Solution Structure of the Octamer Motif in Immunoglobulin Genes via Restrained Molecular Dynamics Calculations†

Klaus Weisz,‡ Richard H. Shafer,‡ William Egan,§ and Thomas L. James*,‡,∥

Departments of Pharmaceutical Chemistry and Radiology, University of California, San Francisco, California 94143-0446, and Biophysics Laboratory, Center for Biologics Evaluation and Research, Food and Drug Administration, Bethesda, Maryland 20892

Received May 21, 1993; Revised Manuscript Received October 15, 1993⊛

ABSTRACT: The solution structure of the DNA decamer d(CATTTGCATC)·d(GATGCAAATG), comprising the octamer motif of immunoglobulin genes, is determined by restrained molecular dynamics (rMD) simulations. The restraint data set includes interproton distances and torsion angles for the deoxyribose sugar ring which were previously obtained by a complete relaxation matrix analysis of the two-dimensional nuclear Overhauser enhancement (2D NOE) intensities and by the quantitative simulation of cross-peaks in double-quantum-filtered correlated (2QF-COSY) spectra. The influence of torsion angles and the number of experimental distance restraints on the structural refinement has been systematically examined. Omitting part of the experimental NOE-derived distances results in reduced restraint violations and lower *R* factors but impairs structural convergence in the rMD refinement. Eight separate restrained molecular dynamics simulations were carried out for 20 ps each, starting from either energy-minimized A- or B-DNA. Mutual atomic root-mean-square (rms) differences among the refined structures are well below 1 Å and comparable to the rms fluctuations of the atoms about their average position, indicating convergence to essentially identical structures. The average refined structure was subjected to an additional 100 ps of rMD simulations and analyzed in terms of average torsion angles and helical parameters. The B-type duplex exhibits clear sequence-dependent variations in its geometry with a narrow minor groove at the T_3·A_3 tract and a large positive roll at the subsequent TG·CA step. This is accompanied by a noticeable bend of the global helix axis into the major groove. There is also evidence of significant flexibility of the sugar–phosphate backbone with rapid interconversion among different conformers.

Sequence-specific recognition of DNA by proteins is an important component of many biological processes. There is increasing evidence that local conformational variations in DNA, especially DNA bending, may play an important role in the regulation of gene expression. Mapping of bent DNA sequences to the *Escherichia coli* genome has shown that most bent sequences are located very close to transcriptional start sites (Tanaka et al., 1991) and may directly activate transcription (Lilley, 1991). The octamer sequence ATGCAAAT which is present in both the promoter and enhancer region of immunoglobulin (Ig) genes is an important regulatory element for their tissue-specific expression. This consensus sequence is recognized by various transcription factors of the POU domain family. The conserved DNA binding domain (POU domain) of these trans-activators is a bipartite structure and can be subdivided into two regions, the POU-specific box and the POU homeo box. The POU subdomains were found to have different recognition sequences overlapping the "left half" and "right half" of the consensus sequence (Verrijzer et al., 1992) and are both responsible for the high-affinity, sequence-specific binding to the octamer element (Sturm & Herr, 1988; Garcia-Blanco et al., 1989).

During the past decade nuclear magnetic resonance (NMR) spectroscopy has made major contributions in providing three-dimensional structures of biopolymers in solution (Wüthrich, 1986). Combined with refinement methods such as distance geometry, restrained molecular mechanics or restrained molecular dynamics calculations, information based on NOE intensities and three-bond coupling constants has been successfully utilized in the determination of the solution structure of various oligonucleotides associated with biologically important DNA sequences. With the growing number of data disclosing subtle, sequence-dependent variations in the DNA double-helix, the question arises as to how accurately structures can be determined by NMR. Since NMR-derived structural information relies on short-range interactions, higher-order structure above the base pair level such as global DNA bending puts even greater demands on number and accuracy of the experimental data. Several studies have addressed this issue and came to sometimes quite different conclusions (Gronenborn & Clore, 1989; Lane, 1990a,b; Metzler et al., 1990; Ulyanov et al., 1992). While some authors reported reasonably well-defined structures with discernible local structural variations (Gronenborn & Clore, 1989), others have questioned the reliable determination of sequence-dependent conformations in the NMR-refined duplexes (Metzler et al., 1990). Of course, the precision and accuracy of the structural refinement will depend on the quality of the experimental restraint set, and much effort has been directed toward a proper interpretation of the NMR data in the past. Full relaxation matrix calculations allow for spin diffusion effects and thus provide more accurate interproton distances from 2D NOE cross-peak intensities (Borgias & James, 1988; Boelens et al., 1989). Similarly, methods for a more reliable extraction of vicinal scalar couplings have been reported which are designed to simplify cross-peak patterns in *J*-resolved 2D NMR spectra

† This work has been supported by USPHS Grants CA27343 and GM39247 awarded by the National Institutes of Health.
* Author to whom correspondence should be addressed.
‡ Department of Pharmaceutical Chemistry, UCSF.
§ Food and Drug Administration.
∥ Department of Radiology, UCSF.
⊛ Abstract published in *Advance ACS Abstracts*, December 15, 1993.

0006-2960/94/0433-354$04.50/0

(Griesinger et al., 1985; Müller, 1987) or allow for the quantitative simulation of the multiplet fine structure (Widmer & Wüthrich, 1986). Another limitation on the accuracy of the NMR data arises from averaging processes due to the existence of multiple exchanging conformations in solution and rapid internal motions. These have been treated explicitly using either specific motional models or order parameters but require additional parameters which are generally difficult to obtain, thus precluding a more detailed analysis (Lane, 1990a; Koning et al., 1991).

In this paper we report the solution structure for the DNA decamer duplex d(CATTTGCATC)·d(GATGCAAATG) comprising the Ig gene octamer motif. Chemical shift assignments as well as the determination of interproton distances and deoxyribose pseudorotational parameters from homonuclear 2D NOE and 2QF-COSY[1] spectra were reported previously (Weisz et al., 1992). These are used as restraints in the present molecular dynamics refinement. Given the exceptionally complete experimental data set with a total of 398 distance and 100 torsion angle restraints, the influence on the refinement process of omitting part of the data is evaluated. The converged structure is analyzed in terms of the time average of the various conformational and helical parameters and their fluctuations in a final restrained molecular dynamics simulation.

MATERIALS AND METHODS

In the following, nucleotides of the decamer duplex are numbered according to

strand 1 5′-C1—A2—T3—T4—T5—G6—C7—A8—T9—C10
strand 2 G20-T19-A18-A17-A16-C15-G14-T13-A12-G11-5′

Sample Preparation and NMR Measurements. Preparation of the DNA sample and NMR measurements have been described in detail (Weisz et al., 1992). A pure absorption double-quantum-filtered COSY spectrum and phase-sensitive 2D NOE spectra with mixing times of 80, 140, and 200 ms were recorded in ^2H_2O at 25 °C for d(CATTTGCATC)·d(GATGCAAATG). In order to observe exchangeable amino and imino protons, additional 2D NOE experiments with 140- and 200-ms mixing times were acquired in H_2O using a $13\bar{3}\bar{1}$ excitation pulse (Hore, 1983).

NMR Analysis. Methods for the extraction of *J*-couplings and interproton distances from 2D NMR experiments were described in a previous paper (Weisz et al., 1992). Proton coupling constants for the deoxyribose were obtained by the quantitative simulation of 2QF-COSY cross-peaks with SPHINX and LINSHA (Widmer & Wüthrich, 1986). An optimized Karplus equation was subsequently used to evaluate pseudorotational parameters from the vicinal *J*-couplings for the various sugar rings (Rinkel & Altona, 1987).

Interproton distances were determined from the measured 2D NOE intensities by a complete relaxation matrix analysis with the program MARDIGRAS (Borgias & James, 1990; Liu et al., 1992). Intensities involving exchangeable protons were corrected for the nonlinear excitation profile of the reading pulse and for the effect of fast exchange with the solvent (*vide infra*) (Liu et al., 1993). MARDIGRAS was performed for the three mixing times of 80, 140, and 200 ms and standard B-DNA, energy-minimized B-DNA, and energy-minimized A-DNA as starting structures to yield nine sets of proton–proton distances.

Generation of NMR Restraints. Restraints for rMD and rMIN calculations comprise data that were obtained from the *J*-coupling and 2D NOE intensity analysis. In addition to the NMR-derived sugar torsion angles and interproton distances, hydrogen bond distances and angles were also included in the restraint data set to maintain Watson–Crick hydrogen bonding during the simulations. The restraint term introduced into the potential energy function to enforce the experimental restraints takes the form of a flat-well potential:

$$\begin{aligned} E_{\text{restr}} &= 2k(r_2 - r_1)(r_1 - r) + k(r_2 - r_1)^2 && \text{if } r < r_1 \\ &= k(r_2 - r)^2 && \text{if } r_1 \le r < r_2 \\ &= 0 && \text{if } r_2 \le r \le r_3 \\ &= k(r - r_3)^2 && \text{if } r_3 < r \le r_4 \\ &= 2k(r_4 - r_3)(r - r_4) + k(r_4 - r_3)^2 && \text{if } r_4 < r \end{aligned} \quad (1)$$

Here, k is a restraint force constant, and r represents either a distance or angle for distance restraints or for torsion and bond angle restraints, respectively. This pseudoenergy term is zero between lower and upper bounds (r_2 and r_3, respectively). Beyond this target range, the potential function rises parabolically with force constants k, until for even wider lower and upper margins (r_1 and r_4, respectively) the penalty function continues linearly to avoid exceedingly high energies for larger violations. The width of the flat and parabolic parts and the force constant of this pseudopotential were adjusted according to the accuracy and type of the experimental data.

Averaged deoxyribose torsion angles ν were calculated from the previously determined pseudorotational parameters assuming rapid exchange between N- and S-type conformers (Weisz et al., 1992). Based on the estimated accuracy of the various sugar pucker parameters, uncertainties in each individual torsion angle $\Delta\nu$ were determined. Upper and lower bounds for the flat part of the force field were subsequently set to $\nu \pm {}^1/_2\Delta\nu$, resulting in an average flat-well width of 7°. The parabolic part was limited by $r_1 = \nu - \Delta\nu$ and $r_4 = \nu + \Delta\nu$. For all 100 endocyclic torsion angle restraints arising from the 20 different sugar moieties, the same force constant k_{tors} was employed.

Distance restraints were established using three MARDIGRAS distance sets which were calculated from NOE intensities of the three mixing times and energy-minimized B-DNA as a starting model. Since energy-minimized B-DNA was shown to be the best starting structure in the MARDIGRAS calculations and closest to the target geometry (Weisz et al., 1992), distances resulting from this initial structural model are expected to exhibit the highest accuracy. For any particular proton–proton distance, the average value d and standard deviation Δd from the three distance sets was calculated. The target range was set to $d \pm \Delta d$, but a minimum well width of 0.4, 0.2, or 0.1 Å was maintained in case of 1, 2, or 3 available distances, respectively. Fast exchange with water causes a loss in NOE intensity for amino and imino protons and may therefore result in overestimated MARDIGRAS distances. To improve distance restraints involving exchangeable protons, interproton distances were recalculated with a new version of MARDIGRAS which allows for proton exchange (Liu et al., 1993). By employing exchange rates as high as 20 s^{-1} for A·T imino and 4 s^{-1} for G·C imino and amino protons, calculated imino and amino distances are likely to be

[1] Abbreviations: 2QF-COSY, double-quantum-filtered correlated spectroscopy; 2D, two dimensional; NOE, nuclear Overhauser enhancement; rms, root-mean-square; rMD, restrained molecular dynamics; rMIN, restrained minimization.

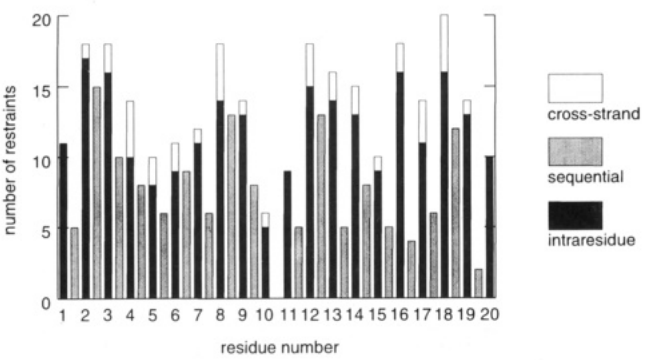

FIGURE 1: Distribution of NOE-derived interproton distances in the DNA duplex d(CATTTGCATC)·d(GATGCAAATG). Note that cross-strand restraints are given for each of the two connected residues, thus amounting to twice their actual number.

underestimated for the solution conditions employed and can be used together with the noncorrected distances to establish lower and upper bounds, respectively.

The average flat-well width over all distance restraints was 0.25 Å. Beyond the flat well, the upper parabolic part of the restraint energy was limited by $r_4 = r_3 + 2$ Å, but no limit was set for the lower parabolic side, i.e., $r_1 = 0$. Since the flat-well width does not necessarily account for individual uncertainties in MARDIGRAS-derived distances and cross-peak intensities, distances were classified into three categories depending on the estimated accuracy of the extracted peak volumes and corresponding force constants suggested by MARDIGRAS. These were subsequently assigned 100%, 70%, and 40% of the maximum force constant k_{dist}. Note that k_{dist} and k_{tors} apply to distance and angle restraint forces, respectively, and hence their values are not directly comparable with each other.

Restraints for hydrogen bonds in G·C and A·T base pairs were added by using crystallographic data (Saenger, 1984). Target distances between proton donor and acceptor in G·C base pairs were set to 2.81 Å $\leq d$(GO6, CN4) $\leq$ 3.01 Å, 2.84 Å $\leq d$(GN1, CN3) $\leq$ 3.05 Å, and 2.76 Å $\leq d$(GN2, CO2) $\leq$ 2.96 Å; for A·T base pairs, lower and upper limits were 2.72 Å $\leq d$(AN1, TN3) $\leq$ 2.92 Å and 2.85 Å $\leq d$(AN6, TO4) $\leq$ 3.05 Å. For all hydrogen bond restraints, the parabolic part extended to $r_1 = 0$ and $r_4 = r_3 + 0.3$ Å with force constants adjusted to 90% of k_{dist}. Flat angles between the three atoms forming a hydrogen bond were set to 170°–190° with parabolic regions continuing on either side for 20°. Force constants for the hydrogen bond flat angle restraints were assigned values of 20% of k_{tors}; the soft potential permits a high flexibility in orientation between complementary bases.

Overall, the restraint set consisted of 100 torsion angle restraints, 398 distance restraints, and an additional 48 restraints defining the Watson–Crick hydrogen bonds of the DNA duplex. The 398 NMR-derived distances are comprised of 240 intraresidue, 140 sequential, and 18 cross-strand distances. Their distribution along the molecule is shown in Figure 1. Clearly, the number of experimental restraints varies with residue, but there are no particularly ill-defined regions and even terminal nucleotides seem to be reasonably well characterized by the restraint set. Moreover, all residues except for the terminal nucleotides C1, G11, and G20 have valuable cross-strand connections to nucleotides of the complementary strand.

Model Building. The coordinates for canonical A- (Arnott & Hukins, 1972) and B-DNA (Arnott & Hukins, 1973) with all hydrogen atoms explicitly included were generated by use of the NUCGEN and EDIT modules of the modeling program package AMBER Version 4.0 (Pearlman et al., 1991). Hexahydrated Na^+ counterions were added at a distance of 5 Å to the phosphorus atom of the phosphate groups to neutralize any negative charges. Energy minimization was performed on these initial models with AMBER using a distance-dependent dielectric constant to mimic bulk solvent effects. The force field parameters of the AMBER force field have been reported previously (Weiner et al., 1986).

Energy minimization started by steepest descent for the initial 200 cycles followed by conjugate gradient methods. The structures were refined until the energy gradient with respect to atomic coordinates was lower than 0.1 kcal·mol^{-1}·Å^{-1}. The final energy-minimized A- and B-DNA coordinates were subsequently used as starting structures in the restrained molecular dynamics calculations.

Restrained Molecular Dynamics. All restrained molecular mechanics/dynamics calculations were carried out on a SUN SPARC2 workstation with the SANDER module of AMBER Version 4.0 which is specifically designed for refinements using NMR-derived data sets. As described above, the restraint pseudoenergy terms which are incorporated into the AMBER force field have the form of a flat well with a flat, parabolic, and linear part. The relative weights of the restraints are allowed to gradually vary during the course of a simulation by multiplying the force constants k_{dist} and k_{tors} of the parabolic regions with an appropriate scaling factor.

Molecular dynamics simulations were carried out for 20 ps *in vacuo* with a time step of 1 fs. All atoms within a 30-Å radius were included in nonbonded interactions. Bond lengths involving hydrogen atoms were kept fixed with the SHAKE algorithm (Ryckaert et al., 1977), and translational and rotational motions were removed every 100 steps. Starting from energy-minimized standard B-DNA, initial velocities were assigned to a Maxwellian distribution at 300 K. The system was subsequently kept at this temperature throughout the rMD run by weak temperature coupling to a heat bath. During the first 200 steps, no restraint forces were applied. From steps 200 to 2000, force constants were slowly built up to final values of k_{dist} = 20 kcal·mol^{-1}·Å^{-2} for the best-defined distance restraints and to k_{tors} = 60 kcal·mol^{-1}·rad^{-2} for the deoxyribose torsion angle restraints. To overcome local energy barriers, a modified rMD protocol was employed for the energy-minimized A-DNA starting model. During the first 2000 steps, the system was gradually heated up to 900 K and equilibrated at this temperature for the next 8000 steps. From step 10 000 to step 12 000, the temperature was finally adjusted to 300 K for the remainder of the rMD simulation. Following the temperature protocol, force constants were gradually increased to a maximum of k_{dist} = 100 kcal·mol^{-1}·Å^{-2} and k_{tors} = 300 kcal·mol^{-1}·rad^{-2} during the first 2000 steps and kept constant throughout the high-temperature period. During the cooling period from step 10 000 to step 12 000, restraint force constants were subsequently reduced to their final values of k_{dist} = 20 kcal·mol^{-1}·Å^{-2} for NMR-derived distances and k_{tors} = 60 kcal·mol^{-1}·rad^{-2} for the sugar torsion angles. For each rMD run of 20 ps, coordinate sets were recorded each 0.2 ps. The last 25 coordinate sets of the final 5 ps were averaged and subjected to final restrained minimization.

Molecular Mechanics. Following coordinate averaging over a rMD time period or of various rMD runs, resulting structures were subjected to restrained energy minimization using SANDER. The restraint force field used at the end of the rMD simulations was also employed during energy minimization.

Simulation of 2D NOE Intensities. 2D NOE cross-peak intensities for structures obtained during rMD refinement were calculated by the complete relaxation matrix algorithm

CORMA (Keepers & James, 1984; Borgias & James, 1988). CORMA also enables a quantitative comparison of simulated and experimental cross-peak intensities by the calculation of various numerical indices (*vide infra*). As in the previous MARDIGRAS calculations (Weisz et al., 1992), an overall isotropic correlation time of 1.6 ns for the DNA duplex was utilized.

Structural Analysis and Display. Analysis of the conformational and helicoidal parameters was carried out with the programs Curves (Lavery & Sklenar, 1988, 1989) and Dials and Windows (Ravishanker et al., 1989), which are especially suited for the characterization of irregular nucleic acids. Dials and Windows is adapted to the analysis of dynamics trajectories and utilizes the Curves algorithm for defining the structure as a function of time. All structures were displayed using MidasPlus (Gallo et al., 1989) on a Silicon Graphics Iris workstation.

RESULTS AND DISCUSSION

Molecular Dynamics and Mechanics Calculations. As pointed out (Stolarski et al., 1992; Schmitz et al., 1992), distance and torsion angle restraints may be inherently inconsistent in rMD refinements. Both structural parameters reflect the time and ensemble average over all exchanging conformations and are not necessarily compatible with a single rigid structure. This is particularly true for torsion angles of the deoxyribose where sugar repuckering occurs. Moreover, due to the nature of NOE intensities and proton–proton scalar couplings, the two types of restraints are subject to different types of averaging and thus may lead to contradicting structural information. To examine the influence of torsion angle restraints on the structural refinement in more detail, we have performed rMIN and rMD calculations with and without torsion angle restraints according to the established protocols for energy-minimized A- and B-DNA starting structures. The results obtained can be summarized as follows: (i) in a restrained energy minimization utilizing only torsion angle restraints, both energy-minimized A- and B-DNA are driven to the target deoxyribose conformation with restraint violations close to zero; (ii) the final deoxyribose pucker for the nonterminal nucleotides as enforced by the torsion angle restraints represents a low-energy conformation and closely reflects the major S -conformation P_S obtained by analyzing proton coupling constants (Weisz et al., 1992); (iii) average distance violations decrease upon incorporating torsion angle restraints in a rMD refinement, implying that both types of data are mutually compatible; (iv) additional torsion angle restraints are important in guiding an A-DNA starting structure to the global minimum. These findings suggest that due to the predominance of one conformer the employed deoxyribose torsion angles constitute valuable structural determinants in the present rMD refinement. However, care must be taken in case of conformationally less homogenous nucleotides with higher populations of minor conformers. For these residues, the average torsion angles will increasingly deviate from the exchanging low-energy conformations, and their use as restraints in a standard refinement may lead to erroneous results.

To cover more conformational space, a total of eight 20-ps rMD simulations with different random number seeds were performed for each of the two starting geometries. Coordinate sets for the last 5 ps of the rMD runs were averaged and subjected to restrained energy minimization to remove averaging artifacts. Restraint pseudopotentials for energy minimization were the same as those used during the final picoseconds of the rMD simulation and included terms for interproton distances, torsion angles, and hydrogen bonds. Besides maintaining Watson–Crick base-pairing during the molecular dynamics run, the hydrogen bond restraints do not affect the final structure as was verified by removing hydrogen bond distances and flat angle restraints for the last few picoseconds in one of the rMD simulations and the subsequent restrained energy minimization. Excluding terminal base pairs, the success rate of finding the global minimum using the above protocols and different random number seeds was close to 100% for both the A- and B-DNA starting structure. Moreover, terminal nucleotides easily converged starting with energy-minimized B-DNA. However, only about 60% of the A-DNA derived structures also showed complete convergence for the duplex termini, and these were subsequently used for further evaluation.

Table 1: Energies and Restraint Deviations for Energy-Minimized A- and B-DNA Starting Models and Structures Obtained during rMD Refinement[a]

structure[b]	E_{pot}[c] (kcal·mol^{-1})	E_{dist}[d] (kcal·mol^{-1})	E_{tors}[d] (kcal·mol^{-1})	Δ_{dist}[e] (Å)	Δ_{tors}[e] (deg)
A-DNA	−1071	8213	535	0.892	43.3
B-DNA	−1136	1370	95	0.310	8.2
rMD-A	−981	253	2	0.136	0.6
rMD-B	−1006	260	1	0.138	0.5
rMD-final	−998	255	2	0.138	0.5

[a] Force constants used for the restraint potential are k_{dist} = 20 kcal·mol^{-1}·Å^{-2} and k_{tors} = 60 kcal·mol^{-1}·rad^{-2}. [b] As defined in the text. [c] Without restraint energies. [d] Sum of distance or torsion angle violation energies as defined by eq 1 in the text. [e] Average torsion angle or distance deviation from the target bounds: $r - r_3$ for $r > r_3$ and $r_2 - r$ for $r < r_2$. See the text for definitions.

Mutual rms deviations within a family of four structures derived from either A- or B-DNA are well below 1 Å. Such deviations are comparable to the magnitude of atomic rms fluctuations during the rMD simulation and generally imply successful convergence (Nilsson et al., 1986). In fact, molecular fluctuations in the present force field lead to rms differences of up to 0.4 Å between successive 5-ps averages in a single converged molecular dynamics run. Thus, one cannot expect smaller rms deviations between independently refined models. Structures of the B-DNA family exhibit slightly lower rms deviations with an average of 0.38 Å compared to 0.54 Å in the A-DNA family. Combining all eight rMD structures, mutual rms differences do not exceed 0.9 Å. Structures rMD-A and rMD-B were obtained by coordinate averaging and subsequent restrained energy minimization of the four members in each family. These were again averaged and energy-minimized with restraints to give the final structure rMD-final.

General Characterization of the rMD-Refined Structures. Energies, deviations from the target values, and rms differences are important parameters for assessing the quality of the rMD-refined structures. In Table 1, energies and average restraint violations are listed for the energy-minimized A- and B-DNA starting models and the refined structures rMD-A, rMD-B, and rMD-final. Clearly, restraint violation energies and deviations are drastically improved for the refined structures compared to starting A- and B-DNA. Torsion angle violations are close to zero for all of the refined structures. Nevertheless, torsion angle violation energies E_{tors} comprise less than 1% of the total violation energy for the rMD structures and less than 10% for the starting structures, demonstrating that no bias is put on the torsion angles during the rMD simulations.

The potential energy E_{pot} for the final structure is increased by 70–140 kcal·mol^{-1} compared to energy-minimized A- and B-DNA and close to E_{pot} of rMD-B. The slightly smaller

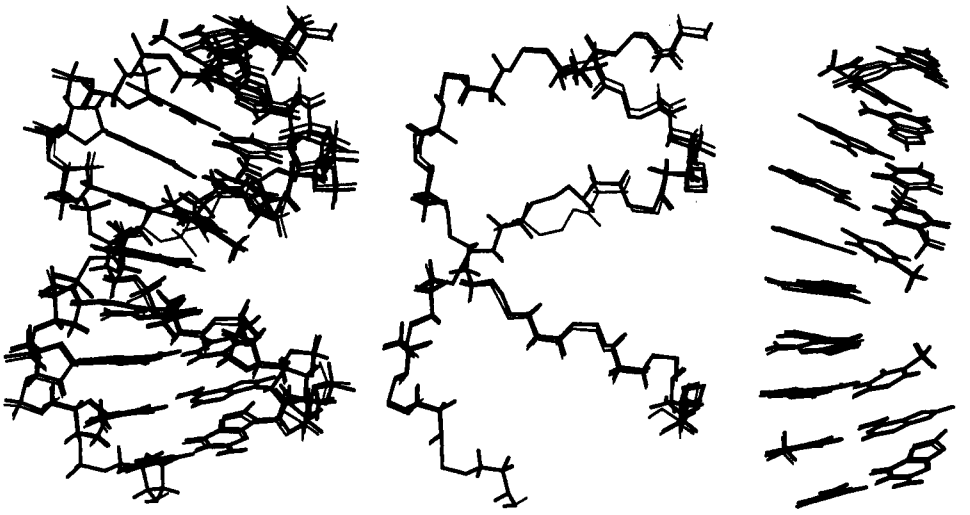

FIGURE 2: Best fit superposition of the two refined structures rMD-A and rMD-B showing all atoms (left) and only atoms of the backbone (center) and the bases (right). Bold lines represent structure rMD-A.

Table 2: Atomic rms Differences (Å) between Energy-Minimized A- and B-DNA Starting Structures and Structures Obtained during rMD Refinement

structure[a]	A-DNA	B-DNA	rMD-A	rMD-B	rMD-final
A-DNA		5.14	3.53	3.71	3.73
B-DNA			2.71	2.49	2.65
rMD-A				0.76	0.50
rMD-B					0.69
rMD-final					

[a] As defined in the text.

distance violations for rMD-A compared to rMD-B and rMD-final is paid by a higher potential energy. However, energies and restraint violations are very similar for all of the refined structures. Successful convergence from different starting geometries is also reflected in the atomic rms deviations, reported in Table 2 for the various models. The rms difference of 5.14 Å between the two starting structures, energy-minimized A- and B-DNA, is reduced to 0.76 Å between the refined structures rMD-A and rMD-B. Visual inspection of their best fit superposition, shown in Figure 2, reveals that the two duplexes have converged essentially to the same geometry even at the termini. However, average rms differences calculated exclusively for the DNA backbone atoms are 1.1 Å and thus significantly larger than those found for the base or deoxyribose moieties with rms deviations = 0.50 and 0.34 Å, respectively. A less well defined DNA backbone is generally found in NMR refinements and can be at least partially attributed to the limited number of experimental data directly related to the backbone geometry. Atomic rms deviations of the final structure rMD-final with the substructures rMD-B and rMD-A are 0.69 and 0.50 Å, respectively. Because all three rMD structures, notwithstanding minor variations in the backbone conformation, are basically equivalent, only rMD-final was subsequently used for a more detailed structural analysis.

As noted, inclusion of torsion angle restraints was beneficial for convergence from an A-DNA starting structure. However, the torsion angle restraints appear to have only a minor role in defining the final structure rMD-final. A further 100-ps simulation using only the distance restraints revealed very little change (0.34 Å rms deviation) from the structure obtained with all restraints.

In another approach to assess the quality of the refined structural models, NMR parameters, i.e., vicinal J-couplings and 2D NOE cross-peak intensities, were calculated for the final geometry and compared with the original experimental data. Proton–proton coupling constants for the individual deoxyribose sugar puckers of the structure were derived from tables in the paper of Rinkel and Altona (1987). In Figure 3a, average rms deviations between calculated and experimental J-couplings are plotted for each nucleotide of the final structure as well as for energy-minimized A- and B-DNA. Compared to A-DNA, rms deviations of B-DNA are found to be significantly smaller, highlighting the increased importance of sugar torsion angles as structural determinants in an A-DNA refinement. For the refined structure, an admixture of a considerable amount of minor conformer leads to extensive averaging of the deoxyribose geometry for the terminal residues and, as a consequence, to increasingly deviant J-couplings. Fortunately, for most of the nucleotides one conformer strongly dominates, and an average rms difference as low as 0.8 Hz for rMD-final compares favorably with the corresponding value of 1.3 Hz found for the B-DNA starting structure. Since furanoses in rMD-final fit the target sugar pucker (*vide supra*), this rms difference should be close to the lower limit which can be expected for a DNA solution structure after standard refinement.

Various numerical indices have been used as a measure to evaluate the overall fit of calculated and experimental 2D NOE intensities such as a difference index (Suzuki et al., 1986; Zhou et al., 1987) or a residual index similar to the R value used in crystallography (Gupta et al., 1988; Baleja et al., 1990a). To allow for the sixth-root dependence between distances and NOE intensities and to appropriately weigh weaker cross-peaks which are of particular importance in defining the overall structure, a modified sixth-root R factor R_1^x has been defined (Thomas et al., 1991):

$$R_1^x = \frac{\sum_i |a_o(i)^{1/6} - a_c(i)^{1/6}|}{\sum_i |a_o(i)^{1/6}|} \quad (2)$$

The summation usually runs over all observed NOE cross-peak intensities a_o and the corresponding calculated values a_c for the input structure.

We have calculated R_1^x values for the two starting structures and rMD-final with the program CORMA. CORMA calculates NOE cross-peak intensities for a particular structure by solving its complete relaxation matrix and also allows for the determination of various numerical indices against an

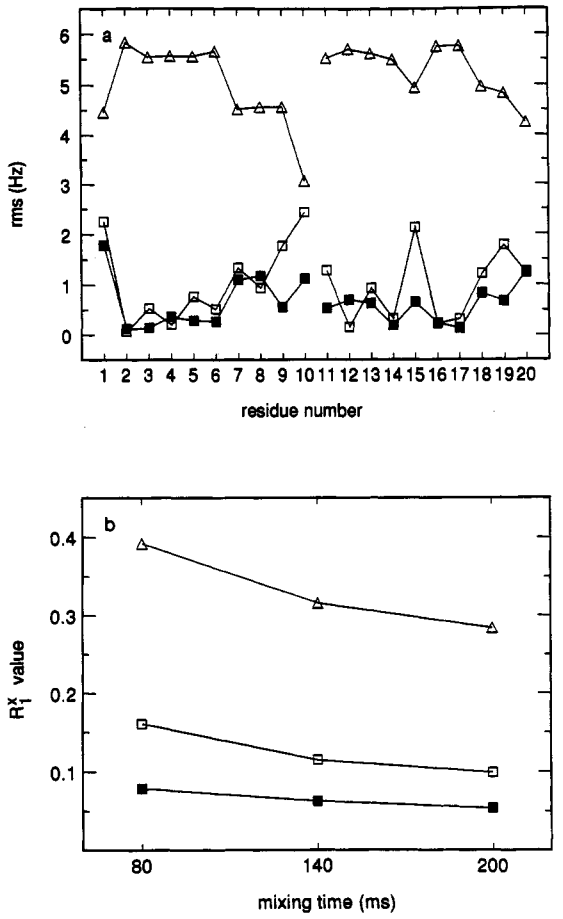

FIGURE 3: (a) Rms deviation between proton-proton deoxyribose coupling constants from the quantitative simulation of 2QF-COSY cross-peaks (Weisz et al., 1992) and calculated for energy-minimized A-DNA (triangles), energy-minimized B-DNA (open squares), and the final rMD structure rMD-final (filled squ ares) as a function of residue. (b) Sixth-root residual index R_1^x as a function of mixing time for energy-minimized A-DNA (triangles), energy-minimized B-DNA (open squares), and the final rMD structure rMD-final (filled squares). For a definition of R_1^x, see eq 2.

experimental NOE data set. As can be seen from Figure 3b, energy-minimized B-DNA fits the target conformation better than the A-DNA starting model which exhibits the highest R_1^x values. However, the fit with the experimental data is further improved for the rMD-refined structure. For all three structural models, residual indices slightly decrease with increasing mixing times. As was also found in previous studies (Schmitz et al., 1991, 1992; Stolarski et al., 1992), R_1^x values calculated only from intraresidue cross-peak intensities are significantly smaller than those calculated from interresidue NOE intensities for all structural models (data not shown). In the case of rMD-final and a mixing time of 140 ms, R_1^x (intra) was found to be 0.050 compared to 0.084 for R_1^x (inter). Clearly, interresidue distances are more sensitive to conformational variations than are intraresidue distances.

With an average of about 20 experimental distance restraints per nucleotide, our present restraint data set comprises a large number of NOE-derived distances. To test how the number of these restraints affects the refinement process, up to 50% of the distances were randomly removed from the full set of data. Similar numbers of restraints (i.e., 10 per nucleotide) are not uncommon in cases of larger DNA fragments or duplexes with highly repetitive sequences, where peak overlap severely hampers the extraction of a large number of cross-peak intensities. Using the established protocols, restrained molecular dynamics simulations were subsequently performed with these reduced restraint sets. As shown in Figure 4 for B-DNA derived models, distance restraint violations as well as potential energies for the refined structures tend to decrease upon randomly omitting experimental restraints. Moreover, R_1^x values improve more or less monotonically as a function of removed distances. These findings are in marked contrast to model calculations where parts of a simulated 2D NOE data set were used as input for subsequent structural refinements (Kaluarachchi et al., 1991). To approximate inaccuracies in experimental cross-peak volumes, a Gaussian distribution of errors was added to the cross-peak intensities in that study. The simulations showed smaller distance deviations and R factors with respect to the target geometry for structures refined against a larger restraint set. Clearly, adding random noise to the data does not reflect the real situation in that it allows for inaccuracies in integration but does not account for the effects of internal motions or conformational averaging. These contribute to more systematic inconsistencies within the restraint set and prevent a single structure from accommodating all restraints at a time. Consequently, releasing some of the restraints will confer more flexibility to the structure and therefore facilitate a closer fit with the remaining restraints without distorting the geometry. Of course, information is lost by omitting data, resulting in less defined and thus less accurate structures. This becomes evident in increasing rms differences to the structure refined against the full data set (see Figure 4c). Surprisingly however, omitting up to 30% of the distances only results in rms deviations of less than 0.8 Å, i.e., rms deviations are still within the range of atomic rms fluctuations and indicate convergence to the same structure. It should be mentioned that similar results were obtained with A-DNA as a starting structure. However, upon removal of more than 30% of the distances, the reduced restraint forces failed to guide the molecule to the global minimum, indicative of high energies and rms deviations for the resulting structure.

Obviously, R values are only of limited value when assessing the accuracy of a refined model. A small restraint set can be easily overfit to a low R value, although the structure is only poorly defined as indicated by a poor convergence from different starting geometries. The same is true when comparing individual R_1^x values for different nucleotides within a DNA duplex. Recently, a free R value has been described as a reliable and unbiased indicator of the accuracy of structures in crystallographic refinements (Brünger, 1992). A set of observed reflections is partitioned into a test set and working set, the latter being used in the modeling process while the test set is used for calculating the free R value. Since both sets are independent, R^{free} is expected to be less prone to overfitting than R. Of course, this strategy requires a data set which is redundant to some degree, therefore precluding its implementation in most cases of NMR refinement.

Encouraged by our findings that structural convergence was even achieved without using part of the total NOE data set, free R_1^x values for rMD refined structures were determined from a test set comprising 20% of the NOE distances which were omitted in the preceding refinement. As seen in Figure 5, standard R_1^x values tend to gradually improve with increasing restraint force fields, reflecting decreasing NOE distance restraint violations. In contrast, free R_1^x values increase with force constants above 20 kcal·mol^{-1}·Å^{-2}, suggesting that the data are overfit by too high restraint forces. Clearly, despite considerable fluctuations due to the limited number of data, $R_1^{x,free}$ values may be helpful in setting up optimal force constants and in assessing the quality of the refined structures apart from any energetic considerations.

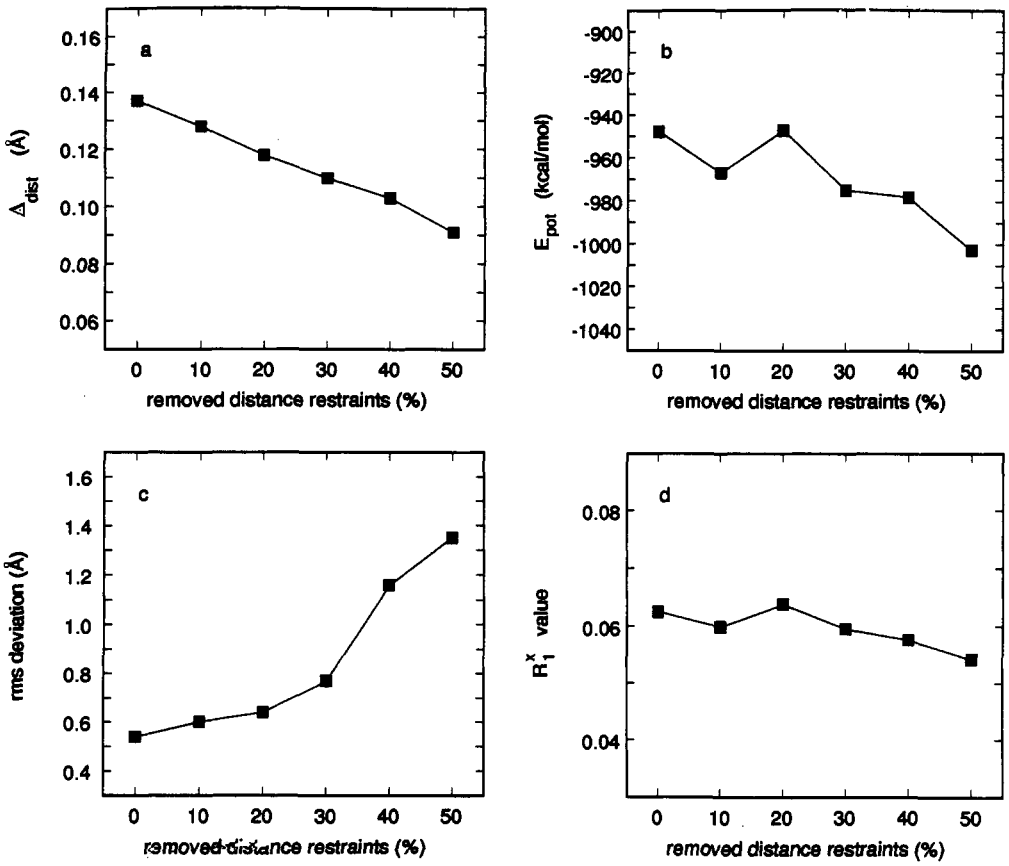

FIGURE 4: Refinements via rMD starting from energy-minimized B-DNA as a function of the percentage of randomly omitted distance data: (a) average deviation from the distance target bounds Δ_{dist}, (b) potential energy E_{pot} excluding any restraint energy terms; (c) rms deviation between structur es refined against the full and partial data set; (d) sixth-root residual index R_1^x calculated for a 140-ms mixing time. Note that the rms deviation for the full data set, i.e., with 0% of the distances removed, reflects the difference between two independently refined structures with different initial velocities. For details of the rMD protocol, see the text.

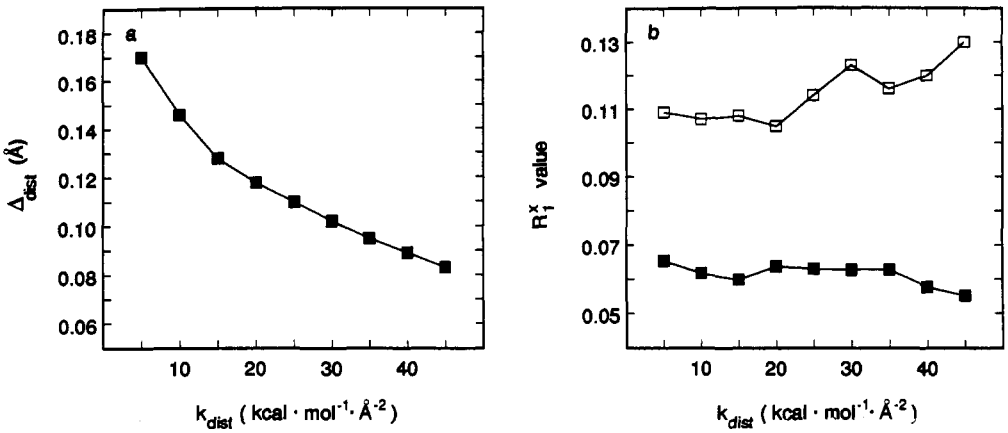

FIGURE 5: RMD refinements as a function of restraint force constant k_{dist}. The energy-minimized B-DNA starting structure was refined against an experimental data set with 20% of the distance restraints randomly removed: (a) average deviation from the distance target bounds Δ_{dist}; (b) sixth-root residual index R_1^x calculated for a 140-ms mixing time a gainst the restraint set used during refinement (filled squares) and against the distances omitted in the refinement process (open squares). For details of the rMD protocol, see the text.

Structural Analysis

The minimized structure rMD-final, shown in Figure 6a, reflects a well-defined, low-energy conformation. However, as illustrated in Figure 6b, considerable atomic fluctuations occur during a molecular dynamics simulation, and these must be taken into account when interpreting structural details of the molecule. Thus, some structural elements exhibit smaller time-dependent fluctuations and appear to be better defined than others. We have therefore subjected the structure rMD-final to an additional 100-ps rMD simulation at 300 K using the previously employed restraint force field. The resulting trajectories were subsequently analyzed in terms of average values and standard deviations for parameters exhibiting continuous, monomodal distributions. These are discussed below.

Backbone Torsion Angles. Due to the limited number of protons along the DNA backbone, backbone torsion angles are only poorly defined by NOE restraints. Direct information on the backbone conformation is largely confined to NOE cross-peaks of H5′ and H5″ protons which are strongly overlapped in most cases. Fortunately, a total of 36 distances involving H5′/H5″ protons helped to restrain the backbone angles in the present rMD calculations.

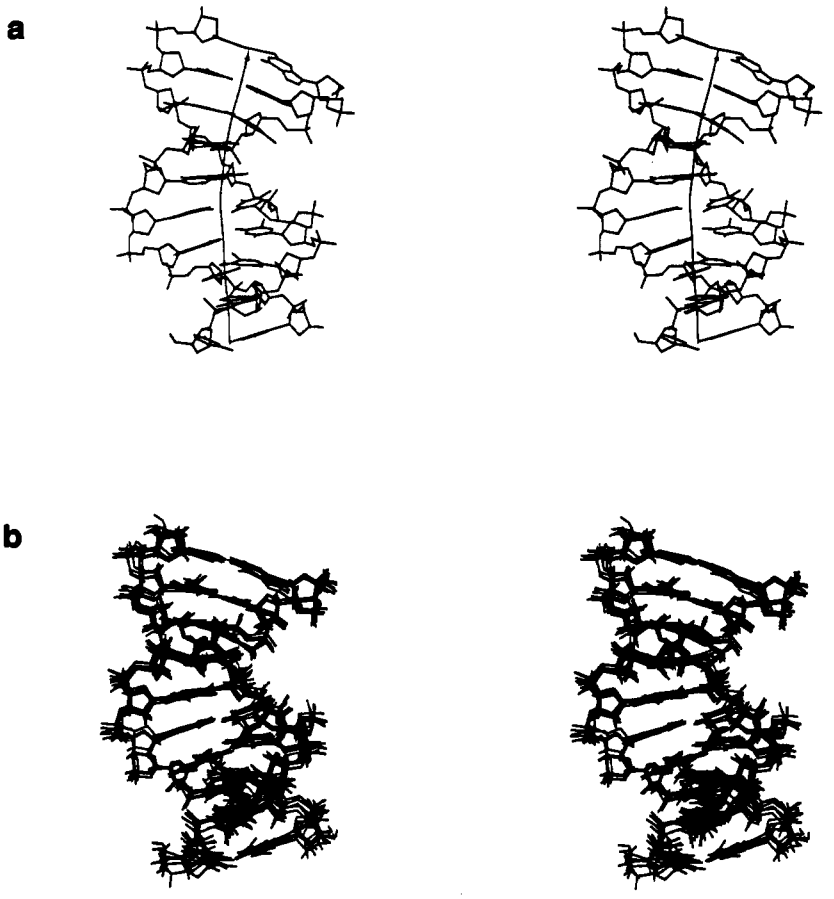

FIGURE 6: Stereoview of (a) rMD-final with its global helix axis and (b) of the best fit superposition of five snapshots taken in 1-ps time intervals during the final 100 ps of restrained molecular dynamics simulations. The view is into the major groove, with residue C1 at the bottom left; hydrogen atoms in panel a are omitted for clarity.

Table 3: Average Backbone Torsion Angles, Glycosidic Torsion Angle χ, and Pseudorotation Phase Angle P with Standard Deviations for a 100-ps Restrained Molecular Dynamics Trajectory of rMD-final[a]

nucleotide	α (deg)	β (deg)	γ (deg)	δ (deg)	ϵ (deg)	ζ (deg)	χ (deg)	P (deg)
C1				112 ± 7	−154 ± 7	−82 ± 7	−144 ± 8	134 ± 14
A2	−73 ± 8	−175 ± 7	59 ± 7	135 ± 5	169 ± 6	−84 ± 7	−102 ± 7	162 ± 7
T3	−70 ± 8	177 ± 6	63 ± 6	113 ± 5	−178 ± 6	−87 ± 6	−127 ± 6	119 ± 6
T4	−72 ± 9	174 ± 7	63 ± 8	114 ± 6	173 ± 7	−84 ± 7	−126 ± 6	118 ± 6
T5	−61 ± 9	171 ± 7	60 ± 7	118 ± 7	−174 ± 7	−89 ± 8	−117 ± 7	123 ± 5
G6	148 ± 15[b]	−173 ± 8	163 ± 11[b]	127 ± 6	−176 ± 7	−89 ± 7	−121 ± 7	151 ± 6
	−82 ± 16[c]		61 ± 14[c]					
C7	−93 ± 14	−177 ± 8	62 ± 10	109 ± 5	171 ± 6	−85 ± 7	−138 ± 7	115 ± 8
A8	−64 ± 8	−178 ± 7	59 ± 5	125 ± 5	−177 ± 6	−80 ± 6	−121 ± 6	154 ± 9
T9	−65 ± 7	166 ± 5	56 ± 5	102 ± 5	−174 ± 5	−81 ± 6	−129 ± 6	110 ± 6
C10	−74 ± 10	168 ± 6	57 ± 9	88 ± 6			−138 ± 7	71 ± 7
G11				137 ± 7	180 ± 6	−94 ± 7	−105 ± 8	165 ± 7
A12	−81 ± 9	−167 ± 7	55 ± 6	126 ± 5	178 ± 6	−88 ± 6	−107 ± 7	155 ± 7
T13	−65 ± 7	168 ± 5	56 ± 6	106 ± 5	−179 ± 5	−91 ± 6	−131 ± 6	119 ± 5
G14	−75 ± 9	177 ± 6	62 ± 8	132 ± 6	−161 ± 7	−79 ± 6	−138 ± 7	153 ± 6
C15	−88 ± 11	177 ± 7	57 ± 9	116 ± 7	−174 ± 7	−104 ± 9	−107 ± 7	123 ± 7
A16	−79 ± 10	171 ± 7	64 ± 7	131 ± 6	−177 ± 6	−100 ± 8	−110 ± 7	152 ± 6
A17	−67 ± 9	176 ± 7	63 ± 8	129 ± 7	−179 ± 7	−101 ± 8	−113 ± 7	155 ± 7
A18	−79 ± 9	174 ± 7	70 ± 7	122 ± 5	−177 ± 6	−80 ± 6	−123 ± 7	145 ± 9
T19	−63 ± 8	163 ± 5	57 ± 7	103 ± 6	−174 ± 7	−82 ± 7	−136 ± 7	110 ± 7
G20	−74 ± 11	174 ± 7	58 ± 8	109 ± 7			−128 ± 8	117 ± 10
A-DNA[d]	−75	−152	45	83	178	−47	−154	13
B-DNA[e]	−46	−146	36	156	155	−96	−98	192

[a] Backbone torsion angles are defined as (IUPAC-IUB Joint Commision on Biochemical Nomenclature, 1983): $P(i)$–α–O5′–β–C5′–γ–C4′–δ–C3′–ϵ–O3′–ζ–$P(i+1)$ [b] Major conformer. [c] Minor conformer. [d] Canonical A-DNA (Arnott & Hukins, 1972). [e] Canonical B-DNA (Arnott & Hukins, 1973).

As seen from Table 3, the backbone torsion angles do not vary significantly with residue and are largely constrained to a preferred orientation, the only exception being α and γ torsion angles of residue G6. In accordance with theoretical and experimental data on oligodeoxyribonucleotide duplexes, torsion angles α favor the –synclinal (–sc, –gauche) range, β angles are locked in the antiperiplanar (ap, trans) domain, and γ torsion angles are confined with only a relatively narrow dispersion to the +synclinal (+sc, +gauche) region. The δ angles which are directly related to the endocyclic ν_3 torsions and thus to sugar puckering are distributed in the +anticlinal (+ac) range, i.e., between the values adopted by the canonical

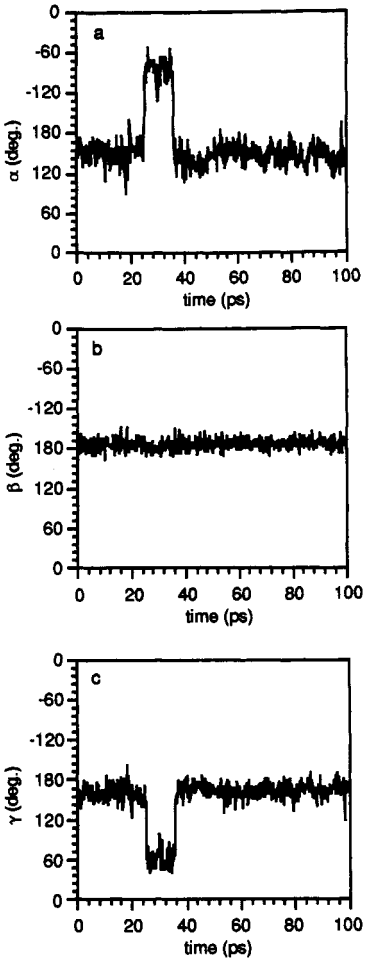

FIGURE 7: Time-dependent fluctuations of the (a) α, (b) β, and (c) γ backbone torsion angles for residue G6 during the final 100 ps of restrained molecular dynamics simulations.

forms of A- and B-DNA. At the 3'-side of the deoxyribose phosphate backbone, only the energetically favored and most common B_1 state is observed with ϵ angles in the ap (trans) region and ζ angles in the –anticlinal/–synclinal (–ac/–sc) boundary range characteristic of canonical B-DNA. As indicated by the larger standard deviation of torsion angles α, rotational mobility about the P–O5' ester linkage seems to be slightly less restricted compared to the other backbone angles. However, despite the limited number of direct experimental restraints, backbone torsion angles generally exhibit no excessively large fluctuations during the molecular dynamics run, implying that they are further constrained by the indirect restraints and the empirical force field.

For the torsion angles α and γ of residue G6, a bimodal distribution involving two distinct conformational states is observed. As seen in Figure 7, α and γ undergo a strongly coupled, reversible crankshaft-type motion during the rMD simulation from trans to the more common –gauche and +gauche conformation, respectively. These correlated jump processes keep the length of the sugar–phosphate backbone approximately the same and maintain regular base stacking in the duplex. Corresponding conformational transitions are frequently observed in free (Ravishanker et al., 1989) and time-averaged (Schmitz et al., 1993) molecular dynamics simulations highlighting the inherent flexibility of the DNA backbone. The backbone perturbations seem to extend to the 3'-linked residue C7 which exhibits a monomodal, yet relatively broad distribution in its corresponding torsion angles. Interestingly, a similar all-trans conformation was also observed between the central CG step in a crystal structure of a self-

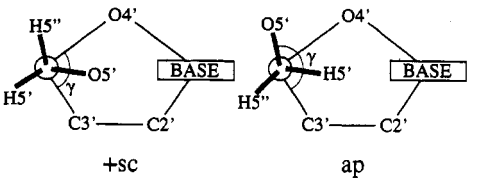

FIGURE 8: Orientations about the C4'–C5' bond of a nucleotide with torsion angles in the +synclinal (+sc) and antiperiplanar (ap) range.

complementary A-DNA octamer and its 3'-methylene phosphonate analog (Heinemann et al., 1987, 1991).

To examine the present backbone restraints in more detail, NMR-derived distances involving H5'/H5" protons were compared to the distances expected for the γ (+sc) and γ (ap) orientation found for residue G6. As shown in Figure 8, all substituents are in energetically favorable, staggered positions for both geometries, but interproton distances to H5'/H5" differ considerably. Thus, intranucleotide distances for a gauche–trans–gauche conformation are found to be d(H8/H6, H5') > 4.2 Å and d(H3', H5") ~ 3.6 Å. Rotation around C5'–C4' into a trans position moves the H5' and H5" protons much closer to base and H3' protons, respectively, decreasing those distances to d(H8/H6, H5') ~ 3.4 Å and d(H3', H5") < 3.0 Å. In contrast, experimental values range from 3.5 Å < d(H8/H6, H5') < 4.2 Å and 2.9 Å < d(H3', H5") < 3.4 Å. Apparently, all NMR-derived H5'/H5" distances systematically adopt values between those expected for the pure gauche–trans–gauche and all-trans conformers indicating conformational averaging in a rather flexible backbone. As seen in Figure 8 also, distances between base and H5" protons as well as H3' and H5' protons are mostly unaffected by this conformational transition; corresponding NOE-derived distances closely match those expected for the two staggered conformations. Since cross-peak intensities are weighted by distance (r^{-6} or r^{-3}, depending on frequency of interconverting conformers), experimental NOEs are biased toward the conformation with the smaller distances, and the refined structure could be locked in a minor, albeit significant, conformation, in this case the all-trans geometry.

Sequence-specific variations of ^{31}P chemical shifts also suggested different sugar–phosphate backbone geometries in B-DNA double helices (Gorenstein et al., 1988; Schroeder et al., 1989). According to theoretical and experimental studies, ^{31}P resonances in a ζ(t)α(g$^-$) or ζ(g$^-$)α(t) phosphodiester conformation are expected to appear at lower field while a ζ(g$^-$)α(g$^-$) conformer is associated with an upfield chemical shift (Gorenstein, 1984). Interestingly, pyrimidine (3'–5') purine steps are generally found to have a ^{31}P chemical shift more downfield than expected indicating increased flexibility with possible mixtures of gauche and trans phosphodiester conformations for these sequences (Ott & Eckstein, 1985a,b; Gorenstein et al., 1988).

Sugar Pucker and Glycosidic Torsion Angles. Values for the pseudorotational phase angle P, and thus the sugar puckering modes, differ significantly for the nonterminal purine and pyrimidine nucleotides. Purine residues adopt a C2'-endo deoxyribose conformation with phase angles of pseudorotation varying between 145° and 162°. In contrast, pyrimidine nucleotides prefer smaller phase angles in the range between 110° and 123° characteristic of a C1'-exo conformer.

The glycosidic torsion angle χ adopts a wide range of values in the anti domain from –102° to –144°. It is well known that χ is correlated with sugar pucker (Fratini et al., 1982; Ulyanov et al., 1992). Upon a decrease in the pseudorotational phase angle, χ tends to shift from –ac toward an ap conformation because of steric interactions between base and sugar.

Table 4: Average Helical Parameters with Standard Deviations for a 100-ps Restrained Molecular Dynamics Trajectory of rMD-final[a]

base pair	shear (Å)	stretch (Å)	stagger (Å)	buckle (deg)	propeller twist (deg)	opening (deg)
C1·G20	0.2 ± 0.4	−0.1 ± 0.2	−1.0 ± 0.5	19 ± 9	−19 ± 11	−6 ± 4
A2·T19	−0.2 ± 0.3	−0.4 ± 0.2	0.3 ± 0.4	12 ± 6	−31 ● 9	−4 ± 4
T3·A18	0.4 ± 0.3	−0.3 ± 0.1	0.4 ± 0.4	−7 ● 7	−36 ± 8	1 ± 4
T4·A17	0.3 ± 0.3	−0.2 ± 0.2	0.1 ± 0.3	−14 ± 6	−18 ± 8	−4 ± 4
T5·A16	0.0 ± 0.3	−0.3 ± 0.1	−0.3 ± 0.4	−8 ± 6	−27 ± 8	−9 ± 4
G6·C15	−0.6 ± 0.3	−0.4 ● 0.2	−0.2 ● 0.4	−6 ± 6	−5 ± 8	−12 ± 4
C7·G14	0.8 ± 0.3	−0.4 ± 0.2	−0.2 ± 0.4	0 ± 7	3 ± 8	−6 ± 3
A8·T13	0.3 ± 0.3	−0.4 ± 0.2	0.1 ± 0.3	−5 ± 6	−17 ± 9	−5 ± 3
T9·A12	0.2 ± 0.3	−0.4 ± 0.2	0.2 ● 0.4	−11 ± 7	−20 ± 8	−7 ± 3
C10·G11	0.4 ± 0.3	−0.3 ● 0.2	0.0 ± 0.5	−18 ± 8	−17 ● 9	−7 ± 4
A-DNA[b]	0.0	−0.5	0.2	0	14	−5
B-DNA[c]	0.0	0.0	0.1	0	4	−4

base pair step	shift (Å)	slide (Å)	rise (Å)	tilt (deg)	roll (deg)	twist (deg)
C1A2·T19G20	−0.3 ± 0.2	−0.1 ± 0.2	3.7 ± 0.3	0 ± 4	5 ± 5	40 ± 3
A2T3·A18T19	0.1 ± 0.2	−0.3 ± 0.2	3.5 ± 0.2	0 ± 4	−3 ± 5	36 ± 2
T3T4·A17A18	−0.1 ± 0.2	0.1 ± 0.2	3.1 ± 0.2	4 ± 4	−2 ± 6	41 ± 2
T4T5·A16A17	−0.3 ± 0.2	−0.3 ± 0.2	3.0 ± 0.2	3 ± 4	−9 ± 5	35 ± 3
T5G6·C15A16	0.1 ± 0.3	0.1 ± 0.2	2.7 ± 0.3	−2 ± 4	20 ± 6	36 ± 3
G6C7·G14C15	−0.2 ± 0.2	−0.3 ● 0.2	2.9 ± 0.3	−7 ± 4	8 ± 5	39 ± 2
C7A8·T13G14	−0.1 ± 0.2	−0.1 ± 0.2	3.2 ± 0.2	6 ± 4	−4 ± 5	37 ± 2
A8T9·A12T13	−0.3 ± 0.2	−0.5 ± 0.1	3.1 ● 0.2	3 ± 4	−1 ± 5	30 ± 2
T9C10·G11A12	0.1 ± 0.2	0.0 ± 0.1	3.0 ± 0.3	4 ± 3	6 ± 4	39 ± 2
A-DNA[b]	0.0	0.0	2.6	0	0	33
B-DNA[c]	0.0	0.0	3.4	0	0	36

[a] Individual parameters follow the conventions developed at the Cambridge meeting on DNA curvature (Dickerson et al., 1989) and are defined with respect to a unique global helix axis. [b] Canonical A-DNA (Arnott & Hukins, 1972). [c] Canonical B-DNA (Arnott & Hukins, 1973).

Although less obvious than for the pseudorotational angles, systematic differences in χ between pyrimidines and purines are indeed observed for the present DNA duplex. While glycosidic torsion angles of pyrimidine nucleotides are distributed around an average of −129°, χ values of purine nucleotides are clustered around −117°, more reminiscent of standard B-DNA.

It should be mentioned that the helical arrangement of bases as well as the backbone and glycosidic torsion angles are not markedly affected upon removing torsion angle restraints in an additional 100-ps rMD simulation. As expected, average δ and pseudorotational phase angles which are directly defined by the torsion angle restraints are more noticeably altered, yet in most cases do not drift beyond the fluctuations observed in a fully restrained rMD simulation. Larger deviations are found for the terminal nucleotides, again indicating that in case of residues with high populations of minor conformers average torsion angles fail to be adequate structural descriptors and, consequently, results must be interpreted with care.

Helicoidal Parameters. Intra-base-pair and inter-base-pair parameters are listed in Table 4 together with their standard deviations for the final 100-ps restrained molecular dynamics trajectory. The translational intra-base-pair parameters shear, stretch, and stagger are generally small, although for the latter, end effects with a value as high as −1.0 Å at C1·G20 are evident. The neighboring base pairs G6·C15 and C7·G14 exhibit a rather large shear in opposite directions of −0.6 and 0.8 Å, respectively. With G6 shifted toward the minor groove and C7 toward the major groove within the base pair, shearing maximizes base-stacking interactions for this dinucleotide dimer. Note that the stretch, which exhibits negative values throughout the decamer, is subjected to smaller amplitude fluctuations during the rMD run than the shear and stagger parameters.

End effects are also manifested in the large buckling for the C1·G20 and C10·G11 termini. These base pairs have positive and negative buckles, respectively, in both cases bending the center of the base pairs away from the center of the double helix. Overall, average buckle angles display a sinusoidal modulation along the duplex with decreased amplitudes toward the inner part of the molecule.

Considerable negative propeller twist with values as high as −36° for T3·A18 is observed for the AT base pairs in the decamer. These large negative propeller twist angles are directly reflected in the 2D NOE cross-peak intensities between imino and adenine H2 protons in H_2O (Weisz et al., 1992). For the propeller-twisted AT base pairs in the T_3·A_3 tract of the duplex, shorter interstrand distances (~3.4 Å) are expected from the thymine imino proton to H2 of adenine base-paired to its 5′-linked neighbor and larger distances (~4.8 Å) to H2 of adenine base-paired to its 3′-linked neighbor, resulting in the medium and low intensity cross-peaks observed in the 2D NOE spectra, respectively. In contrast, both central GC base pairs are only slightly propeller-twisted and essentially flat given their small buckle. Except for T3·A18, base pair opening values are negative, implying opening toward the minor groove, the most prominent base pair being G6·C15 with an opening angle of −12°. Smaller opening angles with a concomitant narrowing of the minor groove are observed for the ATT·AAT tract. The narrow minor groove generally associated with A_n·T_n structures (Fratini et al., 1982; Lipanov & Chuprina, 1987) is also reflected in the interstrand distances between adenine H2 and the H1′ proton of the nucleotide base-paired to its 5′-linked neighbor (Kintanar et al., 1987; Ulyanov et al., 1993). Corresponding cross-peaks in the 2D NOE spectra for the run of three consecutive adenines decrease in intensity from the nucleotide pair A18–T4 to A17–T5 and vanish for A16–G6 (Weisz et al., 1992). Accordingly, H2–H1′ distances of the refined structure increase, implying a gradual widening of the minor groove in the 5′-direction along the A_3-tract as reported previously (Nadeau & Crothers, 1989; Ulyanov et al., 1993).

The inter-base-pair parameters shift and slide describe linear x- and y-displacements between two base pairs. Values for the decamer duplex tend to be slightly negative for both parameters, i.e., looking down the helix axis from the C1·G20 terminus, successive base pairs are displaced toward the minor groove and the 3′–5′ strand. A more negative slide observed for all purine–pyrimidine compared to pyrimidine–purine steps in the decamer helps to relieve interstrand purine–purine clashes in the major and minor groove (Calladine, 1982). The rise value is generally in the range of 3.0–3.5 Å and thus close to a standard B-form structure. However, a reduced rise of less than 3.0 Å is observed at the two central G·C base pairs where DNA bending occurs (*vide infra*). Again, end effects with a maximum rise of 3.7 Å are evident for the first base pair step. Tilt and roll angles for the central trimer TGC·GCA deviate significantly from the rest of the duplex. Whereas for the nonterminal flanking base pair steps moderately high positive tilt and negative roll values are observed, tilt and roll simultaneously change sign for the TGC central sequence, indicating local perturbations of the structure. In particular, large positive roll angles between steps T5G6 and, to a lesser extent, G6C7 result in an opening toward the major groove and a concomitant local expansion of the minor groove. For pyrimidine–purine steps, a negative propeller twist leads to unfavorable contacts in the minor groove between purines of the complementary strands. Relieving this steric clash, e.g., by introducing a large roll as observed for the base pair dimer T5G6·C15A16, increases the interstrand distance between

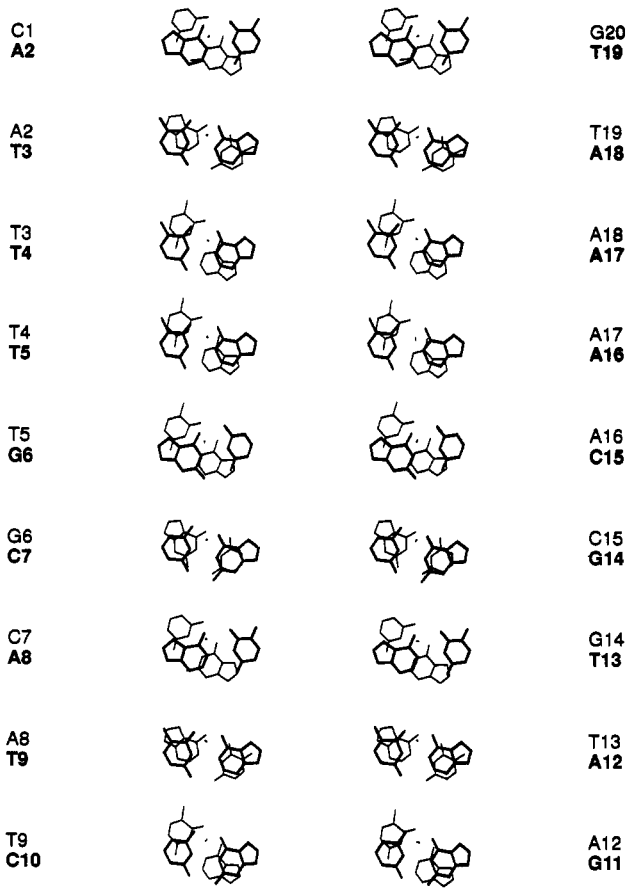

FIGURE 9: Stereoviews of the nine individual base pair steps viewed down the global helix axis. Bases in strands 1 and 2 are numbered along the left and right, respectively. Points indicate where the global helix axis pierces the base pair planes.

A16 H2 and G6 H1′ protons to about 5 Å, compatible with the disappearance of a corresponding cross-peak in the experimental 2D NOE spectra.

Except for the underwound A8T9·A12T13 step, values for helical twist range from 35° to 41° and clearly point to a B-like structure. There is good agreement to calculations of twist angles in B-DNA which are based on various crystallographic and solution data and predict a decreased helical twist for AT steps and a larger than average twist angle for GC and GA steps (Kabsch et al., 1982). Moreover, inspection of several NMR-refined structures suggested a highly conserved central twist within tetranucleotide sequences (Mauffret et al., 1992). Indeed, GC twist angles of 39° (Baleja et al., 1990a) and 37° (Mauffret et al., 1992) found for a TGCA tract agree with a value of 39° in the present study.

Base Stacking and DNA Curvature. The observed stacking pattern between bases is typical for B-type double helices with decreasing intrastrand base overlap in the order purine (3′–5′) pyrimidine, purine (3′–5′) purine, and pyrimidine (3′–5′) purine (Figure 9). The central G6C7·G14C15 base pair dimer exhibits a particularly extensive overlap of the six-membered purine ring with the pyrimidine ring of the 3′-linked residue. This increased stacking interaction is primarily achieved by shear dislocations in opposite directions within the two base pairs and is further maximized by small buckle and propeller twist angles which manifest essentially flat base pair planes (*vide supra*). In contrast, pyrimidine (3′–5′) purine steps show no significant intrastrand base overlap, and, for the C7A8·T13G14 dimer, even partial interstrand overlap between the two purine rings, reminiscent of A-DNA, is observed.

The low stacking energy of pyrimidine (3′–5′) purine steps favors the large roll angle toward the major groove for the T5G6·C15A16 dimer. Large positive roll angles and a concomitant bending of the helix, particularly at TG·CA steps, have been observed frequently and suggest a pronounced lability at this base pair step (Baleja et al., 1990b; Nilges et al., 1987a,b; Mujeeb et al., 1993). The central palindromic sequence TGCA displays little helical symmetry, however, and the C7A8·T13G14 dinucleotide does not exhibit this feature. In accord with energy minimization predictions (Ulyanov & Zhurkin, 1984), AT and TT/AA sequences exhibit a small negative roll toward the minor groove. The roll and tilt parameters reported in Table IV also indicate greater fluctuations in the roll angle, supporting a DNA bending anisotropy with a preference for bending toward a groove (Ulyanov & Zhurkin, 1984, Zhurkin et al., 1991). It is well documented that A_n·T_n tracts, especially for $n \geq 4$, are bent in free DNA (Diekmann, 1986; Koo et al., 1986). The origin of this bend was attributed either to "AA wedge" angles at each dinucleotide step (Trifonov, 1985) or to junctions between structurally dissimilar helices (Levene & Crothers, 1983). Apparently, the AT_3·A_3T tract of the decamer has a small intrinsic curvature due to successive negative roll values at the AT and TT/AA junctions. However, overall bending of the duplex is clearly dominated by the large positive roll and concomitant axis tip toward the major groove for the central T5G6·C15A16 and to a lesser extent G6C7·G14C15 dimer. Because axis tip angles at these two neighboring base pair steps are almost in phase and not completely counterbalanced by out-of-phase bends in the flanking sequences, the global helix axis is noticeably bent in its center (see Figure 6). The static overall curvature of the decamer amounts to 17° when measured by the angle between the two outer helical axis segments.

The $\alpha(g^-)\beta(t)\gamma(g^+)$ to $\alpha(t)\beta(t)\gamma(t)$ transition between nucleotides T5 and G6 does not significantly affect the helical arrangement of the bases. Noticeable changes are observed for twist and roll angles. The $\alpha(t)\beta(t)\gamma(t)$ backbone causes the structure to unwind by about 5° while increasing its roll angle by 7° at the T5G6·C15A16 step. However, this larger roll toward the major groove is compensated by changes in roll for the two neighboring base pair steps with no net change in global axis bending due to the all-trans conformation. Rather, the backbone conformational transition merely redistributes the global axis curvature via a more localized kink as opposed to a smoother bend in the standard $\alpha(g^-)\beta(t)\gamma(g^+)$ conformer. This kink differs markedly from the geometries which have been proposed for kinked helices. Model building studies led to the conclusion that rotating only torsion angle γ from the usually preferred +sc range into ap imposes a kink toward the minor groove of the DNA double helix (Crick & Klug, 1975). In a subsequent study which was based on crystallographic data for DNA–drug complexes, kinking toward the major groove was proposed to be the result of a change in the normal C2′-endo deoxyribose sugar puckering to a mixed sugar puckering pattern of the type C3′-endo (3′–5′) C2′-endo (Sobell et al., 1977). In both cases, the kink was also associated with partial unwinding and unstacking of adjacent base pairs on either side of the hinge.

Interestingly, the conformational behavior of the T5G6·C15A16 step is accompanied by an unusual upfield shift of the C15 H1′ proton resonating at 5.43 ppm at 25 °C (Weisz et al., 1992). A similar upfield-shifted cytosine H1′ resonance of a T_nG·CA_n junction exhibits a pronounced temperature dependence (Ulyanov et al., 1993). Moreover, a growing number of studies suggest dynamically and structurally unique

TG·CA dimers (Timsit et al., 1991; Donlan & Lu, 1992). NMR studies have demonstrated that thymine imino protons in GTG sequences exhibit exceptionally fast hydrogen exchange rates (Cheung et al., 1984). Selection of DNA sequences with anomalously low gel-electrophoretic mobility revealed the importance of the dinucleotide base step TG·CA in DNA bending (Beutel & Gold, 1992). And finally, gel retardation anomalies of curved DNA fragments were interpreted in terms of reversible kinks that occur specifically at TG·CA stacks and relieve the effects of curvature (McNamara et al., 1990). Clearly, the backbone conformational transition outlined above constitutes a possible mechanism for this kink formation, thus emphasizing the impact of backbone flexibility on local DNA structure.

CONCLUSION

The present study shows that homonuclear 2D NOE and coupling constant data combined with molecular dynamics simulations can provide well-defined solution structures of a DNA fragment. The unusually large number of accurately determined NMR data enables reliable determination of sequence-dependent local structural variations in the decamer duplex. In a standard rMD refinement, however, limitations arise from the nature of the experimental restraints. In the presence of internal motions, NMR data strictly reflect time averages that cannot be matched simultaneously by a rigid structure and result in discrepancies of some interproton distances. Consequently, adding more restraints to better define the structure leads to increased restraint violations and *R* factors for the final geometry, thus limiting their usefulness in evaluating the refinement process. In principle, conformational flexibility could be incorporated into the interpretation of NOE data via sophisticated models (Keepers & James, 1982, 1984) or even simplified models (Lipari & Szabo, 1982); however, we currently suffer from a paucity of experimental information describing internal motions in DNA.

Generally, local motions in DNA duplexes are restricted and affect NOE-derived distances only to a minor extent. However, in addition to sugar repuckering the present analysis suggests dynamic equilibria within the sugar–phosphate backbone involving an extended $\alpha(t)\beta(t)\gamma(t)$ backbone conformation. This is in accord with previous experimental relaxation as well as computational studies on DNA duplexes (James, 1984; Keepers & James, 1982; Keepers et al., 1982). Considering the apparent flexibility of the phosphate backbone, the often-used approach of constraining backbone torsion angles to standard A- and B-form rotamer conformations in structural refinement may be inappropriate. On the other hand, the occurrence of an all-trans conformer, strongly suggested by the available NMR data at the pyrimidine–purine step T5G6, may simply result from the distance-weighted averaging of NOE cross-peak intensities in case of multiple conformations. So NOE data can be biased toward the $\alpha(t)\beta(t)\gamma(t)$ conformer which exhibits smaller NOE-observable interproton distances. In fact, an all-trans geometry should have a higher energy and merely constitute a minor conformer. Heteronuclear ^{1}H–^{31}P correlated spectra could be helpful in supplementing the information available for the backbone torsion angles, but less populated conformers may remain unnoticed in averaging of scalar couplings in this case.

The most striking feature of the present decamer duplex is the large roll angle observed between the central base pair dimer T5G6·G14C15, which is accompanied by a significant bend of the global axis into the major groove. This pyrimidine–purine dimer exhibits significant backbone flexibility supporting previous reports of unusual behavior for this labile dinucleotide. TG steps have also been described as major factors in DNA bending and kink formation, which are believed to be crucial structural elements of various *in vivo* processes such as regulation of gene expression. In fact, TG/CA dinucleotides appear to be strongly overrepresented in consensus protein-binding sites (Trifonov & Brendel, 1986). Interestingly, a similar palindromic sequence CATG is part of the heptamer element in the immunoglobulin promoter region which is also recognized by the octamer transcription factors (Kemler et al., 1989). Possible binding modes of transcription factors to regulatory DNA sequences are revealed by the crystal structure of a CAP–DNA complex which exhibits two 40° kinks between TG·CA base pairs on each side of the dyad axis (Schultz et al., 1991). It can be hypothesized that the central TG step in the octamer sequence provides a flexible hinge on protein binding which enables rearrangements to better accommodate proteins in a tight complex. Localized kinks formed by conformational changes may help to maximize DNA–protein interactions and are thus likely to be stabilized in the complex. Indeed, electrophoretic analyses and ligase-mediated cyclization experiments suggested the induction of a 37° bend toward the major groove in the octamer upon binding of several POU proteins (Verrijzer et al., 1991). However, in contrast to the present results, no intrinsic curvature was found for the free DNA. This might be attributable to the relatively small bending angle which makes anomalous electrophoretic mobility difficult to detect without any further amplification of the inherent bend. It should also be pointed out that interaction of octamer factors with DNA probably involves more complex recognition patterns. This is indicated by sequence requirements for binding of the Oct-1 POU domains (Verrijzer et al., 1992). While the canonical octamer sequence represents the optimal binding site for Oct-1, an A to T substitution at corresponding position 16 of the present decamer had a moderately negative effect on binding affinity which dropped to about 70%. Also, in addition to the core octamer motif, flanking sequences were found to be essential for high-affinity binding of the Oct-1 POU domain. Thus, DNA recognition and binding by POU domain proteins are likely to rely on several structural motifs with the centrally located TG·CA step being one of the potential factors governing these processes.

ACKNOWLEDGMENT

We thank Drs. U. Schmitz, N. Ulyanov, and A. Mujeeb for many helpful discussions. This research was supported in part by the Pittsburgh Supercomputing Center Grant No. 1 P41 RR06009 from the NIH National Center for Research Resources.

REFERENCES

Arnott, S., & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun. 47*, 1504.
Arnott, S., & Hukins, D. W. L. (1973) *J. Mol. Biol. 81*, 93.
Baleja, J. D., Germann, M. W., van de Sande, J. H., & Sykes, B. D. (1990a) *J. Mol. Biol. 215*, 411.
Baleja, J. D., Pon, R. T., & Sykes, B. D. (1990b) *Biochemistry 29*, 4828.
Beutel, B. A., & Gold, L. (1992) *J. Mol. Biol. 228*, 803.
Boelens, R., Koning, T. M. G., van der Marel, G. A., van Boom, J. H., & Kaptein, R. (1989) *J. Magn. Reson. 82*, 290.
Borgias, B. A., & James, T. L. (1988) *J. Magn. Reson. 79*, 493.
Borgias, B. A., & James, T. L. (1990) *J. Magn. Reson. 87*, 475.
Brünger, A. T. (1992) *Nature 355*, 472.
Calladine, C. R. (1982) *J. Mol. Biol. 161*, 343.
Cheung, S., Arndt, K., & Lu, P. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3665.

Crick, F. H. C., & Klug, A. (1975) *Nature 255*, 530.
Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., Lavery, R., Nelson, H. C. M., Olson, W. K., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., von Kitzing, E., Wang, A. H.-J., & Zhurkin, V. B. (1989) *J. Biomol. Struct. Dyn. 6*, 627.
Diekmann, S. (1986) *FEBS Lett. 195*, 53.
Donlan, M. E., & Lu, P. (1992) *Nucleic Acids Res. 20*, 525.
Fratini, A. V., Kopka, M. L., Drew, H. R., & Dickerson, R. E. (1982) *J. Biol. Chem. 257*, 14686.
Gallo, K., Huang, C., Ferrin, T. E., & Langridge, R. (1989) *Molecular Interactive Display and Simulation (MIDASplus)*, University of California, San Francisco.
Garcia-Blanco, M., Clerc, R. G., & Sharp, P. A. (1989) *Genes Dev. 3*, 739.
Gorenstein, D. G. (1984) *Phosphorus-31 NMR. Principles and Applications* (Gorenstein, D. G., Ed.) Academic, New York.
Gorenstein, D. G., Schroeder, S. A., Fu, J. M., Metz, J. T., Roongta, V., & Jones, C. R. (1988) *Biochemistry 27*, 7223.
Griesinger, C., Sørensen, O. W., & Ernst, R. R. (1985) *J. Am. Chem. Soc. 107*, 6394.
Gronenborn, A. M., & Clore, G. M. (1989) *Biochemistry 28*, 5978.
Gupta, G., Sarma, M. H., & Sarma, R. H. (1988) *Biochemistry 27*, 7909.
Heinemann, U., Lauble, H., Frank, R., & Blöcker, H. (1987) *Nucleic Acids Res. 15*, 9531.
Heinemann, U., Rudolph, L.-N., Alings, C., Morr, M., Heikens, W., Frank, R., & Blöcker, H. (1991) *Nucleic Acids Res. 19*, 427.
Hore, P. J. (1983) *J. Magn. Reson. 55*, 283.
IUPAC-IUB Joint Commission on Biochemical Nomenclature (1983) *Eur. J. Biochem. 131*, 9.
James, T. L. (1984) in *Phosphorus-31 NMR. Principles and Applications* (Gorenstein, D. G., Ed.) pp 349–400, Academic, New York.
Kabsch, W., Sander, C., & Trifonov, E. N. (1982) *Nucleic Acids Res. 10*, 1097.
Kaluarachchi, K., Meadows, R. P., & Gorenstein, D. G. (1991) *Biochemistry 30*, 8785.
Keepers, J. W., & James, T. L. (1982) *J. Am. Chem. Soc. 194*, 929.
Keepers, J. W., & James, T. L. (1984) *J. Magn. Reson. 57*, 404.
Keepers, J. W., Kollman, P. A., Weiner, P. K., & James, T. L. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 5537.
Kemler, I., Schreiber, E., Müller, M. M., Matthias, P., & Schaffner, W. (1989) *EMBO J. 8*, 2001.
Kintanar, A., Klevit, R. E., & Reid, B. R. (1987) *Nucleic Acids Res. 15*, 5845.
Koning, T. M. G., Boelens, R., van der Marel, G. A., van Boom, J. H., & Kaptein, R. (1991) *Biochemistry 30*, 3797.
Koo, H.-S., Wu, H.-M., & Crothers, D. M. (1986) *Nature 320*, 501.
Lane, A. N. (1990a) *Biochim. Biophys. Acta 1049*, 189.
Lane, A. N. (1990b) *Biochim. Biophys. Acta 1049*, 205.
Lavery, R., & Sklenar, H. (1988) *J. Biomol. Struct. Dyn. 6*, 63.
Lavery, R., & Sklenar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 655.
Levene, S. D., & Crothers, D. M. (1983) *J. Biomol. Struct. Dyn. 1*, 429.
Lilley, D. M. J. (1991) *Nature 354*, 359.
Lipanov, A. A., & Chuprina, V. P. (1987) *Nucleic Acids Res. 15*, 5833.
Lipari, G., & Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546.
Liu, H., Thomas, P. D., & James, T. L. (1992) *J. Magn. Reson. 98*, 163.
Liu, H., Kumar, A., Weisz, K., Schmitz, U., Bishop, K. D., & James, T. L. (1993) *J. Am. Chem. Soc. 115*, 1590.
Mauffret, O., Hartmann, B., Convert, O., Lavery, R., & Fermandjian, S. (1992) *J. Mol. Biol. 227*, 852.
McNamara, P. T., Bolshoy, A., Trifonov, E. N., & Harrington, R. E. (1990) *J. Biomol. Struct. Dyn. 8*, 529.
Metzler, W. J., Wang, C., Kitchen, D. B., Levy, R. M., & Pardi, A. (1990) *J. Mol. Biol. 214*, 711.
Müller, L. (1987) *J. Magn. Reson. 72*, 191.
Mujeeb, A., Kerwin, S. M., Kenyon, G. L., & James, T. L. (1993) *Biochemistry* (in press).
Nadeau, J. G., & Crothers, D. M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 2622.
Nilges, M., Clore, G. M., Gronenborn, A. M., Brünger, A. T., Karplus, M., & Nilsson, L. (1987a) *Biochemistry 26*, 3718.
Nilges, M., Clore, G. M., Gronenborn, A. M., Piel, N., & McLaughlin, L. W. (1987b) *Biochemistry 26*, 3734.
Nilsson, L., Clore, G. M., Gronenborn, A. M., Brünger, A. T., & Karplus, M. (1986) *J. Mol. Biol. 188*, 455.
Ott, J., & Eckstein, F. (1985a) *Biochemistry 24*, 2530.
Ott, J., & Eckstein, F. (1985b) *Nucleic Acids Res. 13*, 6317.
Pearlman, D. A., Case, D. A., Caldwell, J. C., Seibel, G. L., Singh, U. C., Weiner, P., & Kollman, P. A. (1991) AMBER 4.0, University of California, San Francisco.
Ravishanker, G., Swaminathan, S., Beveridge, D. L., Lavery, R., & Sklenar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 669.
Rinkel, L. J., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621.
Ryckaert, J. P., Ciccotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
Schmitz, U., Pearlman, D. A., & James, T. L. (1991) *J. Mol. Biol. 221*, 271.
Schmitz, U., Sethson, I., Egan, W. M., & James, T. L. (1992) *J. Mol. Biol. 227*, 510.
Schmitz, U., Ulyanov, N. B., Kumar, A., & James, T. L. (1993) *J. Mol. Biol.* (in press).
Schroeder, S. A., Roongta, V., Fu, J. M., Jones, C. R., & Gorenstein, D. G. (1989) *Biochemistry 28*, 8292.
Schultz, S. C., Shields, G. C., & Steitz, T. A. (1991) *Science 253*, 1001.
Sobell, H. M., Tsai, C.-C., Jain, S. C., & Gilbert, S. G. (1977) *J. Mol. Biol. 114*, 333.
Stolarski, R., Egan, W., & James, T. L. (1992) *Biochemistry 31*, 7027.
Sturm, R. A., & Herr, W. (1988) *Nature 336*, 601.
Suzuki, E., Pattabiraman, N., Zon, G., & James, T. L. (1986) *Biochemistry 25*, 6854.
Tanaka, K., Muramatsu, S., Yamada, H., & Mizuno, T. (1991) *Mol. Gen. Genet. 226*, 367.
Thomas, P. D., Basus, V. J., & James, T. L. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1237.
Timsit, Y., Vilbois, E., & Moras, D. (1991) *Nature 354*, 167.
Trifonov, E. N. (1985) *CRC Crit. Rev. Biochem. 19*, 89.
Tifonov, E. N., & Brendel, V. (1986) *GNOMIC—A Dictionary of Genetic Codes*, VCH Verlagsgesellschaft, Weinheim, Germany.
Ulyanov, N. B., & Zhurkin, V. B. (1984) *J. Biomol. Struct. Dyn. 2*, 361.
Ulyanov, N. B., Gorin, A. A., Zhurkin, V. B., Chen, B.-C., Sarma, M. H., & Sarma, R. H. (1992) *Biochemistry 31*, 3918.
Ulyanov, N. B., Sarma, M. H., Zhurkin, V. B., & Sarma, R. H. (1993) *Biochemistry 32*, 6875.
Verrijzer, C. P., van Oosterhout, J. A. W. M., van Weperen, W. W., & van der Vliet, P. C. (1991) *EMBO J. 10*, 3007.
Verrijzer, C. P., Alkema, M. J., van Weperen, W. W., Van Leeuwen, H. C., Strating, M. J. J., & van der Vliet, P. C. (1992) *EMBO J. 11*, 4993.
Weiner, S. J., Kollman, P. A., Nguyen, D. T., & Case, D. A. (1986) *J. Comput. Chem. 7*, 230.
Weisz, K., Shafer, R. H., Egan, W., & James, T. L. (1992) *Biochemistry 31*, 7477.
Widmer, H., & Wüthrich, K. (1986) *J. Magn. Reson. 70*, 270.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Zhou, N., Bianucci, A. M., Pattabiraman, N., & James, T. L. (1987) *Biochemistry 26*, 7905.
Zhurkin, V. B., Ulyanov, N. B., Gorin, A. A., & Jernigan, R. L. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 7046.