Insights into the Dynamic Nature of DNA Duplex Structure via Analysis of Nuclear Overhauser Effect Intensities[†]

Marco Tonelli and Thomas L. James*

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143-0446

Received April 21, 1998; Revised Manuscript Received June 4, 1998

ABSTRACT: Sequence-dependent structures of DNA duplexes in solution can be reliably determined using NMR data if care is taken to determine restraint bounds accurately. This entails use of complete relaxation matrix methods to analyze nuclear Overhauser effect (NOE) spectroscopic cross-peak intensities, yielding accurate distance restraints. NMR studies of various DNA duplexes have suggested that there may be some limited internal motions. First, it is typically not possible to reconcile all vicinal proton coupling constants in deoxyribose rings with a single conformer. In addition, with the increased accuracy of interproton distance measurements afforded by the complete relaxation matrix algorithm MARDIGRAS, we find inconsistencies in certain distances which can most readily be ascribed to limited conformational flexibility, since conformational averaging is nonlinear. As base−sugar interproton distances depend on both sugar pucker and glycosidic torsion angle χ, motion involving these structural variables should be reflected by experimental data. Possible motional models have been considered to account for all of the data for three DNA duplexes. Analysis of intraresidue base−sugar interproton NOE bounds patterns suggests a motional model with individual sugars in equilibrium between S (2′-*endo*) and N (3′-*endo*) conformations, with S being preferred. As sugar repuckering is correlated with changes in glycosidic torsion angle χ, different sugar conformers imply different values for χ, but this is insufficient to account for all data. A two-state jump between anti and syn glycosidic conformers was considered, but it was incapable of accounting for all data. However, a model with restricted diffusion (rocking) about the glycosidic bond in addition to sugar repuckering was capable of accommodating all experimental data. This motional model is in qualitative agreement with experimental ^{13}C relaxation-derived order parameter values in a DNA duplex.

The nuclear Overhauser effect (NOE)[1] has been widely used to investigate the three-dimensional (3D) structure of biologically relevant macromolecules in solution. In a typical two-dimensional NOE (2D NOE) experiment, the dipole moments of two hydrogen atoms within a distance of ca. 5 Å will couple, yielding an off-diagonal peak in the spectrum. To a first approximation, the intensity of this cross-peak is inversely proportional to the sixth power of the distance between the two atoms involved. NOE intensities are often used to calculate distances between hundreds to thousands of pairs of hydrogen atoms; these distances are then used to restrain the structure during structural refinement procedures.

NOE intensities may also be affected by conformational fluctuations that occur during the time of the NMR experiment. In the case of exchange faster than the largest chemical shift difference of a particular proton in the various exchanging conformers (typically >1000 Hz), only one set of NOE peaks is observed in the spectrum. The intensity of these peaks is a weighted average of the intensities given by the atoms in each of the conformers. Complicating the situation, the manner in which NOE intensities are averaged depends (a) on the relative population of conformers, (b) on the rate of exchange between conformers, which can be alternatively viewed as internal motion, and (c) on the angular fluctuations if the exchange is extremely fast.

For purposes of structure determination, a single rigid conformation is assumed, so a single distance corresponding to each NOE intensity is naturally derived. The averaging in any case is obviously not a simple geometric average of distances in the different conformers. When the internal motion is faster than the overall correlation time for molecular tumbling, the measured distance is subject to $\langle r^{-3} \rangle$ averaging and depends as well on angular dispersion; in this case, the apparent distance between a pair of nuclei is generally slightly weighted toward the shortest of the distances in the various conformers if the internal motion (conformational exchange) is not isotropic (*1*), but the apparent distance is identically equal to the geometric mean distance, weighted by conformer populations, if by chance the conformational exchange entails isotropic jumps (*2*). However, if jumps between the individual conformers are slow compared to the reciprocal of the overall correlation

[†] This work was supported by Grant GM39247 from the National Institutes of Health.

* To whom correspondence should be addressed. Phone: (415) 476-1916. FAX: (415) 502-4690. E-mail: james@picasso.ucsf.edu.

[1] Abbreviations: 1D NMR, one-dimensional NMR; 2D NMR, two-dimensional NMR; 2D NOE, two-dimensional nuclear Overhauser effect; 2QF-COSY, double-quantum-filtered correlation spectroscopy; 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; S^2, order parameter.

S0006-2960(98)00905-2 CCC: $15.00
Published on Web 07/30/1998

time but fast compared to the relaxation time, the apparent distance is subject to $\langle r^{-6} \rangle$ averaging over individual proton positions, so the apparent distance may be strongly distorted from the geometric average position.

It will be recognized that a Karplus relationship between vicinal scalar coupling constants and torsion angles is also not linear, so torsion angles are not geometrically averaged by conformational fluctuations. Even with perfect data and with careful analysis, with such nongeometric averaging, the individual distance and torsion angle restraints will be internally inconsistent. However, such inconsistencies encode information about the ongoing conformational equilibrium and, in principle, could be used to study it.

To infer anything from sets of inconsistent distances derived from NOE intensities, it is essential that they be determined with high accuracy. Our distances are calculated by use of the program MARDIGRAS (*3*−*5*), which is based on complete relaxation matrix analysis. Consequently, MARDIGRAS is able to account for network relaxation and multispin effects and, hence, to calculate proton−proton distances with greater accuracy than the commonly employed two-spin or isolated spin-pair approximation.

Previous NMR studies conducted on DNA duplexes show that certain classes of distances are inconsistent with others (*6*). Indeed, some are inconsistent with the duplexes possessing structure in either the B or A family of structures, i.e., those which one would imagine are energetically most reasonable. We have also observed that a single conformer will rarely account for all coupling constants measured via detailed simulation and fitting, but reasonable fits are obtained with a two-state model representing a rapid interconversion between S- and N-type sugar puckers (*7*, *8*).

The current investigation examines possible causes for these inconsistencies, focusing on the likelihood that internal motions are responsible. In our studies, we will consider inconsistencies involving intraresidue base-to-sugar distances. Consequently, we will be concerned with flexibility involving (a) sugar pucker, which is described by the phase angle P and the amplitude Φ of pseudorotation, and (b) rotation about the glycosyl bond C1′−N (described by the torsion angle χ) within the same residue (Figure 1). This entails a relatively simple subset of conformational space, whose members can be almost fully described by two variables, P and χ angles, with the sugar pucker amplitude Φ being allowed to vary only within a narrow range. Furthermore, from an NMR point of view, it is a well-defined system, with 6−9 hydrogen atoms depending on which residue we are studying. Interactions among these atoms give NOE cross-peaks that can be used for our structural studies.

Two principal low-energy sugar conformations have been described in nucleotide structures: C3′-*endo* with $0° < P < 36°$ (in the "north" of the P circle, or N) and C2′-*endo* with $144° < P < 190°$ ("south", or S). As for the χ angle, crystallographic (*9*) and NMR data (*10*) indicate that the base can adopt two orientations relative to the sugar moiety, called syn (χ angle values 190−270°) and anti (40−90°). A good correlation between the glycosidic angle χ and the torsional angle δ, which is strongly correlated with the sugar pucker, has also been demonstrated (*11*).

Since base−sugar interproton distances depend on both sugar pucker and χ torsion angle, any motion involving these

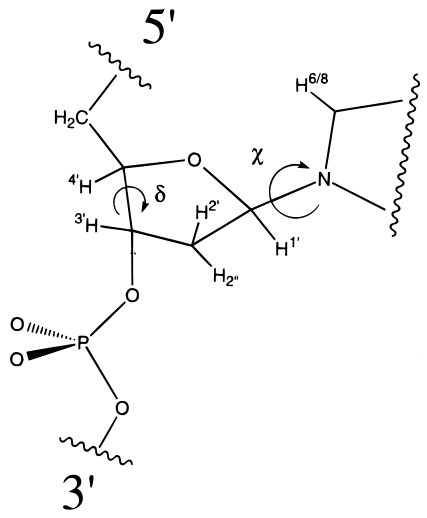

FIGURE 1: A portion of a deoxyribose nucleotide with labeling of the base and sugar hydrogen atoms and defining torsion angles δ and χ.

structural variables should also be reflected by the experimental data. Thus, to justify base−sugar interproton distances fully, we need to consider the effect of sugar repuckering and glycosyl bond flexibility simultaneously. This approach offers new insights for DNA structure analysis since, so far, sugar repuckering and χ flexibility have mostly been considered independently.

MATERIALS AND METHODS

Interproton Distances Derived from 2D NOE Spectra. NOE-derived distance bounds from three independent DNA duplexes are considered: (**I**) the Pribnow box octamer with sequence d(GTATAATG)·d(CATTATAC) (*12*); (**II**) the octamer motif found in both the promoter and enhancer regions in immunoglobulin genes which is contained in a DNA decamer with sequence d(CATTTGCATC)·d(GATGCAAATG) (*13*); and (**III**) the σ^K consensus sequence undecamer d(GCATATGATAG)·d(CTATCATATGC) (Tonelli et al., unpublished results).

All proton−proton distance bounds for these three DNA duplexes were calculated from 2D NOE experiments by MARDIGRAS (*3*−*5*) and were previously reported; the protocols used to determine these distances were also described. It will be noted that the bounds for the σ^K consensus sequence are broader. This is due to a very conservative approach used in determination of bounds; indeed, to account for integration error and spectral noise, the RANDMARDI version of the software has been utilized (*5*).

Of all the distances that were calculated and used for structure refinement, only base-to-sugar intraresidue bounds are analyzed in the present paper. In particular, we focus our attention on H6/8−H1′, H6/8−H2′, H6/8−H2″, H6/8−H3′, and H6/8−H4′ intraresidue distances. The H5−sugar proton distances of cytosines and methyl−sugar proton distances in thymine residues manifest a behavior similar to that of H6/8−sugar but will not be discussed further in this paper.

Molecular Models. Distances corresponding to those measured from 2D NOE data were obtained from various molecular models, including canonical DNA structures (A-

and B-DNA). These canonical models were generated by AMBER (*14*). The canonical models were also used as starting models for MARDIGRAS calculations and structure refinement as described in the original papers.

To generate a larger pool of conformers capable of accommodating most of the experimental NMR data, we utilized molecular dynamics with weighted time-averaged restraints (MD-tar). MD-tar permits a broader search of conformational space in a restrained molecular dynamics (rMD) simulation, as the calculation does not demand that the measured value of the distance be satisfied exactly at every single step of the simulation as in standard rMD calculations (*15*). MD-tar simulations of the octamer Pribnow box have been described already (*16*). The MD-tar simulations result in a very large number of structures, each of which may fit a different subset of the experimental interproton distance restraints. Monte Carlo simulations were also carried out on a DNA dimer with sequence d(CA)·d(TG) with the program DNAminiCarlo (*17*, *18*). No restraints were used in the Monte Carlo simulations in order to produce some structures broadly covering conformational space. The expanded pool of conformers is used in the present study to explore the conformational space accessible to duplex DNA— to examine whether the reproducibly inconsistent experimental distances can be reconciled by conformational exchange.

RESULTS AND DISCUSSION

In this section we analyze the interproton distance bounds for three different DNA duplexes previously investigated in our laboratory. It is evident in Figure 2 that the bounds for the σ^K consensus sequence undecamer (duplex **III**) were determined in an extremely conservative fashion (Tonelli et al., unpublished results), taking into account numerous potential errors ranging from uncertainties in overall rotational correlation time to peak integration and spectral noise contributions (*5*). We will formulate two plausible motional models entailing motions about the glycosyl bond to explain the data. Further examination of the NOE bounds and MD-tar simulation data will lead us to favor one of these models. This inclination will be further strengthened by analysis of the pool of structures generated by Monte Carlo calculations.

Base-to-H2′ and Base-to-H3′ Distances Derived from 2D NOE Data Indicate Sugar Repuckering Motions. First, let us consider H6/8−H2′ and H6/8−H3′ NOE distance bounds. These distances are dramatically affected by sugar pucker in A- and B-DNA: (a) H6/8−H2′ is about 1.7 Å shorter in B-DNA, and (b) H6/8−H3′ is about 1.4 Å shorter in A-DNA.

The effect of sugar pucker on these distances has been investigated by several groups (*10*, *18*−*21*). Lane has proposed the presence of a fast N/S repuckering equilibrium to rationalize NOE data (*19*, *20*). In fact, by comparison of experimental NOE data to canonical A- and B-DNA models, it is evident that these distances cannot be reconciled with a single conformation: base−H2′ bounds are consistently somewhat longer than standard B-DNA but not too close to A-DNA (Figure 2b), while base−H3′ distance bounds are consistently shorter than found in B-DNA and closer to A-DNA (Figure 2d) than are base−H2′ bounds. This apparent contradiction can be justified by the presence of a conformational equilibrium with fast exchange between N and S sugar puckers, with S being the favored conformation, and considering that the experimentally derived distances are biased toward the conformation with the shortest distance. Base−H2′ bounds are consequently longer and similar to standard B-DNA distances (S sugar pucker) because the major conformer has S pucker and a shorter base−H2′ distance; base−H3′ bounds, on the other hand, are shorter than B-DNA and further shifted toward A-DNA, because the N pucker has the shortest base−H3′ distance. As noted earlier, the presence of a S/N sugar repuckering equilibrium is also suggested by scalar coupling-based analysis (*7*, *22*, *23*), and LaPlante et al. have proposed a repuckering process to account for ^{13}C chemical shift data in DNA duplexes (*24*).

Finally, we were able to reproduce base−H2′ and base−H3′ experimental data by back-calculation of the NOE intensities using a mixture of A- and B-DNA model structures (Donati and James, unpublished results) using the algorithm in CORMA (*1*). We note also that we have developed another method, termed PARSE, that generates an ensemble of conformers with an assessment of the probability of each conformer, using cross-relaxation rates simulated by CORMA, by consideration of a much wider pool of conformers which includes structures resembling A- and B-DNA (*6*). We have found in simulated data sets that PARSE is able to select the few correct interconverting conformers from a much wider pool (hundreds) of conformers (*25*).

Similar considerations can be made for base−H2″ and base−H4′ NOE bounds. However, these distances seem to be affected by other factors as well, so we will discuss them later in the paper.

Base-to-H1′ Distances Are Shorter Than in either A- or B-DNA: Two Possible Motional Models. Since the intraresidue H6/8−H1′ distance is a direct function of the glycosidic torsion angle, we expect the value of this distance bound to reflect any motion about the glycosidic bond. Canonical DNA structures, A and B, have similar H6/H8−H1′ distances (3.6−3.7 Å for pyrimidines, 3.8−3.9 Å for purines in both A- and B-DNA), even with χ torsion angles differing by about 60°. Surprisingly, NOE bounds are consistently shorter than either A- and B-DNA distances (Figure 2a). This observation is also in agreement with previously reported data (*26*).

The simplest explanation is that the H6/8−H1′ distances are being systematically underestimated because of errors in the MARDIGRAS treatment. This results from the strong contribution of spin diffusion to these particular NOE intensities via the H2′ and H2″ pathway, especially at higher mixing times. However, it has been shown in model simulations (*5*) that the error resulting from spin diffusion in the average distances calculated by MARDIGRAS is small and is compensated by a wider scatter of individual distances, such that the correct distances are always found within the calculated bounds. The same type of conclusions can be drawn from analysis of the bounds calculated by MARDIGRAS at four different mixing times (70, 130, 200, and 270 ms) for the σ^K consensus sequence (data plots not shown). Taking into account the size of the molecule, we expect the NOE intensities to be affected negligibly by spin diffusion at 70 ms, while being dominated by spin diffusion at 270 ms. It was observed that, while the average distance value decreases by about 0.1 Å going from 70 to 270 ms, the scatter

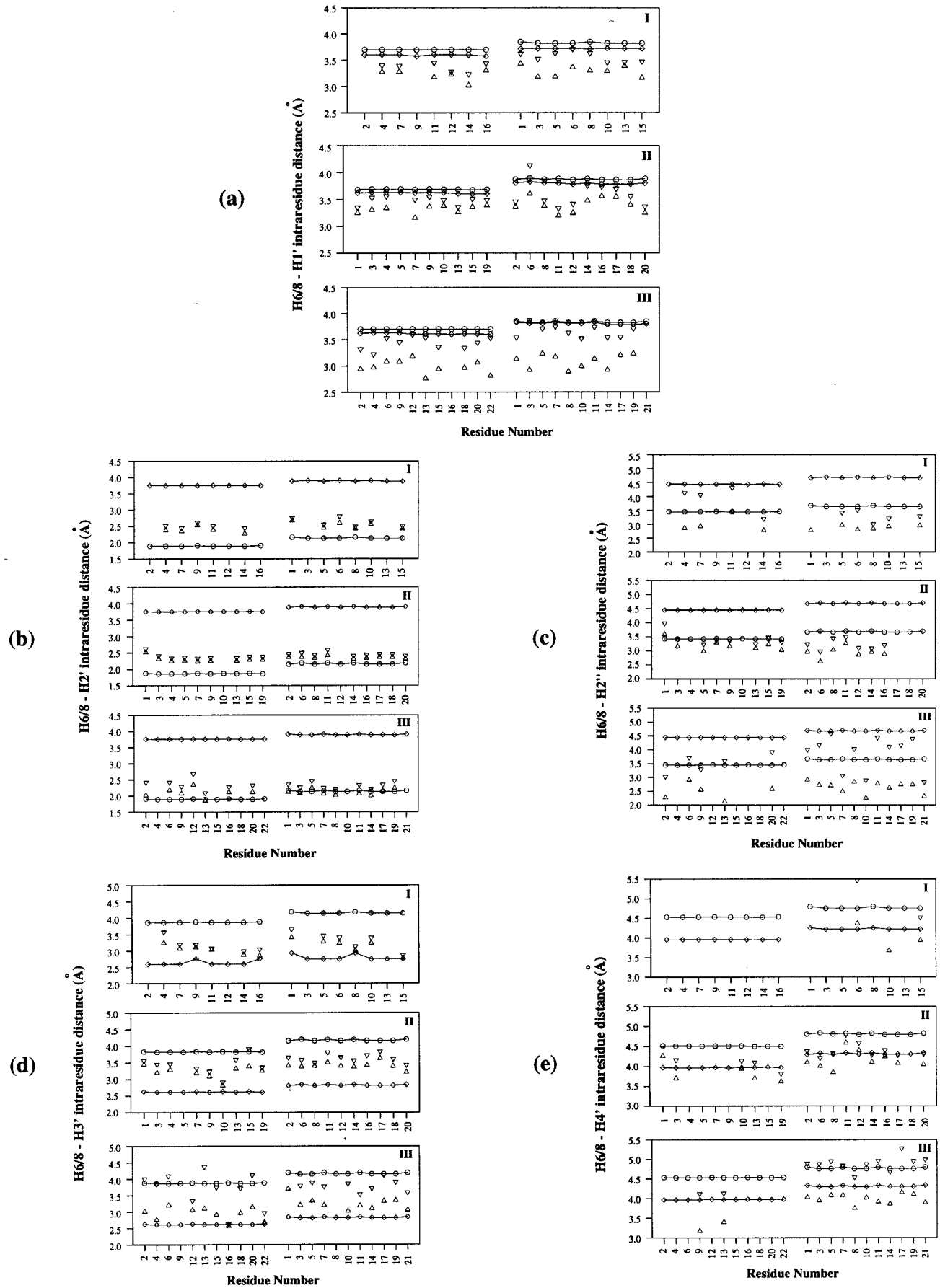

FIGURE 2: NOE distance bounds (lower △, upper ▽) calculated by MARDIGRAS and the corresponding distances from standard A-DNA (◇) and B-DNA (○) for the pribnow box octamer d(GTATAATG)·d(CATTATAC) (**I**), immunoglobulin gene decamer d(CATTTGCATC)·d(GATGCAAATG) (**II**), and σ^K consensus sequence undecamer d(GCATATGATAG)·d(CTATCATATGC) (**III**). H6/8−H1′ (a), H6/8−H2′ (b), H6/8−H2″ (c), H6/8−H3′ (d), and H6/8−H4′ (e) intraresidue bounds and standard distances are shown.

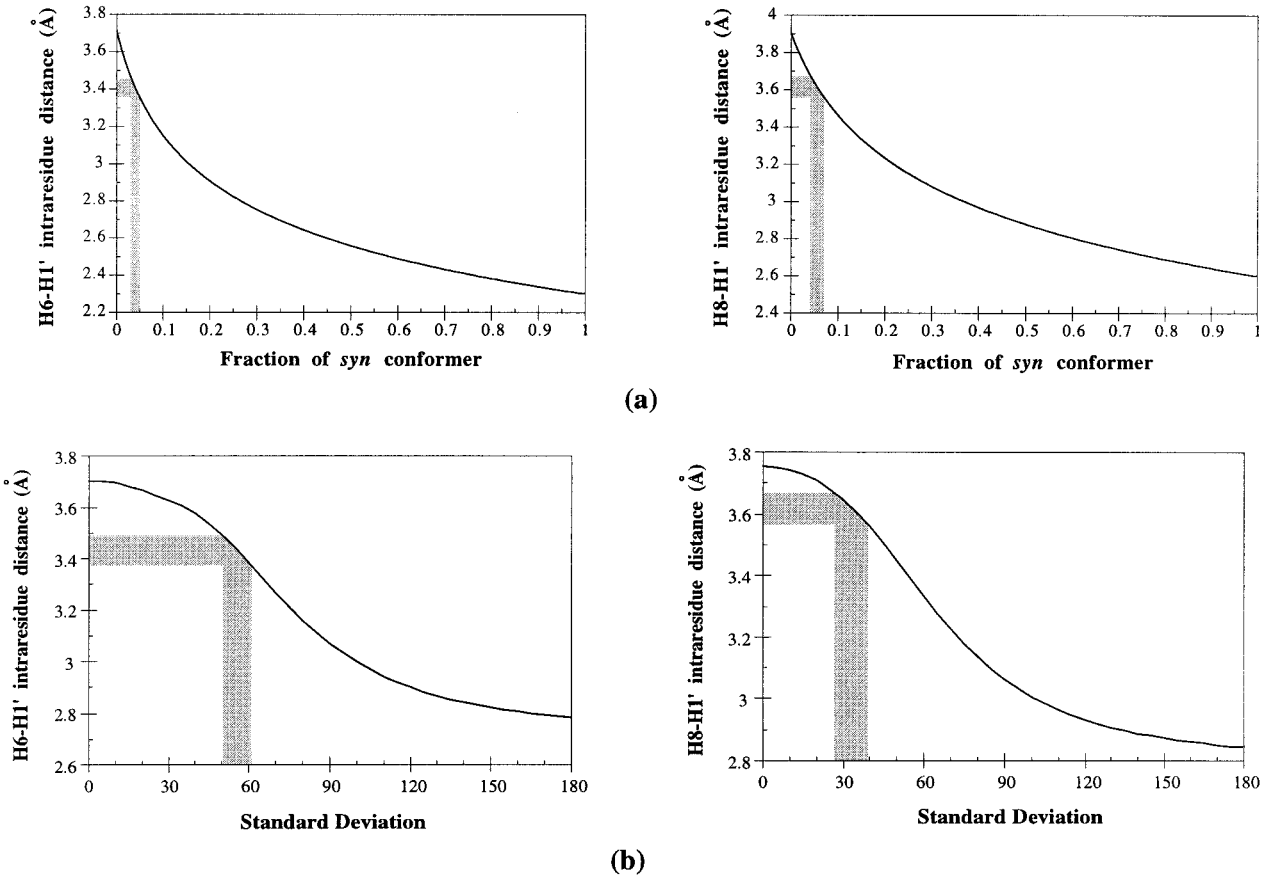

FIGURE 3: (a) Dependence of calculated H6/8−H1′ distances on the fraction of syn conformer present in equilibrium with the anti conformer. The distance value read at 0 fraction of syn conformer represents the actual H6/8−H1′ distance of the anti conformer, and vice versa, the value read at 1 fraction of syn is the actual distance of the syn conformer (no syn/anti equilibrium occurring). (b) Dependence of calculated H6/8−H1′ distances on the standard deviation assuming a normal distribution of the χ torsion angle about an average value of 235° for pyrimidines and 255° for purines. The distance value read at 0 standard deviation represents the H6/8−H1′ distance when no restricted rotational motion of χ is occurring. In both (a) and (b), distances were calculated assuming sixth root averaging of distances for a single conformer. The shaded areas represent the ranges of fraction of syn conformer (a) and standard deviation (b) that are required to match the experimental base−H1′ NOE upper bounds of all nonterminal residues in all three duplexes.

of individual distances increases. In the end, this small decrease was well within the bounds ultimately used in structural calculations.

The second explanation considers the possibility that a dynamic process occurring at the χ angle is responsible for short H6/H8−H1′ distances. It has been observed that intraresidue base−H1′ distances change sinusoidally with torsion angle χ (*10*), such that the distance in canonical A- and B-DNA lies on either side of the top of a sinusoidal curve. Thus, any fluctuation of the χ angle from the value of standard A- and B-DNA must lead to a shorter H6/H8−H1′ distance. On this basis, two plausible internal motion models have been proposed to explain the experimental data (*26*): (1) the jump diffusion model, which describes the effect of a jump between two stable conformations (syn and anti); (2) the restricted diffusion model, which allows restricted rotation about the glycosidic bond between the angles $\chi \pm \Delta\chi$, where χ is a value in the anti conformational region. In the first model, the syn conformer has a much shorter base−H1′ distance (about 1.3 Å shorter) than the anti. In the restricted diffusion model, rocking about the glycosyl bond produces conformers with shorter base−H1′ distance. In either case, because of the distance weighting of NOE intensities, the NOE-derived distances will be shorter than either A- or B-DNA, in accord with experimental data.

Moreover, neither of these two models is physically unreasonable. Experimental data from X-ray crystallography and NMR indicate that both anti and syn conformers can be present in oligonucleotides. On the other hand, Drew et al. (*27*) reported that in many cases the sugar ring appears to be rocking about the glycosidic bond. Establishing whether one or the other of these motions is occurring in our DNA duplexes in solution is our task in this paper.

Assuming that distances are averaged as the sixth root by fast motion (vide supra), (1) for the two-state jump diffusion model, we calculated the fraction of the syn conformer that needs to be present at equilibrium in order to explain the experimental data, and (2) for the restricted diffusion model, assuming a normal distribution of the χ torsion angle with an average value of 255° for purines and 235° for pyrimidines (typical of B-DNA), we calculated how broad this distribution should be in order to shorten base−H1′ distances by the observed amount. We found that the experimental NOE data can be satisfied by (1) less than 10% of syn conformer at equilibrium for either purines or pyrimidines (Figure 3a) or (2) a normal distribution with a standard deviation of 28−40° for purines and even higher for pyrimidines (50−60°) (Figure 3b).

While the percent of syn conformer calculated is in agreement with what has been previously estimated (*26*), the

standard deviation values we found are a bit larger than expected, especially for purines [Gaudin et al. (*26*) found that a $\Delta\chi$ of 28° could best explain their relaxation data]. This may be the result of oversimplification in our calculations. We reserve discussion of this issue in more detail later in the paper. At this point let us just conclude that analysis of the H6/8−H1′ NOE distances alone does not allow us to favor unequivocally one model over the other.

Consideration of All Intraresidue Base−Sugar Distances. So far, we have only analyzed H6/H8−H2′, H6/H8−H3′ distances and H6/H8−H1′ distances independently. Now, let us consider all the base-to-sugar distances, including distances to H2″ and to H4′ that we have not discussed yet. While a base−H1′ distance depends directly only on the glycosidic torsion angle, the other base−sugar proton−proton distances are determined by both the χ angle and sugar pucker.

Intraresidue distances to H2″, like those to H2′, are shorter in B-DNA than in A-DNA. We might therefore expect the corresponding NOE bounds to behave alike. However, if we look carefully at the experimental values, we notice that, while base−H2′ bounds are mostly in agreement or longer than in B-DNA (Figure 2b), base−H2″ bounds are consistently shorter than in standard B-DNA (Figure 2c). In a similar way, comparing base−H3′ with base−H4′ distances, we note that they are both longer in B-DNA than in A-DNA, but the NOE bounds for base−H4′ are systematically shorter and closer to A-DNA than are base−H3′ bounds (Figure 2d,e).

Sugar repuckering, which accounted for base−H2′ and base−H3′ bounds observations, cannot fully explain what we see with bounds to H2″ and to H4′. There must be some other variable that is affecting base−sugar distances differently. This causes bounds to H2″ and H4′ to be shorter than those to H2′ and H3′, respectively. Since the only other structural variable directly affecting these distances is the torsion angle χ, motion about this angle must be responsible for these observations.

We conclude that any motion about the glycosyl bond that is responsible for short base−H1′ bounds is also shortening base−H2″ and base−H4′ distances, while it is not affecting base−H2′ and base−H3′ distances. Of the two previously proposed motional models involving the χ angle, we will find out that only one can explain this behavior.

Feasible Internal Motional Models: Analysis of MD-tar Trajectories. To help explain the observed NOE bounds, we analyzed structures generated during MD-tar simulations. The trajectories we used were run on the Pribnow box octamer duplex (*16*). In MD-tar simulations, the experimental restraints and the associated penalty function are monitored as a running average with exponential weighting to emphasize more recent steps during an rMD simulation. As the interproton distance restraints are enforced as a running average rather than at every snapshot as in the usual rMD simulations, MD-tar has proved to explore conformational space more efficiently. This allows MD-tar to account better for any molecular flexibility encoded in the experimental data which cannot be satisfied by a single conformation.

First, we examined how the χ torsion angle and the pseudorotation angle P change during the simulation. In agreement with the original paper (*16*), we observed that (1) torsion angle χ appears to be more dispersed compared to what has been seen in regular rMD, (2) the syn conformer at the glycosidic bond is only found in terminal residues and in one subterminal residue (adenine 15), (3) the sugar rings are jumping between N and S conformations, with the S conformer dominant, (4) the change in P is associated with a smaller, but consistent, change in the χ torsion angle value. This is in agreement with the previously observed correlation between the glycosidic torsion angle and the backbone/sugar torsion angle δ (*11*), since δ is strongly related to the sugar pucker. Recapitulating, for individual residues during MD-tar simulations two interchanging conformations with different P and χ values exist. These are characterized by sugar puckers in the N and S regions. The syn conformer is populated only for the terminal residues of the duplex.

Finally, in our aim to explain the observed NOE bounds, we plotted base−sugar interproton distances against the glycosidic torsion angle value for each snapshot of the trajectory. This allows us to see directly any dependence of the distances on variations of χ. Figure 4 shows all base−sugar distances vs χ angle plots for adenine 5. It is evident for those distances that change significantly between N and S sugar puckers (base H8−H2′, H8−H2″ and H8−H3′ in particular), two interchanging populations corresponding to conformers with S and N sugar pucker. The two curves shown in the plots represent how the distances change with the χ angle in standard A and B adenine residues. These curves were obtained simply by measuring the H8−sugar distances while "manually" changing χ starting with two standard A and B adenine residues. This entire range is, of course, physically not very realistic; however, it helps in understanding the plots.

Now, let us reconsider the two proposed motional models of motion about the χ angle: the jump diffusion model and the restricted diffusion model. Using the plots in Figure 4, we will try to justify the NOE patterns observed for all H8−sugar distances using first one model and then the other. To simplify our analysis when trying to explain the effect of χ flexibility, we will focus our attention on the major population with S sugar pucker. The minor N conformer is used to analyze the effect of sugar repuckering.

The jump diffusion model describes the effect of a jump between syn and anti conformers. These structures have preferred χ regions from 190° to 270° (anti) and from 40° to 90° (syn) (*10*). The MD-tar structures (Figure 4) reflect the experimental H8−H1′ NOE bounds (Figure 2a) which are shorter than in A- or B-DNA, both being in the anti region. The syn conformer has H8−H1′ distances about 1.3 Å shorter than the anti conformer. Thus, the presence of small amounts of a syn conformer in equilibrium with the anti conformer can explain the observed short values. Base H8−H2′, H8−H2″, H8−H3′, and H8−H4′ bounds are directly affected by sugar pucker as well as χ angle. Sugar repuckering with S/N equilibrium can partially explain the experimental NOE values. However, we noticed that H8 bounds to H2″ and to H4′ are relatively shorter when compared to both A- and B-DNA than are bounds to H2′ and to H3′, respectively (see Figure 2). We attributed this difference to χ flexibility (vide supra). In the syn conformer, all these distances have about the same value as in the anti conformer, or in some cases, they are longer. Thus, the jump

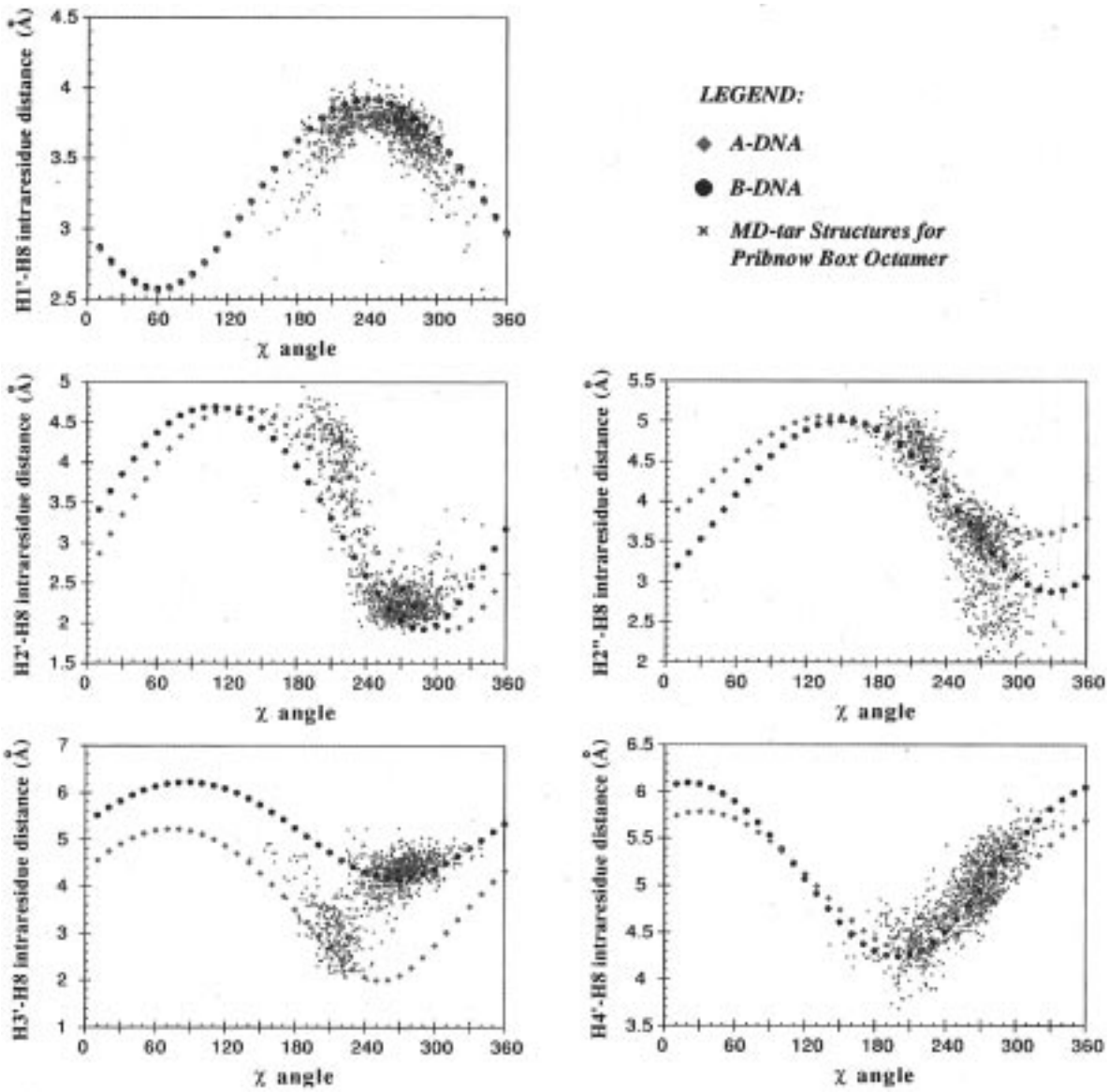

FIGURE 4: Correlation between H8−sugar proton distances and the glycosidic torsion angle for adenosine 5 in the Pribnow box octamer duplex d(GTAT**A**ATG)·d(CATTATAC) obtained from MD-tar simulations (~900 structures, green). The plots also show how H8−sugar proton distances change with χ angle in adenosine residues with either S (blue) and N (red) sugar pucker. These curves were obtained simply by measuring the base H8−sugar proton distances while manually changing χ starting with two standard A and B adenosine residues.

diffusion model does not explain shortening of H8 NOE bounds to H2″ and H4′ relative to those to H2′ and H3′, respectively.

The restricted diffusion model allows restricted diffusion about the glycosidic bond between the angles $\chi \pm \Delta\chi$, with an average χ value of about 255° for adenines with S sugar pucker. This model can explain the observed base H8−H1′ short bounds, since deviations of χ from the average value decrease the distance to H1′. Base H8−H2′ and H8−H3′ distances fall in a valley of the sinusoidal curve, while H8−H2″ and H8−H4′ distances lie on a slope of the curve. Consequently, experimental distances of H8 to H2′ and to H3′ will increase when χ oscillates around the center position. Instead, distances to H2″ and to H4′ will increase when χ changes upward and decrease when χ moves downward the slope. Since NOE distance bounds are biased toward short values, we conclude that rocking about the glycosidic bond *can* justify short H8−H2″ and H8−H4′ bounds, while bounds to H2′ and H3′ are little affected. Obviously, the same arguments hold for H6 protons in pyrimidines as for purine H8.

Thus, of the two proposed models for χ flexibility, only the restricted diffusion model, combined with S/N sugar repuckering, is able to justify the observed base−sugar bounds patterns.

Analysis of a Pool of Conformers Generated by Monte Carlo Calculations. So far, we have analyzed structures being generated by MD-tar simulations. One could argue that biases affecting this method may also prejudice our conclusions. Consequently, we decided to analyze a pool of conformers generated by the Monte Carlo module of the DNAminiCarlo program. Unrestrained energy minimization calculations were previously carried out on a two-base-pair DNA fragment with sequence d(CA)·d(TG). Improved sampling of conformational space was achieved by applying the "scanning procedure" described elsewhere (*17*, *18*).

This pool of conformers was analyzed in a fashion similar to that of the MD-tar trajectories described in the previous

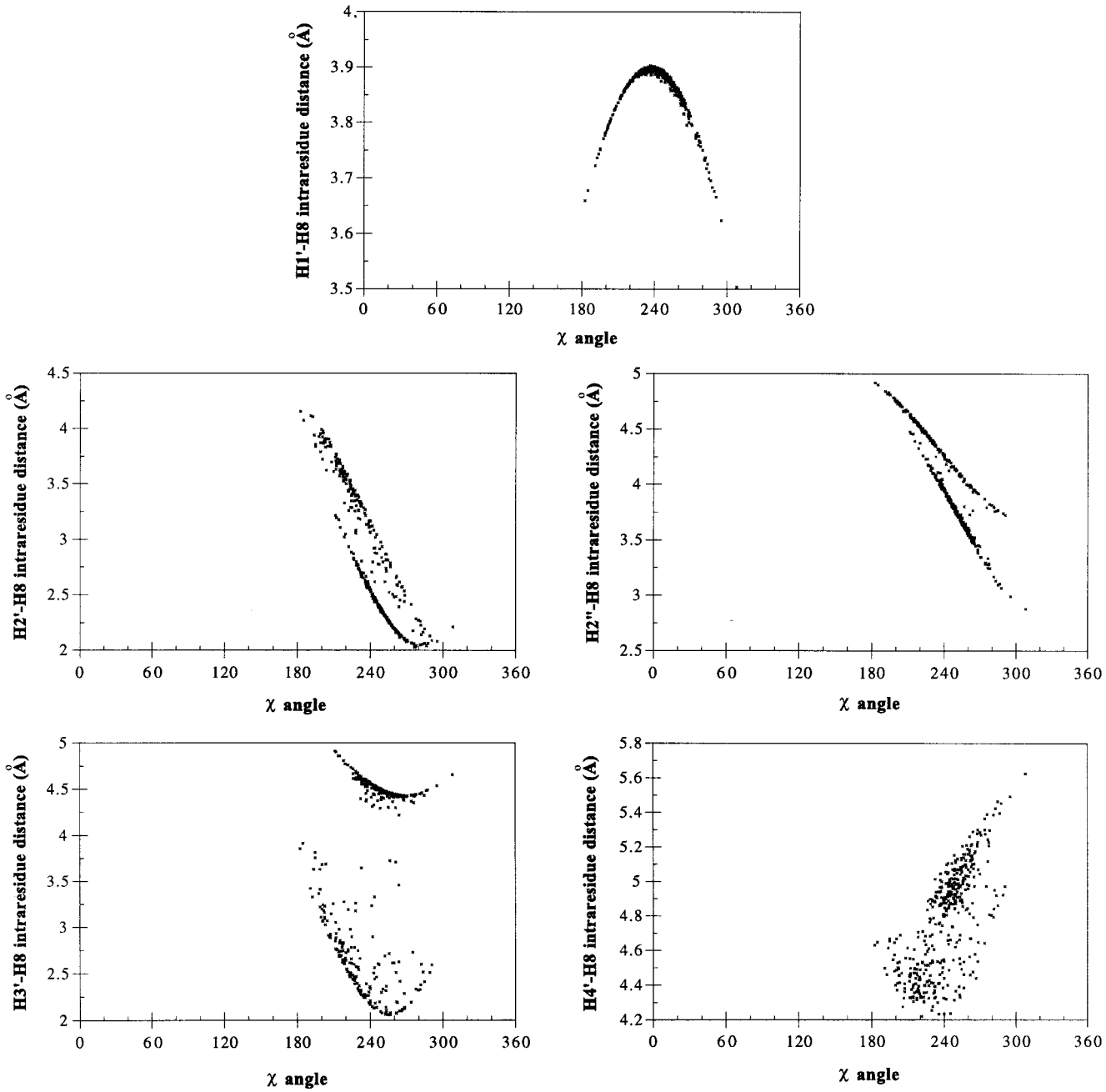

FIGURE 5: Correlation between H8−sugar proton distances and the glycosidic torsion angle for adenosine 2 in the dimer duplex d(CA)·d(TG) obtained from unrestrained energy minimization performed by the Monte Carlo modules of the DNAminiCarlo program. Covering of conformational space was achieved by applying the "scanning procedure" that has been described elsewhere (*17, 18*).

section. Figure 5 shows the dependence of intraresidue base−sugar distances on the χ angle value for the adenine residue. As for MD-tar, it is easy to notice the presence of two populations corresponding to conformers with S and N sugar pucker.

Two main differences can be pointed out: (1) the torsion angle χ is shifted toward smaller values, and (2) there is a narrower range of distances in the pool of conformers generated by the Monte Carlo method. The shift in the χ angle value probably results from the NOE restraints being applied during MD-tar but not during DNAminiCarlo calculations. In particular, short H8−H2′ and H8−H2″ bounds are pulling the χ torsion angle to higher values during MD-tar simulations in order to satisfy the restraints. Second, because of the narrower ranges of distances, the two populations corresponding to N and S sugar conformers are more clearly defined in the DNAminiCarlo plots. This may appear somewhat surprising if we consider that the Monte Carlo simulations were run free of experimental restraints. However, it can be easily understood since, in DNAminiCarlo, sugar pucker amplitudes are allowed to change only within a small range, while there are no restrictions in MD-tar simulations. In fact, in the MD trajectories, it is easy to find conformations with unrealistic sugar pucker amplitudes. These are often the result of inconsistencies in the NOE data enforced during the simulation.

With the differences being taken into account, the DNAminiCarlo plots support the same type of conclusions we draw from analysis of MD-tar simulations. Specifically, the existence of a S/N sugar pucker equilibrium in tandem with restricted rotation about the glycosidic bond best explains the NOE patterns.

CONCLUSIONS

A Motional Model for Duplex DNA: Sugar Repuckering and Glycosidic Torsion Angle Flexibility. Analysis of intraresidue base−sugar interproton NOE bounds patterns suggests a motional model with S/N sugar repuckering, with S being the preferred conformer. Associated with repuckering, there is a small, but consistent, change in the glycosidic torsion angle χ. In addition to this, the χ torsion angle fluctuates around an average value, as described by a restricted diffusion model. Thus, sugar repuckering accounts for (a) base H8/H6−H2′ and H8/H6−H2″ bounds which are longer than in typical B-form DNA conformation and (b) H8/H6−H3′ and H8/H6−H4′ bounds which are shorter than in standard B-DNA and closer to A-DNA. On the other hand, restricted rotation about the glycosidic torsion angle justifies (a) H8/H6−H1′ bounds shorter than in both A- and B-DNA and (b) H8/H6−H2″ and H8/H6−H4′ bounds relatively shorter than for H8/H6−H2′ and H8/H6−H3′ bounds, respectively, when compared to standard A- and B-DNA values.

Even if we cannot exclude the presence of small amounts of syn conformer at equilibrium with the anti conformer in our analysis, it is clear that the jump diffusion model cannot fully explain the experimental data. Hence, this model is not required.

We pointed out earlier that the standard deviation estimated for the restricted diffusion model is broader than what has being previously estimated. However, our calculations were performed assuming a single Gaussian distribution centered about the χ angle value typical of S sugar pucker. Our study also suggests the presence of a smaller percent of N conformer with shifted χ torsion angle. The reasoning for variations in H8/H6−H1′ distances with restricted χ angle diffusion on the S conformer also applies to the N conformer. Moreover, both MD-tar and DNAminiCarlo simulations suggest a more scattered distribution of the χ angle in the N conformer. Thus, the presence of a small percent of N conformer with a broader χ angle distribution should further diminish the measured distance. Of course, this implies a narrower distribution for the χ angle in the S conformer. By repeating our calculations considering two χ angle distributions corresponding to the N and S conformers, we could estimate the size of the distributions, but this would entail nearly as many parameters as observables.

Finally, we need to consider the effect that random rotational molecular motion has on NOE intensities. In the model-free approach of Lipari and Szabo (*31*) the information on internal motions is specified by two model-independent quantities: an order parameter S^2 which is a measure of the spatial restriction of the motion and an effective correlation time τ_e which is a measure of the rate of the random internal motion. Kojima and James (unpublished results), using the model-free analysis on six relaxation parameters measured on a ^{13}C-labeled DNA duplex d(CATTTGCATC)·d(GATGCAAATG), estimated internal motions in the picosecond time scale (10−40 ps), with an order parameter $S^2 \approx 0.8$ for C−H vectors in both bases and sugars. Using a simple wobble-in-a-cone model (*28*) gives some physical perspective to an order parameter; a value of 0.8 means that a C−H vector can rapidly reorient up to an angle of 44°. Similar relaxation data have also been described by other groups (*26, 29*).

Existence of such a fast internal motion, faster than the overall tumbling time of the duplex, will reduce cross-relaxation rates and, consequently, NOE intensities. Model-free analysis indicates that motional averaging of distances will change from a $\langle r^{-6} \rangle$ dependence to a $(S^2 \times \langle r^{-6} \rangle)$. Thus, distances calculated from NOE intensities will be overestimated if we do not take into account the contribution from fast internal motion. However, our intensities, calculated by MARDIGRAS, were normalized against fixed distances (cytosine H5−H6 and thymine methyl−H6 distances). Since all C−H vectors in DNA were found to have order parameters around 0.8, we expect all proton−proton pairs to have similar order parameter values as well. Normalization of the NOE intensities should minimize any contribution from fast internal motion as long as the internal motions are comparable for all proton pairs (*30*).

Studying molecular flexibility through analysis of proton−proton distance values offers an advantage, compared with the heteronuclear relaxation parameter studies alone, of yielding a realistic picture of the type of motion in the context of known structural information. Our model describes two types of motion: restricted rotation about an average position of the glycosidic torsion angle χ and N/S sugar repuckering associated with a shift of the average χ value by about 60°. It is reasonable to argue that these two motions occur on a different time scale, with sugar repuckering being slower, since it involves transitions between different energy minima. If this is the case, then the value of τ_e, found by the model-free analysis of relaxation parameters, should reflect the time scale of the fastest motion, i.e., χ angle fluctuations. While the time scale for sugar repuckering is not truly known, it is likely to be in the nanosecond range and, consequently, to be difficult to distinguish from overall molecular tumbling but to have nonnegligible effects on cross-relaxation rates.

In recent years, methodology development has permitted us to determine quite well the time-average, sequence-dependent structure of double helical nucleic acids. Key to this has been the use of interproton distances determined from two-dimensional nuclear Overhauser effect spectroscopy using the complete relaxation matrix approach embodied in the software MARDIGRAS. However, with the increasing number of studies conducted in our laboratory, it has become clear that certain classes of distances are inconsistent with others in each of the structures studied. Indeed, some are inconsistent with the duplexes possessing structure in either the B or A family of structures, i.e., those which one would imagine are energetically most reasonable.

The current investigation has examined possible causes for these inconsistencies, focusing on the likelihood that internal motions are responsible. It has been demonstrated that repuckering of the deoxyribose ring, which scalar coupling-based spectroscopy results suggest occurs, cannot reconcile all the NOE data. However, restricted rotational diffusion about the glycosidic bond, together with sugar repuckering, does account for the NOE data, indicating that this is a general phenomenon as it has been observed in each of the DNA duplexes studied in our laboratory. A more complete description of the dynamic nature of the duplex accounting for repuckering and glycosidic bond diffusion will require an analysis using more sophisticated modeling and additional data.

ACKNOWLEDGMENT

We are grateful to Dr. Chojiro Kojima for helpful discussions. Use of the facilities of the Pittsburgh Supercomputing Center for some calculations and the UCSF Computer Graphics Laboratory, supported by NIH Grant RR01081, is gratefully acknowledged.

REFERENCES

1. Keepers, J. W., and James, T. L. (1984) *J. Magn. Reson. 57*, 404−426.
2. LeMaster, D. M., Kay, L. E., Brünger, A. T., and Prestegard, J. H. (1988) *FEBS Lett. 236*, 71−76.
3. Borgias, B. A., and James, T. L. (1990) *J. Magn. Reson. 87*, 475−487.
4. Liu, H., Kumar, A., Weisz, K., Schmitz, U., Bishop, K. D., and James, T. L. (1993) *J. Am. Chem. Soc. 115*, 1590−1591.
5. Liu, H., Spielmann, H. P., Ulyanov, N. B., Wemmer, D. E., and James, T. L. (1995) *J. Biomol. NMR 6*, 390−402.
6. Ulyanov, N. B., Schmitz, U., Kumar, A., and James, T. L. (1995) *Biophys. J. 68*, 13−24.
7. Rinkel, L. J., and Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621−649.
8. Schmitz, U., and James, T. L. (1995) *Methods Enzymol. 261*, 3−44.
9. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer, New York.
10. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
11. Fratini, A. V., Kopka, M. L., Drew, H. R., and Dickerson, R. E. (1982) *J. Biol. Chem. 257*, 14686−14707.
12. Schmitz, U., Sethson, I., Egan, W., and James, T. L. (1992) *J. Mol. Biol. 227*, 510−531.
13. Weisz, K., Shafer, R. H., Egan, W., and James, T. L. (1994) *Biochemistry 33*, 354−366.
14. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W., and Kollman, P. A. (1995) *J. Am. Chem. Soc. 117*, 5179−5197.
15. Torda, A. E., Scheek, R. M., and van Gunsteren, W. F. (1989) *Chem. Phys. Lett. 157*, 289−294.
16. Schmitz, U., Ulyanov, N. B., Kumar, A., and James, T. L. (1993) *J. Mol. Biol. 234*, 373−389.
17. Zhurkin, V. B., Lysov, Y. P., Florentiev, V. L., and Ivanov, V. I. (1982) *Nucleic Acids Res. 11*, 1811−1830.
18. Ulyanov, N., Gorin, A. A., Zhurkin, V. B., Chen, B.-C., Sarma, M. H., and Sarma, R. H. (1992) *Biochemistry 31*, 3918−3930.
19. Lane, A. N. (1990) *Biochim. Biophys. Acta 1049*, 189−204.
20. Lane, A. N. (1990) *Biochim. Biophys. Acta 1049*, 205−212.
21. Wijmenga, S. S., Mooren, M. M. W., and Hilbers, C. W. (1993) in *NMR of Nucleic Acids; from Spectrum to Structure* (Roberts, G. C., Ed.) pp 217−288, IRL Press, Oxford.
22. Schmitz, U., Zon, G., and James, T. L. (1990) *Biochemistry 29*, 2357−2368.
23. Lane, A. N. (1993) *Prog. NMR Spectrosc. 25*, 481−505.
24. LaPlante, S. R., Zanatta, N., Hakkinen, A., Wang, A. H.-J., and Borer, P. N. (1994) *Biochemistry 33*, 2430−2440.
25. Ulyanov, N. B., Mujeeb, A., Donati, A., Furrer, P., Liu, H., Farr-Jones, S., Konerding, D. E., Schmitz, U., and James, T. L. (1998) in *Determination of Structural Ensembles from NMR Data: Conformational Sampling and Probability Assessment* (Leontis, N. B. and SantaLucia, J. J. Eds.) pp 181−194, American Chemical Society, Washington, DC.
26. Gaudin, F., Paquet, F., Chanteloup, L., Beau, J. M., Nguyen, T. T., and Lancelot, G. (1995) *J. Biomol. NMR 5*, 49−58.
27. Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., and Dickerson, R. E. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2179−2183.
28. London, R. E. (1980) in *Intramolecular Dynamics of Proteins and Peptides as Monitored by Nuclear Magnetic Relaxation Measurements* (Cohen, J. S., Ed.) pp 1−69, John Wiley & Sons, New York.
29. Borer, P. N., LaPlante, S. R., Kumar, A., Zanatta, N., Martin, A., Hakkinen, A., and Levy, G. C. (1994) *Biochemistry 33*, 2441−2450.
30. Kumar, A., James, T. L., and Levy, G. C. (1992) *Isr. J. Chem. 32*, 257−261.
31. Lipari, G., and Szabo, A. (1982) *J. Am. Chem. Soc. 104*, 4546−4559.

BI980905D