasible dinucleotide steps. Structures which meet the spatial criteria for hairpin loop formation are selected from the distributions and subsequently minimized using all-atom molecular mechanics. Both d(CT$_n$G) and d(CA$_n$G) oligomers, where n = 3, 4, or 5, are modeled. These sequences are chosen because of the large number of published NMR and thermodynamic studies on DNA hairpins containing thymine or adenine residues. The minimized three-dimensional hairpin loop structures are compared with one another as well as analyzed in terms of available experimental data. The computational approach provides the first detailed analysis of DNA hairpin loop structure in terms of a multistate conformational model. Investigation of the minimized conformations reveals several interesting structural features. First, hairpin loops of the same sequence adopt several distinctly different conformations, as opposed to minor variants of the same equilibrium structure, that could potentially interconvert in solution. Second, in contrast to double-helical nucleic acids, the hairpin loop models exhibit hydrophobic and hydrophilic surfaces. The different disposition of hydrophobic groups in loops versus duplexes could modulate both protein–nucleic acid interactions and nucleic acid self-associations. Third, perpendicular aromatic interactions of loop residues are observed in many of the computed hairpins. This sort of interaction might be important in the stabilization of non-hydrogen-bonded nucleic acid secondary and tertiary structures. The predicted structural features in the models help, in addition, to account for the unusual thermodynamic properties of DNA hairpin loops. Comparison of the theoretically-generated NOEs in different structures further reveals that very different molecular structures and interactions can, in principle, produce the same NOEs. The multistate description suggested by this observation differs from the conventional interpretation of DNA solution structure in terms of the fluctuations about a single preferred chain conformation. There is not necessarily only one set of closely related structures consistent with the observed data.

The unique three-dimensional structures associated with single-stranded DNA may play a significant role in biological processes. The hairpin and/or cruciform structures formed by palindromic sequences, for example, occur at or near loci of biological function (Streisinger et al., 1966; Müller & Fitch, 1982). The potential importance of the single-stranded loops in such structures has stimulated a variety of physical and chemical studies (Haasnoot et al., 1983, 1986; Marky et al., 1983; Hilbers et al., 1985; Wemmer et al., 1985; Hare & Reid, 1986; Ikuta et al., 1986; Orbons et al., 1986; Roy et al., 1986; Erie et al., 1987, 1989; Chattopadhyaya et al., 1988, 1990; Pramanik et al., 1988; Senior et al., 1988; Wolk et al., 1988; Xodo et al., 1988, 1991; Blommers et al., 1989, 1991; Germann et al., 1990; Paner et al., 1990, 1992; Rentzeperis et al., 1991; Boulard et al., 1991; Amaratunga et al., 1992; Doktycz et al., 1992; Kang et al., 1992). Comprehension of the data accumulated in these experiments calls for realistic models of the single-stranded loops. The large conformational flexibility of the nucleic acid backbone together with the spatial constraints required for hairpin loop formation, however, makes this a difficult theoretical problem. Not surprisingly, only a few models of nucleic acid hairpins based on first principles have appeared in the literature (Marky & Olson, 1982, 1987; Haasnoot et al., 1986; Garcia et al., 1988; Harvey et al., 1988).

As part of a program to understand the thermodynamic and structural consequences of incorporating single-stranded DNA fragments into double-stranded chains, we have carried out computer modeling studies of DNA hairpins using a combination of Monte Carlo chain simulations and molecular mechanics energy minimizations. We use the Monte Carlo methods to generate representative conformations of single-stranded DNA oligonucleotides of arbitrary sequence (Erie, 1990; Erie et al., 1992). We then select the conformations that meet the three-dimensional spatial criteria for hairpin loop formation and minimize the resulting oligomeric structures using all-atom molecular mechanics. We study the d(CT$_n$G) sequence, where n = 3, 4, or 5, in an effort to comprehend recent thermodynamic and spectroscopic data on DNA dumbbells, i.e., double-helical DNA with two symmetrically spaced single-stranded thymine-containing hairpin loops (Erie et al., 1987, 1989). In addition, we model the corresponding d(CA$_n$G) fragments in order to investigate the effect of base sequence on DNA loop structure. We choose the T$_n$ and A$_n$ sequences because of the large number of published NMR and thermodynamic studies on DNA hairpins containing thymine or adenine residues (Haasnoot et al., 1983, 1986; Hilbers et al., 1985; Hare & Reid, 1986; Ikuta et al., 1986; Pramanik et al., 1988; Senior et al., 1988; Wolk et al., 1988; Xodo et al., 1988, 1991; Blommers et al., 1989, 1991;

† This work was funded by USPHS Grants GM20861, GM23509, GM31483, and GM34469 and by the Charles and Johanna Busch Memorial Fund of Rutgers University.

* To whom correspondence should be addressed.

‡ Present address: Institute of Molecular Biology, University of Oregon, Eugene, OR 97403.

0006-2960/93/0432-436$04.00/0

Germann et al., 1990; Paner et al., 1990; Rentzeperis et al., 1991; Boulard et al., 1991). We analyze the minimized secondary structures of the theoretically-predicted hairpin loops in terms of the available experimental data and compare the various models against one another. To our knowledge, the computational approach taken here provides the first detailed structural analysis of DNA hairpin loop structure in terms of multistate conformational alternatives. Unlike conventional energy minimization approaches, there is no preconceived notion of the overall folding of the DNA introduced in these calculations. The input data are simply a set of low-energy dinucleotide steps consistent with the observed conformational properties of small molecules. The present study is complementary to the mononucleotide "buildup" and optimization schemes introduced in two recent NMR investigations of DNA hairpin loop structure (Boulard et al., 1991; Blommers et al., 1991). The structures predicted here offer new insights into the spectroscopic and thermodynamic properties of nucleic acid hairpin molecules in solution.

METHODS

We take advantage of the distributions of conformations of short single-stranded oligonucleotides recently generated with Monte Carlo methods from the constituent mononucleotides (Erie, 1990; Erie et al., 1992). From the distributions we select chain conformations with the potential to form hairpin loop structures and then refine these structures by all-atom potential energy minimization. We investigate the d(CT$_n$G) and d(CA$_n$G) sequences, where n = 3, 4, or 5, selecting and refining the potential hairpin loop structures using the procedures outlined below.

Monte Carlo Chain Generation. A distribution of 10^6 oligonucleotide chains is generated using a Monte Carlo buildup procedure. Specific chain conformations are constructed from energetically favorable arrangements of the constituent mononucleotides. The individual dinucleotide steps are identified using a torsion angle minimizer (Erie et al., 1992). The glycosyl and acyclic backbone torsions of the dimers are allowed to vary, while the sugar rings are held fixed in one of the two preferred puckered forms (C2′- or C3′-endo). A total of 108 conformationally distinct states per dimer are considered in this first stage of minimization. The acyclic torsion angles within 5 kcal/mol of the global minimum of the resulting optimized states are then allowed to vary by ±10° in an effort to estimate the breadth of the different local minima. The energies of a total of 2187 (3^7) angle combinations are examined per local conformational minimum. Finally, the energies of the dinucleotide states are scaled so that the populations of C2′-endo and C3′-endo sugar puckers match those found in NMR solution studies. This last step is necessitated by limitations in the theoretical methods to predict DNA sugar puckering accurately (Erie et al., 1992). The conformer populations of the individual acyclic torsion angles in the composite dimer ensembles are found to be in good agreement with the distributions of backbone conformations deduced from NMR coupling constants and the frequencies of glycosyl conformations in X-ray crystal structures, suggesting that the low-energy states are reasonable. The low-energy dimer forms (consisting of 150–325 conformational states per dimer step) are next used as variables in a Monte Carlo algorithm which generates the conformations of single-stranded d(CX$_n$G) chains, where X = A or T and n = 3, 4, or 5. Chain configurations are constructed by overlapping the common terminal sugar–base ends of successive dimer units (e.g., the Ys of dXpdY and dYpdZ dimers). Fortunately, the abrupt changes in the sugar–base geometry associated with this approximation have limited effect on the local "fold" of the chain backbone (Erie et al., 1992). The chains are built sequentially from the 5′-end using random numbers to select the conformations of the overlapping dinucleotide units. Since the conformations of the dinucleotide starting states appear to be quite reasonable, the conformations of the oligonucleotides generated from these states are also thought to be reasonable.

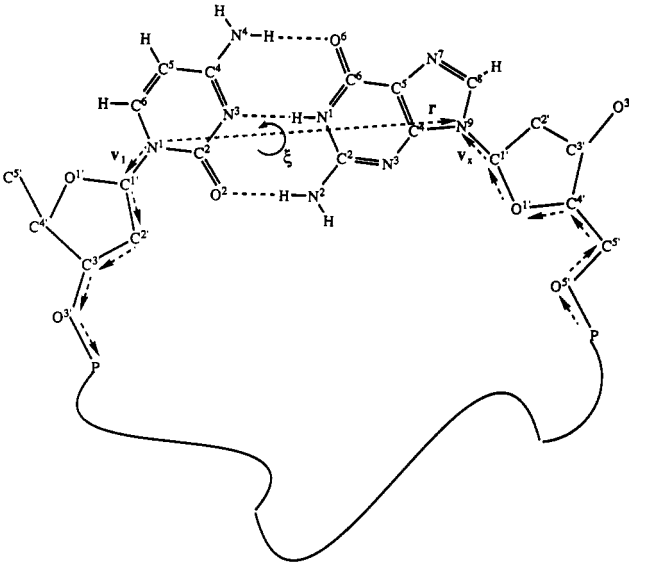

FIGURE 1: Schematic illustration of a DNA hairpin loop highlighting the CG base pair which closes the structure, the first and last bond vectors $\mathbf{v}_1$ and $\mathbf{v}_x$ of the chain model, the end-to-end vector $\mathbf{r}$, and the pseudo torsion angle ξ. The dashed arrows indicate the direction of the chain backbone.

Selection of Chains for Hairpin Formation from Monte Carlo Distributions. Oligonucleotide chains with the potential to form hairpin loops must adopt geometries that accommodate base pair formation between the first and last residues of the chain. To satisfy this constraint in the d(CX$_n$G) fragments, three requirements must be met. First, the end-to-end vector $\mathbf{r}$ connecting the N1 atom of the 5′-terminal C and the N9 atom of the 3′-terminal G must have the proper magnitude and direction for standard Watson–Crick base pairing. Second, the first and last chemical bonds of the fragment (i.e., N1–C1′ of the 5′-terminal C and C1′–N9 of the 3′-terminal G) must exhibit the necessary angular orientation for base pairing. Third, the resulting terminal CG base pair must be planar.

The parameter $\mathbf{r}$ (see Figure 1) is calculated from the chemical bond lengths, valence angles, and torsion angles of the DNA fragment using standard methods of polymer chain statistics (Flory, 1969; Erie, 1990; Erie et al., 1992). A coordinate system is introduced for this purpose with the origin at N1, the x-axis lying along the N1–C1′ glycosyl bond of the 5′-terminal residue (i.e., the first chemical bond vector of the fragment), and the y-axis in the plane determined by atoms C2, N1, and C1′. The positive direction of the y-axis is chosen to make an acute angle with the C2 to N1 chemical bond vector, and the z-axis is positioned to make a right-handed coordinate system. The coordinates, in angstrom units, of the N1–N9 end-to-end vector of a standard Watson–Crick base pair (Arnott & Hukins, 1972) in this reference frame are (−5.5, −7.2, 0.1). In the search for hairpin structures, each component of the observed end-to-end vector $\mathbf{r}$ is constrained to values within ~10% of this reference.

The angular orientation requirement for base pairing is determined by the scalar product Λ of unit vectors along the first and last bond vectors $\mathbf{v}_1$ and $\mathbf{v}_x$ of the chain fragment (Figure 1). The angle between these bond vectors in a standard Watson–Crick base pair is approximately 103°. Consequently, in this study Λ is constrained to values between 0.0 and –0.7, corresponding to angles between 90° and 134°. The third constraint for loop formation is treated at a later stage in the modeling (see below).

Identification of hairpin structures also requires a reliable estimate of nucleic acid conformational preferences. Limitations in the energy minimization of the constituent dinucleotides (Erie, 1990; Erie et al., 1992) may exclude hairpin loop structures from Monte Carlo samples of oligonucleotide conformations. The 5 kcal/mol cutoff energy previously used to identify conformations accessible to individual dinucleotide steps in the Monte Carlo simulations of d(CX$_n$G) oligomers is very stringent and is thought to rule out several potentially important states with energies only sightly higher than this arbitrary limit (Erie, 1990; Erie et al., 1992). The local bending of these states is critical for the formation of the overall 180° turn of the hairpin loop. The energy maps for the dinucleotide residues containing thymines appear to be incomplete since no such structures are found in the Monte Carlo samples of d(CT$_n$G) fragments. The limited flexibility at the thymine dimer steps prevents the oligomers from meeting the displacement and orientational constraints (e.g., $\mathbf{r}$ and Λ) of loop closure (see discussion of Table I below). Aside from the arbitrary cutoff which eliminates several chain reversal and bent conformations expected to populate hairpin loops, it is well-known that single-stranded thymine chains are highly flexible, adopting nonhelical structures (Cantor & Schimmel, 1980; Saenger, 1984) and frequently occurring in DNA hairpin loops (Haasnoot et al., 1983, 1986; Hilbers et al., 1985; Hare & Reid, 1986; Ikuta et al., 1986; Erie et al., 1987, 1989; Pramanik et al., 1988; Senior et al., 1988; Wolk et al., 1988; Xodo et al., 1988, 1991; Blommers et al., 1989; Germann et al., 1990; Paner et al., 1990; Rentzeperis et al., 1991). The available NMR structures of thymine-containing hairpins reveal a number of different folding arrangements (Haasnoot et al., 1986; Hare & Reid, 1988; Senior et al., 1988; Blommers et al., 1991; Boulard et al., 1991), although the data are consistently interpreted in terms of the fluctuations about a single equilibrium structure. The few available crystallographic examples of single-stranded thymine-tract DNAs, including two different hairpin structures with four thymine residues (Chattopadhyaya et al., 1988, 1990; Kang et al., 1992), are also suggestive of the potential mobility of the single-stranded loops. Aside from the very different spatial arrangements adopted in the two solid-state hairpin structures, the temperature factors of the thymine bases are significantly larger than those of the bases in the attached helical stems.

Although the Monte Carlo simulations produce very few potential hairpin loop conformations for d(CT$_n$G) sequences (Erie, 1990; Erie et al., 1992), they do reveal a number of potential hairpin structures for adenine-containing chain fragments (Table I). To study a broader range of thymine loop conformations, d(CT$_n$G) structures have been constructed using the d(CA$_n$G) backbones found by Monte Carlo sampling to meet the displacement and orientational criteria for hairpin loop formation. This approximation is based on the following arguments: (1) several of the dCpdT, dTpdT, and dTpdG dimer steps between 5 and 10 kcal/mol of their global minima are similar to the 0–5 kcal/mol states of dCpdA, dApdA, andApdG, respectively; (2) the predicted conformations of dinucleotides containing adenine residues are similar to those found in nucleic acid structures determined by NMR and X-ray crystallography; (3) because single-stranded adenine-containing chains tend to be more stacked than single-stranded thymine-containing chains in solution (Cantor & Schimmel, 1980; Saenger, 1984), the T-containing chains may be equally, if not more, flexible at the chain backbone level; (4) thymine residues are found to be more flexible than adenine residues in NMR investigations of hairpin loops containing four unpaired bases (Senior, 1988; Senior et al., 1988; Blommers et al., 1989). The use of the A-containing backbone conformations for constructing T-containing hairpin structures is further justified by the fact that the accepted Monte Carlo chains are also subjected to molecular mechanics minimization which relieves unfavorable local conformations. In fact, a few of the backbone torsion angles in the preliminary hairpin structures undergo significant changes upon minimization (see Results).

Initial Chain Generation. The exocyclic backbone torsion angles, the sugar ring conformations, and the glycosyl torsion angles are recorded for the Monte Carlo chains that meet the first two criteria for hairpin formation (i.e., $\mathbf{r}$, Λ). The torsion angle information is then used in conjunction with a set of standard bond lengths and valence bond angles [detailed in Erie et al. (1992)] and the base sequence to generate the full coordinates of the d(CA$_n$G) and d(CT$_n$G) fragments.

Since the Monte Carlo method of chain generation does not account for the long-range excluded volume of the molecules, some of the chains that meet the loop formation criteria may adopt conformations with atoms in distant parts of the chain in steric conflict. The nonbonded contacts of all atoms in the preliminary set of hairpin loop arrangements have therefore been checked. Since each residue is locally minimized, only the nonbonded contacts between atoms in different residues are determined. In addition, because a limited number of such nonbonded contacts can usually be relieved by energy minimization (Lipton & Still, 1988), all chains with fewer than 15 unacceptable nonbonded contacts are retained. For simplicity, the hard-sphere nonbonded contact limit is taken to be 2.0 Å for all atom parts.

The chains that meet all of the above criteria must conform to a third constraint which assures that the terminal Watson–Crick base pair and the glycosyl bonds of the bases forming the base pair lie approximately in the same plane. This requirement can be met by limiting the value of the pseudo torsion angle ξ, defined by the orientation of the initial (N1–C1′) and terminal (C1′–N9) glycosyl bonds with respect to the N1–N9 end-to-end vector $\mathbf{r}$, to values within ±45°. This constraint is determined by the relationship

$$\cos \xi = \frac{(\mathbf{v}_1 \times -\mathbf{r})\cdot(\mathbf{v}_x \times \mathbf{r})}{|(\mathbf{v}_1 \times -\mathbf{r})||(\mathbf{v}_x \times \mathbf{r})|} \geq \frac{\sqrt{2}}{2} \qquad (1)$$

where $\mathbf{v}_1$ and $\mathbf{v}_x$ are the initial and terminal bond vector, respectively, and $\mathbf{r}$ is the end-to-end vector (Figure 1).

The chains that conform to the above three constraints for hairpin formation have been visually inspected on an Evans and Sutherland PS390 graphics workstation using the DOCK molecular modeling package (Stodola et al., 1988). Representative examples are chosen among chains that adopt similar three-dimensional conformations. Chains in which loop residues intersect the duplex stem of the hairpin are deemed unacceptable and are not simulated in these studies.

Oligomer Minimization. The unfavorable nonbonded contacts that may exist in the spatially acceptable Monte Carlo hairpin structures are relieved with a standard all-atom

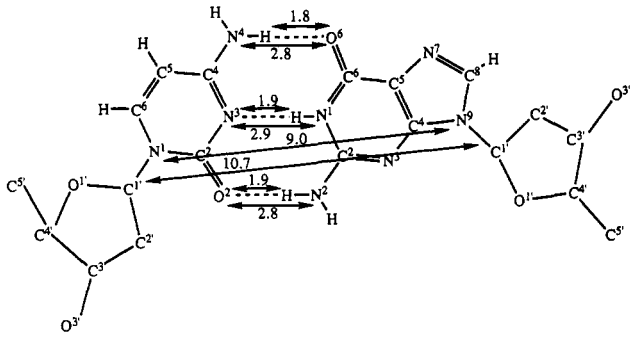

FIGURE 2: Diagram of the terminal CG base pair of a d(CX$_n$G) fragment showing the eight distance constraints employed in energy minimization.

molecular mechanics minimization procedure. Prior to energy minimization, the glycosyl torsion angles of the thymine residues are constrained to the anti range (i.e., the 2-carbonyl of the ring is pointed away from the pentose; see below for torsion angle conventions). Syn conformations of pyrimidines (where the carbonyl is directed above the sugar ring) are sterically forbidden by hard-sphere contacts arising from the choice of fixed bond lengths and valence angles in the Monte Carlo structures (Erie et al., 1992). Accordingly, any glycosyl torsions found to be in the syn conformation in the starting adenine residues are rotated by 180° in the thymine units. All morphologically unique chains with reasonable starting structures are then minimized in Cartesian space using the IMPACT software package (Bassolino et al., 1988) with the "all-atom" parameter set of Weiner et al. (1986). Distances between the base pair hydrogen-bonding atoms as well as those between the terminal C1′···C1′ and N1···N9 atom pairs are constrained to ideal values to preserve chain closure during minimization. To limit propeller twisting, both heavy atom and hydrogen atom distances of the terminal bases are constrained. An ideal CG base pair meeting these distance constraints is depicted in Figure 2. The minimized structures resulting from the energy calculations are relieved of all unfavorable nonbonded contacts.

After minimization, the roll, tilt, and twist angles and the vectorial translations (e.g., shift, slide, and rise) between successive bases are determined and then used in the conformational analysis of hairpin loop structures (Babcock et al., 1989; Babcock & Olson, 1992). The propeller twist angle between the planes of the base pair which closes the loop also is calculated. The role and tilt angles are components of the net bend angle between adjacent residues, the former describing the bending about the long axis and latter the bending about the short pseudodyad axis of a Watson–Crick base pair. In single-stranded DNA, there are no such base pairs; local coordinate axes can, nevertheless, be determined from the position that would be ideally occupied by the complement to a given base (see below). The net bending angle is determined by the scalar product of the normals to the bases. The twist angle is defined as the rotation of a base with respect to the normal vector of its predecessor, while the vectorial translations are the spatial displacements of the coordinate origins of consecutive bases within the assumed reference frame. Shift is the displacement in the x-direction, slide the displacement in the y-direction, and rise the displacement in the z-direction (see below). The coordinate systems are chosen in accordance with the guidelines set at the 1988 EMBO workshop on nucleic acid structure (Diekmann, 1988, 1989; Sarma, 1988; Dickerson, 1989a,b). The origin is positioned along the dyad axis of a theoretical base pair, with the x-axis directed into the major groove and the y-axis parallel to and displaced by 2.2 Å from the N1···N9 vector of an ideal base pair. The positive direction of the y-axis is chosen to point toward the reference base, and the z-axis (i.e., the normal) is defined to form a right-handed coordinate system.

Table I: Number of d(CX$_n$G) Monte Carlo Chains Meeting Specified Hairpin Loop Closure Criteria[a]

	chain closure criteria[b]			
sequence	r	r	r + Λ	r + Λ + ξ + S
d(CA$_3$G)	212 461	607	40	8
	160 892	609	31	
d(CT$_3$G)	114 202	0	0	0 (8)[c]
	56 052	69	0	
d(CA$_4$G)	223 602	384	25	15
	166 318	362	28	
d(CT$_4$G)	65 653	9	7	6 (36)[c]
	18 338	30	0	
d(CA$_5$G)	244 135	3513	813	11
	203 716	2797	355	
d(CT$_5$G)	25 064	6	0	0 (10)[c]
	3 010	5	3	

[a] Total sample size for each chain fragment is 10^6. The first entry for each sequence refers to oligomers built by discarding the 3′-end (e.g., dY) of the 5′-dXpdY-3′ dimer steps and the second entry to oligomers built by discarding the 5′-end (e.g., dX). [b] r = distance only; r = volume element; r + Λ = volume element + orientation; r + Λ + ξ + S = volume element + orientation + hydrogen bonding + steric. [c] Thymine-containing loop models constructed using potential hairpin folds identified on the basis of r + Λ in the Monte Carlo samples of the corresponding d(CA$_n$G) fragments in parentheses.

The torsion angles of the chain backbone, the glycosyl torsion linking the sugar and base, and the phase angle and amplitude of the sugar ring pseudorotation in the minimized structures are also determined. The backbone torsion angles are defined as 0° in the planar cis conformation and are assigned positive values for right-handed rotations. Angles in the C4′–C3′–O3′–P–O5′–C5′–C4′ bond sequence are defined in terms of two conventions: one (Sundaralingam, 1969; Sundaralingam et al., 1973) which emphasizes the chemistry and structural symmetry of the sugar–phosphate backbone (ψ', ϕ', ω', ω, ϕ, ψ) and the other (Seeman et al., 1976) which focuses on the directionality of the P···P units (α, β, γ, δ, ϵ, ζ). The glycosyl torsion angle χ is measured with respect to cis arrangements of the O1′–C1′–N9–C8 atoms in purines and the O1′–C1′–N1–C6 atoms in pyrimidines. The anti conformation is defined where $\chi = 30 \pm 90°$ and the syn conformation where $\chi = 210 \pm 90°$. [It should be noted that χ is sometimes defined with respect to the trans arrangements of these atoms (Saenger, 1984), resulting in angular values that differ by 180°.] The pseudorotation parameters of the sugar rings are determined using the standard equations of Altona and Sundaralingam (1972).

RESULTS AND DISCUSSION

In the sections that follow, the theoretically-predicted hairpin loops are presented. The structures are compared against one another and, when possible, with respect to structures proposed from NMR experiments. The computer-generated structures also are discussed in terms of available thermodynamic data.

Monte Carlo Chains Prior to Minimization. The number of chains from the Monte Carlo samples that meet the preliminary criteria for hairpin loop formation (i.e., r, Λ, ξ, and long-range steric contacts) is listed as a function of base sequence in Table I. Data are presented for oligomers constructed with two chain buildup routines (Erie et al.,

1992): one by discarding the 3′-residues (e.g., dY) of successive 5′-dXpdY-3′ dimer steps and the other by discarding the 5′-residues (e.g., dX). The necessity for building thymine-containing loops from the d(CA_nG) fragments is obvious from the number of accepted structures. Interestingly, the fraction of chains that meet the hairpin loop criteria ($\sim 10^{-5}$) is consistent with the expected loop closure probabilities of short DNA chains (Scheffler et al., 1968). Similar numbers of chains are accepted for the three- and five-membered loops of both sequences; however, there are more four-membered loops accepted than either three- or five-membered loops. Furthermore, although the same dimeric building blocks are employed, there are twice as many T_4 loops as A_4 loops found to meet the geometric criteria for hairpin loop formation.

Visual inspection of the selected chains reveals that several of the potential hairpin conformations are sterically unacceptable and others are not unique. In fact, only one structural class of each of the three- and five-membered loops and two classes of each of the four-membered loops are found to be both sterically acceptable and morphologically unique. Consequently, only one representative conformation from each class has been chosen for minimization. The structures which are not reported are related to those presented below by small fluctuations of individual chain torsions. Not unexpectedly (since they are derived from the same starting structures), the chain conformations with optimum hairpin geometries in each morphological class are identical for the A_n and T_n loops. The thymine-containing chains, however, can adopt one arrangement sterically forbidden to the corresponding adenine-containing sequences (see below). The larger number of total hairpin structures for the T- versus A-containing loops in Table I also reflects these steric effects. For simplicity, the loops are discussed in terms of a single representative structure rather than as an ensemble of closely related fluctuating states.

Loop Structures. Color-coded space-filled representations of the energy-refined hairpin loop structures are presented in Figures 3–5. Four views, differing by consecutive rotations of 90° about the vertical axis of the figure, are shown for each structure (see legend to Figure 3). The backbone carbon atoms are colored yellow, the aromatic carbon atoms white, the phosphorus atoms violet, the nitrogen atoms blue, and the oxygen atoms red. In addition, superimposable black and white wire frame representations of the loops are shown, in stereo, in Figures 6–8.

Since the methods employed in this study generate the full coordinates of the hairpin loops, the distances between protons that are close enough to produce NOE cross peaks can be calculated and the resulting spatial connectivity patterns determined. The theoretical NOEs derived below are based on an outer proton···proton cutoff distance of 5 Å in the energy-minimized d(CT_nG) and d(CA_nG) sequences. The theoretical connectivity patterns produced from proton–proton distances in different bases as well as from sugar H1′–base proton distances are reported in Table II. Only the most commonly observed NOE distances are included in the theoretical connectivity patterns. In addition, the roll, tilt, twist, and bend angles, the vectorial translations (e.g., shift, slide, rise) between consecutive bases, and the propeller twist of the terminal CG base pair are compiled in Table III. The exocyclic backbone torsion angles, the glycosyl rotations, and the pseudorotation parameters of the rings are listed in Table IV. The latter two sets of data are used below to describe the stacking patterns and structural features that give rise to the theoretical NOEs. The contributions to the potential energy of the different minimized loop structures are included in

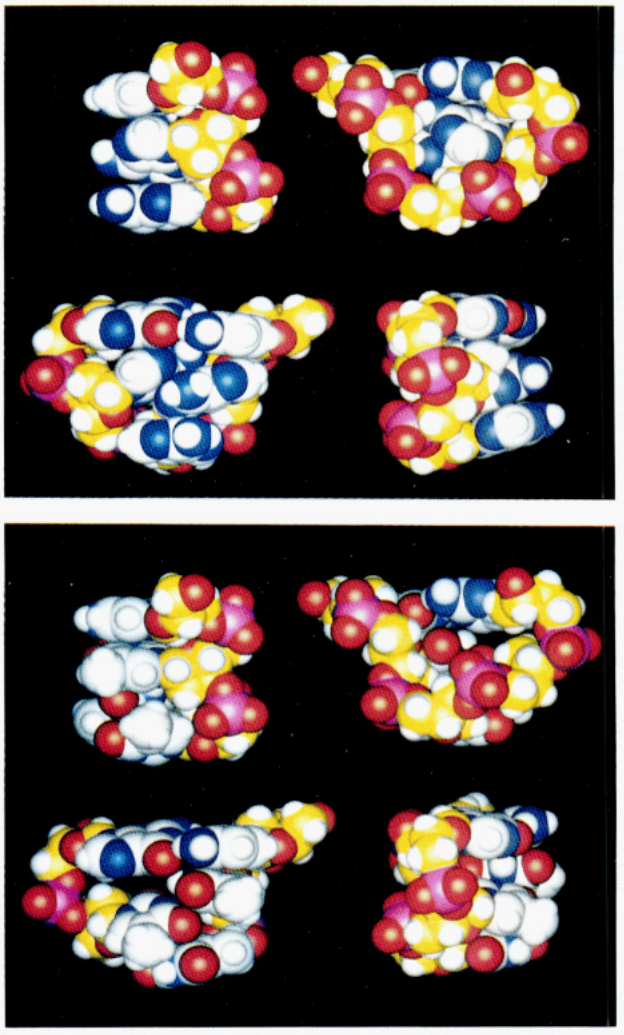

FIGURE 3: Color-coded space-filling representations of energy-minimized (a, top) d(CA_3G) and (b, bottom) d(CT_3G) hairpin loops. Four views of each molecule are shown: (i) upper left, a side view with the cytosine residue in front; (ii) upper right, a face view obtained by rotation of 90° about the vertical axis from (i) with the cytosine on the left; (iii) lower right, view rotated 180° from (i); (iv) lower left, view rotated 180° from (ii). That is, each view is rotated by 90° about the vertical axis from the previous view in a clockwise direction on the photograph. Color code: backbone C, yellow; aromatic C and H, white; P, violet; N, blue; O, red. Structures were rendered on an Evans and Sutherland PS390 graphics workstation using the DOCK molecular modeling package (Stodola et al., 1988).

Table V as estimates of feasibility. Not included in these data are potentially important interactions of the single-stranded loop with the attached helical stem and the surrounding solvent. The computer-generated structures are compared with one another and, when possible, with the structures deduced from NMR investigations (Haasnoot et al., 1983, 1986; Hare & Reid, 1986; Pramanik et al., 1988; Senior et al., 1988; Wolk et al., 1988; Xodo et al., 1988; Blommers et al., 1989; Piotto et al., 1990; Boulard et al., 1991). Atomic coordinates of the theoretically-generated hairpin loops are available as supplementary material.

Computer-Generated Hairpin Loops. (A) d(CA_3G) versus d(CT_3G). Inspection of the space-filling and stick representations in Figures 3 and 6 and the derived structural data in Table III and IV reveals that the conformations of the energy-minimized d(CT_3G) and d(CA_3G) models are fairly similar at the morphological level. In both structures, the bases in the loops are stacked parallel to the terminal base pairs that close the loops. The first two loop residues (A2, A3 or T2,

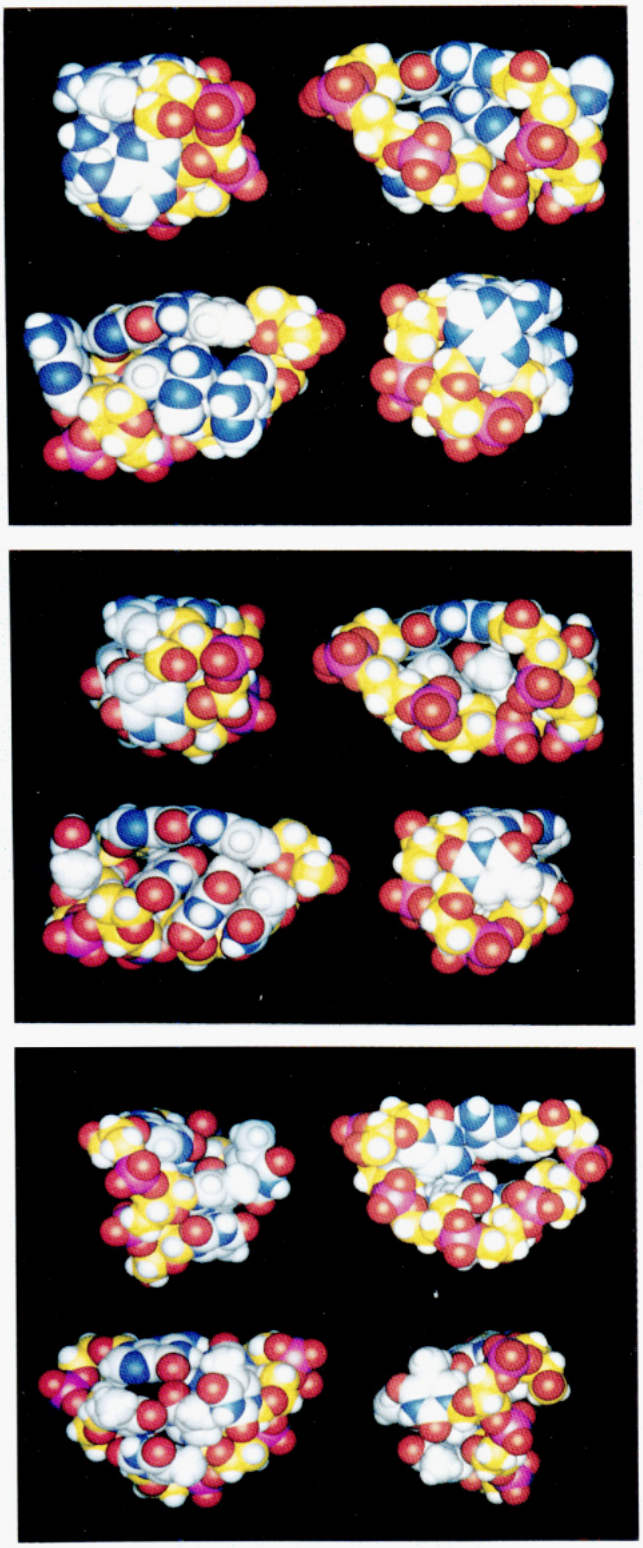

FIGURE 4: Color-coded space-filling representations of energy-minimized (a, top) d(CA_4G), (b, middle) d(CT_4G)-1, and (c, bottom) d(CT_4G)-2 hairpin loops. See legend to Figure 3.

T3) are stacked in a helical fashion on C1 (noted in boldface in Figure 6), and the turn of the loop is made between residues 3 and 4. These features are reflected in large changes in the tilt and bend angles as well as by large displacements in the base translation (compare the base morphological parameters relating residues 3 and 4 in Table III with the structure in Figures 3 and 6). The notable differences (>60°) in torsion angles and sugar puckers between the A_n and T_n loops of the same size are shown in boldface in Table IV. The structures

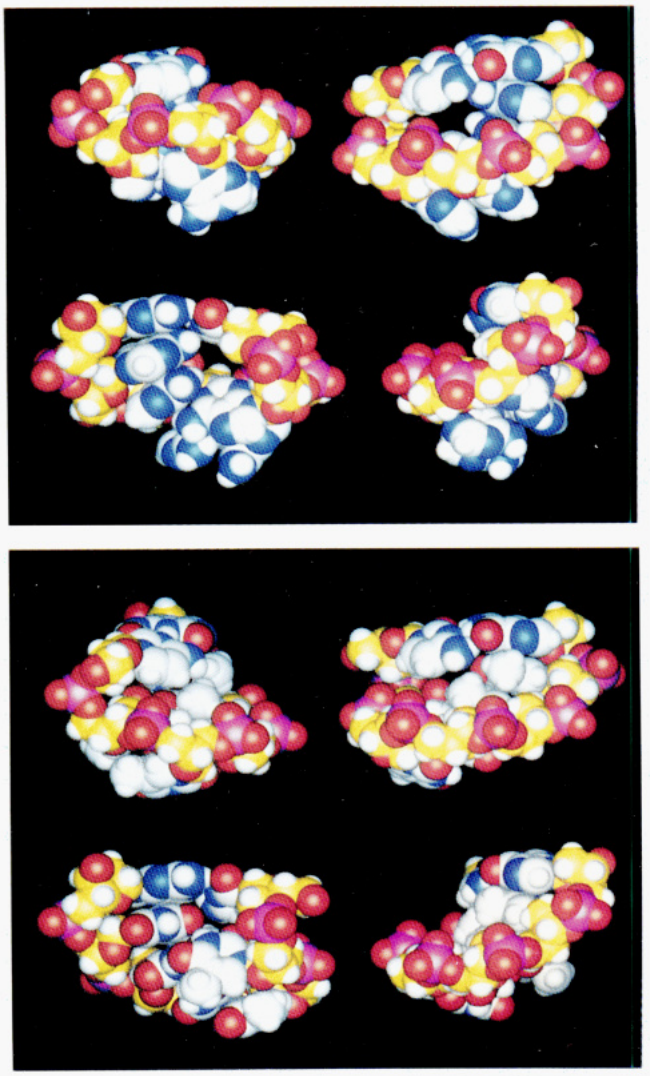

FIGURE 5: Color-coded space-filling representations of energy-minimized (a, top) d(CA_5G) and (b, bottom) d(CT_5G) hairpins loops. See legend to Figure 3.

of the A_3 and T_3 loops conform, in part, to the empirical model of loop formation proposed by Haasnoot et al. (1986) in which the helical geometry is extended from the 3′-end of the attached stem duplex [i.e., the cytosine base of the d(CX_3G) fragments studied here] by two or three bases into the hairpin with one or two bases folding across the minor groove to close the loop (at G in this example). It is important to note that the 3′-end of the double-helical stem is the start or 5′-end of the X_3 loop (i.e., 5′-XpXpX-3′). Loop residues that stack on the 3′-end of the helical stem fold in the normal 5′- to 3′-direction of the nucleic acid chain. The unpaired bases in the computed structures, however, are found to stack above the so-called β-face (Rose et al., 1980) of the terminal CG base pair rather than above the α-face as assumed in the Haasnoot et al. (1986) model. (In this convention, the atoms of the bases are given their usual IUPAC numbers with the α-face of the ring being that in which the progression from atom 1 to atom 2 to atom 3 around the ring is clockwise and the β-face that in which the progression is counterclockwise. The convention has the advantage that purines and pyrimidines in Watson–Crick base pairs have a common orientation, with their α-surfaces in the same direction, so the α-face of the base pair is unambiguously identified.) More specifically, the α-face of a Watson–Crick CG base pair is the side which shows the glycosyl linkages below the base rings and the

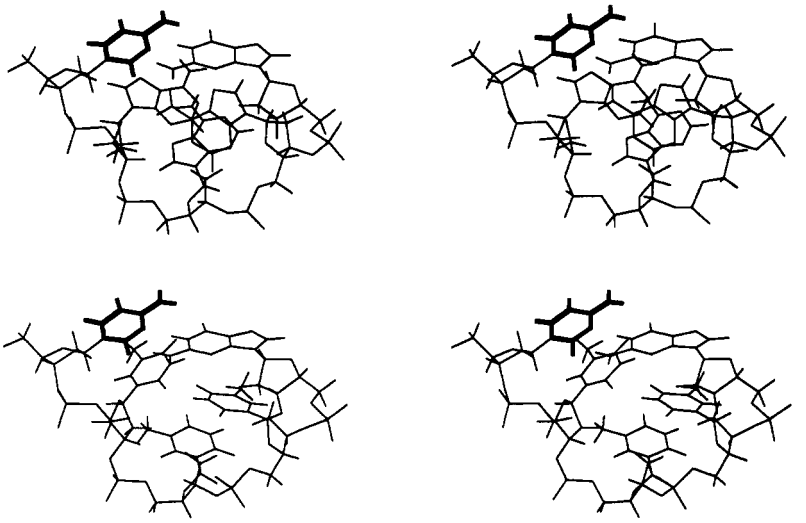

FIGURE 6: Stereo wire frame representations of (a, top) d(CA_3G) and (b, bottom) d(CT_3G) hairpin loops in a common coordinate frame embedded in residue C1 (noted in boldface).

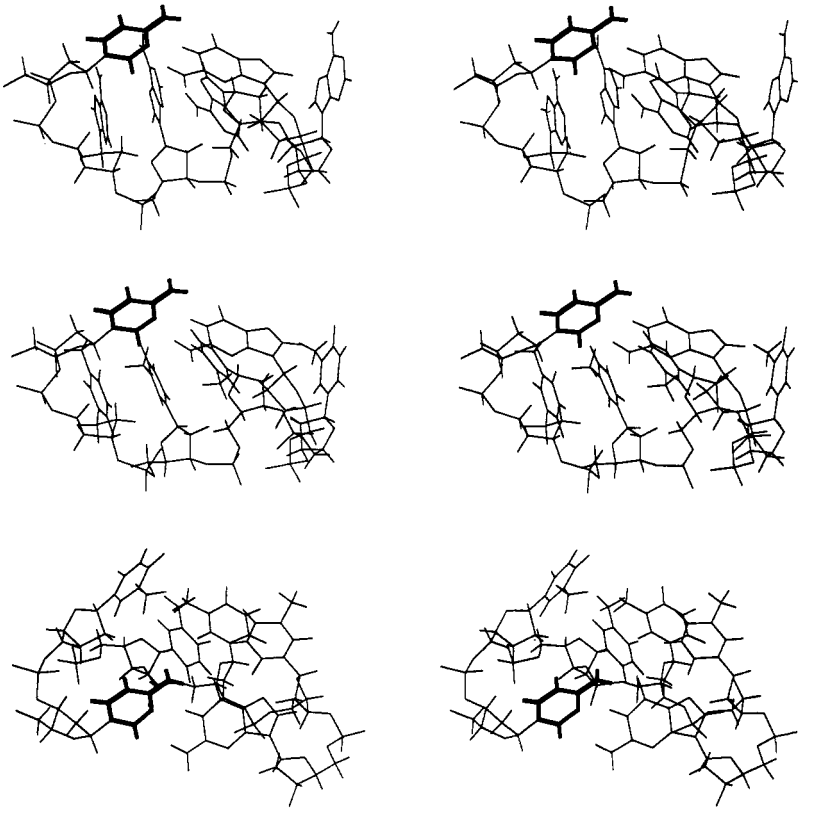

FIGURE 7: Stereo wire frame representations of (a, top) d(CA_4G), (b, middle) d(CT_4G)-1, and (c, bottom) d(CT_4G)-2 hairpin loops in a common coordinate frame embedded in residue C1 (noted in boldface).

cytosine lying to the left of the pseudodyad axis. This means that the single-stranded bases in the computed d(CX_3G) hairpins fall below the plane of the printed page if the terminal base pair is represented as C·|·G with the dots denoting the hydrogen bonding and the vertical bar the pseudodyad.

The principal morphological difference between the two d(CX_3G) structures involves tertiary base interactions. The T4 residue in d(CT_3G) partially intercalates between T2 and T3, whereas the A2 residue in d(CA_3G) partially intercalates between A3 and A4. These structural differences result in twist angles of opposite signs between residues X3 and X4 (Table III) in the two loops. The differences are manifested at the chain backbone level primarily by changes in the O3′–P (ω'/ζ) and P–O5′ (ω/α) phosphodiester backbone torsions (see boldface numbers in Table IV). These rotations alter the relative positions of the flanking sugars and bases while keeping the phosphate positions approximately fixed. A simple change in the "order" of stacking residues 2 and 4 will produce these two different conformations. For example, if A2 were "stacked" above A4 instead of below it (in the reference frame of Figures 3 and 6), then d(CA_3G) would have essentially the same base–base orientations as in d(CT_3G). If residues 3 and 4 could interconvert in this manner in solution, the resulting conformational equilibrium would significantly complicate the analysis of NOE measurements. This point is discussed further in the sections that follow.

Despite their relative overall structural similarities, d(CA_3G) and d(CT_3G) have very different theoretical NOE connectivity patterns (Table II). In d(CA_3G), there are predicted sequential aromatic base NOEs between C1···A2, A2···A3,

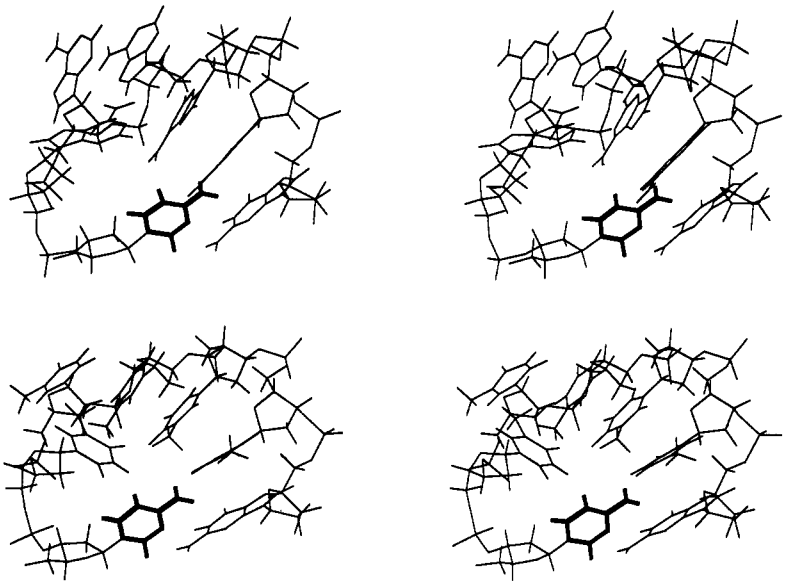

FIGURE 8: Stereo wire frame representations of (a, top) d(CA_5G), (b, bottom) d(CT_5G) hairpin loops in a common coordinate frame embedded in residue C1 (noted in boldface).

and A4···G5, and sugar–base NOEs between C1···A2 and A4···G5. In addition, there are two nonsequential theoretical base–base NOEs between A2···A4. These nonsequential NOEs are particularly interesting, as they are indicative of tertiary base–base interactions in the loop. In d(CT_3G), sugar–base and base–base NOEs are predicted only between residues C1···T2 and T2···T3 (Table II). The lack of theoretical NOEs between residues 3 and 4 in both sequences is probably a result of the reversal of chain direction at this point (see Figures 3 and 6). The only theoretical NOEs that the two structures have in common are those between the first two residues and the theoretical base–base NOE between residues 2 and 3. This lack of consistency in the theoretical NOEs is apparently due to (1) the differences in size of the A and T bases and the locations of the nonexchangeable protons on the different chemical residues and (2) the differences in base stacking between residues 2 and 4. It is therefore important to be cautious when comparing NOE connectivity patterns for different sequences, since very different theoretical NOE connectivity patterns can be consistent with qualitatively similar chain conformations.

(B) d(CA_4G) versus d(CT_4G). Prior to minimization, two distinct morphological classes of d(CA_4G) and d(CT_4G) structures appear to be acceptable. Upon minimization, however, only one of the d(CA_4G) structures is found to minimize to a reasonable final state. Two of the adenine rings in the other starting conformation intersect and disrupt the primary chemical structure of the oligomer upon molecular mechanics minimization. This problem is not found in the corresponding four-membered thymine loop [hereafter referred to as d(CT_4G)-2]. The observation is not too surprising given that adenine bases are significantly larger than thymine bases and thus are more likely to encounter steric contacts. In addition, the steric conflicts in the disallowed A_4 loop involve residues A3 and A5 with glycosyl torsions initially in the anti range. The difference in minimization of the A_4 versus T_4 loops is therefore not related to the variation of glycosyl torsions in certain thymine residues from syn to anti conformations prior to minimization (see Methods above). Since only one of the A_4 loops can be minimized, a comparison is made between this loop and the corresponding T_4 loop [hereafter referred to as d(CT_4G)-1]. Color-coded space-filling and superimposable stereo stick representations of these two loops are shown in Figures 4 and 7, respectively.

In both d(CA_4G) and d(CT_4G)-1, the loop residue stacking is perpendicular to the terminal CG base pair. Residues X2, X3, and X4 are aligned above one another in the loop in a partially disordered helical array with a local twist at the T2–T3 and T3–T4 steps of roughly 55° and the local twist at the A2–A3 and A3–A4 steps equal to 37° and 25°, respectively. In d(CA_4G), the glycosyl torsion χ of the A4 residue is syn, while in d(CT_4G)-1 the T4 χ is anti (Table IV). The theoretical NOEs are apparently a reflection of this structural difference. There are theoretical base–base and sugar–base NOEs between T3 and T4 but only a weak (>4 Å) theoretical base–base NOE between A3 and A4 (Table II). Structural differences in the chain backbone in the two loops are manifested primarily through "crankshaft" rotations about the O5′–P and C5′–C4′ bonds of residues 3 and 4 (noted in boldface in Table IV), while less significant changes (<60°) are observed in other torsion angles. These backbone differences may stem from A4 being in the syn conformation and/or from steric hindrances due to the larger adenine ring. To convert the backbone of the A_4 loop to that of the T_4 loop, the P–O5′ (ω/ζ) and C5′–C4′ (ψ/γ) torsions in both residues 3 and 4 must flip from g^-g^+ conformations to g^+g^- arrangements (Table IV). The g^+g^- angle combination of these torsions is quite rare in solid-state and solution studies (Olson, 1982), and references cited therein). This unusual conformation might possibly be stabilized in the d(CT_4G)-1 structure by the close proximity of the O2 atom of T4 and the amino groups of the terminal CG base pair as well as by perpendicular stacking interactions involving residues T3, T4, and G6 (Figures 4b and 7b). Perpendicular interactions between small aromatic rings have been previously predicted to be stabilizing features of molecular conformation (Karlström et al., 1983; Pawliszyn et al., 1984; Hunter & Sanders, 1990; Linse, 1992) as well as observed in numerous crystal structures, including those of benzene (Cox et al., 1958; Piermarni et al., 1969), several proteins (Burley & Petsko, 1985; Singh & Thornton, 1985; Fowler & Moore, 1988), various minor groove binding drug–DNA complexes (Teng et al., 1988; Larsen et al., 1989), and the hairpin loop of d(CGCGCGT$_4$CGCGCG) (Chattopadhyaya et al., 1988, 1990). In addition, two recent hairpin

Table II: Theoretical NOEs for d(CX$_n$G) Sequences[a]

d(CA$_3$G)	base–base	H6	C1	H8	A2	3.0	
		H2	A2	H8	A3	4.2	
		H2	A2	H2	A3	4.5	C1 A2 A3 A4 G5
		H8	A4	H8	G5	4.3	
	sugar–base	H1′	C1	H1′	A2	2.5	
		H1′	C1	H8	A2	3.1	C1 A2 A3 A4 G5
		H1′	A4	H8	G5	5.0	
	nonsequential	H2	A2	H8	A4	4.0	
		H2	A2	H2	A4	4.6	C1 A2 A3 A4 G5
d(CT$_3$G)	base–base	H6	C1	H6	T2	4.2	
		H6	C1	H7A	T2	3.5	C1 T2 T3 T5 G5
		H6	T2	H7A	T3	3.2	
	sugar–base	H1′	C1	H1′	T2	2.6	
		H1′	C1	H6	T2	3.6	C1 T2 T3 T5 G5
		H1′	T2	H6	T3	5.0	
		H1′	T2	H7A	T3	3.6	
d(CA$_4$G)	base–base	H6	C1	H8	A2	4.0	
		H8	A2	H8	A3	3.6	
		H2	A2	H2	A3	4.9	C1 A2 A3 A4 A5 G6
		H2	A3	H8	A4	4.2	
		H2	A5	H8	G6	3.5	
	sugar–base	H1′	C1	H8	A2	2.2	
		H1′	A2	H1′	A3	4.9	C1 A2 A3 A4 A5 G6
		H1′	A2	H8	A3	4.0	
		H1′	A5	H8	G6	4.6	
	nonsequential	H1′	C1	H8	A3	4.7	C1 A2 A3 A4 A5 G6
d(CT$_4$G)-1	base–base	H6	C1	H6	T2	4.3	
		H6	C1	H7A	T2	2.5	
		H6	T2	H6	T3	4.7	
		H6	T2	H7C	T3	2.5	
		H6	T3	H6	T4	4.9	C1 T2 T3 T4 T5 G6
		H6	T3	H7C	T4	2.8	
		H6	T4	H6	T5	4.4	
		H6	T4	H7A	T5	4.9	
		H7A	T5	H8	G6	4.9	
	sugar–base	H1′	C1	H6	T2	3.3	
		H1′	C1	H7A	T2	4.1	
		H1′	T2	H6	T3	4.3	
		H1′	T2	H7C	T3	4.7	C1 T2 T3 T4 T5 G6
		H1′	T3	H6	T4	3.9	
		H1′	T3	H7A	T4	4.4	
		H1′	T4	H7A	T5	4.3	
	nonsequential	H6	C1	H7C	T3	3.7	
		H1′	C1	H7C	T3	2.9	C1 T2 T3 T4 T5 G6
		H7B	T4	H8	G6	4.3	
d(CT$_4$G)-2	base–base	H7A	T2	H7B	T3	3.9	
		H7B	T2	H7B	T3	3.7	
		H6	T3	H6	T4	4.7	C1 T2 T3 T4 T5 G6
		H6	T3	H7A	T4	2.6	
		H7A	T3	H7A	T4	3.5	
	sugar–base	H1′	T2	H6	T3	4.2	
		H1′	T2	H7A	T3	4.2	
		H1′	T3	H6	T4	2.3	C1 T2 T3 T4 T5 G6
		H1′	T3	H7C	T4	2.5	
		H1′	T5	H8	G6	4.4	
	nonsequential	H7C	T3	H7B	T5	4.9	C1 T2 T3 T4 T5 G6
d(CA$_5$G)	base–base	H6	C1	H2	A2	5.0	
		H8	A2	H8	A3	4.8	
		H2	A2	H8	A3	4.9	
		H8	A3	H8	A4	4.6	C1 A2 A3 A4 A5 A6 G7
		H2	A3	H8	A4	4.4	
		H8	A4	H8	A5	4.4	
		H8	A5	H8	A6	4.3	
		H2	A5	H2	A6	4.9	
		H8	A6	H8	G7	4.0	
	sugar–base	H1′	A2	H8	A3	2.1	
		H1′	A3	H8	A4	3.5	
		H1′	A4	H8	A5	4.4	C1 A2 A3 A4 A5 A6 G7
		H1′	A5	H8	A6	4.3	
		H1′	A6	H8	G7	4.8	
	nonsequential	H2	A2	H8	A4	2.9	
		H2	A2	H8	A5	4.2	C1 A2 A3 A4 A5 A6 G7
d(CT$_5$G)	base–base	H6	T2	H7A	T3	2.5	
		H7C	T2	H7A	T3	4.0	
		H6	T3	H7C	T4	3.4	
		H7C	T3	H7C	T4	4.2	
		H6	T4	H6	T5	4.3	

Table II (Continued)

	H6	T4	H7A	T5	2.6	
	H7A	T4	H7C	T5	2.4	C1 T2 T3 T4 T5 T7 G7
	H6	T5	H6	T6	4.5	
	H6	T5	H7C	T6	2.5	
	H7B	T5	H7B	T6	2.3	
	H6	T6	H8	G7	3.7	
	H7A	T6	H8	G7	4.2	
sugar–base	H1′	T2	H6	T3	3.5	
	H1′	T2	H7C	T3	2.9	
	H1′	T2	H7C	T4	2.2	
	H1′	T3	H6	T4	4.2	
	H1′	T3	H7C	T4	4.6	
	H1′	T4	H6	T5	4.0	C1 T2 T3 T4 T5 T7 G7
	H1′	T4	H7A	T5	4.4	
	H1′	T5	H6	T6	3.9	
	H1′	T5	H7C	T6	4.4	
	H1′	T6	H1′	G7	4.6	
	H1′	T6	H8	G7	4.0	
nonsequential	H6	C1	H7A	T4	5.0	C1 T2 T3 T4 T5 T7 G7
	H6	T2	H7B	T4	4.4	

[a] NOEs are based on proton–proton distances less than 5 Å. Sequential base–base distances are listed first, followed by sugar–base NOEs and nonsequential NOEs. The five columns of data include the names of the interacting protons with residue numbers and the distances, in angstrom units, of the specified pairs. The spatial connectivity patterns for each type of NOE are shown to the right.

Table III: Bending and Twisting Angles and Vectorial Translations

			tilt (deg)	roll (deg)	twist (deg)	bend (deg)	shift (Å)	slide (Å)	rise (Å)	distance (Å)	propeller twist (deg)
d(CA_3G)	C1	A2	23.8	−12.4	17.2	26.7	0.4	−0.9	−5.2	5.2	
	A2	A3	−10.6	13.3	50.7	16.5	0.3	−1.5	−3.0	3.3	
	A3	A4	−164.5	23.2	−46.0	148.2	−3.4	−5.7	6.0	9.0	
	A4	G5	−17.2	−36.5	63.7	38.2	4.9	1.1	−2.5	5.6	
	C1	G5									−12.6
d(CT_3G)	C1	T2	−14.6	4.8	17.7	15.3	0.5	−0.5	−2.5	2.6	
	T2	T3	29.5	−13.8	45.9	31.7	0.9	−0.9	−7.0	7.1	
	T3	T4	−154.0	57.8	20.3	160.0	−1.6	−2.8	5.7	6.6	
	T4	G5	−13.2	8.1	26.2	15.4	−0.5	0.5	−4.2	4.3	
	C1	G5									−20.1
d(CA_4G)	C1	A2	72.4	−44.6	58.3	80.4	2.1	0.2	−8.9	9.1	
	A2	A3	1.0	2.3	36.9	2.5	1.4	−0.8	3.1	3.5	
	A3	A4	−161.7	16.8	24.9	156.7	−0.3	−2.9	7.5	8.0	
	A4	A5	44.3	37.9	71.3	54.2	1.1	−1.1	−7.8	7.9	
	A5	G6	−64.3	14.3	37.3	64.6	0.9	−2.9	6.6	7.3	
	C1	G6									−10.6
d(CT_4G)-1	C1	T2	84.2	−9.9	38.2	82.8	0.6	−0.8	−11.0	11.0	
	T2	T3	6.5	5.1	57.4	7.9	2.9	−0.6	3.0	4.2	
	T3	T4	10.0	14.4	54.0	16.9	2.6	0.2	3.0	3.9	
	T4	T5	−54.5	140.7	−79.9	123.5	−10.1	−0.1	4.2	10.9	
	T5	G6	−92.5	17.5	15.7	93.7	−0.4	−4.7	7.4	8.8	
	C1	G6									−7.4
d(CT_4G)-2	C1	T2	87.0	−131.9	41.8	146.2	8.5	−1.7	−6.5	10.8	
	T2	T3	−112.5	45.1	22.5	119.8	−2.5	−3.5	9.0	10.0	
	T3	T4	31.1	−1.1	59.7	29.7	2.5	−0.3	−6.2	6.7	
	T4	T5	−159.1	62.8	2.8	170.8	−1.4	−2.6	3.7	4.7	
	T5	G6	17.7	−34.6	12.3	38.8	−3.0	−0.5	−5.2	6.1	
	C1	G6									−33.5
d(CA_5G)	C1	A2	25.3	−109.2	16.1	111.5	6.9	4.8	1.2	8.5	
	A2	A3	43.1	−9.5	48.0	42.8	0.2	−1.0	−6.2	6.2	
	A3	A4	−10.4	8.3	32.4	13.1	−0.3	−1.9	3.6	4.1	
	A4	A5	−16.0	−5.8	57.2	16.3	3.6	−0.9	5.4	6.5	
	A5	A6	6.1	10.1	51.2	11.4	2.6	−0.2	3.0	4.0	
	A6	G7	10.9	12.4	43.9	16.1	3.1	−0.2	2.9	4.2	
	C1	G7									−16.0
d(CT_5G)	C1	T2	9.0	−95.4	−33.4	93.9	2.4	2.3	3.9	5.2	
	T2	T3	60.9	−38.1	47.1	69.5	0.6	2.2	−9.0	9.2	
	T3	T4	11.0	−14.1	44.6	17.4	0.9	−1.2	2.7	3.1	
	T4	T5	1.8	11.8	47.4	11.6	2.4	−0.7	3.4	4.2	
	T5	T6	7.4	15.1	38.9	16.5	1.5	−0.2	3.1	3.5	
	T6	G7	3.6	13.5	47.2	13.6	3.3	0.2	3.2	4.6	
	C1	G7									−24.6

loop structures deduced from NMR show at least one loop residue that lies in the minor groove perpendicular to the stem base pair (Blommers et al., 1991; Boulard et al., 1991). It is also noted that the unpaired bases in d(CA_4G) and d(CT_4G)-1 lie above the β-face of the CG base pair. As noted below, this positioning may be related to the helical character of the duplex stem.

One of the most interesting features of the d(CA_4G) and d(CT_4G)-1 loops is the position of both A5 and T5 perpendicular to the CG base pair but on the exterior of the loop,

Table IV: Exocyclic and Glycosyl Torsion Angles (in deg) for d(CX$_n$G) Sequences[a]

		ϕ'/ϵ C3′–O3′	ω'/ζ O3′–P	ω/α P–O5′	ϕ/β O5′–C5′	ψ/γ C5′–C4′	ψ'/δ C4′–C3′	χ C1–N1 (N9)
d(CA$_3$G)	C1						144.9	–4.5
	A2	–176.0	–93.7	–95.6	71.8	176.7	72.0	–104.5
	A3	–152.9	**–79.6**	**–85.1**	178.5	56.4	77.0	**–136.4**
	A4	170.0	**50.8**	**155.8**	157.9	58.3	134.5	**23.4**
	G5	158.7	**30.3**	170.8	166.0	54.0	127.9	–9.6
d(CT$_3$G)	C1						143.8	–1.1
	T2	–172.7	–91.3	–95.2	72.9	179.5	91.3	–130.8
	T3	173.3	**31.7**	**149.9**	159.1	55.2	128.9	**38.1**
	T4	–163.8	**–88.5**	**–84.2**	179.7	56.4	140.8	**–87.9**
	G5	–143.2	**–54.2**	180.0	–176.3	59.4	135.5	–2.9
d(CA$_4$G)	C1						111.9	25.1
	A2	–56.4	85.2	50.5	–176.4	167.5	132.0	46.9
	A3	–177.8	–83.1	**–58.7**	173.2	**47.3**	132.3	54.1
	A4	–175.7	–93.6	**–67.4**	176.2	**40.0**	89.1	**–112.1**
	A5	–77.7	–43.4	–174.1	155.7	64.0	136.0	46.2
	G6	–174.5	–72.9	–74.5	170.7	64.4	122.3	**–133.9**
d(CT$_4$G)-1	C1						135.7	31.7
	T2	–66.5	91.6	58.5	–179.9	169.8	146.3	42.0
	T3	–170.3	–124.2	**44.9**	147.2	**–47.6**	168.5	94.6
	T4	–178.3	–126.1	**43.4**	160.2	**–61.2**	120.4	**76.8**
	T5	–82.9	–58.7	–171.9	171.6	62.0	140.0	50.8
	G6	–175.6	–80.0	–64.9	–177.8	60.4	112.6	**–46.5**
d(CT$_4$G)-2	C1						72.4	30.0
	T2	–155.1	57.2	–136.4	–179.8	–81.6	148.5	61.8
	T3	–154.7	–87.5	–93.5	–168.9	59.3	143.3	–13.2
	T4	–179.8	–107.4	–142.3	–176.5	55.4	138.4	35.6
	T5	–172.6	–100.2	–89.3	–174.2	62.9	134.0	–120.1
	G6	–132.4	–57.7	–165.1	–168.3	60.3	136.0	23.9
d(CA$_5$G)	C1						104.0	37.7
	A2	**–73.0**	**–40.0**	**–70.6**	173.6	–64.7	159.4	–27.7
	A3	–169.2	–92.8	–155.2	–172.8	63.7	136.8	48.1
	A4	–175.1	–94.9	–81.8	–172.1	52.0	136.6	51.9
	A5	–176.1	–81.9	**–69.8**	176.3	**60.1**	128.9	46.6
	A6	–172.1	–83.2	**–77.4**	176.2	**53.6**	139.4	73.3
	G7	179.7	–83.3	–76.4	179.5	52.4	128.1	81.3
d(CT$_5$G)	C1						81.4	38.2
	T2	**–176.1**	**57.1**	**–159.9**	163.6	–26.8	171.8	–35.1
	T3	–162.5	–144.6	–150.4	–172.3	65.4	145.4	24.7
	T4	–168.6	–85.3	–89.7	–179.0	62.7	143.5	69.0
	T5	–176.6	–125.2	**42.5**	158.2	**–47.4**	166.1	89.2
	T6	–175.1	–127.3	**42.3**	161.8	**–52.1**	166.1	99.9
	G7	172.5	–86.9	–73.6	176.0	58.8	128.7	73.0

[a] The torsion angles are listed in degrees with the planar cis conformation defined as 0°. The first torsion angle in the chain is ψ' of the 5′ cytosine (C1). Notable differences in torsion angles between the A$_n$ and T$_n$ loops of the same size are highlighted in boldface.

Table V: Energy Components of Optimized d(CX$_n$G) Loop Structures

	DNA loop						
energy[a] (kcal/mol)	d(CA$_3$G)	d(CT$_3$G)	d(CA$_4$G)	d(CT$_4$G)-1	d(CT$_4$G)-2	d(CA$_5$G)	d(CT$_5$G)
bond	2.285	2.489	3.286	4.399	4.113	3.518	4.628
angle	23.765	23.017	26.303	33.533	27.683	29.614	37.773
dihedral	55.743	55.867	62.634	66.603	69.460	69.948	79.024
LJ-1,4	25.460	24.240	30.625	29.215	29.078	36.132	34.066
EL-1,4	–14.093	–12.520	–16.446	–14.320	–14.440	–19.002	–16.399
LJ	–52.663	–44.040	–62.006	–46.069	–42.652	–73.895	–61.183
EL	3.005	5.677	8.364	6.631	4.262	3.057	2.129
HB	–1.940	–1.904	–1.699	–1.590	–1.715	–1.846	–1.690
E(total)	41.563	52.825	51.062	78.404	75.790	47.526	78.347
rms grad	1.007	0.904	1.344	1.158	1.949	0.900	0.838

[a] Energy contributions are based on the following standard potential energy terms (Bassolino et al., 1988): bond = bond length (l) stretching = $\sum_{\text{bonds}} K_l(l - l_0)^2$; angle = valence angle (θ) bending = $\sum_{\text{angles}} K_\theta(\theta - \theta_0)^2$; dihedral = intrinsic torsional (ϕ) potential $\sum_{\text{dihedrals}} K_\phi[1 + \cos(n\phi - \gamma)]$; LJ = Lennard-Jones 6–12 potential = $\sum_{i<j}(A_{ij}/r_{ij}^{12} - B_{ij}/r_{ij}^{6})$; EL = electrostatic potential = $\sum_{i<j}(q_iq_j/\epsilon r_{ij})$; HB = hydrogen bonding = $\sum_{\text{H-bonds}}(C_{ij}/r_{ij}^{12} - D_{ij}/r_{ij}^{10})$; rms grad = root-mean-square gradient = $(1/3N)\langle \nabla E_{\text{total}} \cdot \nabla E_{\text{total}} \rangle$.

as if a three-membered loop were more favorable. This unusual loop structure is very similar to the recently reported solution structure of the T$_3$A hairpin in which the second thymine of the loop is external and folded into the minor groove of the attached helical stem (Blommers et al., 1991). The external positioning of X5 bases in the current models produces weak theoretical NOEs between T4 and T5 but no theoretical NOEs between A4 and A5 (Table II). The lack of theoretical NOEs between A4 and A5 is apparently related to the syn orientation of the A4 residue as described above (Figures 4a and 7a and Table IV). In addition, there is a weak theoretical base–base NOE but no theoretical sugar–base NOE between T5 and G6

in d(CT_4G)-1. On the other hand, there are both theoretical aromatic and base–sugar NOEs between A5 and G6 in d(CA_4G). These differences are again related to the different sizes of the thymine and adenine bases as well as to the locations of the interacting protons. Significantly, A5 protrudes above the CG base pair, lying in a orientation where it could make contact with the stem of a longer hairpin. By contrast, the smaller thymine ring does not protrude above the base pair and is therefore less likely to interact with the hairpin stem. These results will be discussed below in reference to hairpin thermodynamics.

All of the theoretical NOEs presented above involve protons in adjacent bases; however, NOEs are also predicted between nonsequential residues in the d(CA_4G) and d(CT_4G)-1 loops. Both theoretical base–base and sugar–base NOEs are observed between C1···T3, and a weak sugar–base theoretical NOE is observed between C1···A3. Significantly, the NOEs between C1···A3 and C1···T3 are the result of perpendicular base–base interactions as opposed to the usual parallel base stacking. The differences in intensity of the theoretical NOEs in the A_4 and T_4 loops are apparently due to the differences in location of the protons on adenine and thymine bases in the two theoretical structures, not to the positions of the bases which are quite similar in the two structures. While a weak theoretical sugar–base NOE between T4 and G6 is found, there is no theoretical NOE between A4 and G6. This difference is likely due to the rotation of A4 away from A3 as discussed above. It should be noted, however, that A4 is still in close contact with G6. The T4···G6 NOE is of particular interest since it suggests an interaction of the loop residues with the helical stem and a potential factor governing the stability of a hairpin in solution.

(*C*) *d(CT_4G)-1 versus d(CT_4G)-2.* As noted above, two different morphological structural classes of d(CT_4G) loops, designated d(CT_4G)-1 and d(CT_4G)-2, are found in this work. Whereas the bases in d(CT_4G)-1 are stacked perpendicular to the terminal CG base pair, those in d(CT_4G)-2 (with the exception of T2) are stacked parallel to the CG base pair (compare Figures 4b and 7b versus Figures 4c and 7c). The latter stacking scheme is similar to that proposed by Haasnoot et al. (1986), albeit somewhat disordered with bases T2 and T3 aligned approximately perpendicular to one another. Consistent with this lack of order, only a few theoretical NOEs are present in the d(CT_4G)-2 model (Table II). The T_4 loop is also spatially displaced in d(CT_4G)-2 relative to d(CT_4G)-1. The thymines in d(CT_4G)-1 are positioned above the β-face (Rose et al., 1980) of the CG base pair. In contrast, both faces of the CG base pair plane are exposed in d(CT_4G)-2 with the thymine loop located along the major groove edge of the base pair (in the vicinity of atoms such as C5 and N6 of cytosine and C5, O6, and N7 of guanine). Only the latter structure is discussed below, since the conformation of the former loop already has been described.

The theoretical base–base and sugar–base NOEs in d(CT_4G)-2 involve protons in the T2···T3 and T3···T4 residues (Table II). The T2···T3 theoretical NOEs are produced by perpendicular aromatic interactions, while those involving T3 and T4 are the result of parallel base stacking. The overlap of residues T3 and T4, however, is somewhat limited by the partial insertion of the T5 base (Figure 7). Nevertheless, there are strong T3···T4 theoretical NOEs in d(CT_4G)-2 owing to the close proximity of the thymine methyl groups to one another and to the backbone (Table II). By contrast, there is no theoretical NOE between residues T4 and T5 despite their rather significant stacking interaction. The C1′–C1′ translation of the T4 and T5 bases is comparable to that between adjacent bases in B-DNA. The normals of the bases, however, are antiparallel to one another (Table III). The lack of a theoretical NOE between T4···T5 is attributed to the reversal of the chain direction at this position. The two thymines stack on top of one another but with opposing base orientations. The change in direction of the chain is accompanied by a 3° twist and a 171° bend of residues T4 and T5.

It is interesting that theoretical base–base and sugar–base NOEs involving T2···T3 and T3···T4 interactions are observed in both d(CT_4G)-1 and d(CT_4G)-2. Significantly, the relative orientations of the bases in these two structures are quite different. The three thymines are stacked in a helical array in d(CT_4G)-1, whereas in d(CT_4G)-2, T2 is nearly perpendicular to T3, with a bending angle of 120°, and T3 is stacked above T4, with a bending angle of 30°. The fact that two drastically different conformations of comparable energy (Table V) can produce the same NOEs is noteworthy. While dramatic changes in overall structure have been found in NMR studies of different DNA sequences (Metzler et al., 1990), the present calculations show how a single chemical entity can adopt markedly different spatial forms with the same interproton interactions. The many solution structures typically reported from the distance geometry (DG) analysis of multidimensional NOE spectra are essentially fluctuations about a single equilibrium state. The overall morphology of neighboring bases is comparable from one solution to the next, unlike the very different arrangements of thymines 2–4 found here for the d(CT_4G) hairpin loop. Our results simply raise the question as to whether the set of structures describing a DNA hairpin in solution in unique. It is, of course, also possible that the observed solution properties represent an equilibrium mixture of very different states, none of which accounts individually for the NMR distances. The analysis of such a dynamic system, however, is beyond the scope of the present calculations.

(*D*) *d(CA_5G) versus d(CT_5G).* Figures 5 and 8 are space-filling stereo representations of the predicted structures of the d(CA_5G) and d(CT_5G) hairpin loops. The configurations of these two chains are very similar, as is apparent from the computed bending and twisting angles of the two sequences (Table III). The loop residues are stacked in a helical array on the 3′-side (5′-end of the stem) with the first loop residue (T2 or A2) on the 5′-side (3′-end of the stem) effecting the reversal of chain direction required for loop formation. The stacking is different in two respects from that found above in the three-membered loops. First, the bases are stacked above the α- rather than the β-face of the terminal CG base pair, suggesting that loop size and loop structure may be intimately related. In addition, the five-membered loops do not conform to the hypothesis of loop formation proposed by Haasnoot et al. (1986). According the Haasnoot et al. (1986), the bases should be stacked on the 5′-side of the loop [e.g., above the cytosine in the d(CX_5G) fragment] with one or two bases folding across the minor groove to close the structure. By contrast, in the predicted five-membered hairpin loops, the bases are stacked on the 3′-side of the loop (e.g., above the terminal guanine base) with one base folding at the 5′-end of the loop. This prediction also suggests that there may be a number of different hairpin loop conformations.

Both d(CX_5G) structures found here are highly ordered, as is apparent from the high degree of base stacking in Figures 5 and 8 and as is reflected by several theoretical NOEs. Sequential NOEs are observed between all residues except X1 and X2 which effect the turn in chain direction. The only

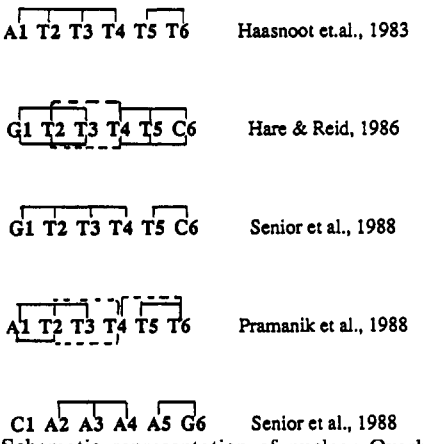

FIGURE 9: Schematic representation of nuclear Overhauser enhancement (NOE) effects experimentally observed in DNA hairpins. Base–sugar NOEs are shown by solid lines above the sequence, base–base NOEs by solid lines below the sequence, and nonsequential NOEs by dashed lines.

significant morphological difference between d(CA_5G) and d(CT_5G) is the partial intercalation of A2 between A4 and A5 in the former structure and the perpendicular orientation of T2 with respect to T4 and T5 in the latter one. Significantly, this is the same sort of perpendicular interaction also observed in the d(CX_4G) loops. As noted above, perpendicular interactions of aromatic amino acids are thought to be important stabilizing forces in proteins (Burley & Petsko, 1985, 1988; Fowler & Moore, 1988). It is also possible that perpendicular interactions between aromatic nucleic acid bases (particularly the pyrimidines) may contribute to the stabilization of nonhelical DNA secondary structures, as well as to certain drug–DNA interactions.

Experimental versus Theoretical Loop Structures. Hairpin loop connectivities determined from various NMR experiments of T_4 and A_4 single-stranded loops are reported in Figure 9. These data can be compared with the theoretically-predicted NOE patterns associated with the computed d(CX_4G) hairpin loops. In comparing the connectivity patterns of the computer models with those obtained from NMR, it is important to keep in mind the following points. First, the theoretical loops have a different base pair at the stem–loop junction from the experimental systems (CG in the former case versus GC or AT in the latter case). Because the energy-minimized conformations of the constituent dinucleotides are dependent on sequence (Erie, 1990; Erie et al., 1992), this change is likely to affect the loop structure. Second, the theoretical loops are static models that do not take into account local fluctuations of the bases, which may affect the experimental NOE data. However, if there is more than a single theoretical structure that meets the criteria for hairpin formation and if these structures are morphologically similar, it may be possible to obtain some information about local fluctuations in loop geometry by comparison of our static structures. Although we have minimized only one structure from each morphological class of hairpin folding, our Monte Carlo analysis produces many closely related structures (Table I) which could also be examined in detail. Third, in contrast to the theoretically predicted NOEs, the experimental NOEs may be influenced by spin diffusion, especially when a full range of mixing times have not been studied. Finally, some theoretically-predicted NOEs may not be observed experimentally because of overlapping resonances in the NMR spectra.

A previous study of the two d(CGAACGX_4CGTTCG) hairpins, where X = A or T (Senior, 1988; Senior et al., 1988), provides experimental NOE data of direct relevance to the current computational predictions. Interestingly, the loops of these two oligomers produce different NOE spatial connectivity patterns. Moreover, the experimental observations are consistent with the configurations of the d(CX_4G) loops predicted in this study. The actual connectivity patterns for the theoretically-predicted loops, however, differ in detail from the experimental patterns (Table II and Figure 9). Nevertheless, in both cases there are fewer NOEs observed for the A_4 than the T_4 loop. As noted in the above discussion of the d(CA_4G) and d(CT_4G) structures, the different NOE patterns may reflect different positions of the NOE protons on the adenine and thymine bases and/or different overall loop conformations. In addition, dynamic factors can influence the experimental NOEs.

Most of the available thermodynamic and structural characterization of DNA hairpin loops is found in chains with four unpaired thymines (Haasnoot et al., 1983; Hare & Reid, 1986; Pramanik et al., 1988; Senior et al., 1988; Wolk et al., 1988; Blommers et al., 1989; Xodo et al., 1991; Amaratunga et al., 1992; Doktycz et al., 1992; Paner et al., 1992). As evident from the NOE connectivity patterns in Figure 9, the proposed loop connectivities are dependent on both stem sequence and structure. Two different structural pictures have developed from the experimental data. One is a general model of loop folding based on a combination of NMR and thermodynamic data (Haasnoot et al., 1983, 1986). The other is an atomic level description derived from distance geometry (DG) and/or conformational energy analysis (Hare & Reid, 1986; Boulard et al., 1991). Since, the former model has been discussed above, only the detailed loop structures are compared to our results below.

Comparison with Distance Geometry Structures. The structure of the d(CGCGT_4CGCG) hairpin deduced from a distance geometry analysis of NOE data (Hare & Reid, 1986) is closer to the d(CT_4G)-2 model described above (parallel stacking) than the d(CT_4G)-1 structure (perpendicular stacking). Specifically, the bases in the DG hairpin loop are stacked parallel to the GC base pair that closes the loop. The principal differences between the energy-based d(CT_4G)-2 model and the DG structure lie in the first two loop residues. In the DG loop, T2 is stacked on G1, whereas in d(CT_4G)-2, T2 is oriented perpendicular to C1 on the exterior of the loop. In addition, T3 is stacked below T4 (in the reference frame of Figure 4) in the DG loop, whereas T3 is stacked above both T4 and T5 in d(CT_4G)-2. Due to these discrepancies, it is not surprising that the NOEs predicted in this study are different from those in the loop structure derived directly from experimental measurements (Table II and Figure 9). As noted above, an exact comparison between the DG and computer-generated structures cannot be made since they are closed by different base pairs (GC versus CG). Significantly, in both the DG and the d(CT_4G)-2 structures, base–base and sugar–base NOEs are observed between T2 and T3. These NOEs, however, result from very different interactions. Bases T2 and T3 are oriented in a perpendicular fashion in the d(CT_4G)-2 loop, while they are stacked in a parallel fashion in the DG model.

The wide conformational sampling introduced in the recently reported NOE distance-constrained conformational energy analysis of the d(CGTGGATCGTTTCGATCCGAG) hairpin reveals a highly unusual solution structure (Boulard et al., 1991). While this loop conformation is quite different from the T3 loop generated in the present study, the overall conformation of the hairpin is of the same general type as the d(CT_4G)-1 loop reported here. That is, the unpaired bases

Table VI: Thermodynamic Data for Hairpins with Four-Membered Loops

	sample	$\Delta H°$ (kcal/mol)	T_m (°C)
I	d(CGAACGT$_4$CGTTCG)[a]	49	67.0
II	d(CGAACGA$_4$CGTTCG)[a]	34	63.0
III	d(CGAACG)-d(CGTTCG)[a]	39	
IV	d(GGATACT$_4$GTATCC)[b]	41[c]	58.3
V	d(GGATACA$_4$GTATCC)[b]	41[c]	55.3
VI	d(GAATTCT$_4$GAATTC)[d]	38	56.6
VII	d(GAATTCA$_4$GAATTC)[d]	33	51.4

[a] These data were taken from Senior et al. (1988). The thermodynamic data were determined calorimetrically in 100 mM NaCl and 10 mM phosphate buffer, pH 7.0. [b] These data were taken from Paner et al. (1990). The thermodynamic parameters were derived from UV melting studies using a van't Hoff analysis. The melting experiments were performed in 100 mM NaCl and 10 mM phosphate buffer, pH 6.5.[c] The enthalpy values were calculated from the $(d\theta/dT)_{max}$ listed in Table II of Paner et al. (1990) using a two-state van't Hoff analysis. These transitions are not two state, and the melting of the T_4 hairpin deviates from two-state behavior to a greater extent than the A_4 hairpin. Consequently, these are minimum values for the enthalpies. The actual enthalpy of the T_4 hairpin should be greater than that of the A_4 hairpin. [d] These data were taken from Germann et al. (1990). The thermodynamic parameters were derived from UV melting curves using a two-state van't Hoff analysis. The melting experiments were performed in 200 mM NaCl, 0.1 mM EDTA, and 50 mM phosphate buffer, pH 7.0.

in the hairpin are oriented perpendicularly to the base pair closing the loop. In addition, the first residue in the loop lies in the minor groove much like the T5 residue of d(CT$_4$G)-1. This solution structure further supports the suggestion that hairpin loops can adopt many different overall conformations and that these conformations can be affected by the stem sequence.

Loop Structure and Thermodynamics. In the sections that follow, various features of the loop structures are discussed in an effort to correlate specific structural features of the computationally predicted loops with thermodynamic data obtained on similar hairpins (Hilbers et al., 1985; Haasnoot et al., 1986; Ikuta et al., 1986; Erie et al., 1987, 1989; Senior et al., 1988; Xodo et al., 1988, 1991; Blommers et al., 1991; Germann et al., 1990; Paner et al., 1990, 1992; Rentzeperis et al., 1991; Amaratunga et al., 1992).

(A) Loop Size, Sequence, and Stability. Several recent publications provide thermodynamic and/or NMR data on DNA hairpins (Senior et al., 1988; Xodo et al., 1988; Germann et al., 1990; Paner et al., 1990) and double hairpins (Erie et al., 1987, 1989; Benight et al., 1988; Rentzeperis et al., 1991) that contain four or five adenine or thymine residues in loops closed by GC or CG base pairs. Table VI lists selected thermodynamic data for several hairpins with four-membered loops (designated below by Roman numerals). Since each hairpin has been studied in different buffers, at different salt concentrations, and using different experimental techniques (spectroscopy and calorimetry), a direct quantitative comparison of the data is not possible. However, as described below, trends in the data can be examined. Since structure and thermodynamics are intimately related, it is useful to analyze the structures of the theoretically generated hairpin loops in terms of the available thermodynamic data on hairpins. For this purpose the enthalpic contribution to the overall free energy is particularly useful, since changes in van der Waals interactions manifest themselves primarily through changes in the enthalpy.

Several features of the X_4 loops are consistent with the thermodynamic data presented in Table VI. First, although two morphological classes of potential hairpin loop structures are found in the Monte Carlo samples of A_4-containing chains, only one class is able to minimize to a reasonable final state. Both classes of loops are found to minimize to reasonable final states, however, if thymine bases are introduced in place of the adenines. This result is consistent with the suggestion that hairpins with four thymine residues in the loops would be more stable than those with four adenosine residues in the loops, although the numbers for the T-containing loops are not a direct result of the Monte Carlo simulations. Inspection of the data in Table VI reveals that, in all cases, the hairpins with four thymine loop residues (I, IV, and VI) are thermally and enthalpically more stable that the corresponding hairpins with four adenine loop residues (II, V, and VII).

Second, in both the d(CA$_4$G) and d(CT$_4$G)-1 structures (Figures 4a,b and 7a,b), the last residue in the hairpin loops, X5, lies perpendicular to the CG base pair on the exterior of the loop. Moreover, the A5 in d(CA$_4$G) protrudes above the CG base pair in such a manner that it could make contact with a longer hairpin stem. This feature is not present in d(CT$_4$G)-1, due, at least in part, to the smaller size of the thymine ring relative to the adenine ring. This structural difference is consistent with the experimental observation that T_4 loops appear to be thermodynamically and enthalpically more stable than the A_4 loops. For example, if the A5 residue in the d(CA$_4$G) structure were to disrupt the duplex stem of a hairpin, a reduction in enthalpy of the hairpin relative to that of the isolated duplex might be observed. By contrast, since the T5 in d(CT$_4$G)-1 does not interact with the duplex stem, one would not expect to observe a reduction in the enthalpy of the hairpin relative to the isolated duplex stem. According to the data in Table VI, the hairpin with four thymine residues in the loop is enthalpically stabilized relative to its isolated stem duplex (I vs III). By contrast, the hairpin with four adenine residues in the loop (II) is enthalpically destabilized relative to its isolated stem duplex (III). Similarly, in a recent investigation (Blommers et al., 1989), hairpin loops (closed by an AT base pair) containing four or six adenine residues are found to be enthalpically destabilized relative to those containing four or six thymine residues, respectively. Interactions like those of the A5 residue with the terminal base pair in d(CA$_4$G), which are not present in the corresponding d(CT$_4$G)-1 structure, could be related to the destabilization of hairpins containing adenosine loops. The theoretically-generated T_4-loop structures, however, provide little insight into the observed increase in enthalpy of the thymine-containing hairpin relative to the stem duplex (I vs III) (Senior et al., 1988). The thymine residues in the current conformational model do not interact with the helical stem. In this connection, calorimetric studies on three dumbbell-shaped, double-hairpin structures, d(TTCCT$_n$GGAATTC-CT$_n$GGAA where n = 4 or 5 (Erie et al., 1987, 1989) and d(GCGCT$_5$GCGCGTACT$_5$GTAC) (Rentzeperis et al., 1991), suggest that there are no enthalpic contributions from the single-stranded thymine loops. In addition, recent investigations of several DNA dumbbells sharing the same 16 base pair duplex core and having 2–14 thymines in the loop show no enthalpic contribution of the thymine residues to the overall melting enthalphy (Amaratunga et al., 1992; Paner et al., 1992). The observations in the double-hairpin structures may arise from the low self-stacking enthalpy of thymine. The different thermodynamic consequences of the T_4 loops in I versus those in the 24-mer dumbbell [e.g., d(TTCCT$_4$-GGAATTCCT$_4$GGAA)] also may reflect differences between single- and double-hairpin structures and/or differences between the stem duplex structures. The latter suggestion is supported by several investigations (Ikuta et al., 1986; Xodo

et al., 1988, 1991; Blommers et al., 1989) which indicate that the enthalpic contributions of thymine residues in the loops are dependent on both stem sequence and loop size; however, their contribution is always small. Our observation of two very different geometries for the T_4 loops of a single sequence may be relevant to the thermodynamic findings. The different structures may lead to different thermodynamic consequences, helping to rationalize the different thermodynamic behavior of different hairpin structures with loops of four thymines.

The available thermodynamic data on hairpin structures with five thymine residues or five adenine residues in the loop present a less consistent picture than the data for the X_4 loops listed in Table VI. In one case (Germann et al., 1990), a hairpin with five thymine residues in the loop is thermally and enthalpically more stable than the corresponding hairpin with five adenine residues in the loop. In another case (Xodo et al., 1988), two hairpins containing five loop residues of either adenine or thymine have very similar thermodynamic profiles, with transition enthalpies similar to that of the corresponding isolated stem duplex. The thermodynamic data in both these studies are derived indirectly from UV melting curves using a van't Hoff analysis and assuming two-state melting behavior. Such an assumption may not be appropriate for these hairpin structures (Senior et al., 1988; Paner et al., 1990) and can artificially influence the thermodynamic data. Our theoretically-generated loop structures, however, are generally consistent with the findings of Xodo et al. (1988), suggesting that, in contrast to loops of four residues, both thymines and adenines are easily accommodated in loops of five residues. Specifically, our theoretical results reveal that both $d(CT_5G)$ and $d(CA_5G)$ structure exhibit similar geometries with the loop bases forming single-helical arrays (Figures 5 and 8).

The ligation properties of the dumbbell-shaped double-hairpin molecules also are consistent with the suggestion that four-membered loops cannot accommodate the bases as well as five-membered loops (Erie et al., 1987, 1989). Specifically, the 24-mer dumbbell $d(TTCCT_4GGAATTCCT_4GGAA)$ with four thymines residues in each loop is resistant to enzymatic ligation, whereas the 26-mer dumbbell $d(TTCCT_5GGAATTCCT_5GGAA)$ with five thymine residues in each loop is readily ligated into a covalently closed, single-stranded circle. It has been proposed that the loop of four thymine residues may distort the adjacent stem duplex, causing the 5′- and 3′-termini to be out of phase and thereby inhibit ligation (Erie et al., 1987, 1989). The successful chemical synthesis of the above T_4 dumbbell (Ashley & Kushlan, 1991), while apparently contrary to this structural interpretation, is consistent with it in the sense that the yield of the cyclization reaction for this T_4 dumbbell is quite low relative to the yields of dumbbells with different stem sequences. The dumbbells with longer stems exhibit significantly greater ligation efficiencies, a result which is consistent with the suggestion that hairpin loop structure may affect local base pairing. The theoretical $d(CT_4G)$-2 and $d(CT_5G)$ loop structures, however, are consistent with the earlier structural interpretation. First, the terminal base pair in the $d(CT_4G)$-2 loop is distorted with a large propeller twist, whereas the terminal base pair in the $d(CT_5G)$ is relatively undistorted with only a small propeller twist (Figure 6 and Table III). Moreover, the overall spatial disposition of the thymines in the five-membered loop is also more favorable than that in the four-membered loops. Specifically, it is easier to dock the $d(CT_5G)$ structure onto a B-DNA duplex than either $d(CT_4G)$-1 or $d(CT_4G)$-2. The ability to form the hairpin stem as a B-DNA helix is related to the ease of ligation of the nicked dumbbell, the central core of which presumably adopts this structure in solution. As noted above, the T_5 loop is located above the α-face of the terminal CG base pair of $d(CT_5G)$ while the T_4 loops are found on either the β-side or along the major groove edge of the terminal base pair. Steric clashes of the four-membered loops with the B-DNA helical stem above the β-face of CG would prevent the formation of the ligated structure. Indeed, it is easier to dock the $d(CT_4G)$-1 loop against a Z-DNA helix where the stem residues lie on the α-side of the CG base pair (Olson et al., 1983; Harvey, 1983). Moreover, the stems of DNA hairpin loops are not limited to ordinary B-type helices, as demonstrated by recent crystallographic data (Chattopadhyaya et al., 1988, 1990; Kang et al., 1992).

The theoretically-predicted loop structures derived in this study further suggest that "phasing" may play an important role in hairpin loop structure. Specifically, we find that three single-stranded residues are sufficient for loop formation (Figures 3 and 4). Furthermore, in loops with four single-stranded residues, we find that one base is oriented perpendicular to the CG base pair on the outside of the loop, as if a loop of three residues were more favorable. It is possible that small DNA hairpin loops are more favorable when there is an odd rather than an even number of bases in the loop (e.g., three and five bases versus four bases in the loop). That is, the bases that close the loop might adopt the proper orientation for base pairing, or be "in phase" if there is an odd number of bases in the loop. This suggestion is supported by a recent study on hairpin structures with three, four, and five adenine or thymine residues in the loops (Germann et al., 1990). Specifically, the hairpins with three residues in the loops are the most stable and the least affected by loop sequence. Furthermore, the loop structure derived from a recent NMR investigation (Piotto et al., 1990) of a hairpin with two 3′-thymine dithiophosphate residues followed by two adenine residues in the loop is consistent with the above proposal. The proposed TTAA-loop structure is similar to that of the $d(CT_4G)$-2 model, with three of the four residues folding to form the loop and the second thymine loop residue, which is flanked by the dithiophosphate groups, lying on the outside. It is also possible that preferred loop size may be tied to the helical character of the hairpin stem. As noted above, it is easier to dock the five-membered loops of this study on B-DNA and the three- and four-membered loops on Z-DNA. Interestingly, in this regard the helical stem in the recent X-ray crystal structure of $d(CGCGCGT_4CGCGCG)$ (Chattopadhyaya et al., 1988, 1990) is a Z-DNA helix and that in $d(GGGGT_4GGGG)$ (Kang et al., 1992) is a four-stranded guanine tetrad.

(*B*) *Loop Structure, Solvation, and Electrostatics.* Inspection of the different views of the DNA hairpins in Figures 3–5 reveals a striking feature that is present in the majority of the theoretical loop structures. The carbon atoms in these structures are color-coded so as to differentiate the aromatic (white) and backbone (yellow) carbons. Interestingly, in both the loops with perpendicular stacking and those with parallel stacking, the bases are "stacked" on a single side of the structures, with the backbone on the opposite side. The loops therefore appear to possess a hydrophobic (aromatic bases) and a hydrophilic (sugar–phosphate backbone) face. In duplex DNA, such hydrophilic and hydrophobic faces are not observed, but rather a very symmetric distribution of hydrophilic and hydrophobic residues on the surface is seen. Consequently, the local solvent environment around a DNA hairpin might be very different from that surrounding double-

Table VII: Phosphate–Phosphate Distances in d(CX$_n$G) Hairpin Loops[a]

	1	2	3	4	5	6	⟨P···P⟩
			d(CA$_3$G) = d(Cp^{1}Ap^{2}Ap^{3}Ap^{4}G)				
1		6.6	11.1	14.1			9.6 ± 3.4
2			7.0	12.5			
3				6.5			
			d(CT$_3$G) = d(Cp^{1}Tp^{2}Tp^{3}Tp^{4}G)				
1		6.6	10.4	14.5			9.7 ● 3.5
2			6.9	12.8			
3				6.7			
			d(CA$_4$G) = d(Cp^{1}Ap^{2}Ap^{3}Ap^{4}Ap^{5}G)				
1		7.7	13.3	17.4	14.4		9.8 ± 4.2
2			6.9	11.7	8.2		
3			5.0	5.9			
4					7.2		
			d(CT$_4$G)-1 = d(Cp^{1}Tp^{2}Tp^{3}Tp^{4}Tp^{5}G)				
1		7.7	13.6	17.5	14.6		9.7 ± 4.3
2			6.7	11.4	8.2		
3				4.8	5.4		
4					7.1		
			d(CT$_4$G)-2 = d(Cp^{1}Tp^{2}Tp^{3}Tp^{4}Tp^{5}G)				
1		6.6	13.1	15.9	18.5		11.6 ± 4.5
2			7.0	11.8	16.3		
3				6.8	13.1		
4					6.7		
			d(CA$_5$G) = d(Cp^{1}Ap^{2}Ap^{3}Ap^{4}Ap^{5}Ap^{6}G)				
1		6.7	11.8	16.2	18.1	17.7	12.6 ± 4.8
2			6.9	13.0	17.2	19.3	
3				6.8	12.7	16.6	
4					6.9	12.6	
5						7.0	
			d(CT$_5$G) = d(Cp^{1}Tp^{2}Tp^{3}Tp^{4}Tp^{5}Tp^{6}G)				
1		6.7	12.7	16.6	17.5	17.3	12.6 ± 4.6
2			6.9	12.5	16.2	18.5	
3				7.2	13.0	17.5	
4					6.8	12.8	
5						6.9	
			B-DNA = d(Cp^{1}Xp^{2}X)·d(p^{3}Xp^{4}Xp^{5}G)				
1		6.6	12.7	18.7	17.7		13.1 ± 4.9
2			6.6	17.7	15.6		
3				15.6	13.2		
4					6.6		

[a] Distances are reported in angstroms. Boldfaced numbers associated with individual matrix elements correspond to the specified phosphate groups of each sequence.

helical DNA. It is further possible that there might be hydrophobically restricted water (such as that found at the surface of proteins) on the side of the loops where the aromatic bases are stacked. Calculations that predict water structure around a DNA hairpin could be very illuminating. It is likely that the distribution of water around the loops may differ from that around the double helix. Significantly, in the recent crystallographic investigation of the Z-DNA d(CGCGCGT$_4$-CGCGCG) hairpin, the T_4 loop is involved in tight intermolecular packing with many water molecules of hydration thereby excluded (Chattopadhyaya et al., 1990). Of the 70 water molecules located in the crystal structure, most are associated with the sugar–phosphate backbone Z-DNA helical stem. The hydrophilic and hydrophobic "faces" of hairpin loops may thus be important in the interactions of nucleic acids with themselves as well as with proteins. Such interactions could influence the nucleation of hairpin folding in nucleic acid chains or modulate the binding of proteins to nucleic acids [i.e., binding of transfer RNA (tRNA) to the ribosome].

In addition to altering the local solvent environment, the arrangement of bases and backbone atoms in hairpins may affect the electrostatic interactions of the highly negatively charged nucleic acid residues (Troll et al., 1986; Conrad et al., 1988; Jayaram et al., 1989; Sharp et al., 1990). According to measurements of a clay model of DNA in an electrolyte tank (Troll et al., 1986), the geometry of the macromolecule determines the electrostatic interactions on its surface. Recent calculations of the electrostatic potentials around yeast phenylalanine and aspartate tRNAs by numerical solution of the nonlinear Poisson–Boltzman equation (Sharp et al., 1990) demonstrate this difference. The calculations account for both electrolyte screening effects and the difference in polarizability between the solvent and the nucleic acid. The dielectric constant of the tRNA is assumed to be 4 D and that of the solvent 80 D. Significantly, using this approach, a negative potential is found to surround all of the tRNA except the anticodon loop, where there is a "hole" in the potential surface. This lack of negative potential at the anticodon loop is attributed both to the geometry of the phosphodiester backbone and to the geometry of the low dielectric region (i.e., the bases) (Sharp et al., 1990). The structure of the anticodon loop in the tRNAs is very similar to the d(CX$_5$G) loops generated in this study, with the single-stranded residues stacked parallel to and above the α-face of the terminal base pair of the helical stem. Consequently, it is not unreasonable to suggest that DNA hairpin loops may also exhibit unusual electrostatic properties which cannot be accounted for using a simple Coulombic potential. Similar calculations of the electrostatic potential, however, have been performed on other RNA loop structures in which the lack of electronegative potential at the loop is not observed (Sharp et al., 1990). It is thus not clear whether this feature is a general property of DNA and RNA hairpin loops.

Experimental data accumulated to date also suggest that the DNA hairpin loop regions do not contribute to the salt dependence of the helix-to-coil transition (Erie et al., 1987, 1989; Marky and Breslauer, unpublished results; Rentzeperis et al., 1991). Accordingly, the single-stranded loops in these structures may be electrostatically different from the DNA duplex regions. The phosphate separations in the hairpin loops derived in this study, however, are not sufficiently large to account for such a difference in electrostatic behavior. This feature can be seen from inspection of the data in Table VII, which lists all phosphate–phosphate (P–P) distances in the predicted models together with those of a three base pair ideal B-DNA helix (Arnott & Huckins, 1972). The phosphate groups in the duplex are numbered so that the fragment resembles a broken hairpin loop with four unpaired bases. Average P···P spacing in each structure is also included. As evident from the table, the phosphates in the optimized loop structures are closer, on average, to one another than they would be if they were in a double helix. If only Coulombic interactions were important, the hairpin loops should make a greater contribution to the electrostatics of order–disorder transitions than is observed in double-hairpin structures (Erie et al., 1987, 1989; Rentzeperis et al., 1991). This apparent disparity may simply reflect a different solvent environment around the hairpin loops compared with the double helices. These solvation differences may even affect long-range electrostatic interactions. It would thus be very interesting to calculate the electrostatic potential around the theoretically-predicted hairpin loops taking into account the dielectric discontinuity.

CONCLUSIONS

The broadly based conformational search carried out here not only reveals novel hairpin loop structures but also suggests

the possibility of several very different structures consistent with the solution properties of such molecules. Our approach using a combination of Monte Carlo simulation and molecular mechanics energy minimization is complementary to new buildup algorithms introduced to interpret the NMR solution structure of DNA. In contrast to conventional molecular modeling carried out in terms of some sort of optimization in Cartesian coordinate space, these procedures sample a much broader range of chain conformations. Consequently, we observe loop morphologies that might otherwise be missed and that have not been proposed previously. Loop conformations generated using NMR distance constraints in conjunction with DG and/or molecular mechanics or dynamics are dependent on and are very similar to the assumed starting conformations. The starting states are usually based on standard B- or A-DNA helical structures and, with two notable exceptions (Boulard et al., 1991; Blommers et al., 1991), fall in the same general morphological class. The range of structures resulting from the calculations essentially correspond to room temperature fluctuations of a single equilibrium structure. By contrast, we predict loop conformations (for the same chain sequence) that show similar physical properties (e.g., NOE distances, energy, etc.) but that differ drastically from one another, and that bear little resemblance to A- or B-like conformations. Furthermore, the DG and molecular mechanics/dynamics methods in current use for determining NMR solution structures search for a single conformation that fits the observed experimental data. These methods exclude the possibility of multiple conformations (with very different three-dimensional shapes) that fit the data equally well, coexisting in solution or an equilibrium mix of conformational states which individually do not necessarily match the observed data. The current study suggests the former possibility but is not geared to address the latter question. The suggestion that more than one loop conformation may exist in solution is supported by the recent crystallographic investigations of the Z-DNA d(CGCGCGT$_4$CGCGCG) and the telomeric d(GGGGT$_4$GGGG) hairpins (Chattopadhyaya et al., 1988, 1990; Kang et al., 1992). Specifically, the *B*-factors for the loop residues are significantly greater than those for the double-helical residues. The authors attribute this result to either static disorder (multiple conformations) and/or thermal motions. There is also information from NMR relaxation measurements (Williamson & Boxer), 1989a,b; Ikuta & Wang, 1990) and EPR studies (Spaltenstein et al., 1989) of the existence of apparently large-scale internal motions (on the microsecond time scale) in single-stranded loops. Significantly, the Monte Carlo method employed in this work generates multiple conformations and provides a starting point for analyzing NMR data in terms of the equilibria and fluctuations of very different three-dimensional structures. New computational methodology is required, however, to account for the dynamics of a multistate transition.

Inspection of the wide variety of hairpin conformations generated in this study also reveals a striking feature that is present to some degree in all structures. Specifically, the aromatic bases stack on a single side of the loop with the backbone carbons on the opposite side. The various loop conformations therefore possess a hydrophilic and a hydrophobic face. This feature, which does not exist in double-helical nucleic acids, may have biological significance. For example, hydrophobic interactions with the nonpolar face of the loop may provide a driving force for certain protein–nucleic acid interactions as well as for nucleic acid self-associations. In addition, separation of the hydrophobic and hydrophilic domains within a loop may impart unusual electrostatic properties to hairpin loops compared to double-helical DNA regions. This possibility would be consistent with recent theoretical computations of the electrostatic properties of tRNAs (Sharp et al., 1990) as well as with the thermodynamic behavior of DNA dumbbells (Erie et al., 1987, 1989; Rentzeperis et al., 1991). The differential electrostatic properties of loop and duplex regions could modulate both nucleic acid–nucleic acid and protein–nucleic acid interactions.

In addition to the unusual arrangements of the hydrophilic and hydrophobic residues in the DNA loops, we also observe perpendicular interactions between the aromatic bases (especially thymines) in the hairpins. Similar perpendicular interactions are observed between thymine residues in the loop of a recently crystallized Z-DNA hairpin (Chattopadhyaya et al., 1988, 1990) as well as in two different hairpin loop structures deduced from NMR (Blommers et al., 1991; Boulard et al., 1991). In the crystal structure, both inter- and intramolecular perpendicular contacts are present. Significantly, these types of interactions have been proposed to be important in the stabilization of protein structures (Burley & Petsko, 1985, 1988; Fowler & Moore, 1988) and small molecules (Karlström et al., 1983; Pawliszyn et al., 1984; Hunter & Sanders, 1990; Linse, 1992). It is possible that perpendicular interactions betwen the nucleic acid bases could also stabilize non-base-paired secondary and/or tertiary structures of DNA and RNA.

ACKNOWLEDGMENT

We thank Professor Barry Honig and Dr. Kim Sharp for informative discussions of hairpin loop structure and electrostatics, Ms. Marla Babcock for the use of several programs to calculate local geometric parameters, Dr. A. R. Srinivasan for assistance in molecular graphics, and Professor Ronald Levy for the use of the IMPACT molecular mechanics package. All calculations were carried out at the Rutgers Center for Computational Chemistry, which was supported in part by the New Jersey Center for Advancement of Biotechnology and Medicine.

SUPPLEMENTARY MATERIAL AVAILABLE

Seven tables giving the atomic coordinates of the theoretically-generated d(CA$_3$G), d(CT$_3$G), d(CA$_4$G), d(CT$_4$G)-1, d(CT$_4$G)-2, d(CA$_5$G), and d(CT$_5$G) hairpin loops (28 pages). Ordering information is given on any current masthead page.

REFERENCES

Altona, C., & Sundaralingam, M. (1972) *J. Am. Chem. Soc. 94*, 8205–8212.

Amaratunga, M., Snowden-Ifft, E., Wemmer, D. E., & Benight, A. J. (1992) *Biopolymers 32*, 865–879.

Arnott, S., & Huckins, D. W. L. (1972) *Biochem. Biophys. Res. Commun. 47*, 1504–1509.

Ashley, G. W., & Kushlan, D. M. (1991) *Biochemistry 30*, 2927–2933.

Babcock, M. S., & Olson, W. K. (1992) in *Computation of Biomolecular Structures: Achievements, Problems, and Perspectives* (Jovin, T., & Soumpasis, D. M., Eds.) Springer-Verlag, Berlin (in press).

Babcock, M. S., Srinivasan, A. R., & Olson, W. K. (1989) *Book of Abstracts: Sixth Conversation in Biomolecular Stereodynamics*, p 9, Albany, NY.

Bassolino, D., Hirata, F., Kitchen, D. B., Kominos, D., Pardi, A., & Levy, R. M. (1988) *Int. J. Supercomp. Appl. 2*, 41–61.

Benight, A. S., Schurr, J. M., Flynn, P. F., Reid, B. R., & Wemmer, D. E. (1988) *J. Mol. Biol. 200*, 377–399.
Blommers, M. J. J., Walters, J. A. L. I., Haasnoot, C. A. G., Aelen, J. M. A., van der Marel, G. A., van Boom, J. H., & Hilbers, C. W. (1989) *Biochemistry 28*, 7491–7498.
Blommers, M. J. J., van de Ven, F. J. M., van der Marel, G. A., van Boom, J. H., & Hilbers, C. W. (1991) *Eur. J. Biochem. 201*, 33–51.
Boulard, Y., Gabarro-Arpa, J., Cognet, J. A. H., Le Bret, M., Buy, A., Téoule, R., Guschlbauer, W., & Fazakerley, G. V. (1991) *Nucleic Acids Res. 19*, 5159–5167.
Burley, S. K., & Petsko, G. A. (1985) *Science 229*, 23–28.
Burley, S. K., & Petsko, G. A. (1988) *Adv. Protein Chem. 39*, 125–189.
Cantor, C. R., & Schimmel, P. R. (1980) *Biophysical Chemistry*, Chapters 6 and 22, W. H. Freeman and Co., San Francisco.
Chattopadhyaya, R., Ikuta, S., Grzeskowiak, K., & Dickerson, R. E. (1988) *Nature 334*, 175–179,
Chattopadhyaya, R., Grzeskowiak, K., & Dickerson, R. E. (1990) *J. Mol. Biol. 211*, 189–210.
Conrad, J., Troll, M., & Zimm, B. H. (1988) *Biopolymers 27*, 1711–1732.
Cox, E. G., Cruickshank, D. W. J., & Smith, J. A. S. (1958) *Proc. R. Soc. London, Ser. B 247*, 15–21.
Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C.-S., von Kitzing, E., Wang, A. H.-J., & Zhurkin, V. B. (1989a) *J. Biomol. Struct. Dyn. 6*, 627–634.
Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., Tung, C.-S., von Kitzing, E., Wang, A. H.-J., & Zhurkin, V. B. (1989b) *Nucleic Acids Res. 17*, 1797–1803.
Diekmann, S. (1988) *J. Mol. Biol. 208*, 787–791.
Diekmann, S. (1989) *EMBO J. 8*, 1–4.
Doktycz, M. J., Goldstein, R. F., Paner, T. M., Gallo, F. J., & Benight, A. S. (1992) *Biopolymers 32*, 849–864.
Erie, D. A. (1990) Ph.D. Thesis, Rutgers University, New Brunswick, NJ.
Erie, D. A., Sinha, N., Olson, W. K., Jones, R. A., & Breslauer, K. J. (1987) *Biochemistry 26*, 7150–7159.
Erie, D. A., Olson, W. K., Jones, R. A., Sinha, N., & Breslauer, K. J. (1989) *Biochemistry 28*, 268–273.
Erie, D. A., Breslauer, K. J., & Olson, W. K. (1992) *Biopolymers* (in press).
Flory, P. J. (1969) *Statistical Mechanics of Chain Molecules*, Chapter 4, Wiley-Interscience, New York.
Fowler, P. W., & Moore, G. (1988) *Biochem. Biophys. Res. Commun. 153*, 1296–1300.
Garcia, A. E., Gupta, G., Sarma, M. H., & Sarma, R. H. (1988) *J. Biomol. Struct. Dyn. 6*, 525–542.
Germann, M. W., Kalisch, B. W., Lundberg, P., Vogel, H. J., & van de Sande, J. H. (1990) *Nucleic Acids Res. 18*, 1489–1498.
Haasnoot, C. A. G., de Bruin, S. H., Berendsen, R. G., Janssen, H. G. J. M., Binnendijk, T. J. J., Hilbers, C. W., van der Marel, G. A., & van Boom, J. H. (1983) *J. Biomol. Struct. Dyn. 1*, 115–129.
Haasnoot, C. A. G., Hilbers, C. W, van der Marel, J. H., Singh, U. C., Pattabiraman, N., & Kollman, P. A. (1986) *J. Biomol. Struct. Dyn. 3*, 843–857.
Hare, D. R., & Reid, B. R. (1986) *Biochemistry 25*, 5341–5350.
Harvey, S. C. (1983) *Nucleic Acids Res. 11*, 4867–4878.
Harvey, S. C., Luo, J., & Lavery, R. (1988) *Nucleic Acids Res. 16*, 11795–11809.
Hilbers, C. W., Haasnoot, C. A. G., de Bruin, S. H., Joorden, J. J., van der Marel, G. A., & van Boom, J. H. (1985) *Biochimie 67*, 685–695.
Hunter, C. A., & Sanders, K. M. (1990) *J. Am. Chem. Soc. 112*, 5525–5534.
Ikuta, S., & Wang, Y.-S. (1990) *Adv. Biophys. Chem. 1*, 125–153.
Ikuta, S., Chattopadhyaya, R., Ito, H., Dickerson, R. E. & Kearns, D. R. (1986) *Biochemistry 25*, 4840–4849.
Jayaram, B., Sharp, K., & Honig, B. (1989) *Biopolymers 28*, 975–993.
Kang, C. H., Zhang, X., Ratliff, R., Moyzis, R., & Rich, A. (1992) *Nature 356*, 126–131.
Karlström, G., Linse, P., Wallqvist, A., & Jönsson, B. (1983) *J. Am. Chem. Soc. 105*, 3777–3782.
Larsen, T. A., Goodsell, D. S., Cascio, D., Grzeskowiak, & Dickerson, R. E. (1989) *J. Biomol. Struct. Dyn. 7*, 477–491.
Linse, P. (1992) *J. Am. Chem. Soc. 114*, 4366–4373.
Lipton, M., & Still, W. C. (1988) *J. Comput. Chem. 9*, 343–355.
Marky, L. A., Blumenfeld, K. S., Kozlowski, S., & Breslauer, K. J. (1983) *Biopolymers 22*, 1247–1257.
Marky, N. L., & Olson, W. K. (1982) *Biopolymers 21*, 2329–2344.
Marky, N. L., & Olson, W. K. (1987) *Biopolymers 26*, 415–438.
Metzler, W. J., Wang, C., Kitchen, D. B., Levy, R. M., & Pardi, A. (1990) *J. Mol. Biol. 214*, 711–736.
Müller, U. R., & Fitch, W. M. (1982) *Nature 298*, 582–585.
Olson, W. K. (1982) in *Topics in Nucleic Acid Structures* (Neidle, S., Ed.) Part 2, pp 1–79, Macmillan, London.
Olson, W. K., Srinivasan, A. R., Marky, N. L., & Balaji, V. N. (1983) in *Structures of DNA, Cold Spring Harbor Symposia in Quantitative Biology*, Vol. 47, pp 229–241, Cold Spring Harbor Laboratory Publications, Cold Spring Harbor, NY.
Orbons, L. P. M., van der Marel, G. A., van Boom, J. H., & Altona, C. (1986) *Nucleic Acids Res. 14*, 4187–4196.
Paner, T. M., Amaratunga, M., Doktycz, M. J., & Benight, A. S. (1990) *Biopolymers 29*, 1715–1734.
Paner, T. M., Amaratunga, M., & Benight, A. S. (1992) *Biopolymers 32*, 881–892.
Pawliszyn, J., Szczęśniak, M. M., & Scheiner, S. (1984) *J. Phys. Chem. 88*, 1726–1730.
Piermarni, G. J., Mighell, A. D., Weir, C. E., & Block, S. (1969) *Science 165*, 1250–1255.
Piotto, M. E., Granger, J. N., Cho, Y., & Gorenstein, D. G. (1990) *J. Am. Chem. Soc. 112*, 8632–8634.
Pramanik, P., Kanhouwa, N., & Kan, L.-S. (1988) *Biochemistry 27*, 3024–3031.
Rentzeperis, D., Karakoz, D. P., & Marky, L. A. (1991) *Biochemistry 30*, 6276–6283.
Rose, I. A., Hanson, K. R., Wilkinson, K. D., & Wimmer, M. J. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 2439–2441.
Roy, S., Weinstein, S., Borah, B., Nickol, J., Appella, E., Sussman, J. L., Miller, M., Shindo, H., & Cohen, J. S. (1986) *Biochemistry 25*, 7417–7423.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Chapters 2, 4, 6, and 13, Springer-Verlag, New York.
Sarma, R. H. (1988) *J. Biomol. Struct. Dyn. 6*, 391–395.
Scheffler, F. E., Elson, E. L., & Baldwin, R. L. (1968) *J. Mol. Biol. 36*, 291–304.
Seeman, N. C., Rosenberg, J. M., Suddath, J. J., Park Kim, J. J., & Rich, A. (1976) *J. Mol. Biol. 104*, 109–144.
Senior, M. M. (1988) Ph.D. Thesis, Rutgers University, New Brunswick, NJ.
Senior, M. M., Jones, R. A., & Breslauer, K. J. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6242–6246.
Sharp, K. A., Honig, B., & Harvey, S. C. (1990) *Biochemistry 29*, 340–346.
Singh, J., & Thornton, J. M. (1985) *FEBS Lett. 191*, 1–6.
Spaltenstein, A., Robinson, B., & Hopkins, P. B. (1989) *J. Am. Chem. Soc. 111*, 2303–2305.
Stodola, R. K., Manion, F. J., Wood, W. P., Jr., & Berman, H. M. (1988) DOCK, The Fox Chase Cancer Center, Philadelphia, PA.
Streisinger, G., Ikada, Y., Emrich, J., Newton, J., Tsugita, A., Terzaghi, E., & Inouye, M. (1966) *Cold Spring Harbor Symp. Quant. Biol. 31*, 77–84.

Sundaralingam, M. (1969) *Biopolymers 7*, 821–860.

Sundaralingam, M., Pullman, B., Saenger, W., Sasisekharan, V., & Wilson, H. R. (1973) in *Conformation of Biological Molecules and Polymers, Proceedings of the Jerusalem Symposium on Quantum Chemistry and Biochemistry* (Bergmann, E. D., & Pullman, B., Eds.) Vol. 5, pp 815–820, Academic Press, New York.

Teng, M.-K., Usman, N., Frederick, C. A., & Wang, A. H.-J. (1988) *Nucleic Acids Res. 16*, 2671–2690.

Troll, M., Roitman, D., Conrad, J., & Zimm, B. H. (1986) *Macromolecules 19*, 1186–1194.

Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case D. A. (1986) *J. Comput. Chem. 7*, 230–252.

Wemmer, D. H., Chou, S. H., & Reid, D. H. (1985) *Nucleic Acids Res. 10*, 3755–3772.

Williamson, J. R., & Boxer, S. G. (1989a) *Biochemistry 28*, 2819–2831.

Williamson, J. R., & Boxer, S. G. (1989b) *Biochemistry 28*, 2831–2836.

Wolk, S. K., Harden, C. C., Germann, M. W., van de Sande, J. H., & Tinoco, I. (1988) *Biochemistry 27*, 6960–6967.

Xodo, L. E., Manzini, G., Quadriofoglio, F., van der Marel, G. A., & van Boom, J. H. (1988) *Biochemistry 27*, 6321–6326.

Xodo, L. E., Manzini, G., Quadriofoglio, F., van der Marel, G. A., & van Boom, J. H. (1991) *Nucleic Acid Res. 19*, 1505–1511.