Nuclear Magnetic Resonance Structure of d(GCATATGATAG)·d(CTATCATATGC): A Consensus Sequence for Promoters Recognized by σ^K RNA Polymerase[†,‡]

Marco Tonelli,[§] Enzio Ragg,[||] Anna Maria Bianucci,[⊥] Krystyna Lesiak,[$] and Thomas L. James*[,§]

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446, Dipartimento di Scienze Molecolari Agroalimentari, Universita degli Studi Milano, 220133 Milano, Italy, Dipartimento di Scienze Farmaceutiche, Universita di Pisa, 56100 Pisa, Italy, and OncorPharm, 200 Perry Parkway, Gaithersburg, Maryland 20877

Received March 2, 1998; Revised Manuscript Received May 4, 1998

ABSTRACT: The three-dimensional structure of d(GCATATGATAG)·d(CTATCATATGC), from the promoter region of a gene regulating sporulation in *Bacillus subtilis* mother cells, was determined utilizing two-dimensional nuclear Overhauser effect (2D NOE) and double-quantum-filtered COSY (2QF-COSY) spectra. To minimize the effect of methods used to obtain restraints and refine structure, several variables were studied. Interproton distance bounds were calculated very conservatively by running the complete relaxation matrix program MARDIGRAS hundreds of times using 2D NOE spectra for exchangeable and for nonexchangeable protons at different mixing times, assuming different overall correlation times and different starting structures. The 435 distance restraints were used with two different structural refinement methods: restrained molecular dynamics (rMD) and restrained Monte Carlo calculations (rMC). Refinement using different procedures and starting structures resulted in essentially the same structure (<0.8 Å rmsd), indicating that the structure is defined by experimental restraints and not the refinement method or variables used. *R* factors indicate the structures fit the experimental NOE data very well. Some helical parameters, notably large negative X displacement, are characteristic of A-DNA, but others are characteristic of B-DNA. As with TG·CA steps in other duplex DNA sequences studied in our laboratory, the two TG·CA steps have a positive roll, with T6-G7 exhibiting the largest, and consequently a bent helix axis. The converged structure represents a time-averaged structure. However, multiple conformations, especially in deoxyriboses, were evident from vicinal coupling constants obtained from quantitative simulations of 2QF-COSY cross-peaks and from persistent inconsistencies in experimental distances due to nonlinear conformational averaging.

Many important biological or pharmacological processes are affected by specific interactions of DNA with other molecules, e.g., proteins, RNA, mutagens, or drugs. It has been observed that such interactions can be significantly affected by changes in the sequence of DNA. Indeed, effects are manifest even when mutated residues are not directly involved in contacting the protein, but the mutation induces conformational distortions in the DNA molecule (*1*). In particular, the three-dimensional arrangement of hydrogen-bonding sites, bases, and helix geometry are structural features of DNA molecules specifically recognized by proteins. For example, specific promoter sequences exhibit 1000−10000 times stronger affinity for transcription factors than do random sequences (*2*). In duplex DNA, major and minor grooves provide the topology and there are sequence-dependent variations in the patterns of hydrogen bond acceptors and donors. There is little distinction in the hydrogen bond pattern in the minor groove, but there is significant variation in the major groove. The major groove also exhibits nonpolar patches in the C5 region of pyrimidines with an obvious difference even between C and T. The strength and kinetics of specific binding of a ligand to a double-stranded DNA depend on the ligand matching these and other contacts in a "direct readout" of the sequence (*3*). However, the sequence-dependent conformation of DNA modifies the three-dimensional pattern of contacts to be made as an "indirect readout" of the sequence. These features range from relatively gross changes in bending, to effects of groove width, and finally to local variations in disposition of contact points. Knowledge about sequence-dependent variations in DNA duplex geometry is essential to understanding fully DNA function in biological processes. There may be some thermodynamic (and possibly kinetic) advantage for a particular DNA free in solution to be in a conformation similar to the one it will possess when it is

[†] This work was supported by Grant GM39247 from the National Institutes of Health. M.T. was supported in part by a fellowship from the University of Pisa.

[‡] The coordinates have been deposited at the Brookhaven Protein Data Bank under accession number 1SKP.

* To whom correspondence should be addressed at the Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-0446. Phone: (415) 476-1916. Fax: (415) 502-4690. E-mail: james@picasso.ucsf.edu.

[§] University of California.

[||] Universita degli Studi Milano.

[⊥] Universita di Pisa.

[$] OncorPharm.

S0006-2960(98)00481-4 CCC: $15.00
Published on Web 08/06/1998

bound. With binding, there is undoubtedly a favorable entropy change upon displacement of counterions, but some of that entropic advantage could be lost by the gain in entropy of the DNA itself upon complexation. If the free DNA structure closely resembles that of the bound state, there would be little entropic price to pay upon binding. In addition, conformational flexibility of the double helix at specific sites may play an important role in recognition by proteins, perhaps by providing loci with a propensity for conformational change which is then manifest upon protein binding.

While the overall folding geometry of proteins and RNA is often the goal of structure studies, when DNA duplexes are being studied, the question that should be addressed is how the conformation varies with sequence. Since these sequence-dependent structural variations are very subtle, the structure must be determined with high resolution and high accuracy. A better understanding of the rules governing sequence-dependent structural heterogeneity in DNA could some day contribute to the "rational design" of new drugs targeted to interact directly with a gene (DNA) rather than a gene product (protein).

Recently, the quality of DNA structures determined via NMR has improved sufficiently to provide an alternative. The methodology for determination of structures with sufficient resolution has not been easily achieved, but the ability to determine an accurate, high-precision structure of nearly any DNA duplex with a length of $\leq$15 base pairs is now possible with sufficient care and effort (*4*). Especially useful NMR experiments are two-dimensional nuclear Overhauser enhancement (2D NOE)[1] spectroscopy and, to a lesser extent, double-quantum-filtered correlated (2QF-COSY) spectroscopy which are used for extracting proton−proton distances and sugar torsion angles, respectively. These NMR-derived parameters are subsequently used as structural restraints in refinement methods such as restrained molecular dynamics (rMD) calculations, which search conformational space to find structures consistent with the experimental data. Other refinement methods are also available; in particular, distance geometry algorithms are commonly used for protein structural refinement, and restrained Monte Carlo (rMC) calculations have occasionally been used.

Improvements in structure determination have also been realized by developing methods for increasing the accuracy and the number of structural restraints, such as, extraction of interproton distances from 2D NOE intensities via a "complete relaxation matrix approach" (*5*) and the determination of sugar torsional angles from proton−proton coupling constants by applying optimized Karplus equations (*6*−*8*). MARDIGRAS (matrix analysis of relaxation for discerning geometry of aqueous structures) is the algorithm we use to calculate reliable distance bounds from NOESY spectra (*9*). On the other hand, accurate coupling constants can be extracted by simulation of COSY cross-peak patterns with the programs SPHINX and LINSHA (*10*).

The DNA sequence d(GCATATGATAG)·d(CTATCATATGC) studied in this paper is found in the promoter region of the gene *gerE* which is expressed in *Bacillus subtilis* mother cells during sporulation (*11*). Gene expression in *B. subtilis* is a hierarchical cascade of at least four temporally controlled gene sets. The first three each contain a regulatory gene that governs expression of the next gene set in the pathway. The σ^K RNA polymerase (carrying transcription factor σ^K) has been shown to transcribe in vitro from *gerE* and at least three other genes involved in sporulation. The promoter regions of these genes each contain sequences similar to CATA− − −TA at about position −10 relative to their transcription start sites, suggesting that this particular sequence is important for recognition of the promoters by σ^K RNA polymerase. A related sequence is also found in the promoter region of genes transcribed by σ^E RNA polymerase. Thus, this undecamer DNA molecule contains the consensus sequence for promoters recognized by RNA polymerases carrying the sporulation σ factors.

This study also continues our investigation of the structural qualities of alternating purine-pyrimidine sequences, especially A- and T-rich sequences. Previous studies of DNA duplexes in our laboratory have produced clear sequence-dependent variations in the geometry which has been statistically analyzed (*12*). Interestingly, two of the sequences studied exhibit a noticeable bend of the global helix axis at each of three TG·CA steps. However, in each of the three cases, the T was preceded by a pyrimidine. The DNA molecule investigated in this paper contains a TG·CA step, with a preceding A, in the middle of the sequence interrupting an alternating series of AT steps. Certainly, the high-resolution structure of this molecule will prove to be valuable in elucidating conformational variations related to DNA sequence.

MATERIALS AND METHODS

Sample Preparation

The two DNA undecamers, d(GCATATGATAG) and d(CTATCATATGC), were synthesized on an Applied Biosystems, Inc., model 380B synthesizer and purified via reverse-phase HPLC as described previously (*13*). The DNA duplex was prepared for NMR by titration of one undecamer in phosphate buffer solution (pH 7.0) with a similar solution of the complementary strand, monitored via 1D NMR. The resulting solution [20 mM sodium phosphate, 0.1 M NaCl, and 0.1 mM EDTA (pH 6.9)] was about 2.5 mM in duplex. For experiments with exchangeable protons, the sample was lyophilized and the dried product dissolved in 90% H_2O/10% ^2H_2O.

NMR Spectroscopy

^{1}H NMR experiments were run at 600 MHz on a Bruker AMX600 spectrometer. All 2D NMR measurements were taken at 25 °C except for 2D NOE experiments in H_2O which were also run at 5 and 10 °C.

For quantitative analysis, four 2D NOE spectra in ^2H_2O were recorded at mixing times of 70, 130, 200, and 270 ms.

[1] Abbreviations: NMR, nuclear magnetic resonance; 1D NMR, one-dimensional NMR; 2D NMR, two-dimensional NMR; 2D NOE, two-dimensional nuclear Overhauser effect; 2QF-COSY, double-quantum-filtered correlation spectroscopy; TOCSY, total correlated spectroscopy; EDTA, ethylenediaminetetraacetate; FID, free induction decay; rMD, restrained molecular dynamics; rMC, restrained Monte Carlo calculations; MD-tar, molecular dynamics with time-averaged restraints; rmsd, root-mean-square deviation; R^x, sixth-root R factor; P, pseudorotation phase angle.

The experiments were carried out using exactly the same conditions without removing the sample from the magnet. Quadrature detection in t_1 was achieved by TPPI (*14*) of the pulse immediately preceding the increment period t_1 in 90° steps along with t_1. Low-power continuous-wave irradiation was applied to the residual water signal during recycle and mixing delays. A spectral width of 6024 Hz was used with the carrier frequency set to the HDO resonance frequency. A total of 1024 FIDs of 4K real data points were collected in t_1 with 32 scans at each t_1 value. The delay between scans was 2.5 s.

2D NOE spectra in H_2O were acquired with a 1−1 echo excitation pulse centered on H_2O. Data were collected for a spectral width of 12 500 Hz, and the excitation maximum was set to the imino proton region at 13.2 ppm. Spectra were recorded with a 158 ms mixing time at 5, 10, and 25 °C. A total of 512 FIDs of 2K real points were collected in t_1 with 64 scans at each t_1 value.

2D NOE spectra were processed using TRIAD, the NMR software within the Sybyl software package of Tripos, Inc. (St. Louis, MO). A Gaussian window function was used for resolution enhancement in both dimensions. Prior to Fourier transformation, the FIDs were zero-filled to give a final 2K × 4K data set with a digital resolution of 1.5 Hz/point in ω_2 and 2.9 Hz/point in ω_1 in the case of the spectra recorded in ^2H_2O. With the larger spectral width, spectra acquired in H_2O had a lower final resolution: 3.1 Hz/point in ω_2 and 6.1 Hz/point in ω_1.

Pure absorption 2QF-COSY data were recorded using the TPPI scheme, with 1024 FIDs of 4K points collected in t_1 and 48 scans for each t_1 delay. Acquisition times were 340 and 85 ms in t_2 and t_1, respectively, with a repetition delay between scans of 2.5 s. The spectral width was set to 6024 Hz. The averaged FIDs were apodized using a squared sine bell function, shifted by 30°, in both dimensions and zero-filled to a final data matrix of 1K × 4K points. Digital resolution of the processed spectrum was 1.47 Hz/point in ω_2 and 5.88 Hz/point in ω_1.

The total correlated spectroscopy (TOCSY) data were acquired by standard methods (*15*, *16*). A 4K × 1K spectrum was acquired with 32 transients at each t_1 value. The spectral width was set to 6024 Hz in both dimensions. This spectrum was processed to yield a 2K × 1K spectrum using a sine-square apodization function shifted by 70° applied in both dimensions prior to Fourier transform.

Measurement of ^{1}H spin−lattice relaxation times T_1 was performed by the inversion−recovery method. Spin−spin relaxation times T_2 were determined using the Hahn spin-echo method. Imino proton exchange rates with bulk water at 5, 10, and 25 °C were determined by analyzing the exponential decay of imino protons, while a spin-lock field was applied selectively to the water signal (*17*, *18*).

^{1}H-detected ^{13}C NMR spectra were acquired at 25 °C on a Varian Unityplus 600 MHz spectrometer equipped with a triple-resonance probe and actively shielded gradient coils. These spectra were all processed and analyzed in the usual way with the locally written programs STRIKER and SPARKY (Day and Kneller, University of California at San Francisco, 1995). The ^{1}H-detected heteronuclear multiple-quantum coherence (HMQC) spectrum (*19*) was recorded using an unlabeled sample and the following parameters: 256 FIDs in the t_1 dimension, 64 repetitions for each FID, 2048 points in the t_2 dimension, and a spectral width of 8000 Hz in each dimension, resulting in about 13 and 53 ppm spanned in the proton and carbon dimensions, respectively.

^{13}C T_1 values were measured using a pulse sequence that includes a double DEPT transfer of magnetization from ^{1}H to ^{13}C and then back again to ^{1}H, after being labeled by T_1 of ^{13}C, for detection (*20*). This pulse sequence had been modified to use gradients instead of homospoil pulses. Due to the low sensitivity of this method on an unlabeled sample like ours, a large number of repetitions were required to obtain a useful signal, precluding two-dimensional experiments. Five 1D spectra, each with 8K data points and an 8000 Hz spectral width, were acquired with relaxation delays of 2, 10, 100, 200, and 500 ms and were used to estimate T_1 values. To minimize experimental error and increase the signal-to-noise ratio, each spectrum was acquired twice and the sum of corresponding spectra used to determine T_1.

^{13}C{^{1}H} NOE experiments were carried out in a similar fashion, with the pulse sequence modified to measure NOE instead of T_1 values. Again, due to the low sensitivity of the method on an unlabeled sample, 1D spectra were obtained rather than 2D spectra. Thus, 1D spectra with 8K data points and an 8000 Hz width were acquired with and without NOE transfer being allowed. Each experiment was repeated twice and the sum of the corresponding spectra used. The ^{13}C-{^{1}H} NOE was determined by the difference of corresponding peak intensities acquired with and without allowing NOE transfer between atoms.

Analysis of NMR Spectra

Model Structures. Several structures were generated to be used as starting models for running MARDIGRAS, CORMA, molecular dynamics simulations, and Monte Carlo calculations. Three different molecular modeling programs were used to generate all-atom models with all hydrogen atoms explicitly included: the module NUCGEN of AMBER version 4.0 (*21*), Sybyl 6.0 (Tripos), and the program DNAminiCarlo (*22*) which allows generation of DNA duplex models by specifying generalized helical parameters. Standard A- and B-form DNA models (*23*, *24*) were created with each of the programs mentioned above. All structures were then energy-minimized by the three programs using their different empirical force fields.

Extraction of Proton−Proton Distance Bounds from 2D NOE Spectra. 2D NOE intensities were measured using the locally written NMR-processing program package SPARKY. Intensities were determined by line fitting the cross-peaks to a Gaussian function and subsequent integration of the theoretical curve (*25*). With the 2K × 4K spectra, good fitting was achieved even for weak cross-peaks, allowing extraction of all intensities solely by line fitting. Final intensities were obtained by averaging corresponding cross-peaks on both sides of the diagonal. In the few cases where the two cross-peak intensities were significantly different, the most reliable was taken. Cross-peaks involving exchangeable protons were integrated from 2D NOE spectra recorded in H_2O and scaled to take into account the excitation profile of the 1−1 echo sequence.

Proton−proton distances were calculated from 2D NOE cross-peak intensities using the complete relaxation matrix approach implemented in the program MARDIGRAS (*26*).

A single correlation time for the whole molecule (isotropic motion) was assumed with no internal motion. For methyl groups, a three-state jump model was used (*27*). The overall isotropic correlation time τ_c was estimated by three independent methods (vide infra). To cover conservatively all uncertainty in the results, three sets of MARDIGRAS runs, with three different correlation time values (2.0, 3.0, and 4.0 ns), were performed. Experimental intensities involving nonexchangeable protons from four 2D NOE spectra acquired with 70, 130, 200, and 270 ms mixing times were used. Prior to the calculation of the distances, the experimental intensities were normalized with the theoretical ones by considering only intensities involving protons at fixed distances, i.e., H5−H6 in cytosines, methyl−H6 in thymines, and H2′−H2″ proton−proton distances. A total of 11 models were used as starting structures to run MARDIGRAS (standard and energy-minimized A- and B-DNA generated by AMBER, Sybyl, and DNAminiCarlo; energy-minimized A-DNA by Sybyl was not used, refer to the model structure section above) for a total of 132 runs.

To calculate distances involving exchangeable protons, a single set of intensities acquired with a 158 ms mixing time was used, together with two starting models and three estimates of correlation time. To account for experimental errors, the RANDMARDI algorithm implemented in the MARDIGRAS software was applied (*9*).

For each proton pair, average values for distances and upper and lower bounds were calculated as well as standard deviations. The corresponding standard deviations were added and subtracted from the average distances, and the resulting values were compared to the average upper and lower bounds. Then, the two values that resulted in the larger width were chosen as upper and lower bounds for structure refinement.

In the same fashion, the program CORMA (*28*) was run to simulate NOE intensities from a model structure and to compare these intensities to the corresponding experimental intensities by calculating conventional *R* factors (*R*) and sixth-root *R* factors (R^x) (*5*). A 3.0 ns overall correlation time value was used for all models.

2QF-COSY Simulation. 2QF-COSY cross-peak patterns for deoxyribose protons were simulated using the programs SPHINX and LINSHA (*10*). A six-nuclei spin system, including H1′, H2′, H2″, H3′, and H4′ protons as well as ^{31}P bound to the 3′-O, was considered. Weak coupling was assumed between all sugar protons except for the strong coupling between H2′ and H2″. Deoxyribose coupling constants were extracted by matching experimental and simulated cross-peak patterns along the better resolved ω_2 dimension (*25*). Then the conformation of the sugar rings was estimated by matching our experimental coupling constants to values calculated assuming a two-state model with rapid *N*−*S* interconversion for the sugar pucker. A graphical method described elsewhere (*8*) was used to find the best match. The proton−proton coupling constants determined by Rinkel and Altona for specific deoxyribose conformers were used (*29*).

Structure Refinement

Two different methods were used for structure refinement: simulation of molecular dynamics with experimental restraints (rMD) and restrained Monte Carlo calculations (rMC).

Generation of Restraints for Structural Refinement. The restraints used during molecular dynamics simulations and Monte Carlo calculations included interproton distances derived from experimental NOE data and Watson−Crick hydrogen bond distances and angles. Sugar torsional angles, from 2QF-COSY analysis, were not used, since the sugar conformation appeared to be well-defined by interproton distance restraints alone. The structural insights derived from scalar coupling data can be used as an independent monitor of the structural results.

The restraint term introduced into the potential energy function to enforce the experimental restraints takes the quintipartite form of a flat-well potential with flat, parabolic, and linear parts. The exact form of the restraint term used by AMBER has been previously described (*30*). Upper and lower bounds from the MARDIGRAS analysis delimit the flat region of the potential. In the flat region, no energy penalty is applied. The force constant for the experimental distance restraints term, k_{NOE}, was set to 20 kcal mol^{-1} Å^{-2} at 300 K. This value was increased when the temperature was raised during rMD and rMC calculations.

Restraints for hydrogen bonds were added for Watson−Crick base pairs. These include distance restraints and flat angles between the three atoms forming the hydrogen bond. The values used are consistent with crystallographic data (*31*) and have been previously used in NMR structure refinement (*30*). Use of Watson−Crick hydrogen bond restraints is especially important for preventing the two strands of the duplex from flying apart when the temperature of the system is raised during simulations. The force constant for Watson−Crick hydrogen bond distance restraints was always lower than that for experimental restraints throughout the simulations.

Restrained Molecular Dynamics Simulations. Restrained molecular dynamics calculations were performed using the module SANDER of AMBER version 4.0 (*21*) on a cluster of Hewlett-Packard 735 computers. Molecular dynamics simulations were performed using energy-minimized A- and B-DNA as starting models. These structures were generated by the NUCGEN module of AMBER and energy-minimized (vide supra). Prior to minimization, large (hydrated) sodium ions (radius of 5.0 Å) were added along the DNA backbone, 5 Å from the phosphorus atoms, to neutralize the phosphate negative charges; they were subsequently free to move. All MD simulations were run in vacuo with optimized parameters previously used to refine duplex structures in our lab (*30*, *32*, *33*).

When we started from B DNA, rMD simulations were run for 50 ps at a constant temperature of 300 K. No restraint forces were applied during the first 200 steps of the run. From step 201 to 1000, the restraint force constant was gradually increased to its final value (20 kcal mol^{-1} Å^{-2}) which was then maintained through the rest of the simulation. When A-DNA was used as the starting structure for MD runs, an annealing protocol was used to overcome local energy barriers. The temperature of the system was gradually raised to 900 K during the first 2000 steps of the run. The system was then kept at this temperature for the next 10 000 steps and gradually cooled to 300 K. The temperature was not changed for the last 10 000 steps of the simulation. The

restraint force constant was increased to 100 kcal mol^{-1} Å^{-2} when the temperature was raised to 900 K and then decreased to its final value of 20 kcal mol^{-1} Å^{-2} when the temperature was lowered to 300 K.

Restrained Monte Carlo Calculations. To assess the dependence of the final geometry on the refinement method used, we also calculated the DNA undecamer structure by restrained Monte Carlo methods with the protocol we previously reported (*34*). The DNAminiCarlo program, run on the Cray C90 computer at the Pittsburgh Supercomputing Center, uses generalized helical parameters as independent variables to generate DNA structures. For our undecamer, we had 10 subsets of six variables defining each of the 10 steps, and 11 subsets of 10 base pair variables, with the sugar puckers defined by the phase angle of pseudorotation *P* (assuming a pucker amplitude of ~35°).

Two DNA structures were generated by DNAminiCarlo with helical parameters typical of standard A- and B-DNA and energy-minimized by the same program. The same Monte Carlo simulation protocol (Markov chain) was used starting from both A- and B-DNA, consisting of a single cycle of simulated annealing. Starting at 100 K, the temperature was gradually raised to 600 K in six steps 100 K apart. The NOE force constant, k_{NOE}, was increased from 5 kcal mol^{-1} Å^{-2} to a maximum value of 80 kcal mol^{-1} Å^{-2} at 500 K and this value maintained when the temperature reached 600 K. The temperature and k_{NOE} were kept constant for 1000 iterations before proceeding to the next fragment of the Markov chain. Each iteration was defined when all independent variables defining the DNA undecamer were changed. The last structure generated during each fragment was used as the starting point for the following one. The molecules were kept at 600 K for 5000 iterations to allow crossing of conformational energy barriers. The system was then gradually cooled to 300 K in three steps of 1000 iterations, each step being 100 K apart. Simultaneously, the force constant was decreased to its final value of 20.0 kcal mol^{-1} Å^{-2}. Finally, the temperature was maintained at 300 K for 10 000 iterations. To ensure proper base pairing, the force constant for the Watson−Crick hydrogen bond term was maintained at half the value used for k_{NOE}, except for the last 10 000 iterations where a value of 5 kcal mol^{-1} Å^{-2} was used.

All DNA-independent variables from the last Markov chain fragment at 300 K were averaged, and the mean values were used to construct the final structure. This was then restraint-minimized with the k_{NOE} set to 20 kcal mol^{-1} Å^{-2}.

Structure Analysis and Display. Values of the atomic root-mean-square deviation (rmsd) between model structures were calculated by the program CARNAL (written by W. S. Ross). Sugar puckers, torsion angles, and helical parameters were calculated with the programs CURVES (*35*) and Dials & Windows (*36*). Values calculated by the two programs were quite similar. All structures were visualized using Midasplus (*37*) and/or Sybyl (Tripos).

RESULTS

Proton Resonance Assignments

For the residues of the undecamer duplex, the following numbering scheme was used:

	1	2	3	4	5	6	7	8	9	10	11	
5'-	G	C	A	T	A	T	G	A	T	A	G	-3'
3'-	C	G	T	A	T	A	C	T	A	T	C	-5'
	22	21	20	19	18	17	16	15	14	13	12	

Nonexchangeable protons were assigned by following intranucleotide and sequential NOE connectivities typical for right-handed DNA (*38*, *39*) in 2D NOE spectra recorded in ^2H_2O. 2QF-COSY, TOCSY, and natural abundance $^1H-^{13}C$ HMQC experiments were used to aid in assignments. Due to severe peak overlap, few H4′ and most H5′−H5″ resonances were not assigned.

Exchangeable protons could be easily assigned by following well-established strategies (*40*). Unfortunately, chemical shift degeneration between H3 in T9 and T15 and in T4 and T18 prevents assignment of most amino protons for A5, A19, A8, and A14. Moreover, the guanine amino proton resonances were not observed under the conditions employed.

Sugar Conformation from Analysis of 2QF-COSY Spectra

The proton homonuclear $^3J_{1'2'}$, $^3J_{1'2''}$, $^3J_{2'3'}$, $^3J_{2''3'}$, and $^3J_{2'2''}$ coupling constants were extracted by simulation of 2QF-COSY cross-peak patterns with the SPHINX and LINSHA programs (*7*, *10*). Unfortunately, due to severe peak overlap, a significant number of residues were not analyzed; A3 and A14, A5, A8, A17, and A19, and T4, T18, and C12 give fully or partially overlapped cross-peaks. Moreover, H2′ and H2″ of C12 are isochronous.

To obtain pseudorotational parameters, which describe the geometry of the sugar moiety, experimental *J* coupling constants were compared with theoretical values using a graphical method (*8*). To match experimental values, we needed to assume a two-state model with rapid *N*−*S* interconversion, with *S* being the major conformer. Table 1 shows the fraction of the *S* major conformer f_S, the pseudorotation angle P_S, and the amplitude Φ_S that gave the best match for each analyzed residue. For the minor *N* conformer, P_N was kept fixed at 9° and sugar pucker amplitudes were assumed to be the same for both *S* and *N* conformers ($\Phi_N = \Phi_S$). Except for the C12 terminal sugar, the fraction of minor *N* conformer was less than 25%. As expected, terminal residues have an increased population of the *N* conformer, undoubtedly due to fraying of the duplex at its ends. Only for T13 was it possible to match the experimental *J* coupling values by using only the *S* conformer ($f_S = 1.0$). Also, for some residues, matches were found with different amplitudes of Φ_S.

In conclusion, our analysis of 2QF-COSY cross-peaks suggests that sugars in DNA duplexes in solution are flexible, undergoing fast exchange between *N* and *S* conformers, with *S* being the favored conformer. Only well-resolved residues have been studied. However, 2QF-COSY pattern similarities suggest that overlapped sugars behave in a manner similar to that of the well-resolved ones.

MARDIGRAS Calculations

To refine the structure of our DNA undecamer at high resolution, we need to calculate accurate and, preferably, precise distance restraints. This is accomplished in our laboratory by using the program MARDIGRAS which solves

Table 1: Phase Angle of Pseudorotation, P_S, Pucker Amplitude, Φ_S, and Fraction, f_S, of the *S*-Type Conformer in d(GCATATGATAG)·d(CTATCATATGC)[a]

residue	f_S	P_S (deg)	Φ_S (deg)
G1	0.85−0.94	175−194	40
T6	0.85−0.9	145−171	30
G7	0.86−0.91	161−179	35
	0.85−0.95	179−194	40
T9	0.85−0.91	139−173	30
	0.79−0.88	130−149	35
A10	0.9−0.95	165−186	40
G11	0.77−0.8	139−163	35
	0.76−0.8	130−145	40
C12	0.45−0.56	163−198	35
	0.5−0.6	169−212	40
T13	0.9−1.0	157−184	35
T15	0.75−0.8	122−133	40
C16	0.8−0.9	155−183	30
	0.65−0.83	139−169	35
T20	0.77−0.85	132−151	35
	0.75−0.8	122−130	40
G21	0.80−0.95	155−175	40

[a] P_N was assumed to be 9°, and $\Phi_N = \Phi_S$ in the analysis.

the complete relaxation matrix, thus incorporating all network relaxation and multispin effects (*26*). To construct a complete intensity matrix, MARDIGRAS first combines experimental intensities with intensities calculated from a starting model by the CORMA algorithm (*28*, *41*). The experimental intensities must be normalized prior to construction of the hybrid matrix.

Previous studies have shown that the initial model used to construct the hybrid matrix has little effect on the distances calculated by MARDIGRAS. However, in our system, we have noticed that some distances depend on starting geometry more than others. Such dependency may reflect ambiguity in calculating the distances under the conditions used, so we decided to run MARDIGRAS using several starting models and to generate distance bounds by combining the results from all MARDIGRAS runs. Then, any ambiguity in distances calculated should be reflected in wider distance bounds. Eleven structures were used as initial models for MARDIGRAS. These structures were constructed using different molecular modeling programs to represent A- and B-DNA conformations; some differences were found between corresponding conformations generated by separate programs, so all models were used to run MARDIGRAS.

Estimation of Overall Correlation Time τ_c. Another important parameter in calculating proton−proton distances is the molecular motional model that is assumed, since motion can affect the value of the NOE intensities. For a hydrated undecamer duplex, the overall molecular tumbling can be considered isotropic and, hence, is well-approximated by a single correlation time, τ_c. The value of τ_c was estimated using various methods.

The simplest method for deriving τ_c is from the proton spin−lattice relaxation time T_1 and spin−spin relaxation time T_2:

$$\tau_c = \frac{2}{\omega_H}\sqrt{\frac{3T_2}{T_1}}$$

where ω_H, the proton Larmor frequency, is given by $2\pi\nu_H$, ν_H being the precession frequency in hertz and assuming that overall tumbling at 600 MHz is in the slow motion limit, where $(\omega_H\tau_c)^2 \gg 1$. T_1 and T_2 values were measured for resolved adenine H2 and H6−H8 protons at 25 °C. The calculated τ_c values are reported in Table 2 together with the corresponding T_1 and T_2 values.

Table 2: T_1, T_2, and τ_c Values for Most H6−H8 and Some H2 Protons

atom	T_1 (s)	T_2 (s)	τ_c (ns)
A14 H8	2.70	0.0432	2.42
A3 H8	2.21	0.0254	2.86
A17 H8	1.92	0.0208	2.95
A5 H8, A19 H8	2.09	0.0190	3.21
A8 H8, A10 H8	2.22	0.0476	2.09
A3 H2	3.60	0.0615	2.34
C2 H6	1.22	0.0833	1.17
C22 H6	1.67	0.0258	2.46
C16 H6	2.23	0.0185	3.36
G7 H8	1.93	0.0465	1.97
T6 H6	1.75	0.0234	2.64
T9 H6	2.69	0.0303	2.89
T13 H6	1.25	0.0308	1.95
T20 H6	1.56	0.0272	2.32

τ_c was also estimated from natural abundance ^{13}C T_1 and $^{13}C\{^1H\}$ NOE values which were calculated from the height of broad overlapped peaks in the H1′ region of 1D spectra at 25 °C. τ_c values were extracted using equations describing the variation of ^{13}C T_1 and $^{13}C\{^1H\}$ NOE with τ_c (*42*), with all constants set to match experimental conditions. The τ_c values, 3.2−4.7 ns estimated from ^{13}C T_1 values and 0.65−1.4 ns from $^{13}C\{^1H\}$ NOE, are not strictly consistent. They both, however, lay in the same range found by the proton T_1 and T_2 method described above. Possible contributions from internal motion were ignored in all calculations.

Finally, it is also possible to utilize the iterative relaxation matrix calculations for an estimation of the correlation time with the assumption that the best fit for the experimental intensities is obtained for the "correct" correlation time. Unfortunately, this endeavor is not straightforward since the necessary normalization for experimental and model intensities obscures that effect. An independent normalization factor that must be kept constant when τ_c is varied is required. This type of normalization is made feasible by adding the total intensities for 2D NOE slices of isolated protons, including the contribution from the diagonal peak, and determining the total intensity for a proton accordingly. Then, normalized theoretical intensities for these isolated proton slices can be compiled with MARDIGRAS using different correlation times. For the undecamer, results for diagonal, H6−methyl, and H2′−H2″ intensities fell into a range between 2 and 4 ns, which is quite similar to τ_c values derived from T_1 and T_2 relaxation parameters. However, inconsistent results were obtained for cytosine H5−H6 cross-peaks, where only the intensities obtained at short mixing times could be reproduced with reasonable correlation times, which were still larger (3−6 ns) than those obtained for the other protons.

In conclusion, the uncertainty in evaluating τ_c by different methods prompted us to run MARDIGRAS calculations using three different values for the overall isotropic correlation time: 2, 3, and 4 ns.

Calculation of Nonexchangeable Proton−Proton Distances. MARDIGRAS calculations were performed on four sets of experimental intensities recorded at four mixing times

Table 3: Total Number of Nonexchangeable and Exchangeable Proton−Proton Distances Calculated by MARDIGRAS[a]

	distance			
τ_m (ms)	total	intraresidue	interresidue	interstrand
70	328	187	134	7
130	409	225	162	14
200	438	237	174	16
270	445	236	185	15
all τ_m[b]	447	239	192	16
158[c]	36	8	8	20

[a] The number of intraresidue, interresidue, and interstrand distances are also shown. [b] Distances between nonexchangeable protons only. [c] Distances involving exchangeable protons extracted from 2D NOE in H_2O.

(70, 130, 200, and 270 ms), using 11 different structures as starting models and with three different correlation time values (2.0, 3.0, and 4.0 ns) for a total of 132 runs. When the results from all these different runs were combined, a total of 447 unique proton−proton distances were obtained. Table 3 shows the number of distances calculated from intensities at each mixing time, specifying how many intraresidue, interresidue, and interstrand distances were found. Due to uncertainty in the stereospecific assignment of H5′ and H5″ protons, all 41 cross-peaks involving these atoms were not used for structure refinement, as well as five cross-peaks implying proton n−methyl ($n + 2$) interactions, due to spin diffusion. Ultimately, a total of 399 distances from nonexchangeable protons, i.e., about 18 per residue, were used for structure determination.

Overlapped Peaks in 2D NOE Spectra. All corresponding nonexchangeable protons in residues A5 and A19 were found to be overlapped as well as all corresponding sugar protons in residues T4 and T18. As a result, no isolated intraresidue cross-peaks for A5 and A19 and only a few T4-A5 and T18-A19 interresidue cross-peaks could be found. This overlap can be justified by noting that residues T4-A5 and T18-A19 are located in the center of a self-complementary segment of the DNA sequence; residues 2−7 and residues 16−21 have the sequence CATATG. Consequently, these residues must have very similar chemical environments, suggesting nearly identical three-dimensional structures. We can safely divide by 2 the volume of overlapped cross-peaks and assign each half to the degenerate proton pairs. By doing this, we were able to obtain 24 experimental restraints, which are included in the above-mentioned 399 nonexchangeable bounds. These restraints are very important (Figure 1), since they belong to a region of the molecule with otherwise few or no restraints. For example, A5 and A19 have no intraresidue restraints if we ignore the overlapped cross-peaks.

With overlapped cross-peaks included, the experimental restraints are well-distributed along our DNA undecamer sequence with 15−20 distance restraints for each residue (Figure 1).

Estimation of Distance Error Bounds for Molecular Dynamics Simulations. With the conservative approach employed (vide supra), the average width of our calculated bounds is 0.77 Å for 447 total distances. For most proton pairs, the bound width increases with the absolute value of the distance; i.e., longer distances have wider bounds. However, the exact width depends on the nature of the two protons involved. The width of our bounds is larger than the average width of bounds previously used to refine DNA structures in our laboratory. The effect of having such wide bounds on structure determination will be discussed later in this paper.

Problems with a Short Repetition Delay. All NOE spectra in ^2H_2O were acquired using a relatively short repetition delay (2.5 s) that does not allow for complete relaxation of all proton longitudinal magnetization between scans. The major effect is that cross-peak intensities will depend differentially upon the delay time, varying most for peaks involving protons with very different T_1 values, such as H2 protons in DNA molecules. The effect of a short repetition delay on NOE intensities of DNA duplexes has been previously discussed (*43*). To test for such an effect, we corrected our intensities with the program SYMM (*43*). The distances calculated from such intensities were in all cases within the original bounds, indicating that our conservative approach already accounted for any error deriving from incomplete relaxation.

MARDIGRAS Calculations Entailing Exchangeable Protons. NOE intensities from 2D spectra recorded in H_2O were corrected to account for the excitation profile of the 1−1 echo water suppression sequence used.

MARDIGRAS calculations were performed with experimental NOE intensities acquired at 10 °C with a 158 ms mixing time, using three different estimates of correlation time (2.0, 3.0, and 4.0 ns) and two different initial models (energy-minimized A- and B-DNA structures generated by AMBER). Nonexchangeable distances were fixed to the values previously calculated from ^2H_2O data; the MARDIGRAS complete hybrid intensity matrix was constructed by combining the normalized experimental exchangeable proton intensities with intensities calculated from initial model distances and "known" nonexchangeable distances. The lower bounds of distances involving exchangeable protons were corrected to account for the loss of magnetization through exchange with the bulk water (*17*, *18*, *33*). To be conservative, the largest determined exchange rate (6.0 s^{-1}) was used for all amino and imino protons. Finally, to account for the contribution to distance error from noise in the spectra and integration errors, the RANDMARDI algorithm was applied (*9*). Upper and lower bounds were chosen from the wide set of distances yielded by the calculations.

A total of 36 new proton−proton distance restraints involving imino and amino protons were calculated (Table 3). These bounds are especially important because they include 20 interstrand restraints (Figure 1).

Comparing Distances Calculated by MARDIGRAS Using Different Overall Correlation Times. Due to some discrepancy in correlation time values estimated using different methods, three different sets of MARDIGRAS runs were performed assuming $\tau_c = 2.0$, 3.0, and 4.0 ns. Since this range is somewhat larger than that previously used in our laboratory, we investigated the effect of correlation time on calculated distances. To summarize, we found the deviation between distances calculated with the different τ_c values is always <10%, usually much less. The MARDIGRAS normalization scheme eliminates most, but not all, of the error introduced by uncertainty in τ_c.

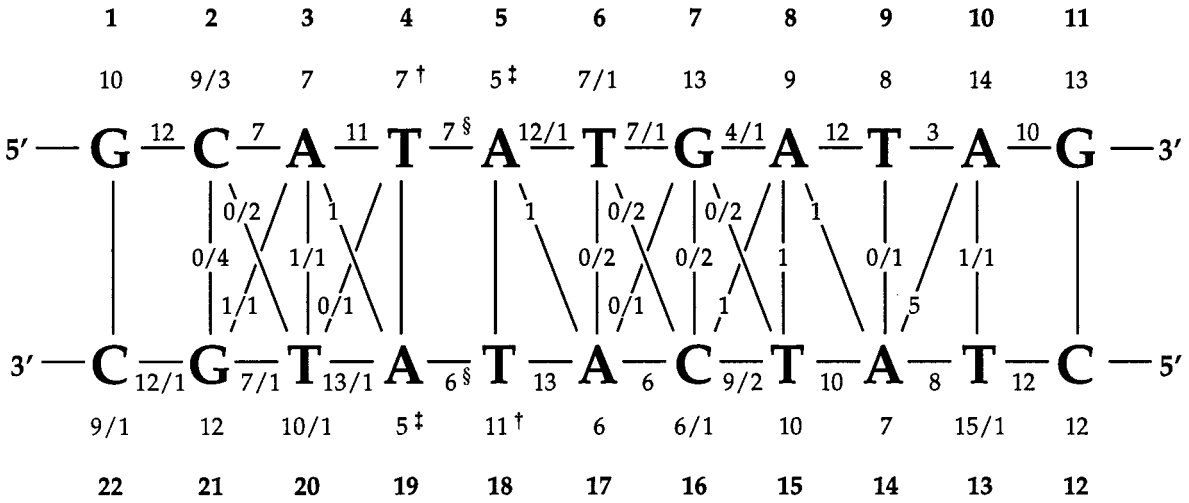

FIGURE 1: Distribution of NOE-derived distance restraints for the DNA undecamer duplex. Bold numerals indicate the sequence order. Plain numerals indicate the number of intraresidue, sequential, and interstrand restraints with placement implying attribution. Where two numbers are reported, the first one gives the number of restraints involving only nonexchangeable protons and the second gives the number involving at least one exchangeable proton. Due to peak overlap, some cross-peaks result in restraints being assigned to two similar residues (see the text for the method of attribution); two (†), four (§), or five (‡) restraints were due to overlap involving residue pairs T4·T18 and A5·A19.

Structure Refinement Results

Restrained Molecular Dynamics Calculations. Ideally, the structure of our DNA undecamer duplex should be fully characterized by the experimental restraints. However, the method chosen for structural refinement may play an important role in defining conformation, especially for regions of molecules which have sparse or conflicting experimental restraints.

We have already discussed how we extracted structural restraints from the experimental data, allowing for experimental error and testing those variables that could affect the final distances. Our final list of experimental restraints consists of 435 sets of proton−proton upper and lower distance bounds (Figure 1). Furthermore, with the large number of MARDIGRAS runs we performed (vide supra), our bounds are, on average, wider than the corresponding bounds previously used in our laboratory to refine DNA duplex structures. The effect of wider bounds on structure calculation will be further investigated here.

Since we already know the general folding of the DNA in solution, i.e., right-handed duplex, we can directly use structure refinement methods that require a "good" starting model. One popular method utilizes molecular dynamics with experimental restraints to search conformational space for structures which are compatible with both the experimental data and a chemical force field. For duplex DNA, we run restrained MD simulations starting from two different models: energy-minimized A- and B-DNA. *Amd1* and *Bmd* are the final structures, shown in Figure 2, obtained from starting A and B models, respectively. It is important to run multiple simulations starting from different models to verify the quality of the experimental restraints; weak or inconsistent bounds may not be able to drive different initial models to the same final structure.

To assess differences between the various structures, we calculated the atomic root-mean-square deviation (rmsd). Three rmsd values were determined (Table 4): all heavy atoms, all heavy atoms except those in terminal residues, and all heavy atoms except terminal and backbone atoms. The refined structures, *Amd1* and *Bmd*, have a much lower rmsd than the corresponding initial models, A- and B-DNA, indicating that convergence to essentially the same structure is achieved. A closer look at the rmsd values reveals that convergence is worse for terminal residues than for the internal ones (compare the first and second values in Table 4). This suggests that experimental bounds acting on terminal residues are not consistent with a single structure, probably due to fraying at the duplex ends. In contrast, the low deviation shown by internal residues suggests that they are defined well by the experimental data. Also, the low rmsd values calculated between *Amd1* and *Bmd* with or without taking into account backbone atoms (the second and third rmsd values) suggest that these structures have very similar backbone geometries. Apparently, even though no experimental data directly restrain backbone atoms (except protons belonging to the deoxyribose ring) and no empirical torsion angle restraints were used, the backbone conformation, surprisingly, does not depend on the initial geometry. Evidently, the geometry of internal bases and sugars, being defined well by NMR bounds, dictates the backbone conformation.

Depending on the initial model, the protocol followed during the MD run may be essential to achieve convergence. Thus, rMD starting from the A model was run with a cycle of simulated annealing. During the trajectory, the temperature reached a maximum of 900 K to allow conformational changes to occur. When we started from B-DNA, the temperature was maintained at 300 K throughout the simulation.

The final geometry also depends on the restraint force constant value, k_{NOE}. This value is empirically chosen to balance the driving force produced by the experimental restraints and the force field. Our goal is to satisfy the experimental data without forcing the molecule into high-energy conformations. To select a k_{NOE} value, we ran several rMD simulations starting from A and B models with increasing k_{NOE} values. The lowest force constant that allowed a good convergence between final structures (20 kcal mol^{-1} Å^{-2}) was chosen. This value was used to calculate *Amd1* and *Bmd* structures.

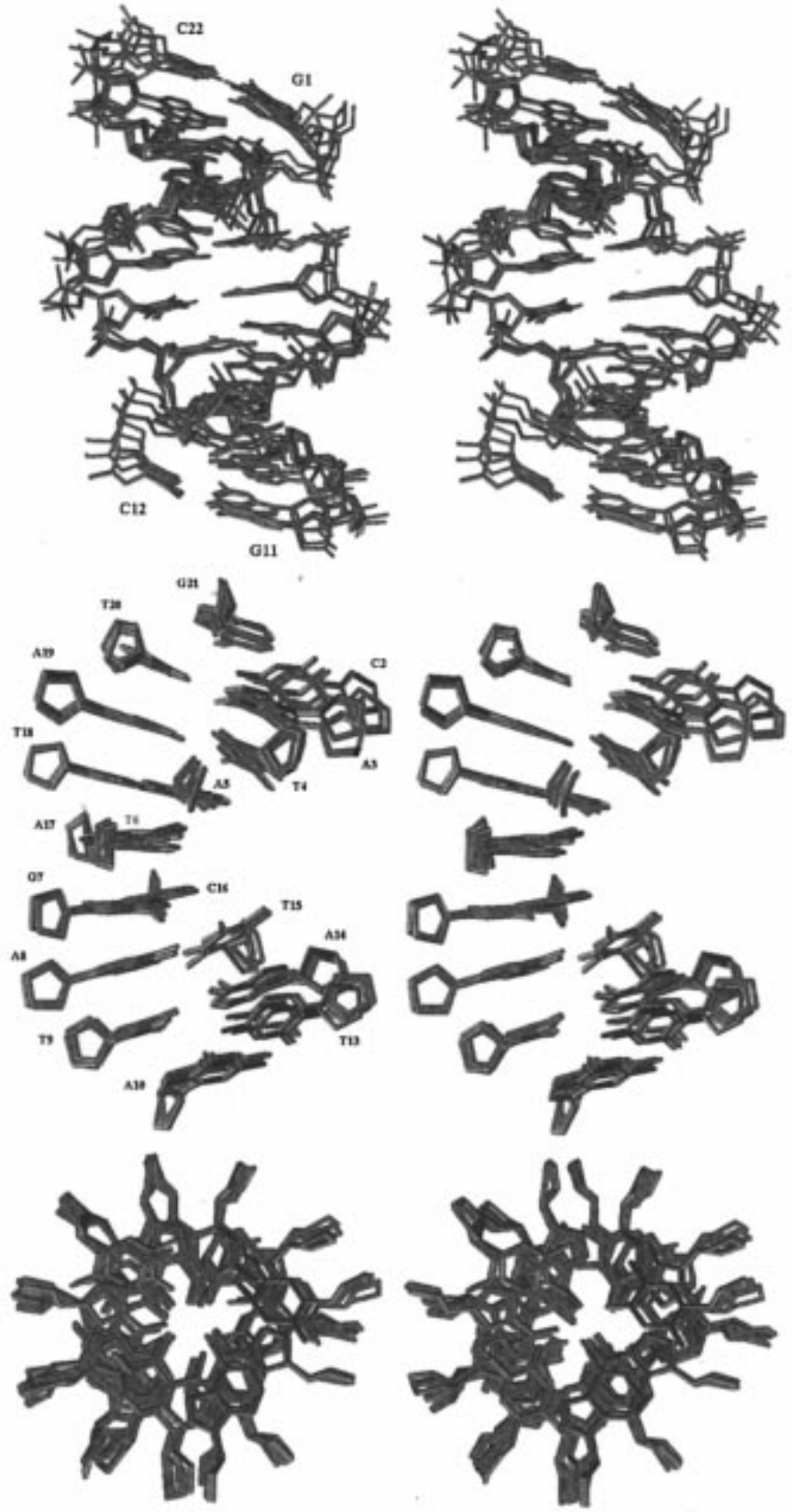

FIGURE 2: Stereoview of four superimposed final structures for the duplex: two models refined by molecular dynamic simulations starting from B- and A-form DNA, *Bmd* and *Amd1*, respectively (green), and two models refined by Monte Carlo calculation starting from B- and A-DNA models, *Bmc* and *Amc*, respectively (red). The top panel displays the whole structures. Only base and sugar heavy atoms of the 20 internal residues are shown in the middle panel. The bottom panel, by showing the superimposed structures along the helical axis, illustrates the A-DNA-like hole in the center of the duplex.

Table 4: Rmsd Values Calculated between Initial and Refined Models[a]

	*mA**	*mB**	*Amd1*	*Amd2*	*Bmd*	*Bmd-f*	*Bmd-n*	*mA*¶	*mB*¶	*Amc*	*Bmc*
	0										
*mA**	0										
	0										
	6.14	0									
*mB**	4.73	0									
	4.31	0									
	2.13	5	0								
Amd1	1.66	3.94	0								
	1.48	3.64	0								
	2.07	4.95	0.7	0							
Amd2	1.65	3.89	0.42	0							
	1.47	3.59	0.29	0							
	2.36	4.73	1.01	1.05	0						
Bmd	1.8	3.75	0.45	0.52	0						
	1.6	3.48	0.34	0.39	0						
	3.27	3.67	2.57	2.41	2.2	0					
Bmd-f	2.61	2.86	2.04	1.97	1.84	0					
	2.36	2.58	1.87	1.82	1.71	0					
	2.16	4.59	1.15	1.11	0.75	2.05	0				
Bmd-n	1.63	3.6	0.62	0.61	0.49	1.74	0				
	1.43	3.35	0.51	0.49	0.41	1.59	0				
	3.77	3.21	3.71	3.62	3.55	2.54	3.27	0			
mA¶	2.83	2.43	2.93	2.88	2.86	2.06	2.67	0			
	2.7	2.16	2.86	2.82	2.8	1.93	2.64	0			
	5.06	1.26	4.17	4.12	3.86	2.9	3.78	2.85	0		
mB¶	3.88	1.06	3.27	3.24	3.06	2.27	2.98	2.16	0		
	3.47	0.96	3.01	2.98	2.85	2.05	2.77	1.89	0		
	2.26	4.39	1.27	1.28	0.88	1.87	0.8	3.15	3.56	0	
Amc	1.66	3.56	0.73	0.76	0.56	1.6	0.6	2.64	2.84	0	
	1.44	3.27	0.66	0.66	0.52	1.46	0.52	2.56	2.61	0	
	2.3	4.27	1.26	1.24	0.77	1.77	0.76	3.11	3.41	0.47	0
Bmc	1.71	3.45	0.77	0.78	0.56	1.55	0.59	2.58	2.74	0.26	0
	1.49	3.19	0.69	0.68	0.52	1.43	0.53	2.52	2.54	0.19	0

[a] For each pair of structures, three rmsd values were determined taking into account all heavy atoms, all heavy atoms except those in terminal residues, and all heavy atoms except terminal residues and backbone atoms. *mA* and *mB* are energy-minimized A- and B-form DNA models generated by AMBER (*) and DNAminiCarlo (¶). Corresponding models generated by the two programs were found to be somewhat different (as can be seen by their rmsd values). *Amd1* and *Amd2* are final models refined by AMBER from *mA** using two different seeds for random number generation (vide infra). *Bmd*, *Bmd-f*, and *Bmd-n* are final structures calculated by AMBER starting from the *mB** model: *Bmd* was obtained by enforcing our experimental NOE restraints, *Bmd-f* without applying any experimental restraints (free MD), and *Bmd-n* using empirically narrowed bounds. *Amc* and *Bmc* are final structures calculated by DNAminiCarlo starting from *mA*¶ and *mB*¶ minimized models, respectively.

Table 5: Potential Energies, E_{pot}, Constraint Energies, E_{cnst}, and Distance Violation Values for Initial and Final Structures[a]

structure	$E_{pot} - E_{cnst}$ (kcal/mol)	E_{cnst} (kcal/mol)	Δ_{avg} (Å)	Δ_{closer} (Å)
*mA**	−1035.471	9464.204	1.086	0.744
*mB**	−1196.288	1347.482	0.500	0.218
Amd1	−1096.467	147.740	0.345	0.075
Amd2	−1067.156	140.544	0.339	0.073
Bmd	−1142.620	158.007	0.345	0.079
Bmd-f	−1306.004	3211.108	0.604	0.321
Bmd-n	−1083.511	431.977	0.316	0.081
mA¶	−226.034	−	0.774	0.481
mB¶	−179.325	−	0.515	0.234
Amc	−47.427	704.450	0.435	0.155
Bmc	−56.071	587.310	0.421	0.144

[a] Two kinds of distance violation values are calculated: to the center position of the bound ranges, Δ_{avg}, and to the closest bound, Δ_{closer}. Energy values are calculated using different force fields in the AMBER (rMD) and DNAminiCarlo (rMC) programs. Hence, the values are not comparable. Refer to the footnote of Table 4 for an explanation of the structure names.

Energies for the converged structures, deducting the restraint energy penalty, are comparable to those of the starting models (Table 5). For each molecule, two values of distance violation are reported. These are calculated to the center position of the experimental bounds range, and to the closest bound, Δ_{avg} and Δ_{closer}, respectively. *Amd1* and *Bmd* give sensibly smaller distance violations than the initial models, especially the A model (Table 5). The large width of our experimental bounds is reflected in the large difference between the Δ_{avg} and Δ_{closer} values.

Even when rMD runs start from the same model, velocities assigned to the atoms at the beginning of the simulation may affect the conformational search. These velocities are randomly generated from a user-selected seed number entered into AMBER. To evaluate the effect of using different seed numbers, we calculated a new structure, *Amd2*, running rMD starting from A-form DNA with a different random seed number. This structure is essentially identical to *Amd1*, as indicated by low rmsd values, as well as by comparable energies and distance violations (Tables 4 and 5).

We need to consider that our wide bounds may not have enough restraining power to drive the structure toward its target conformation, letting the chemical force field dominate. To test for this effect, we ran two independent MD simulations: the first without experimental restraints and the second applying bounds with a reduced width. Both simulations started from B-DNA models using the same protocol as with wide bounds. We observed that free MD yields a final structure (*Bmd-f*) that, judging from rmsd values, energies, and distance violations, is intermediate

Table 6: Sixth-Root *R* Factors (R^x) Calculated by CORMA for Initial and Refined Structures Using Experimental 2D NOE Intensities Acquired at Four Mixing Times[a]

	270 ms mixing time			200 ms mixing time			130 ms mixing time			70 ms mixing time		
structure	intra	inter	global	intra	inter	global	intra	inter	global	intra	inter	global
*mA**	0.160	0.211	0.182	0.183	0.265	0.217	0.206	0.299	0.244	0.272	0.278	0.274
*mB**	0.062	0.093	0.075	0.063	0.100	0.079	0.071	0.122	0.093	0.087	0.118	0.100
Amd1	0.050	0.060	0.054	0.051	0.062	0.056	0.052	0.063	0.057	0.054	0.067	0.060
Amd2	0.051	0.060	0.055	0.053	0.061	0.056	0.053	0.062	0.056	0.055	0.065	0.059
Bmd	0.050	0.060	0.055	0.051	0.063	0.056	0.053	0.064	0.058	0.055	0.068	0.060
Bmd-f	0.075	0.110	0.090	0.074	0.126	0.096	0.079	0.161	0.113	0.104	0.180	0.135
Bmd-n	0.048	0.058	0.052	0.046	0.063	0.053	0.046	0.069	0.056	0.050	0.073	0.059
mA¶	0.118	0.142	0.128	0.122	0.168	0.141	0.130	0.207	0.162	0.139	0.248	0.184
mB¶	0.062	0.097	0.077	0.062	0.110	0.082	0.071	0.136	0.098	0.111	0.147	0.126
Amc	0.068	0.076	0.071	0.068	0.081	0.073	0.068	0.085	0.075	0.116	0.106	0.111
Bmc	0.067	0.073	0.069	0.067	0.076	0.070	0.068	0.081	0.073	0.112	0.100	0.107

[a] Refer to the footnote of Table 4 for an explanation of the structure names.

between A- and B-form DNA but distant from *Amd1* and *Bmd*. On the contrary, the structure calculated with tighter bounds (*Bmd-n*) has features similar to those of the structures calculated with wider bounds (Tables 4 and 5).

Perhaps the most important criterion for evaluating the quality of NMR structures is the estimation of their ability to reproduce the experimental data by calculating residual factors, *R* and R^x. These are determined by comparing the NOE intensities calculated for a particular structure at a given mixing time with the corresponding experimental intensities. This is done with the program CORMA (*28*, *41*). For each structure, we calculated *R* and R^x values individually for NOE spectra acquired at four mixing times (Table 6). *R* factors sensibly improved after rMD refinement. *Bmd*, *Amd1*, and *Amd2* all have comparable *R* factors, lower than those of the starting geometries, especially for A-form DNA. As expected, with narrow bounds, a slight decrease in *R* factors resulted, while free MD yielded a structure (*Bmd-f*) with *R* factors higher than those of B-form DNA. This indicates that the experimental bounds are driving the structure during rMD and not the empirical force field. Consistent with our observation that NOE spectra recorded at intermediate mixing times (130 and 200 ms) gave spectra of better quality, they also had lower *R* factors.

Restrained Monte Carlo Calculations. For final corroboration of the structures calculated by rMD, we used our experimental bounds as restraints with a different structure refinement method, restrained Monte Carlo calculations (rMC). The Monte Carlo subroutines of the DNAminiCarlo program were used for this purpose. With DNAminiCarlo, not only is the conformational search method different than MD, but the force field and the definition of molecular structure are also different from those used in AMBER. In fact, DNAminiCarlo uses helical parameters to describe the structure of DNA duplexes, rather than atomic coordinates.

Again, to test dependence on the initial model, we ran rMC starting from energy-minimized A- and B-DNA models (both generated by DNAminiCarlo). However, in contrast to our rMD simulations, the same protocol was used, consisting of a single cycle of simulated annealing starting from both A- and B-DNA models. The same value of restraint force constant as in rMD simulations, 20 kcal mol^{-1} $Å^{-2}$, was chosen. Two final structures, *Amc* and *Bmc*, were obtained starting from energy-minimized A- and B-DNA, respectively. Rmsd values calculated between *Amc* and *Bmc* are well below 1.0 Å, even when heavy atoms of terminal residues are considered (Table 4). Thus, when we started from A- and B-DNA models, rMC produced even better convergence than rMD. It is possible that the fewer degrees of freedom allowed during rMC simulations result in better convergence. It is most reassuring that rmsd values calculated against rMD structures indicate that Monte Carlo calculations produce a geometry similar to rMD, especially with B-DNA as the starting model. *Amc* and *Bmc* structures are shown in Figure 2 superimposed on the final models refined by rMD (*Amd1* and *Bmd*).

Energies of structures calculated with different force fields cannot be compared directly. Instead, their energies can be compared to those of A- and B-form DNA minimized with the corresponding force fields. Thus, rMC final structures have energies that are about 100 kcal/mol higher than those of both A- and B-form DNA models. As in rMD, distance violations decrease also during Monte Carlo refinement; however, the rMC final structures have values that are about 0.1 Å higher than those of rMD structures (Table 5).

Analyzing *R* factors for rMC structures (Table 6), we noticed the final structures and the initial B-DNA model have similar values, both being considerably smaller than the initial A-DNA model *R* factors. A closer look reveals that intraresidue NOE intensities are better satisfied by the initial B-DNA model, while the interresidue intensities give smaller *R* factors with the final structures. In addition, some starting models may well satisfy intraresidue intensities, since these depend only on sugar pucker and glycosidic torsional angle. Interresidue bounds, instead, are more complex and can be better explained only after restrained conformational search. Because of inconsistencies in NOE intensities (due to experimental error, molecular flexibility, etc.), improved fitting of interresidue restraints may dictate disruption of local geometry, worsening intraresidue *R* factors.

Energies, distance violations, and *R* factors of structures determined by Monte Carlo methods could be improved by performing consecutive cycles of simulated annealing, as we previously described (*34*). However, our goal was to refine DNA models with rMC methods to conformations similar to those found by rMD. Even though the structure refinement method, the force field, and the initial models are different from those used in rMD refinement, we found essentially the same structure. This strongly indicates that the experimental data are responsible for the final geometry and that the structure we calculated represents a good time-averaged representation of our DNA undecamer in solution.

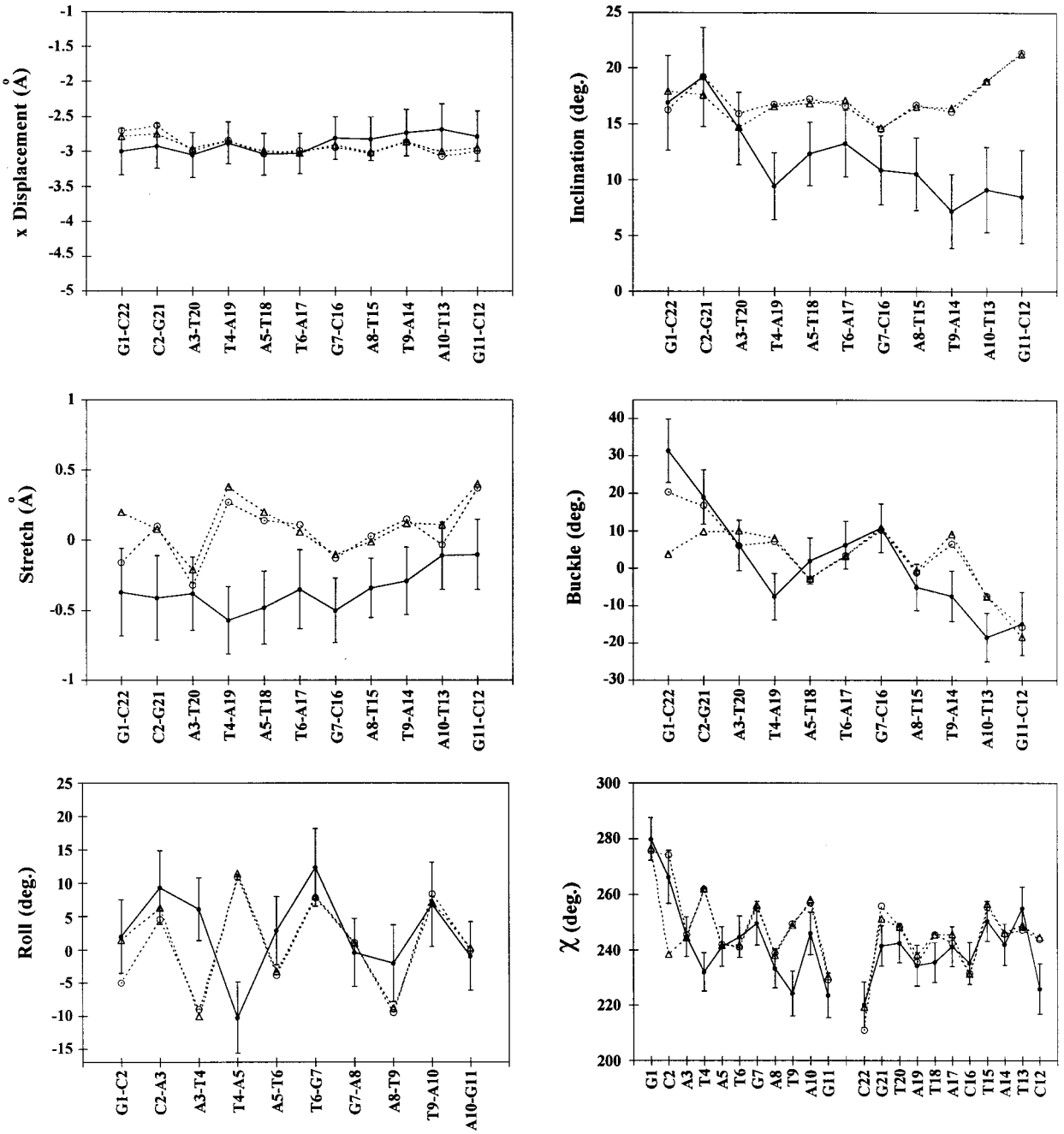

FIGURE 3: Selected helical parameters and glycosidic torsion angle calculated from 100 ps of rMD simulation at 300 K starting from the *Bmd* model (solid curve; average and standard deviation shown) and from the final *Bmc* (○) and *Amc* (△) models.

DISCUSSION

Analysis of Refined Structures. Perhaps the most striking feature of our final structure is the presence of a hole in the center of the DNA duplex when it is viewed along the helix axis (Figure 2, bottom). This quality is observed in canonical A-DNA, but not in B-form DNA, and is reflected in the large negative X displacement of the bases (Figure 3). This finding is consistent with a pattern that has been observed in our laboratory for other DNA sequences. In fact, with improvement of the NMR technique, it has become evident that DNA duplexes in solution deviate somewhat from canonical B-DNA with some features tending to resemble those of the A-DNA conformation. The X displacement and inclination of the bases are among those helical parameters that deviate most toward A-form DNA. The refined model of our undecamer, however, shows features that are closer to A-form DNA than any DNA duplex previously solved in our laboratory. Since our experimental restraints are much wider than those used to solve other duplexes, we considered the possibility that this large bound width could be the reason for the strong A-DNA character we found for our undecamer. To investigate further the determinants of our final conformation, we ran two independent MD simulations: one using experimental bounds that were empirically narrowed and the second without experimental restraints. The final structure, *Bmd-n*, obtained with narrowed restraints, has X displacement and inclination of the bases closer to those of canonical B-DNA, which indeed suggests that our wide bounds are responsible for the A-DNA-like features exhibited by our undecamer. On this basis, one could further argue that the

empirical force field, permitted by wide experimental bounds, is driving the structure toward A-form DNA. This hypothesis, however, is not supported by free MD simulations, which show that the force field actually drives the molecule toward a structure, *Bmd-f*, with more B-DNA-like helical parameters. Moreover, very similar structures, which have some A-DNA characteristics, were also refined by Monte Carlo methods (*Amc* and *Bmc* models), which calculate molecular energies using a force field different from that in AMBER (*22*). Consequently, we cannot positively identify any single factor responsible for the strong A-DNA character shown by our final models. Perhaps, the large width of our bounds causes different types of restraints to become important in defining the DNA structure. At the same time, we cannot rule out the dependence of the final geometry on the particular sequence of our duplex, since we do not know what the effect of wider bounds on other sequences would be.

One important consequence of X displacement and inclination of the bases being closer to those of A-DNA is that the global shape of the molecule resembles that of A-form DNA. This may explain why our final structure, surprisingly, gives lower rmsd values when compared to A- than to B-DNA initial models (see Table 4).

Helical Parameter Differences between Refined rMD and rMC Structures. To investigate how precisely the helical parameters are defined by the experimental bounds, we ran 100 ps of rMD simulations at a temperature of 300 K starting from the structure refined from B-DNA (*Bmd*). The same value of k_{NOE}, 20.0 kcal mol^{-1} Å^{-2}, was used throughout the simulation. At the end, for each helical parameter, we calculated its average and standard deviation values. A first look at the standard deviation of several parameters shows that terminal residues have a broader distribution, suggesting, as expected, that the structure of the duplex is less defined at its ends. A similar procedure was also used to obtain final rMC structures. In Monte Carlo simulations, in fact, the independent helical parameters of all the structures generated during the last Markov chain fragment at 300 K were averaged and the mean values were used to construct the final model. A selection of helical parameters from the 100 ps rMD simulation and for two final rMC structures, *Bmc* and *Amc*, are plotted together in Figure 3.

Considering the low rmsd values we calculated between these structures, especially for internal residues, we would expect helical parameters to be very similar. Surprisingly, while the two rMC models are essentially identical, significant changes are found with rMD helical parameters. In particular, base pair parameters, e.g., opening and stretch, and base parameters, such as inclination and Y displacement, manifest the biggest differences. Other rMC helical parameter differences are mainly contained within the standard deviation of the corresponding rMD parameters. Moreover, parameters that define the geometry of the base pair and steps involving residue T4 show the biggest differences between rMD and rMC structures, while T6, G7, and A8 are among the best-defined residues. To understand this deviation in helical parameters, we need to consider the differences in how the DNA duplex structure and experimental data are handled during rMD and rMC refinements.

First of all, DNAminiCarlo and AMBER, the programs used to run rMC and rMD simulations, respectively, describe DNA duplexes in two different ways. In DNAminiCarlo, the independent helical variables are used to describe duplex geometry, with a single parameter, the pseudorotation phase angle *P*, defining sugar puckers. Given a value of *P*, the sugar conformation with minimum energy is taken. Consequently, base and sugar rings during Monte Carlo simulations are locked into optimal conformations. In contrast, with AMBER, the molecules are defined by the Cartesian coordinates of their atoms. Thus, AMBER allows more degrees of freedom for the structure to satisfy experimental restraints, which explains why models refined by rMD have lower *R* factors and distance violations than those refined by rMC (Tables 5 and 6).

Second, from analysis of our experimental data, we found that some classes of restraints are systematically inconsistent with others and, hence, all restraints cannot be satisfied simultaneously by any single conformation. A more detailed analysis of NOESY and COSY data prompted us to believe that this lack of consistency arises from the flexibility of the sugar moieties and about the glycosidic bonds in DNA duplexes (paper in preparation). Differences in how these incongruities are "assimilated" into the DNA structure may lead to the observed deviations between rMD- and rMC-refined models. Thus, during molecular dynamics simulations, incompatibilities in the experimental bounds can be partially or fully satisfied by forcing base and sugar rings into slightly distorted conformations. On the other hand, in Monte Carlo calculations, the geometry of base and sugar rings cannot be altered to accommodate experimental inconsistencies, since only optimal conformations are permitted. Hence, during rMC runs, locally inconsistent restraints perturb the helical geometry.

Inconsistencies between experimental bounds and empirical hydrogen bond restraints can also contribute to or be responsible for observed deviations of base pair parameters (e.g., opening or stretch). In structures refined by rMD, the bases are often twisted to satisfy experimental bounds and flat Watson−Crick hydrogen bonds, simultaneously. Instead, during Monte Carlo calculations, since bases cannot be bent, hydrogen bonds are somewhat disrupted to enforce experimental restraints. As expected, inspection of rMC and rMD final models reveals that the bases have nonflat geometry in structures refined by rMD, but not in the rMC ones. Also, the pucker amplitude of sugar rings, Φ, reaches unrealistic values during simulations of molecular dynamics, while it is contained within a narrow range (35−40°) in Monte Carlo runs (see Φ in Figure 4).

A third factor leading to differences between rMC and rMD structures is the fact that the helical parameters extracted for rMD structures may not be properly calculated due to deformed base ring geometry. Artifacts in measuring helical parameters may also cause the observed discrepancy between these parameters and rmsd values, i.e., how structures can give rmsd values of <1 Å yet show significant differences in helical parameters. For example, helical parameters of T4·A19 base pairs and T4·A5 steps are quite different between rMD and rMC final models (Figure 3). However, the bases of T4·A5 steps from rMD and rMC overlap with an rmsd value of 0.25 Å (Figure 5). Visually, the only significant difference between these steps is in the geometry of T4, which is rather distorted in rMD structures.

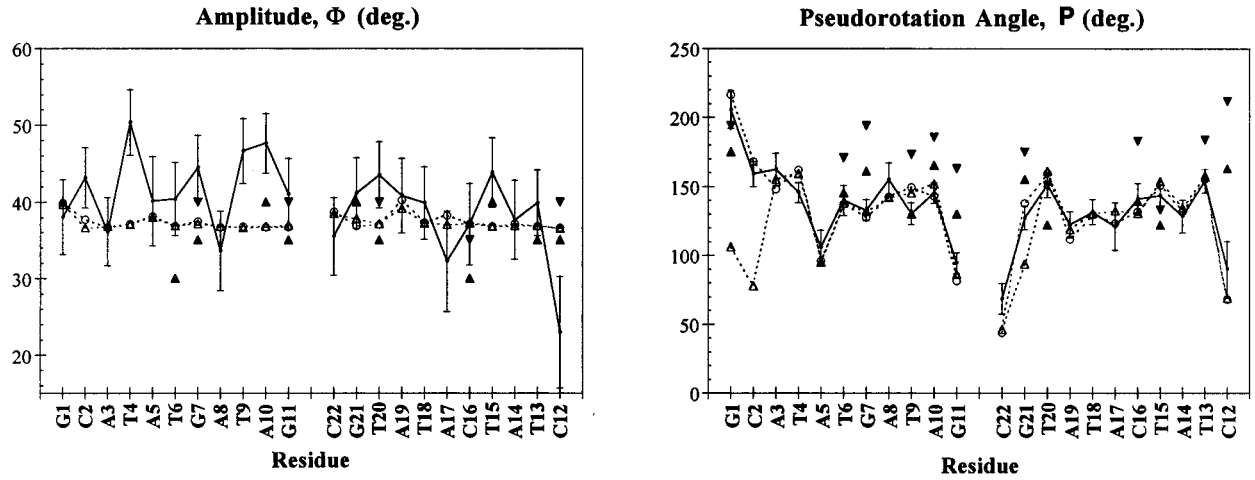

FIGURE 4: Phase angle of pseudorotation *P* and amplitude Φ calculated for sugar moieties of structures generated during 100 ps of rMD simulation at 300 K starting from the *Bmd* model (solid curve; average and standard deviation shown), for the final *Bmc* (○) and *Amc* (△) models refined by rMC calculations, and extracted from 2QF-COSY cross-peak analysis (τ).

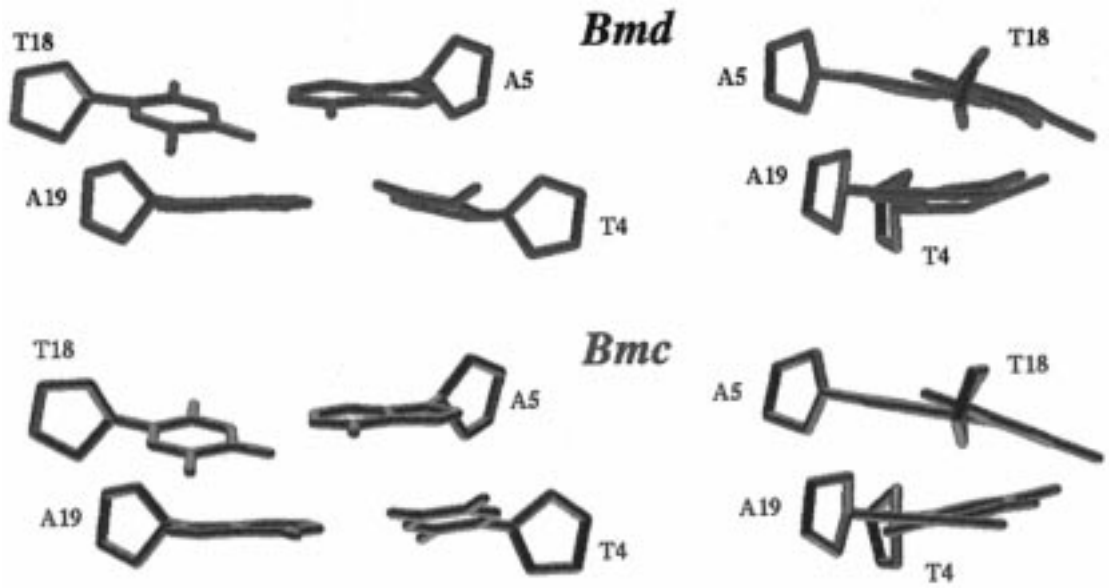

FIGURE 5: Side-by-side views of the T4-A5·T18-A19 base pair step of the *Bmd* (green) and *Bmc* (red) final structures. While the rmsd between these structures is only 0.25 Å (calculated using only the heavy atoms of the bases), they exhibit significant differences in helical parameters (see Figure 3).

We conclude that helical parameter differences between rMD and rMC structures are mainly derived from internally inconsistent experimental bounds. These inconsistencies distort base rings and promote unusual sugar pucker amplitudes in rMD structures, and local helical parameter perturbations in rMC structures refined with DNAminiCarlo. Moreover, helical parameters in rMD final structures may be ambiguously calculated because of nonflat base geometry.

Despite all this, rMD and rMC final models overlap with rmsd values well below 1 Å (considering only internal residues, Table 4), indicating that the structure of our DNA undecamer duplex depends little on the refinement method used and is essentially defined by experimental bounds. The relatively large number of accurate distance bounds per base (15−20 restraints/base) probably compensates for the smaller restraining power of the wider bounds. In other words, a large number of accurate bounds, defined with lower precision, will lead to a high-quality structure. By superimposing corresponding base pair steps of rMC and rMD structures, we found that T6·G7, G7·A8, and A8·T9 steps are better defined and superimpose with lower rmsd values than the overall structures. The conformation of these residues is defined by more experimental bounds than other regions of the duplex (Figure 1), consistent as well with the notion that the experimental data and not the refinement method determine the structure.

Visually Analyzing Final Models. Since we cannot fully rely on analyzing our final structures by their helical parameters, we visually inspected the final models. First, the geometry of terminal base pairs appears to be distorted, probably due to fraying of the duplex at its ends. In particular, terminal and penultimate residues show a high positive buckle at the 5′-end of the duplex and a high negative buckle at the 3′-end (Figure 3). This is also associated with the sugar pucker of terminal residues being shifted toward the *N* conformation (Figure 4) and some unusual backbone torsional angles. Consequently, the geometry of these residues should not to be used for structural analysis.

Looking at internal residues in structures refined by AMBER, we noticed that all thymidine residues have nonplanar base geometry with the methyl group displaced from the base plane toward the 3′-end. The thymine base ring is also oriented with its C5−C6 edge toward the 3′-end as if it is being pulled by the methyl group. Also in adenosine residues, the H8 atom is shifted from the base plane toward the 3′-end, but not as much as the methyl group in thymidines. Moreover, the distortion caused by H8 does not seem to change the orientation of the base ring. In rMC structures, thymine and adenine bases are oriented similarly. As expected, the thymine methyl groups and the adenine H8 atoms remain in the base plane, since DNAminiCarlo does not allow deformation of base geometry. As a result of the orientation of thymine and adenine bases, T·A and A·T base pairs should have negative propeller twist, T·A steps positive roll, and A·T steps close to zero or negative roll.

Guanine bases behave like adenine bases, while cytidines resemble thymidines, with the H5 atom shifted toward the 3′-end. However, the G7·C16 base pair appears flatter than A·T base pairs. Other G·C base pairs in our sequence are either terminal or penultimate, so their structural features are subject to terminal fraying.

T·G steps have positive roll, with T6·G7 exhibiting the largest roll. This observation is in accord with the T·G geometry in other sequences studied in our laboratory. A big positive roll causes the helix axis to be bent. This feature may be typical for pyrimidine-purine steps, such as T·G and T·A.

Sugar and Backbone Conformations. Sugar puckers in the undecamer have *P* values characteristic of the C2′-endo pucker, except for terminal residues, with rMD and rMC structures in good agreement (Figure 4). However, sugar pucker amplitudes, Φ, oscillate within a restricted range in Monte Carlo final models (35−40°), but reach unrealistically high values in the rMD structures (up to 50°, Figure 4). As already mentioned, this is probably a compromise to accommodate inconsistencies in experimental restraints which is allowed during rMD runs by the AMBER program, but not by DNAminiCarlo in rMC calculations.

Figure 4 also shows the sugar conformational parameters, *P* and Φ, that were determined from deoxyribose vicinal coupling constants (Table 1). This information was not used for structural refinement but could be used to cross-validate the final structures. Except for terminal residues, values for P_S, the pseudorotation angle for the major conformer, determined from 2QF-COSY analysis are within 10° of the corresponding values in the refined structure, with the COSY-derived values being systematically larger. The pucker amplitude ranges between 30 and 40°. To calculate the sugar conformation from proton−proton coupling constants, we had to assume a two-state model with the sugar ring undergoing fast *S*−*N* conformational exchange. However, rMD and rMC procedures assume a conformational search for only a single conformation with no internal motion. When conformational exchange is ignored, the sugar rings in the refined structures have lower *P* values.

The backbone conformations of rMC- and rMD-refined structures exhibit similar torsion angles typical of B-DNA. Again, deviations are found in terminal residues and, surprisingly, also for residue T4. This is just another indication of the peculiar nature of this residue.

T4 Is a Special Residue. Further inspection of T4 in rMD structures reveals that the methyl group is out of the base plane in the 3′-direction, pulling the C5 atom of the ring in the same direction. At the same time, the base is still engaged in flat hydrogen bonding with A19. The result is that the T4 base has a rather distorted geometry in the final rMD models (Figure 5). In rMC structures, T4 has a much flatter base with the methyl group in-plane, but the hydrogen bonds to A19 are perturbed (Figure 5). This explains the observed differences between rMD- and rMC-refined structures in most T4·A19 base pair and some T4·A5 step parameters, and indicates that some restraints acting on T4 are not consistent. Whether this is the result of internal flexibility or experimental error, we do not know.

Analysis of Distance Restraint Violations in the Final Structures. An indication of internal inconsistencies in our experimental data comes from analysis of bound violations in final structures. For this purpose, we consider models refined by rMD and rMC separately. With the larger number of degrees of freedom permitted for the structure by AMBER, rMD-refined models give lower distance violations than rMC structures. Also, we noticed that some types of restraints involving sugar protons are consistently violated. In most of these cases, the bounds are shorter than the distances in the final structures. Only a few distances violate the experimental restraint upper bound. Internal flexibility, such as *N*−*S* repuckering of the sugar ring, may be responsible for such systematic bound violations. In fact, NOE intensities are averaged by rapid exchange between two local conformations. Moreover, because of the inverse sixth-root relationship between distances and intensities, estimated distances will be biased toward the conformer with the shortest distance. Even the presence of a small amount (e.g., 10−20%) of a second conformer may significantly affect those experimental bounds with shorter distances in such a minor conformer. Fast *N*−*S* sugar repuckering was also suggested by COSY cross-peak analysis.

However, other experimental bounds, such as H1′−base intraresidue restraints, suggest that sugar repuckering by itself is not enough to fully explain bound violations. H1′−base intraresidue distances in refined models are consistently longer than the corresponding experimental bounds, but this cannot be explained by means of sugar flexibility, since H1′−base intraresidue distances do not directly depend on sugar conformation. A more detailed analysis of experimental distance bounds is being conducted in our laboratory and should provide insights into conformational flexibility which will be the subject of a future paper.

CONCLUSIONS

Determining solution structures involves several steps ranging from extraction of structural information from NMR spectra to refinement of the structure using such restraints. During each step, variables must be selected and assumptions made. The resulting geometry could be affected by the methods and variables chosen, especially if the experimental data poorly define the structure. In our study, we have tried to minimize the effect of the method on the final outcome, and in the process, we have examined the influence of some

of the choices on the resulting structure. Interproton distances and bounds were calculated very conservatively from 2D NOE intensities by running MARDIGRAS hundreds of times using 2D NOE spectra for exchangeable protons and for nonexchangeable protons recorded at different mixing times, assuming different overall correlation times and several different starting structures (vide supra). As a result, the distance bounds we obtained are wider than those for any previous studies conducted in our laboratory, raising the question of whether these bounds would have enough restraining power to yield the correct geometry. For this reason, we decided to refine the structure of our molecule using two different methods: restrained molecular dynamics (rMD) and restrained Monte Carlo calculations (rMC). These methods have different conformational search procedures, different empirical force fields, and even a different way of describing the DNA duplex structure (atomic coordinates vs helical parameters). Nonetheless, rMC and rMD yielded essentially the same structures, with the atomic rmsd being well below 1 Å between final structures. Hence, we are confident that our final structure is defined by experimental restraints and does not depend significantly on the refinement method or variables used. While convergence to the same final conformation, even when different refinement methods are used, indicates that the experimental restraints are defining the structure of our duplex, it does not tell us how well this structure describes the real conformation of the molecule in solution. Comparison of the experimental 2D NOE intensities and intensities calculated for the refined structures via conventional *R* factors and sixth-root *R* factors (R^x) is one good criterion indicating that the converged structures fit the experimental NOE data very well.

Surprisingly, refined rMD and rMC structures exhibiting low rmsds revealed significant differences in helical parameter values. Analysis suggests that these discrepancies depend on the effect of bound inconsistencies on structure calculation by the two different methods. In particular, the larger degrees of freedom permitted during rMD cause sugar and base rings to be distorted in accommodating inconsistencies in experimental data. In contrast, this is not possible in the rMC refinement. It was also interesting that helical parameters would often suggest that two structures differ more than what appears from simple visual inspection. It turns out that base distortions, such as those occurring during MD calculations, may impair extraction of helical parameters. Because of this, our paper does not rely much upon helical parameter analysis.

As noted in previously solved NMR structures of DNA duplexes, some helical parameters deviate from those of canonical B-DNA, taking on some characteristics of A-DNA. This quality is even more pronounced in the refined duplex structure reported here, being most obvious in the large negative X displacement (Figure 2). While this study utilized larger experimental bound widths than previously employed, calculations using empirically narrowed bounds and no bounds in MD calculations could not conclusively demonstrate that the wider bounds were the source of the stronger A character; we cannot exclude the possibility that our particular duplex sequence plays the key role in defining final overall geometry.

As with the three TG steps in other duplex DNA sequences studied in our laboratory (*30*, *32*), the two T·G steps in the undecamer have a positive roll, with T6·G7 exhibiting the largest roll. A big positive roll causes the helix axis to be bent. This is in accord with the results of a normal vector analysis of DNA duplex structures deposited in the Brookhaven Protein Data Bank (*44*). Dickerson found that roll bending occurs almost exclusively at pyrimidine-purine steps. DNA bending at T·G steps has also been held responsible for the unusual electrophoretic mobility shown by duplexes containing such steps (*45*, *46*).

In each of the three TG·CA steps we have previously encountered in solution structures determined in our lab (*30*, *32*), bending was observed, but the T was preceded by a pyrimidine. In the σ^K promoter sequence studied here, we also find roll bending, although the TG·CA step has a preceding A and is in the middle of the sequence interrupting an alternating series of A·T steps. We might speculate that bending in the free solution structure may confer some advantage in binding of the promoter to the σ^K transcription factor.

In conclusion, we successfully determined a unique structure for our DNA duplex. We believe that our model is an accurate representation of the time-averaged structure of the duplex in solution. Further studies are necessary to elucidate its dynamic properties, which are clearly suggested by our experimental data.

ACKNOWLEDGMENT

We thank Carlos González for help during the early stages of this work, He Liu, Uli Schmitz, and Nicolai Ulyanov for useful discussions, and Vladimir Basus for help with running the heteronuclear NMR experiments. Use of the facilities of the Pittsburgh Supercomputing Center for some calculations and the UCSF Computer Graphics Laboratory, supported by NIH Grant RR01081, are gratefully acknowledged.

SUPPORTING INFORMATION AVAILABLE

Two tables listing the proton chemical shifts and vicinal proton coupling constants of d(GCATATGATAG)·(CTATCATATGC) (3 pages). Ordering information is given on any current masthead page.

REFERENCES

1. Koudelka, G. B., Harrison, S. C., and Ptashne, M. (1987) *Nature 326*, 886−889.
2. von Hippel, P. H., Bear, D. G., Morgan, W. D., and McSwiggen, J. A. (1984) *Annu. Rev. Biochem. 53*, 389−419.
3. Harrison, S. C., and Aggarwal, A. K. (1990) *Annu. Rev. Biochem. 59*, 933−969.
4. Schmitz, U., and James, T. L. (1995) *Methods Enzymol. 261*, 3−44.
5. James, T. L. (1991) *Curr. Opin. Struct. Biol. 1*, 1042−1053.
6. Celda, B., Widmer, H., Leupin, W., Chazin, W. J., Denny, W. A., and Wüthrich, K. (1989) *Biochemistry 28*, 1462−1470.
7. Schmitz, U., Zon, G., and James, T. L. (1990) *Biochemistry 29*, 2357−2368.
8. Gonzàlez, C., Stec, W., Kobylanska, A., Hogrefe, R., Reynolds, M., and James, T. L. (1994) *Biochemistry 33*, 11062−11072.
9. Liu, H., Spielmann, H. P., Ulyanov, N. B., Wemmer, D. E., and James, T. L. (1995) *J. Biomol. NMR 6*, 390−402.
10. Widmer, H., and Wüthrich, K. (1986) *J. Magn. Reson. 70*, 270−279.

11. Zheng, L., Halberg, R., Roels, S., Ichikawa, H., Kroos, L., and Losick, R. (1992) *J. Mol. Biol. 226*, 1037−1050.
12. Ulyanov, N. B., and James, T. L. (1994) *Appl. Magn. Reson. 7*, 21−42.
13. Stolarski, R., Egan, W., and James, T. L. (1992) *Biochemistry 31*, 7027−7042.
14. Marion, D., and Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
15. Braunschweiler, L., and Ernst, R. R. (1983) *J. Magn. Reson. 53*, 521−528.
16. Bax, A., and Davies, D. G. (1985) *J. Magn. Reson. 65*, 355−360.
17. Adams, B., and Lerner, L. (1992) *J. Magn. Reson. 96*, 604−607.
18. Liu, H., Kumar, A., Weisz, K., Schmitz, U., Bishop, K. D., and James, T. L. (1993) *J. Am. Chem. Soc. 115*, 1590−1591.
19. Bax, A., and Subramanian, S. (1986) *J. Magn. Reson. 67*, 565−569.
20. Nirmala, N. R., and Wagner, G. (1988) *J. Am. Chem. Soc. 110*, 7557−7558.
21. Pearlman, D. A., Case, D. A., Caldwell, J. C., Seibel, G. L., Singh, U. C., Weiner, P., and Kollman, P. A. (1990) *AMBER*, version 4.0, University of California, San Francisco.
22. Gorin, A. A., Ulyanov, N. B., and Zhurkin, V. B. (1990) *Mol. Biol. (Engl. Transl.) 24*, 1036−1047.
23. Arnott, S., and Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun. 47*, 1504−1509.
24. Arnott, S., and Hukins, D. W. L. (1973) *J. Mol. Biol. 81*, 93−105.
25. Weisz, K., Shafer, R. H., Egan, W., and James, T. L. (1992) *Biochemistry 31*, 7477−7487.
26. Borgias, B. A., and James, T. L. (1990) *J. Magn. Reson. 87*, 475−487.
27. Liu, H., Thomas, P. D., and James, T. L. (1992) *J. Magn. Reson. 98*, 163−175.
28. Borgias, B. A., and James, T. L. (1988) *J. Magn. Reson. 79*, 493−512.
29. Rinkel, L. J., and Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621−649.
30. Weisz, K., Shafer, R. H., Egan, W., and James, T. L. (1994) *Biochemistry 33*, 354−366.
31. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer, New York.
32. Mujeeb, A., Kerwin, S. M., Kenyon, G. L., and James, T. L. (1993) *Biochemistry 32*, 13419−13431.
33. Gonzàlez, C., Stec, W., Reynolds, M., and James, T. L. (1995) *Biochemistry 34*, 4969−4982.
34. Ulyanov, N. B., Schmitz, U., and James, T. L. (1993) *J. Biomol. NMR 3*, 547−568.
35. Lavery, R., and Sklenar, H. (1990) *CURVES 3.0. Helical Analysis of Irregular Nucleic Acids*, Laboratory for Theoretical Biology, CNRS, Paris.
36. Ravishanker, G., Swaminathan, S., Beveridge, D. L., Lavery, R., and Sklenar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 669−699.
37. Gallo, K., Huang, C., Ferrin, T. E., and Langridge, R. (1989) *Molecular Interactive Display and Simulation (MIDASplus)*, University of California, San Francisco.
38. Scheek, R. M., Russo, N., Boelens, R., Kaptein, R., and Boom, J. H. v. (1983) *J. Am. Chem. Soc. 105*, 2914−2916.
39. Feigon, J., Leupin, W., Denny, W. A., and Kearns, D. R. (1983) *Biochemistry 22*, 5943−5951.
40. Boelens, R., Scheek, R. M., Dijkstra, K., and Kaptein, R. (1985) *J. Magn. Reson. 62*, 378−386.
41. Keepers, J. W., and James, T. L. (1984) *J. Magn. Reson. 57*, 404−426.
42. Abragam, A. (1961) *Principles of Nuclear Magnetism*, Oxford University Press, Oxford, England.
43. Liu, H., Tonelli, M., and James, T. L. (1996) *J. Magn. Reson. 111B*, 85−89.
44. Dickerson, R. E. (1998) *Nucleic Acids Res. 26*, 1906−1926.
45. Beutel, A. B., and Gold, L. (1992) *J. Mol. Biol. 228*, 803−812.
46. Nagaich, A. K., Bhattacharyya, D., Brahmachari, S. K., and Bansal, M. (1994) *J. Biol. Chem. 269*, 7824−7833.

BI980481N