Solution Structure of a Disulfide Cross-Linked DNA Hairpin

Hong Wang,[†,‡] Erik R. P. Zuiderweg,*[,‡] and Gary D. Glick*[,†,‡]

Contribution from the University of Michigan, Ann Arbor, Michigan 48109-1055

Received November 23, 1994[⊗]

Abstract: The three-dimensional structure of a disulfide-stabilized analog of the Dickerson/Drew dodecamer hairpin has been determined by NMR spectroscopy. The structural ensemble was determined from a set of conformations obtained from distance geometry calculations and optimized using restrained molecular dynamics and simulated annealing. The structures were further refined by iterative relaxation matrix approach calculations to include the effects of spin diffusion. To assess the quality of these structures, a signal-to-noise scaled "R" value has been defined that is based on the ratio of measured and back-calculated NOE cross-peaks. In the absence of motional dynamics, the group of ten structures with the lowest R values represents the group of structures that is most compatible with the NMR data. The stem region of the average structure forms a short helix, and the helical parameters and backbone angles fall within the range of observed values for B-DNA. The disulfide cross-link is not as defined as the stem or the loop because only a few experimental constraints are available for this region of the molecule; however, none of the conformations of the cross-link produced during the SA−IRMA refinement distorts the duplex geometry away from B-like DNA. The first three nucleotides in the loop stack over the 5′-end of the helix and are followed by a sharp turn at residue T8 which closes the loop. This pattern of loop folding is similar to that observed in studies of other DNA hairpins possessing tetranucleotide loops.

Introduction

Biophysical characterization of DNA duplexes possessing single-stranded regions continues to be an area of intensive research. Hairpin stem−loop structures have attracted considerable attention in this regard because palindromic or near-palindromic sequences frequently occur in regions of DNA associated with the control of transcription, translation, and other biological functions.[1−4] Although substantial progress has been made in elucidating the structural and thermodynamic parameters of hairpin formation,[5−31] such investigations can be difficult because self-complementary sequences often dimerize or oligomerize. This problem can be particularly troublesome for NMR and crystallographic experiments which require relatively concentrated samples for analysis.

In contrast to sequences that readily fold into hairpins, some inverted repeat sequences only fold into unique stem−loop structures under narrow ranges of temperature, ionic strength, pH, and DNA concentration. One such example is the hairpin formed by the Dickerson/Drew dodecamer, d(CGCGAATTCGCG)$_2$.[32] In buffers containing high [Na^+] this sequence forms a B-like double helix, and the thermodynamic values associated with the duplex to random coil transition have been measured.[7] However, in buffers containing [Na^+] ≤ 10 mM and with [DNA] ≤ 50 μM, biphasic melting profiles are

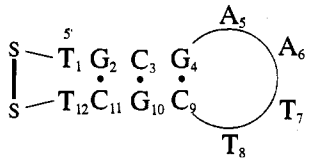
† Department of Chemistry.
‡ Biophysics Research Division.
⊗ Abstract published in *Advance ACS Abstracts,* March 1, 1995.

(1) Müller, U. R.; Fitch, W. M. *Nature* **1982**, *298*, 582−585.
(2) Horwitz, M. S. Z.; Loeb, L. A. *Science* **1988**, *241*, 703−705.
(3) Wells, R. D. *J. Biol. Chem.* **1988**, *263*, 1095−1098.
(4) Bianchi, M. E.; Beltrame, M.; Paonessa, G. *Science* **1989**, *243*, 1056−1059.
(5) 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. *J. Biomol. Struct. Dyn.* **1983**, *1*, 115−129.
(6) Hassnoot, C. A. G.; Hilbers, C. W.; van der Marel, G. A.; van Boom, J. H.; Singh, U. C.; Pattabiraman, N.; Kollman, P. A. *J. Biomol. Struct. Dyn.* **1986**, *3*, 843−857.
(7) Marky, L. A.; Blumenfeld, K. S.; Kozlowski, S.; Breslauer, K. J. *Biopolymers* **1983**, *22*, 1247−1257.
(8) Hilbers, C. W.; Haasnoot, C. A. G.; de Bruin, S. H.; Joorden, J. J.; van der Marel, G. A.; van Boom, J. H. *Biochimie* **1985**, *67*, 685−695.
(9) Wemmer, D. H.; Chou, S. H.; Reid, D. H. *Nucleic Acids Res.* **1985**, *10*, 3755−3772.
(10) Hare, D. R.; Reid, B. R. *Biochemistry* **1986**, *25*, 5341−5350.
(11) Ikuta, S.; Chattopadhyaya, R.; Ito, H.; Dickerson, R. E.; Keams, D. R. *Biochemistry* **1986**, *25*, 4840−4849.
(12) Orbons, L. P. M.; van der Marel, G. A.; van Boom, J. H.; Altona, C. *Nucleic Acids Res.* **1986**, *14*, 4187−4196.
(13) Roy, S.; Weinstein, S.; Borah, B.; Nickol, J.; Appella, E.; Sussman, J. L.; Miller, M.; Shindo, H.; Cohen, J. S. *Biochemistry* **1986**, *25*, 7417−7423.
(14) Erie, D. A.; Sinha, N.; Olson, W. K.; Jones, R. A.; Breslauer, K. J. *Biochemistry* **1987**, *26*, 7150−7159.
(15) Erie, D. A.; Olson, W. K.; Jones, R. A.; Sinha, N.; Breslauer, K. J. *Biochemistry* **1989**, *28*, 268−273.
(16) Chattopadhyaya, R.; Grzeskowiak, K.; Dickerson, R. E. *J. Mol. Biol.* **1990**, *211*, 189−210.
(17) Pramanik, P.; Kanhouwa, N.; Kan, L.-S. *Biochemistry* **1988**, *27*, 3024−3031.
(18) Wolk, S. K.; Hardin, C. C.; Germann, M. W.; van de Sande, J. H.; Tinoco, I. *Biochemistry* **1988**, *27*, 6960−6967.
(19) Xodo, L. E.; Manzini, G.; Quadrifoglio, F.; van der Marel, G. A.; van Boom, J. H. *Biochemistry* **1988**, *27*, 6321−6326.
(20) Xodo, L. E.; Manzini, G.; Quadrifoglio, F.; van der Marel, G. A.; van Boom, J. H. *Nucleic Acids Res.* **1991**, *19*, 1501−1511.
(21) Blommers, M. J. J.; Walter, J. A. L. I.; Haasnoot, C. A. G.; Aelen, J. M. A.; van der Marel, G. A.; van Boom, J. H.; Hilbers, C. W. *Biochemistry* **1989**, *28*, 7491−7498.
(22) Blommers, M. J. J.; Van De Ven, F. J. M.; Van Der Marel, G. A.; Van Boom, J. H.; Hilbers, C. W. *Eur. J. Biochem.* **1991**, *210*, 33−51.
(23) Germann, M. W.; Kalisch, B. W.; Lundberg, P.; Vogel, H. J.; van de Sande, J. H. *Nucleic Acids Res.* **1990**, *18*, 1489−1498.
(24) Paner, T. M.; Amaratunga, M.; Doktycz, M. J.; Benight, A. S. *Biopolymers* **1990**, *29*, 1715−1734.
(25) Rentzeperis, D.; Kharakoz, D. P.; Marky, L. A. *Biochemistry* **1991**, *30*, 6276−6283.
(26) Boulard, Y.; Gabarro-Apra, J.; Cognet, J. A. H.; Guy, A.; Teoule, R.; Guschlbauer, W.; Fazakerley, G. V. *Nucleic Acids Res.* **1991** *19*, 5159−5167.
(27) Holbrook, S. R.; Cheong, C.; Tinoco, I., Jr.; Kim, S.-H. *Nature* **1991**, *353*, 579−581.
(28) Amaratunga M.; Snowden-Ifft, E.; Wemmer, D. E.; Benight, A. J. *Biopolymers* **1992**, *32*, 865−879.
(29) Doktycz, M. J.; Goldstein, R. F.; Paner, T. M.; Gallo, F. J.; Benight, A. S. *Biopolymers* **1992**, *32*, 849−864.
(30) Kang, C. H.; Zhang, X.; Ratliff, R.; Moyzis, R.; Rich, A. *Nature* **1992**, *356*, 126−131.
(31) Blatt, N. B.; Osborne, S. E.; Cain, R. J.; Glick, G. D. *Biochimie* **1993**, *75*, 433−441.
(32) Wing, R.; Drew, H.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Nature* **1980**, *287*, 755−758.

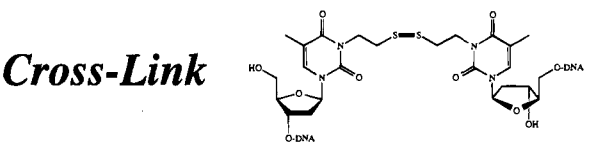
0002-7863/95/1517-2981$09.00/0

obtained. On the basis of concentration-dependent UV thermal denaturation experiments, Breslauer proposed[7] that the first transition (T_m = 33.2 °C) in these biphasic curves defines premelting of the duplex to a hairpin structure while the second transition (T_m = 62.4 °C) represents conversion of the hairpin to a random coil. However, the hairpin formed by initial thermal denaturation of d(CGCGAATTCGCG)$_2$ is inaccessible to structural analysis at high resolution for two reasons. First, this hairpin represents an intermediate along the unfolding pathway of the parent dodecamer, and, although the melting curves show a plateau after the duplex to hairpin transition (at ~40 °C), intermediate states at this point may be populated.[7] Second, neither NMR nor crystallographic measurements are practical at the temperature and DNA concentration where this premelting intermediate is most stable.

We have recently developed a method to stabilize both DNA and RNA with disulfide cross-links that is generally applicable to a range of secondary structures, including duplexes,[33] triplexes,[34] and hairpins.[35,36] We have used this disulfide chemistry to "trap" the hairpin form of the Dickerson/Drew dodecamer (**1**) and have recently provided evidence that the



(1)



cross-link is sterically accommodated by this premelting intermediate analog.[33,37] Here we present the spectral assignments and methodology used to determine the three-dimensional structure of **1**. The NMR structural ensemble was determined from a set of 60 starting structures obtained from distance geometry (DG) calculations, regularized using restrained molecular dynamics (RMD) and simulated annealing (SA), and then taken through three cycles of iterative relaxation matrix approach (IRMA) refinement. To assess the quality of these structures, we have defined a signal-to-noise scaled NMR "*R*" value based on the ratio of measured and back calculated NOE cross-peaks. If effects of dynamics are neglected, we believe that the group of ten structures with the lowest *R* values defined as such represents the spread of structures that is most compatible with the NMR data. The structural features of **1** are discussed and compared to other DNA hairpins containing tetranucleotide loops.

Materials and Methods

1. Sample Preparation. Disulfide cross-linked hairpin **1** was synthesized as previously described by Glick et al.[33] For spectra measured in D_2O, the DNA was lyophilized three times from D_2O to remove residual traces of H_2O and was then dissolved in buffer (50 mM NaCl, 10 mM NaD_2PO_2, pH 7) to a final concentration of 1.5 mM. The spectra measured in H_2O buffer also included 10% D_2O.

2. NMR Spectroscopy. All NMR spectra were measured at 500 MHz on a Bruker AMX 500 instrument at 12 °C unless otherwise stated. The data were transferred to a Silicon Graphics Indigo workstation and were processed using the Felix software (Biosym Technologies, Inc.). Prior to Fourier transformation each data set was apodized with shifted sine-bell window functions and zero filled to twice the original size except where otherwise stated.

To assign the nonlabile protons, DQF–COSY, TOCSY, and NOESY spectra were measured as described below. Briefly, the DQF–COSY spectrum[38,39] was measured in D_2O with presaturation of the residual water resonance. A total of 512 t_1 increments of 2048 complex points were collected with spectral widths in both dimensions of 4504 Hz. A Z-filtered clean TOCSY experiment,[40–43] which suppresses ROESY artifacts using a clean MLEV16 isotropic mixing sequence, was also measured in D_2O. The total spin lock time employed was 75 ms using 10 kHz rf power. The imino and amino proton assignments were made from a two-dimensional (2D) NOESY spectrum (τ_m = 250 ms) acquired in 90% H_2O/10% D_2O with a jump-return read pulse to suppress the water signal.[44]

To analyze the deoxyribose conformations, a phase-sensitive DQF–COSY spectrum was recorded with 1240 t_1 increments, each containing 8192 complex points. The t_2 dimension was zero filled to 16384 points following apodization with a 60°-shifted sine-bell window function, and the t_1 dimension was zero filled to 4096 points after apodization with a 45°-shifted sine-bell function. An E. COSY experiment[45,46] was also performed by combining a DQF–COSY spectrum, measured with 16 scans per t_1 increment, with a TQF–COSY spectrum acquired with 24 scans per t_1 increment. Both spectra were collected with 1518 t_1 increments of 8192 t_2 points over 4500 Hz.

To assign the phosphorous resonances, a nonselective proton-detected $^1H-^{31}P$ heteronuclear correlation experiment[47] was measured with 2048 complex points over a spectral width of 2500 Hz in the proton dimension and 80 data points over a spectral width of 500 Hz in the phosphorous dimension. A modified version of the selective $^1H-^{31}P$ correlation experiment[48] using IBURP shaped pulses[49] was used to determine $^{31}P-H3'$ coupling constants. The spectrum was collected with 2048 complex points in the proton dimension over a 2000 Hz spectral width, 120 points in the phosphorus dimension over a 360 Hz spectral width, and 256 scans per t_1 increment. The data was processed by zero filling the proton dimension in t_2 to 4096 points and zero filling the phosphorus dimension in t_1 to 512 points.

To obtain interproton distance constraints, 2D phase-sensitive NOESY spectra were measured in D_2O (τ_m = 15, 25, 30, 50, 75, 100, 150, and 300 ms) over a single continuous 10 day period without solvent signal suppression. A 7 s recycling delay between transients was used to allow the magnetization to relax to equilibrium. The mixing times were randomly varied by 5 ms for the experiments where τ_m = 15, 25, and 35 ms, and by 10 ms for the experiments with other τ_m values, to suppress zero quantum artifacts.[50] Each spectrum was measured with 512 FIDs of 4096 complex points over a spectral width of 4054 Hz. The data were processed in the t_2 dimension by apodization with 1 Hz

(33) Glick, G. D.; Osborne, S. E.; Knitt, D. S.; Marino, J. P., Jr. *J. Am. Chem. Soc.* **1992**, *114*, 5447–5448.
(34) Goodwin, J. T.; Osborne, S. E.; Swanson, P. C.; Glick, G. D. *Tetrahedron Lett.* **1994**, *35*, 4527–4530.
(35) Glick, G. D. *J. Org. Chem.* **1991**, *56*, 6746–6747.
(36) Goodwin, J. T.; Glick, G. D. *Tetrahedron Lett.* **1994**, *35*, 1647–1650.
(37) Wang, H.; Osborne, S. E.; Zuiderweg, E. R. P.; Glick, G. D. *J. Am. Chem. Soc.* **1994**, *116*, 5021–5022.
(38) Rance, M.; Sorenson, O. W.; Bodenhausen, G.; Wagner, G.; Ernst, R. R.; Wüthrich, K. *Biochem. Biophys. Res. Commun.* **1983**, *117*, 479–485.
(39) Shaka, A. J.; Freeman, R. *J. Magn. Reson.* **1983**, *51*, 169–176.
(40) Rance, M. *J. Magn. Reson.* **1987**, *74*, 557–564.
(41) Griesinger, C.; Otting, G.; Wüthrich, K.; Ernst, R. R. *J. Am. Chem. Soc.* **1988**, *110*, 7870–7872.
(42) Bax, A.; Davis, D. G. *J. Magn. Reson.* **1985**, 65, 355–360.
(43) Braunschweiler, L.; Ernst, R. R. *J. Magn. Reson.* **1983**, *53*, 521–528.
(44) Plateau, P.; Guéron, M. *J. Am. Chem. Soc.* **1982**, *104*, 7310–7311.
(45) Griesinger, C.; Sorenson, O. W.; Ernst, R. R. *J. Am. Chem. Soc.* **1985**, *107*, 6394–6396.
(46) Griesinger, C.; Sorenson, O. W.; Ernst, R. R. *J. Chem. Phys.* **1986**, *85*, 6837–6852.
(47) Sklenar, V.; Miyashiro, H.; Zon, G.; Miles, T.; Bax, A. *FEBS Lett.* **1986**, *208*, 94–98.
(48) Sklenar, V.; Bax, A. *J. Am. Chem. Soc.* **1987**, *109*, 7525–7526.
(49) Geen, H.; Freeman, R. *J. Magn. Reson.* **1991**, *93*, 93–141.
(50) Macura, S.; Huang, Y.; Suter, D.; Ernst, R. R. *J. Magn. Reson.* **1981**, *43*, 259–281.

exponential line broadening followed by zero filling to 4096 complex points. In the t_1 dimension the data were multiplied with a 90° shifted sine-bell function and zero filled to 1024 complex points. The final real matrix size for each spectrum was 4096 × 1024 points.

The volumes of clearly resolved NOESY cross-peaks were obtained by direct integration using the "box method" in the FELIX software. In addition, the cross-peak volumes of several partially overlapped resonances were obtained by simulation using a method similar to that described by Denk et al.[51] Briefly, peak shapes in each dimension of the desired resonance were obtained by extracting the corresponding 1D vectors from non-overlapping peaks belonging to the same spin system. The 1D peak shapes in $\omega 1$ and $\omega 2$ were used to construct a "synthetic peak" by multiplying them in the 2D frequency space. The overlapped peak was then simulated with the normalized synthetic peaks, and the fractional contribution of each peak to the total volume was obtained. The volume of each peak was then calculated from the volume of the overlapping peak using the simulated fraction contribution.

3. Structure Calculations. Structure calculations were performed on a Silicon Graphics 4D360GTX computer using the NMRchitect and DISCOVER (Version 2.9) modules within the Biosym software.

Interproton Distances. Distance constraints for initial structure calculation were derived from initial NOE buildup rates using eq 1

$$r_{ij} = (\sigma_{ref}/\sigma_{ij})^{1/6} r_{ref} \quad (1)$$

where the σ_{ref} and σ_{ij} are the initial buildup rates of the reference peak and target peak, respectively. Initial buildup rates were obtained by fitting the NOE volumes as a function of mixing times to a third-order polynomial of the form $ax + bx^2 + cx^3$ and taking the linear coefficient. The distance of the H5−H6 proton pair of cytosine (2.54 Å) was used as the reference distance for proton pairs that do not involve a methyl group.[52] For proton pairs involving a methyl group, the thymine H6-methyl pseudoatom distance of 2.9 Å was used as reference and another 1 Å distance correction was added to the upper distance boundary. Pseudoatoms were used for H5′ and H5″ protons, which were not stereospecifically assigned.

Torsion Angle Restraints. All the sugar residues, except those of G10 and T12, were restrained to a C2′-endo conformation (south) with a tolerance of ±10° in phase angle and ±8° in pucker angle. Residue G10 was not constrained because pseudorotation analysis was not possible due to resonance overlap. Pseudorotation analysis for T12, which possesses a strongly coupled spin system, indicated that the C2′-endo and C3′-endo (north) conformations are equally populated. Therefore, T12 was also not restrained in the structure calculation (*vide infra*).

The backbone torsion angle ϵ was calculated from the coupling constants between the H3′ protons and the corresponding ^{31}P atom according to the Karplus equation[53,54]

$$J = 15.3 \cos^2 \theta - 6.1 \cos \theta + 1.6 \quad (2)$$

and the relationship

$$\epsilon = \theta - 120 \quad (3)$$

All four solutions of this equation were used in the first phase of the structure calculation. However, only the two energetically feasible angles (antiperiplanar and anticlinal) were produced in the structures and were subsequently used in the structure refinements.

Distance Geometry Calculation. DG calculations were used to generate initial structures for refinement[55] and were performed using the DGII program[56] as implemented in the NMRchitect module of Biosym software. The input distance restraints were smoothed using triangle and sequential-tetrangle algorithms and then were embedded in four dimensions following metrization to improve sampling of conformational space. The embedded structures were majorized and finally optimized using a simple simulated annealing and minimization protocol against the distance error function.

Simulated Annealing and Restrained Molecular Dynamics Calculation. All SA and RMD calculations[57−60] were conducted using the AMBER all-atom force field[61,62] as implemented in the DISCOVER software. The calculations were carried out *in vacuo* with a distance-dependent dielectric constant of the form $4r$. The phosphate charge was reduced to $-0.32e$ to mimic the effects of counterions.[63] A double cutoff distance of 15 Å with 3 Å transition distance was used for nonbonded interactions. The pseudoenergy terms for the NOE-derived restraints (eqs 4−7) possessed the form of a flat well with parabolic sides within a predefined maximum force constant and continued linearly beyond that margin.

$$E_{NOE} = k_u(r - r_u)^2 \quad \text{when} \quad r > r_u \quad (4)$$

$$E_{noe} = 0 \quad \text{when} \quad r_u > r > r_l \quad (5)$$

$$E_{NOE} = k_l(r - r_l)^2 \quad \text{when} \quad r < r_l \quad (6)$$

$$E_{NOE} = k_{max}(r - r_{max}) \quad \text{when} \quad k_u|r - r_u| > K_{max} \text{ or } k_l|r_l - r| > k_{max} \quad (7)$$

Prior to SA, the structures produced by the DG calculations were energy minimized with reduced force constants to alleviate bad contacts. The starting temperature of each structure (300 K) was then quickly raised to 1000 K and kept at that temperature for 15 ps. The NOE restraints and the covalent force constants were scaled faster than the nonbonded interactions to increase the conformational search efficiency (at 300 K the NOE restraint was set at 40 kcal mol^{-1} Å^{-2} and reached a maximum of 1000 kcal mol^{-1} Å^{-2} at 1000 K). The van der Waals radii for all atoms were also scaled by a factor of 0.85. To facilitate crossing of conformational barriers during the high-temperature search, a simple quadratic nonbonded energy term was used in place of the normal Lennard-Jones nonbonded energy term in the simulation.[59] The system was then slowly cooled to 300 K over 10 ps, equilibrated at 300 K for 10 ps, and energy minimized using the NMR constraints and the AMBER force field to a first-derivative root mean square deviation (rmsd) of 0.1 Å^2/kcal.

IRMA and Back-Calculation. IRMA[64,65] was carried out using a hybrid experimental/model relaxation matrix calculation combined with the SA protocol described above for iterative structure calculation. For both IRMA and back-calculation, an overall correlation time of 2.3 ns was employed. This value was estimated by fitting the NOE mixing time series of cytosine H5−H6 proton pairs and their diagonal peaks using the IRMA program. NOE intensities of all the mixing times were weighted equally in the IRMA calculation.

Results

Labile Proton Resonance Assignments. A thermal denaturation study was conducted to examine the stability of the three G·C base pairs in **1**. Between 1−20 °C only three base-

(51) Denk, W.; Baumann, R.; Wagner, G. *J. Magn. Reson* **1986**, *67*, 386−390.
(52) Wüthrich, K. *NMR of Proteins and Nucleic Acids*; John Wiley & Sons, Inc.: New York; 1986.
(53) Lankhorst, P. P.; Haasnoot, C. A. G.; Erkelens, C.; Altona C. *J. Biomol. Struct. Dyn.* **1984**, *1*, 1387−1405.
(54) Gorenstein, D. G. *Methods Enzymol.* **1992**, *211*, 254−286.
(55) Crippen, G. M.; Havel, T. F. *Distance Geometry and Molecular Conformation*; Research Studies Press: Taunton, Somerset, England; 1988.
(56) Havel, T. E. *Prog. Mol. Biol. Biophys.* **1991**, *56*, 43−78.
(57) Zuiderweg, E. R.; Scheek, R. M.; Boelens, R.; van Gunsteren, W. F.; Kaptein, R. *Biochimie* **1985**, *67*, 707−715.
(58) Clore, G. M.; Nilges, M.; Sukumaran, D. K.; Brunger, A. T.; Karplus, M.; Gronenborn, A. M. *EMBO J.* **1986**, *5*, 2729−2735.
(59) Nilges, M.; Clore, G. M.; Gronenborn, A. M. *FEBS Lett.* **1988**, *239*, 129−136.
(60) Schmitz, U.; Pearlman, D. A.; James, T. L. *J. Mol. Biol.* **1991**, *221*, 271−292.
(61) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S., Jr.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765−784.
(62) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. *J. Comput. Chem.* **1986**, *7*, 230−252.
(63) Tidor, B.; Irikur, K. K.; Brooks, B. R.; Karplus, M. *J. Biomol. Struct. Dyn.* **1983** *1*, 231−252.
(64) Boelens, R.; Koning, T. M. G.; Kaptein, R. J. *Mol. Struct.* **1988**, *173*, 299.
(65) Boelens, R.; Koning, T. M. G.; van der Marel, G. A.; Van Boom, H.; Kaptein, R. *J. Magn. Reson.* **1989**, *82*, 290−308.

Table 1. Proton Resonance Assignments for 1[a]

base	H1′	H2′	H2″	H3′	H4′	H5′(H5″)	H5/H5M/H2	H6/H8	imino	amino
T1	5.88	1.88	2.37	4.74	4.13	3.73/3.75	1.81	7.46		
G2	5.97	2.74	2.82	5.01	4.41	4.04/4.13		8.06	13.18	
C3	5.70	1.74	2.21	4.77	4.16	4.13/4.23	5.46	7.33		6.74/8.64
G4	5.63	2.37	2.62	4.92	4.34	3.96/4.18		7.59	13.14	
A5	6.03	2.61	2.69	4.94	4.37	4.18/4.21	7.87	7.93		
A6	5.97	2.57	2.63	4.88	4.33	4.27/4.12		8.00		
T7	5.90	2.07	2.29	4.74	4.29	4.03/4.04	1.43	7.22		
T8	6.23	2.33	2.64	4.67	4.26	4.03/4.18	1.95	7.56		
C9	5.52	2.27	2.46	4.88	4.23	4.08/4.14	5.92	7.70		7.22/8.77
G10	6.03	2.72	2.78	5.07	4.43	4.12/4.17		8.02	13.22	
C11	6.25	1.99	2.25	4.90	4.23	4.16/4.31	5.57	7.42		7.23/8.77
T12	6.29	2.31	2.37	4.56	4.10	4.07/4.16	1.95	7.74		7.23/8.77

[a] The proton chemical shifts are referenced to an external TSP standard.

paired imino protons centered at 13.18 ppm are observed. Above 20 °C these peaks broaden, presumably due to chemical exchange with buffer. Two of these signals are of comparable intensity while the third one, which resonates furthest upfield at 13.14 ppm, is ~30% lower in intensity. These imino protons were assigned from their NOE connectivity to the amino protons of their base-paired partners C11, C9, and C3, which were assigned based on the strong NOE cross-peaks to their associated H5 and H6 protons. The upfield imino proton (13.14 ppm) that has lower intensity corresponds to the G·C base-pair directly adjacent to the loop. As mentioned in the Discussion section, this may arise due to conformational flexibility at the stem–loop interface. Based on the results of these experiments, we measured all spectra at 12 °C to ensure hydrogen bonding in the helical regions while providing maximum chemical shift dispersion.

Nonlabile Proton Resonance Assignments. The nonlabile protons were assigned using DQF–COSY, TOCSY, and long mixing time NOESY spectra as described by Wüthrich.[52] These data are summarized in Table 1. The stereospecific assignments of H2′ and H2″ were based on the relative size of the $J_{1'}$, $J_{1'2''}$, $J_{2'}$, and $J_{2''3'}$ coupling constants obtained from the DQF–COSY spectrum along with the relative intensity of the NOE cross-peaks between the H1′ and H2′/H2″ protons.[52] The H5′ and H5″ protons were assigned from both the H3′–H5′/H5″ NOE connectivities and H3′–phosphorous–H5′/H5″ correlations observed in the ^{31}P–^{1}H COSY spectrum (the H3′ protons assigned prior to this latter correlation). No attempt to stereospecifically assign the H5′ and H5″ protons was made. An expansion of the base to H1′ region of the NOESY spectrum illustrating the quality of the data and the sequential assignments is given in Figure 1. The connectivity network indicates that the G·C stem forms a B-like duplex that is somewhat interrupted at the A5–A6 step and completely broken at the T7–T8 step in the putative loop. A strong NOE cross-peak is observed between the H1′ proton and the H6 base proton of A5, which indicates that this residue adopts a syn-geometry about the glycosidic torsion angle.

The H2 protons of A5 and A6 were easily identified by their characteristically long T1 relaxation times (>2 s). However, their sequential assignments could not be unambiguously confirmed without data from the long mixing time NOESY spectrum measured in D_2O. Specifically, the H2 resonance at 7.87 ppm shows weak NOE cross-peaks to the H2′ and H2″ protons of G4 and was therefore assigned to residue A5. The other H2 proton displays weak cross-peaks to the H2′ and H2″ protons of A5 and methyl protons of T8 and was thus assigned to residue A6.

Pseudorotation States of the Deoxyribose Sugars. Scalar coupling constants and summations of the scalar coupling constants were obtained from both an E. COSY experiment and from the simulation of a high-resolution DQF–COSY spectrum using the program CHORDS.[66] Coupling constants corresponding to $J_{1'2'}$, $J_{1'2''}$, and $J_{2'}$ were obtained from simulation of H1′–H2′, H1′–H2″, and H2′–H3′ cross-peaks, respectively, whereas summations of 2′, 2″, and 3′ proton–proton coupling constants were directly measured from the COSY cross-peaks. These data are summarized in Table 2. The values of $J_{1'}$, $J_{1'2''}$, $J_{2'}$, and $J_{2''3'}$, suggest that, with exception of residues G10 and T12, each of the sugar moieties predominantly adopts a C2′-endo conformation. The pseudorotation states were obtained by fitting each of the coupling constants and summations of coupling constants to a two-state equilibrium (C3′-endo ⇌ C2′-endo) using a grid search of pseudophase angles for every 8° and pucker angles of 28, 36, and 44°.[67,68]

Phosphorus Coupling Constants and ϵ Backbone Torsion Angle. The sequential assignments of phosphorus resonances were obtained from a ^{1}H–^{31}P heteronuclear correlation experiment using the H3′ proton assignments (Figure 2). The ^{1}H–^{31}P coupling constants were then determined from a H3′ selective COSY experiment (Table 3). In our hands the pulse sequence reported by Sklenar[48] produced a streak of strong t_1 noise at the water frequency that interfered with several of the H3′ resonances. To circumvent this problem a 180° pulse was inserted in both the proton and the phosphorus channels to separate evolution of the coupling constants from the chemical shift evolution. In this scheme both 180° pulses can be centered in their respective evolution periods (Figure 3). As seen in Figure 4, this pulse sequence affords excellent water suppression and does not produce the strong noise obtained with the original pulse sequence.

Structure Calculations. To generate an ensemble of structures consistent with the experimental data, possessing the least starting-structure bias, we conducted an extensive search of conformational space that combined DG, high-temperature SA, and IRMA calculations. First, starting structures were generated by DG using the distance restraints obtained from the NOE buildup rates and the torsion restraints derived from the coupling constants analysis. Only those NOE's that had buildup curves without a lag time as judged from mixing time plots were included as restraints. The restraint list consisted of 130 interproton distances, 40 torsion angles for the sugars, and the 11 ϵ backbone torsion angles. In addition, nine hydrogen-bond restraints for the three G·C base pairs were included (no other planarity constraints were introduced). The structures produced using these restraints were optimized by SA to generate starting

(66) Majumdar, A.; Hosur, R. V. *Prog. Nucl. Magn. Reson. Spectrosc.* **1992**, *24*, 109–158.
(67) Rinkel, L.; Wijk, J. van, Huckriede; Altona, C. *J. Biomol. Struct. Dyn.* **1987**, *4*, 621–649
(68) Wijk, J. van, Huchriede, B. D.; Ippel, J. H.; Altona, C. *Methods Enzymol.* **1992**, *211*, 286–306.

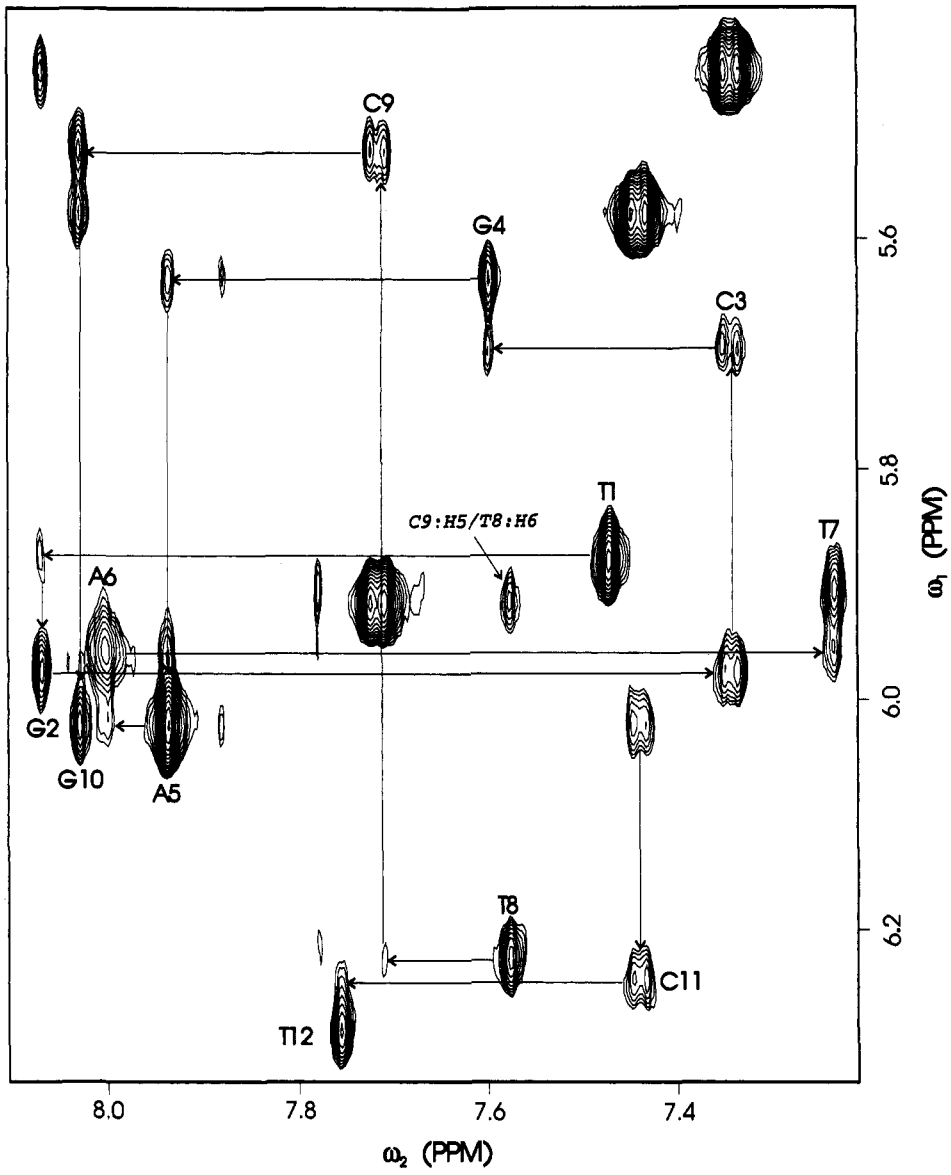


Figure 1. Contour plot of base to H1′ region of NOESY (τ_m = 300 ms) measured in D_2O buffer at 12 °C showing sequential assignments. Note the absence of connectivity at the A5−A6 and T7−T8 steps.

Table 2. Proton−Proton Coupling Constants and Summation of Coupling Constants of the Deoxyribose Sugars

residue	$J_{1'2'}$	$J_{1'2''}$	$J_{2'2''}$	$J_{2'3'}$	$J_{2''3'}$	Σ1′	Σ2′	Σ2″	Σ3′	% south
T1	8.5	5.8	−14.0	5.7	<2[a]	14.4	29.2	21.0	11	80
G2	9.8	5.3			<2	15.4	28.3	21.9	11	90
C3	8.2	6.1	−13.8	7.3	<2	14.3	29.8	22.8	12	85
G4	9.7	5.4	−13.4	5.8	<2	15.7	30.2	22.7		90
A5	8.7	4.8	−13.8	7.0	<2		29.5			75
A6	8.6	5.6			<2	14.6	29.8	20.9	11	18
T7	9.2	5.8	−13.9	5.4	<2	14.9	30.0	20.9	13	85
T8	8.2	6.6	−14	7.2	<2	14.9	31.1	21.7	11	80
C9	9.6	5.2	−13.6	7.2	<2	14.8	28.9	21.2	12	90
G10	9.1	5.8			<2	15.0	b	b	11	
C11	7.5	6.5	−13.8	7.2	<2	14.4	30.5	22.8		75
T12	6.8	6.4	−14.4	6.3	4.5	13.6	c	c	14	50

[a] Based on the absence of H2″−H3′ COSY peak. [b] Overlapped. [c] Strong coupling.

conformations for subsequent hybrid SA−IRMA refinement. A total of 60 starting conformations were generated and analyzed for restraint violations and structure regularity. Of the 60 starting structures, 17 possessed restraint violations > 1 Å or contained impossible covalent geometries and were therefore not considered for further analysis.

The remaining 43 geometries were refined as follows. To begin, a new set of distance restraints were calculated for each structure using IRMA. Only those distances resulting from clearly resolvable cross-peaks in the experimental NOESY spectra were included in this new restraint file. This new restraint file, aside from the torsion angle and hydrogen bond constraints discussed above, consisted of 177 interproton distances, including some additional distances derived from cross-peaks with spin diffusion. We found that these additional restraints helped to improve the match between the back-calculated and experimental spectra.[69] The force constants for these additional restraints were set to half the value of other NOE restraints. Figure 5 shows the NOE connectivities observed between interresidue protons in the loop. Several generic distance restraints with only a lower-bound of 4 Å were added to constrain proton pairs that show strong peaks present in the back calculated spectra at all mixing times, but not present in experimental spectra. Before using these generic lower bound restraints, two conditions were checked. First, we verified that the additional cross-peaks were originated from direct NOE

(69) Banks, K. M.; Hare, D. R.; Reid, B. R. *Biochemistry* **1989**, *28*, 6996−7010.

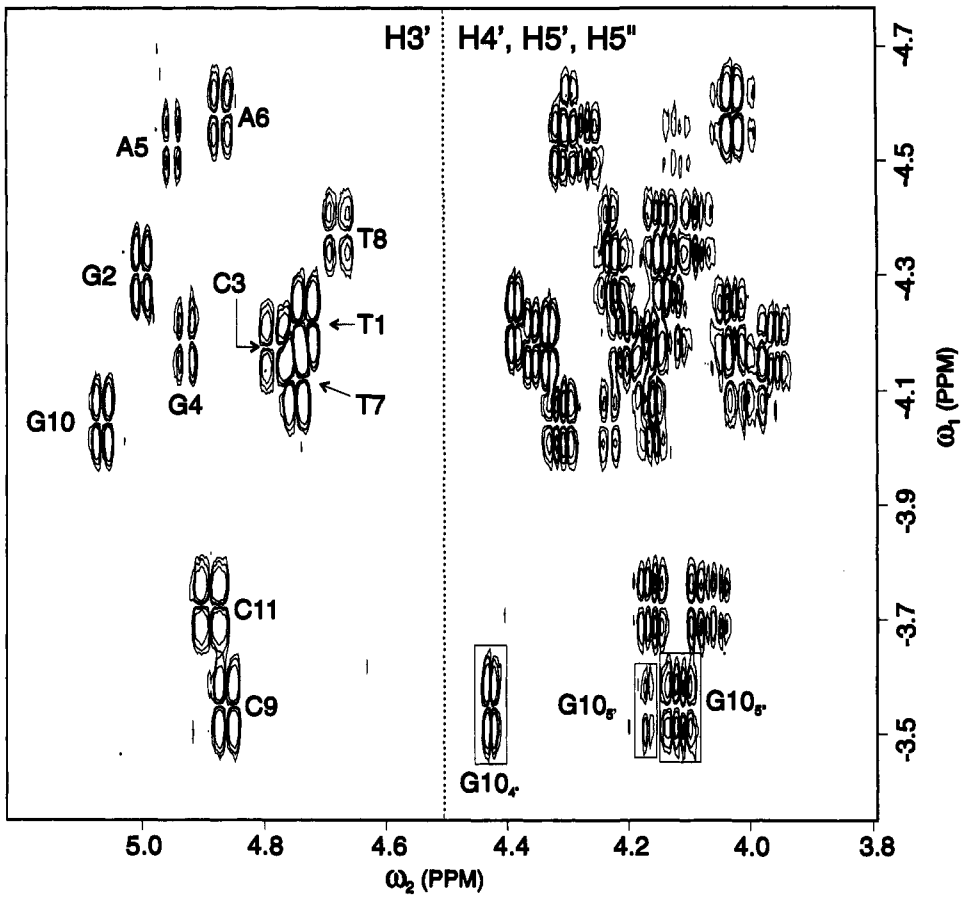


Figure 2. Phosphorus–proton heteronuclear correlation spectrum. Peaks corresponding to the internucleotide H3′($i-1$)–P(i) connectivity are labeled with the nucleotide containing the H3′($i-1$). The intranucleotide H4′(i)–P(i), H5′(i)–P(i), and H5″(i)–P(i) of G10 are labeled to show an example of the sequential connectivity through phosphorus.

Table 3. Phosphorus Chemical Shifts and Phosphorus–H3′ Proton Coupling Constants[a]

residue	chemical shift	coupling constant	residue	chemical shift	coupling constant
T1	−4.20	5.7	T7	−4.09	7.0
G2	−4.28	3.6	T8	−4.35	5.2
C3	−4.16	5.6	C9	−3.53	6.6
G4	−4.16	3.8	G10	−4.03	4.3
A5	−4.51	3.7	C11	−3.70	7.0
A6	−4.56	3.9			

[a] The phosphorus chemical shifts are referenced to an external TMP standard.

transfers between the proton pairs involed rather than through spin diffusion by NOESY calculation using a two-spin approximation. Second, we checked that the protons which give rise to the extra calculated peaks did have NOEs with other protons so that the absence of the cross-peak was not due to local dynamics. The new set of distance restraints together with the base-pair and torsion angle restraints was then used in the SA–IRMA refinement protocol. The entire SA–IRMA refinement was iterated three times to minimize the differences between the experimental and back-calculated spectra.[69]

Each of the final 43 structures possesses a similar overall geometry (Figure 6). The average heavy atom rmsd for superimposing each structure in this ensemble on the average structure is 1.1 Å. The loop and terminal residues are less defined than residues in the stem: the heavy atom rmsd of the stem alone is 0.5 Å, whereas the rmsd of the whole molecule excluding the T1–T12 cross-link is 0.8 Å. To obtain a final conformational ensemble that is most consistent with the experimental data we combined visual comparison of simulated and experimental NOESY spectra along with quantitative R value calculations. Visual comparison was important because it enabled comparison of peaks that could not be included in the R value calculation because of resonance overlap. A comparison of the H6–H1′/H3′ region of the experimental and simulated NOESY spectra is shown in Figure 7. The experimental and simulated spectra are in good overall agreement, with the greatest difference occurring for the H6–H1′ NOE associated with residue T1. As discussed latter, this deviation appears to be caused by local conformational flexibility.

Visual comparison of the experimental and simulated spectra is limited because the fit is predominantly judged on the basis of peak height. Peak height is dependent on both line width and resonance multiplicity, and these parameters are hard to estimate. An R value calculation typically uses peak volumes and is therefore more suited for quantitative comparisons. R values are often calculated following eq 8

$$R = \frac{\sum |A^{\exp} - A^{\sin}|}{\sum A^{\exp}} \tag{8}$$

where $A^{\exp}$ and $A^{\sin}$ denote the experimental and simulated NOE intensities, respectively.[70] The average R value calculated using this formula for the 43 converged structures is 0.14. However, eq 8 overemphasizes strong peaks and is not sensitive to smaller, less intense resonances that often contain long range structural information important for defining global conformation. This problem is circumvented by using a $Q^{1/6}$ factor, where the NOE dynamic range is compressed by taking the sixth-power root of

(70) Gonzalez, C.; Rullmann, J. A. C.; Bonvin, A. M. J. J.; Boelens, R.; Kaptein, R. *J. Magn. Reson.* **1991**, *91*, 659–664.

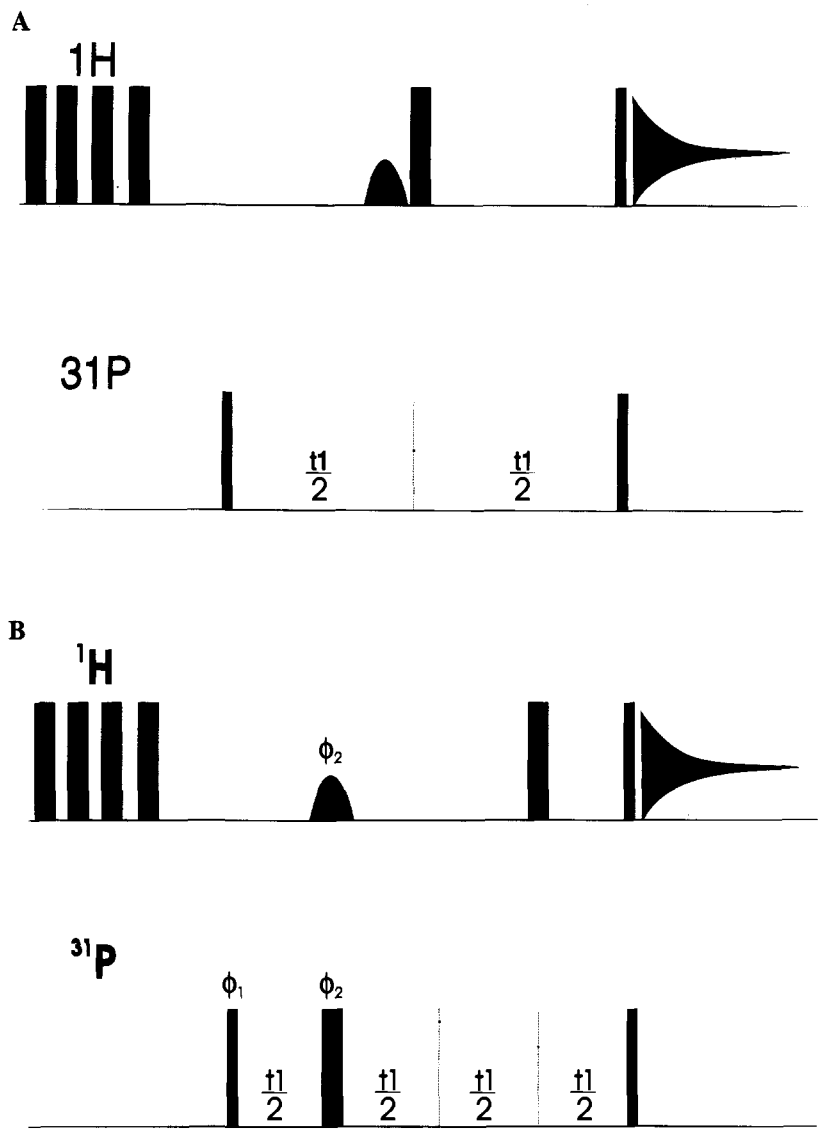


Figure 3. Pulse sequences for the phosphorus–proton selective correlation experiment. (A) Original Sklenar[48] pulse sequence. (B) Modified pulse sequence with all the 180° pulses symmetrically positioned. The shaped pulse is a selective inversion IBURP pulse centered on the H3′ protons. Narrow bars corresponsd to 90° flip angle pulses and wide bars to 180° pulses. A train of 180° proton pulses at the start of the pulse sequence is used for water suppression. All pulses are applied along x axis unless indicated. The phase cycle is as follows: $\phi_1 = 4(x), 4(-x)$; $\phi_2 = x, y, -x, -y$; rec. $= x, -x, x, -x, -x, x, -x, x$.

the value of the observables.[71,72] This definition is obviously based on the analogy of the sixth-power root scaling of distance in the *two-spin* approximation of NOE buildup. Because NOE buildup is not of interest here, we chose to define a model-free R factor in which the ratio of experimental and calculated peak values is assessed. For the accurate structure, this ratio should obviously be equal to one for every pair of calculated and observed NOE cross-peaks, irrespective of their volume. To overcome the extreme sensitivity to poor signal-to-noise ratios for this definition[70] we developed a signal-to-noise weighted R value for the analysis of our structures:

$$R = \frac{\sum w(1-q)}{\sum w} \tag{9}$$

where

$$q = \begin{Bmatrix} A^{exp}/A^{sim} & \text{if } A^{exp} \le A^{sim} \\ A^{sim}/A^{exp} & \text{if } A^{exp} > A^{sim} \end{Bmatrix} \tag{10}$$

and

$$w = 2\left[\frac{A^{exp}}{A^{exp} + A^{no}/2} - 0.5\right] \tag{11}$$

and A^{no} is the rms noise value, which was estimated by integrating some randomly selected noise areas in the spectra. This weighting function ensures that a peak does not contribute to the R value if the signal-to-noise (S/N) ratio is <0.5, but contributes fully for S/N ratios >0.5. Negative weights do not occur since peaks with S/N ≤ 0.5 can not be measured.

Quality of the Structure. The structure determination paradigm used here was designed to extensively sample conformational space for an ensemble that is compatible with the NMR data (i.e., full incorporation of spin-diffusion effects by relaxation matrix calculations). The possibility of starting-structure dependency in IRMA was minimized by carrying out

(71) Whitka, J. M.; Srinivasan, J.; Bolton, P. H. *J. Magn. Reson.* **1992**, *98*, 611–617.
(72) Bonvin, A. M. J. J.; Vis, H.; Breg, J. N.; Burgering, M. J. M.; Boelens, R.; Kaptein, R. *J. Mol. Biol.* **1994**, *236*, 328–341.

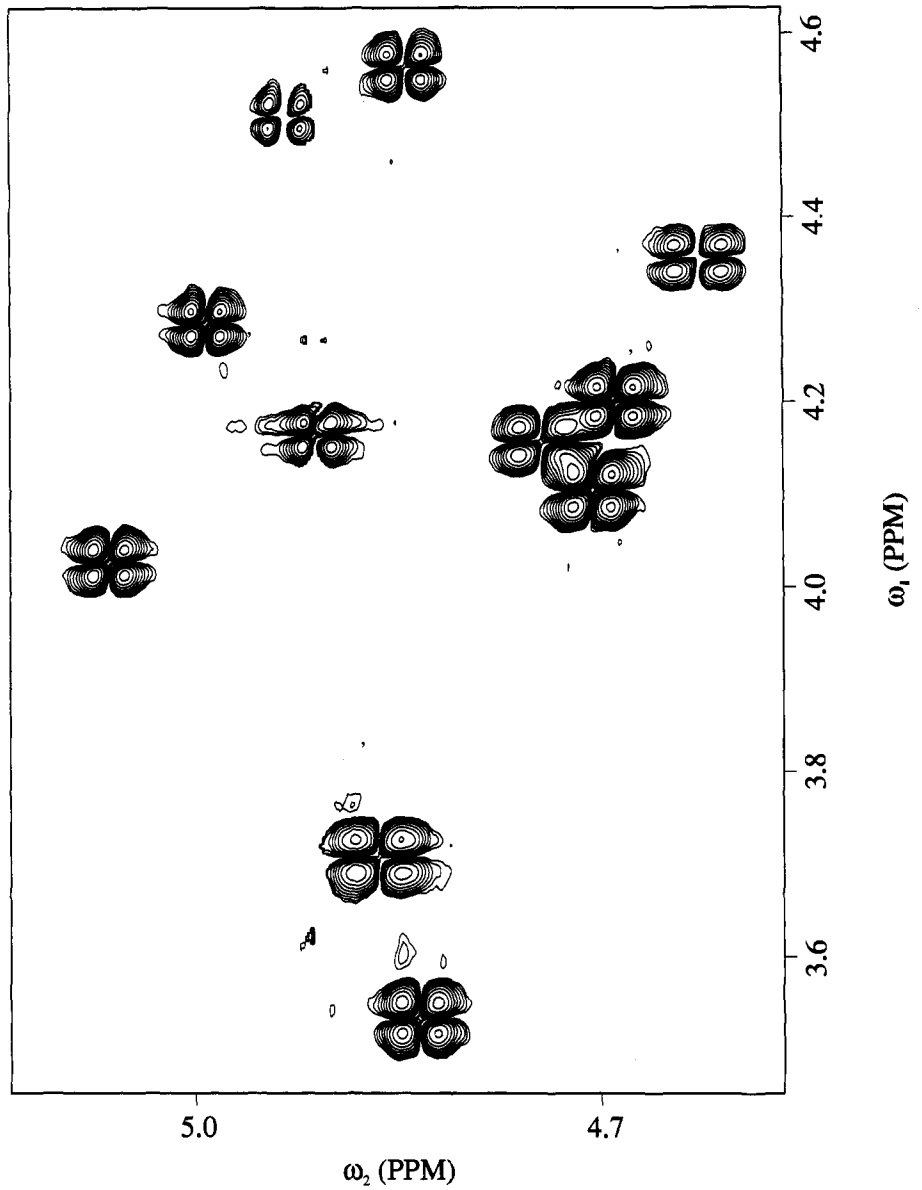


Figure 4. Phosphorus–H3′ proton selective correlation spectrum used to measure phosphorus to H3′ coupling constants. The spectrum was measured using the pulse sequence shown in Figure 3B.

the refinement procedure on many structures in parallel. Further "randomization" of IRMA-driven structures was assured by only using those distances refined by IRMA that corresponded to observable NOE cross-peaks in the IRMA–SA procedure. We found that an IRMA–SA refinement cycle carried out in this way converged after three cycles. The average R value of the 43 refined structures obtained using eq 9 is 0.27 with standard deviation of 0.006. However, the ensemble of the ten structures with the lowest R values defines a tighter group as judged from computer graphics (Figure 8; Tables 4 and 5). Thus, the tightest ensemble appears to correspond to the groups of structures with the best R values, and vice versa. These observations indicate that the R value defined by eq 9 can discriminate between subtle conformational differences and is useful in assessing the quality of the structure.

Setting the threshold of acceptance at these ten structures with roughly equal R values is, however, an arbitrary choice. It would be interesting to use a (semi)independent method as proposed by Brünger[73] to assess the quality ("accuracy") of this NMR-derived DNA structure. In this approach, molecular models based on 90% of the restraint data are tested to predict the randomly chosen remaining 10% of the constraints. The success of this prediction ("cross-validation") is a measure for the quality of the structure. However, this method is limited for evaluation of DNA structures derived from NMR due to the sequential nature of the NOE observables in these molecules; it will be difficult to define a group of NOEs that *can* successfully predict a small test set of remaining NOEs, even in a perfect structure. We therefore appear to be left with the R value based ensemble choice. Moreover, although motional heterogeneity in **1** is limited (the resonances have a uniform line width), the definition of a precise ensemble with or without cross-validation is unreliable in the absence of better dynamical information (e.g., ^{13}C-relaxation[74]) and possibly dynamically averaged restraints.[75,76] Therefore, we believe that the group of ten structures with the lowest R value as computed from an

(73) Brünger, A. T.; Clore, G. M.; Gronenborn, A. M.; Saffrich, R.; Nilges, M. *Science* **1993**, *261*, 328–331.

(74) Palmer, A. G., III; Rance, M.; Wright, P. E. *J. Am. Chem. Soc.* **1991**, *113*, 4371–4380.

(75) Torda, A. E.; Scheek, R. M; Van Gunsteren, W. F. *J. Mol. Biol.* **1990**, *214*, 223–235.

(76) Bonvin, A. M. J. J.; Boelens, R.; Kaptein, R. *J. Biomol. NMR* **1994**, *4*, 143–149.

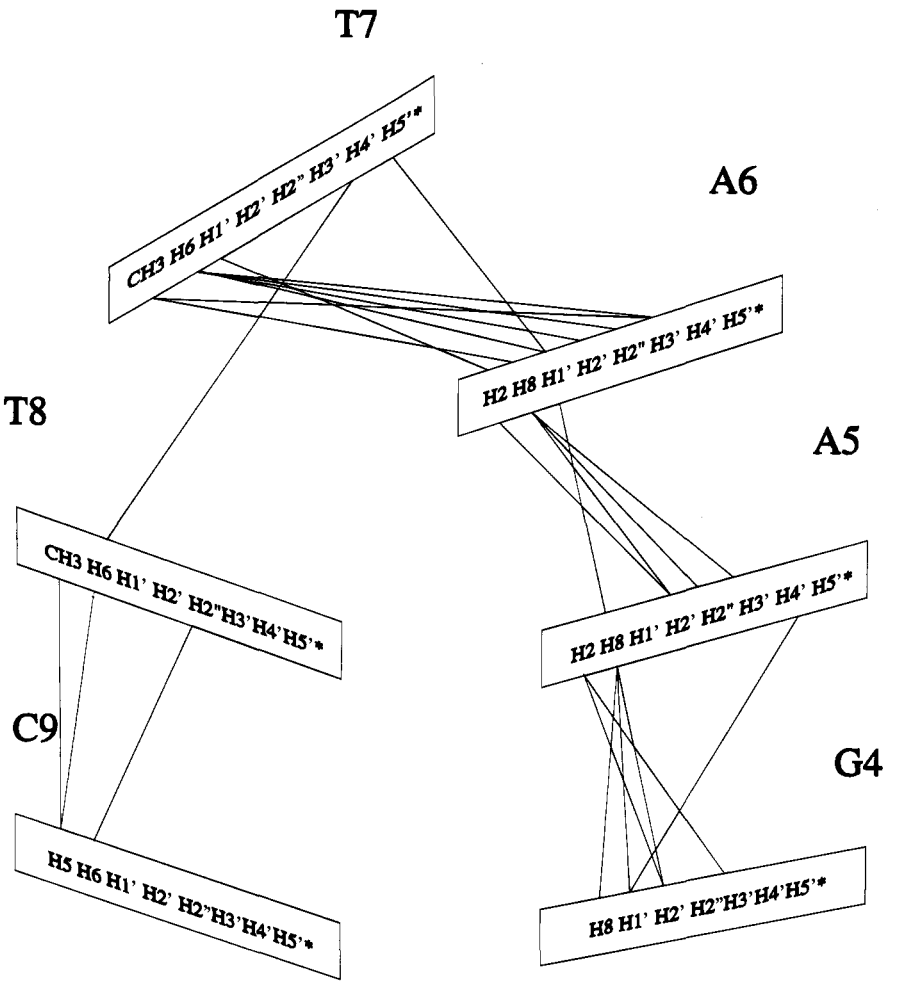


Figure 5. Schematic representation of the NOE connectivities between the interresidue protons in the loop of **1**.

Figure 6. Superimposition of the 43 structures after IRMA refinement.

ensemble-oriented full-relaxation matrix method represents the best precision and accuracy we currently can define. A limited statistical analysis of the properties of this ensemble is given in Table 6.

Discussion

Structural Features: Stem. The stem region of **1** (Figure 8) is the best-defined part of the molecule (rmsd 0.43 Å). The helical parameters listed in Table 5, along with the values for the backbone and deoxyribose torsion angles (Table 4), indicate that the stem duplex of **1** is similar to most other B-like helices.[77] These data indicate that the disulfide cross-link does not disrupt the structure of B-DNA (*vide infra*).

Structural Features: Loop. Three lines of evidence suggest that the first nucleotide in the loop (A5) is moderately flexible. First, the deoxyribose ring of A5 has a higher proportion of the C3′-endo conformation (25%) relative to the sugars of other residues. Second, the line width of the A5 resonances are different than those for the other bases in the loop. Third, while the A5:H8−A5:H1′ distance (~2.6 Å) indicates that A5 is syn about χ, the presence of a weak A5:H8−A5:H3′ NOE cross-peak is somewhat more consistent with a gauche-like conformation for this torsion angle (if χ were anti, unfavorable steric interactions with G4 would occur). However, the distance between H8 and H3′ is nearly 6 Å for a syn glycoside and clearly cannot give rise to this cross-peak. We believe that these data are best accommodated not by a single rigid structure but rather by an equilibrium between at least two conformers, where the syn-geometry is the predominant structure. This finding is consistent with recent theoretical[78] and experimental[79,80] results which demonstrate the importance of analyzing hairpin loop geometry in terms of multistate conformational equilibria.

(77) Saenger, W. *Principles of Nucleic Acid Structure*, Springer-Verlag NewYork, 1984.

(78) Erie, D. A.; Suri, A. K.; Breslauer, K. J.; Jones, R. A.; Olson, W. K. *Biochemistry* **1993**, *32*, 436−454.

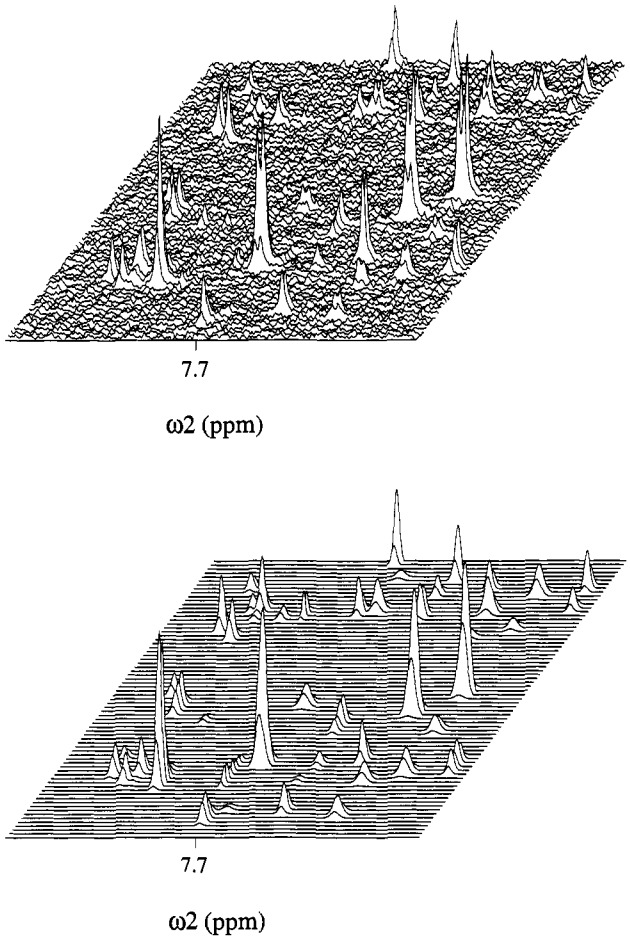


Figure 7. Comparison of the stack plots of the base to 1′ and base to 3′ regions of the experimental NOESY and simulated NOESY spectra (τ_m = 300 ms).

The first three bases in the loop are parallel to the bases on the 5′ end of the stem. In the majority of conformers shown in Figure 6 this trinucleotide segment partially stacks over G4. Yet for a smaller group of structures the entire tetranucleotide loop is rotated relative to the average structure, allowing for additional stacking of A5−T7 over G4. Such motion at the stem−loop interface is caused by variations in the backbone torsion angle ϵ and has recently been observed in NMR studies of d($T_8C_4A_8$).[81] Although this structural variation can be caused by an intrinsic flexibility of the stem−loop interface, there is a lack of NOE constraints between residues T7 and T8 due to a sharp turn between them (see Figure 5). Thus, this "apparent flexibility" may also arise from an insufficient number of experimental constraints that precisely defines the relative orientation of these two domains.

Proceeding further through the loop, there is a sharp turn between residues T7 and T8 which acts to connect T8 to the 3′ end of the stem duplex. The presence of a turn between T7 and T8 is clearly supported by the absence of sequential connectivities for these residues in the NOESY spectra. At the 3′ side of the stem−loop interface, T8 is stacked over both C9 and A5. The four bases in the loop project into the major groove where they are partially shielded from solvent, whereas the sugar-phosphate backbone forms a hydrophilic face and projects toward solvent. Because of the turn, the base at T7 is partially exposed to the solvent with the methyl group pointing away from the major groove. An "amphipathic" geometry like this has recently been predicted for DNA hairpins and may help regulate both protein•DNA interactions and controlled aggregation in solution.[78]

The overall folding pattern of **1** agrees with the empirical model of loop formation proposed by Haasnoot et al.[6] in which helical geometry and stacking is extended from the 3′ end of the stem duplex before making a 180° turn at the 5′ end to close the loop. The backbone angles at the T7−T8 step generating the 180° turn, $\zeta(+)\alpha(+)\beta(t)\gamma(t)$, are similar to those found in other tetraloop hairpins. These data confirm previous studies[80] which suggest that this combination of torsion angles is energetically quite favorable. Base pairing between the first and last loop residues has previously been noted for several DNA hairpins. For example, Blommers observed Hoogsteen hydrogen bonding between T8 and A11 in d(ATCCTAA-TTTA-TAGGAT)[22] and wobble base pairs have been proposed for both d(CGCG-TTTT-CGCG)[10] and d(ATCCTAA-TTTT-TAGGAT).[82] However, steric crowding in the loop coupled with the syn glycoside for A5 precludes such interactions for **1**.

Structural Features: Terminus. Although the terminal residues are constrained by the disulfide bond, three pieces of data suggest that residue T1 is quite flexible. First, the sugar pucker of T1 cannot be represented by a single conformation but only by an equilibrium between several ring conformers (we modeled this as an equilibrium between the C2′-endo and C3′-endo geometries, where both states are equally populated). Second, the NOE buildup rate for the H2′−H2″ cross-peak is significantly smaller than that of the other residues, which suggests a smaller correlation time and increased conformational mobility. Third, the intensities of the H6−H2′ and H6−H2″ NOE cross-peaks indicate that the glycosidic torsion angle for T1 is anti (−90° to 90°) whereas the T1:H6−T1:H1′ NOE intensity is more consistent with a high anti rotomer (−90° to −70°). This latter cross-peak also is poorly reproduced by back-calculation, and the corresponding distance restraint is violated by ≥0.1 Å for each structure in the final conformational ensemble. To determine if this violation was caused by error propagation through other constraints and not by local dynamics, the restraints on neighboring protons were gradually relaxed and the SA−IRMA refinement was repeated. The restraint was still violated in each of these new structures. It is important to note that although this region of the molecule cannot be precisely determined, none of the conformations of the cross-link in the final ensemble disrupts Watson−Crick base-pairing in the stem duplex (Figure 8).

There are a limited number of experimental constraints for the cross-link, and this substructure is the least defined part of the molecule. However, there are several indications that one of the alkylthiol linkers on either the 3′- or 5′-modified thymidine may have a preferred conformation. Both linkers show similar connectivities in the NOESY spectra, yet differences in the COSY and TOCSY multiplet fine structure are observed. On one side of the duplex the COSY and TOCSY patterns are similar to the NOESY multiplet structure which suggests that the coupling constants between the geminal proton pairs are averaged on the NMR time scale. By contrast, the COSY and TOCSY patterns of the other linker show that two of the four geminal coupling constants are small indicating a preferred (gauche) geometry. We have observed similar results in other sequences containing this cross-link which suggest that

(79) Baxter, S. M.; Greizerstein, M. B.; Kushlan, D. M.; Ashley, G. W. *Biochemistry* **1993**, *32*, 8702−8711.
(80) Mooren, M. M.; Pulleyblank, D. E.; Wijmenga, S. S.; van der Ven, F. J.; Hilbers, C. W. *Biochemistry* **1994**, *33*, 7315−7325.
(81) Zhou, N.; Vogel, H. J. *Biochemistry* **1993**, *32*, 637−645.
(82) Blommers, M. J. J.; Haasnoot, C. A. G.; Hilbers, C. W.; van Boom, J. H.; van der Marel, G. A. in *Structure and Dynamics of Biopolymers*, NATO ASI Series E, 133; Nijhoff: Boston, 1987; pp 78−91.

Figure 8. Stereoview of the superimposition of the ten best structures judged from the calculation of R values.

Table 4. Backbone and Glycosidic Torsion Angles, Pseudorotation Phases, and Pucker Values of the Average Structure

residue	α	β	γ	δ	ϵ	ζ	χ	P	ϕ_m
T1			−60	130	−80	−180	−155	142	28
G2	−95	146	53	142	−179	−102	−103	160	36
C3	−67	172	60	107	−168	−92	−130	110	34
G4	−71	171	61	140	−165	−137	−97	150	36
A5	129	−100	179	116	−73	166	34	118	41
A6	71	−94	−174	136	−177	−93	−103	152	36
T7	148	−166	178	132	−85	43	−125	143	30
T8	148	−173	178	115	−170	−83	−125	116	34
C9	−62	168	63	143	−80	161	−83	150	37
G10	75	−150	−175	152	−175	−109	−107	171	39
C11	−75	178	59	115	−154	−69	−116	115	29
T12	−79	−179	56	103		−133	81	21	

Table 5. Helical Parameters of the Stem Region of the Average Structure Calculated by NEWHELIX93[a]

base pair parameters						
code	tip	incl	prop	buck	*X* dsp	*Y* dsp
G2•C11	−7	2	−19	10	−1	2
C3•G10	−5	−1	−11	10	1	2
G4•C9	−1	0	−6	4	2	2

base step parameters						
code	roll	tilt	cup	slide	twist	rise
G2•C11/C3•G10	2	1	0	0	37	3
C3•G10/G4•C9	4	3	−6	1	39	4

[a] Available from the Brookhaven Protein Databank.

the geometry of our disulfide modification is dictated by the conformational preferences of the cross-link itself and not by the sequence to which it is attached. Because the alkylthiol chains do not have NOE connectivities to their respective thymidine bases, it has not been possible to unambiguously determine which linker belongs to which modified residue. To overcome this limitation we are currently preparing a variant of **1** in which the linker on the 3′-end is uniformly deuterated.

Implications for this Cross-Link. Because the hairpin formed by d(CGCGAATTCGCG)$_2$ is not amenable to high-resolution NMR analysis for comparison with **1**, we cannot state

Table 6. Pairwise rmsd Values of the Best Ten Structures Relative to Their Average Structure[a]

structure	all	stem	loop	G2−C11
1	0.74	0.50	0.43	0.65
2	0.65	0.41	0.50	0.49
3	0.67	0.41	0.66	0.55
4	0.74	0.52	0.36	0.49
5	0.63	0.39	0.32	0.38
6	0.52	0.45	0.44	0.46
7	0.74	0.52	0.46	0.58
8	0.63	0.36	0.63	0.58
9	0.50	0.36	0.30	0.45
10	0.50	0.36	0.30	0.45
av	0.63	0.43	0.44	0.51

[a] The heavy atoms of the whole molecule, the stem (G2−G4:C9−C11), the loop (A5−T8), and the whole molecule without the two termini (G2−C11) are compared separately.

with absolute certainty that our disulfide cross-link does not alter DNA geometry. However, we have now examined the 3-D structures of several cross-linked duplexes[33] along with a ground-state DNA hairpin[31,35] and have compared them to their unmodified counterparts. In each of these cases the distance and torsion angle constraints for the wild-type sequences and their disulfide cross-linked analogs are within experimental error, indicating that their conformations are also isomorphous.[83] Taken together, these findings provide good evidence that stabilizing oligodeoxyribonucleotides by incorporating this modification will not cause major structural perturbations. Disulfide cross-linked oligomers such as these therefore should prove generally useful in studies of nucleic acid folding, structure, and function.

Acknowledgment. We thank Dr. A. Majumdar for providing the CHORDS program, R. Cain for helpful discussions, and S. Osborne for preparation of **1**. Supported in part by grants from the NIH and NSF to G.D.G. Partial funding for the 500 MHz NMR spectrometer was provided by the NIH Shared Instrumentation Program.

JA943805D

(83) Cain, R. J.; Zuiderweg, E. R. P.; Glick, G. D. Submitted.