# The Three-Dimensional Structure in Solution (pH 5.8) of a DNA 9-mer Duplex Containing 1,$N^2$-Propanodeoxyguanosine Opposite Deoxyadenosine. Restrained Molecular Dynamics and NOE-Based Refinement Calculations†

Ping Huang and Moisés Eisenberg*

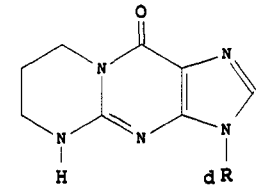
*Department of Pharmacological Sciences, School of Medicine, State University of New York at Stony Brook, Stony Brook, New York 11794-8651*

*Received January 17, 1992; Revised Manuscript Received April 27, 1992*

ABSTRACT: The solution structure of the complementary d(C1-A2-T3-G4-X5-G6-T7-A8-C9)·d(G10-T11-A12-C13-A14-C15-A16-T17-G18) nonanucleotide duplex (designated X·A 9-mer) that contains a 1,$N^2$-propanodeoxyguanosine exocyclic adduct, X5, opposite deoxyadenosine A14 in the center of the helix at pH 5.8 is investigated by use of restrained molecular dynamics followed by NOE-based back-calculation refinement. The molecular dynamics calculation is based on 91 interresidue and 97 intraresidue interproton distance restraints derived from two-dimensional nuclear Overhauser enhancement data on the X·A 9-mer at mixing times of 50 and 250 ms [Kouchakdjian, M., Marinelli, E., Gao, X., Johnson, F., Grollman, A., & Patel, D. J. (1989) *Biochemistry 28*, 5647–5657]. Separate runs start from classical A and B DNA and converge to essentially identical structures (atomic root mean square difference of 0.69 Å). Both structures are B-type DNA in character and satisfy the experimental distance restraints with the rms difference of only 0.001 Å between the calculated and experimental interproton distances. The dynamics behavior of the A → B DNA transition is monitored and analyzed. Our results clearly indicate that the driving force of the convergence is the experimental interproton distance restraints. The molecular dynamics structures are further refined by a back-calculation dynamics which directly minimizes the difference between the observed 2D NOE intensities and those calculated by the full relaxation matrix approach. The fit of the refined structures to the NOE intensities is measured by the NOE *R* value, which is analogous to the crystallographic residual index. These *R* values of the final structures are only 0.17. The refined structures are generally B type, and their convergency improves slightly to an atomic root mean square difference of 0.64 Å, despite relatively large structural shifts (~1 Å) which occur during the back-calculation refinement in both cases. These results suggest that the converged refined structures represent reasonable approximations of the solution structure.

Cyclic nucleic acid adducts, which are induced by environmental pollutants, may play an important role in mutagenesis and carcinogenesis (Singer & Barsch, 1986; Singer & Grunberger, 1983). Recently, biological (Grollman, 1989) and NMR studies (Kouchakdjian et al., 1989, 1990) have focused on 1,$N^2$-propanodeoxyguanosine (**1**) (designated X), embedded in DNA oligodeoxynucleotide duplexes. It has been shown that this adduct is an excellent model for the acrolein modified deoxyguanosine. Like the naturally occurring mutagenic adduct, upon replication, deoxyadenosine is incorporated preferentially opposite it. Chemically, however, X is significantly more stable than the unsaturated exocyclic acrolein adduct. It was also demonstrated by NMR that the structure of the X·A 9-mer duplex (**2**) undergoes a pH-dependent conformational transition at the lesion site with a p$K$ of 7.4. This paper deals with the structure at the acidic pH of 5.8.

The two-dimensional NOESY studies (Kouchakdjian et al., 1989) of the X·A 9-mer duplex at pH 5.8 in aqueous solution establish that X5 adopts a syn orientation while A14 on the opposite strand adopts an anti orientation about the glycosidic bond at the lesion site. NMR reveals that at pH 5.8 the deoxyadenosine base opposite X is protonated at the N1 position and the X5(*syn*)·A14(*anti*) pair (**3**) is stabilized by two hydrogen bonds. It was also demonstrated that the X·A 9-mer forms a right-handed helix with the 1,$N^2$-propano bridge of X5(*syn*) protruding to the major groove.

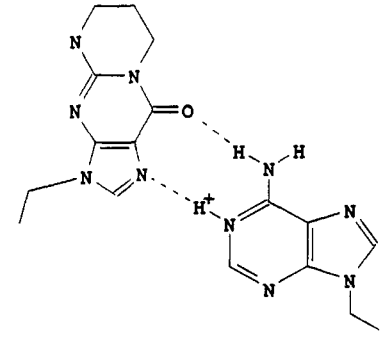



(1)

C1 — A2 — T3 — G4 — X5 — G6 — T7 — A8 — C9

G18 — T17 — A16 — C15 — A14 — C13 — A12 — T11 — G10

(2)




(3)

X(syn) A(anti)

The structure described above was determined by applying a distance-restrained energy minimization to a canonical

† This work was supported by NIH Grants CA47995 and ES0406804.

0006-2960/92/0431-6518$03.00/0 

B-form DNA (Kouchakdjian et al., 1990). That approach yielded a structure which satisfies relatively broad distance restraints derived from 2D NMR data but essentially does not deviate significantly from the starting structure, with the exception of the syn conformation for the propanodeoxyguanosine residue. A more detailed knowledge of the structure of the X·A 9-mer duplex in solution is, however, desirable to understand at the molecular level how the local helix distortions at sites of exocyclic adduct formation exert substantial effects on DNA replication and repair and lead ultimately to mutagenesis and carcinogenesis. It has been widely recognized that molecular dynamics refinements using the interproton distances measured from NOESY spectra as restraints provide a powerful methodology for the elucidation of three-dimensional molecular structures. With this approach, molecular dynamics calculations are performed with a modified potential such that the interproton distances are incorporated in the form of an effective potential term. The energetic contributions of all experimentally determined distances, therefore, are taken into account during the entire course of the dynamics simulation. It has been shown that this method yields final structures with a satisfactory radius of convergence from various starting conformations, and it has been applied to both proteins (Brünger et al., 1986; Clore et al., 1986a) and oligonucleotides (Nilsson et al., 1986; Nilges et al., 1987a,b; Gronenborn & Clore 1989; Baleja et al., 1990). Since the distance data calculated from the volume integrals of the observed NOE cross peaks are only approximate due to spin diffusion effects (Kalk & Berendsen, 1976), the structures derived by distance-restrained molecular dynamics need to be improved. There have been a number of techniques to calculate the intensities of the cross peaks in time-dependent NOESY data sets based on a multispin relaxation analysis for a given structure (Keelers & James, 1984; Marion et al., 1987; Lefevre et al., 1987). The structures resulting from restrained molecular dynamics can therefore be refined in an iterative manner so as to minimize the difference between the two sets of NOEs, experimentally determined and theoretically calculated. Although in principle one might use NOEs directly in a long molecular dynamics run, thereby eliminating the step of making rather approximate distance determinations, the computational time required is very long and impractical. The major drawback of the use of relaxation matrix in NMR refinement is the high computational cost since the diagonalization necessary to calculate intensities from a conformation is time-consuming (Nilges et al., 1991). Moreover, it is unclear whether structures determined exclusively by back-calculating NOEs directly in long molecular dynamics simulations would be necessarily "more correct" than those produced by using approximate distances, followed by refinement with NOEs (Baleja et al., 1990).

In this paper, we have chosen to study the X·A 9-mer duplex (**2**) using distance-restrained molecular dynamics, followed by NOE-based refinement calculations. The interproton distances were derived from the NOESY spectra at mixing times of 50 and 250 ms (Kouchakdjian et al., 1989). Even though the estimated distances are defined by broad lower ($\Delta_-$ = 0.5–1.2 Å) and upper ($\Delta_+$ = 0.8–1.2 Å) bounds, convergence is achieved (rms deviation of 0.69 Å) starting from two different initial structures, namely, classical A- and B-type DNA (rms deviation of 5.18 Å), after a high-temperature (450 K) dynamics (to allow crossing of structural energy barriers) followed by a lower-temperature (300 K) equilibration. Analysis of the dynamics behavior in the transition of A DNA to a B-type structure, as well as the analysis of energies, demonstrates that the driving force of this convergence is the experimental interproton distance restraints although the distance bounds are relatively large. We follow the dynamics with a back-calculation refinement procedure (Nilges et al., 1991) which incorporates the difference between observed and calculated 2D NOE intensities as an effective energy term and the analytical derivatives of NOESY intensities with respect to atomic coordinates (Yip & Case 1989). The fits to the experimental 2D NOE intensities are improved significantly after back-calculation refinements. This can be measured by the NOE $R$ value which corresponds to the residual index commonly used in crystallography (Stout & Jensen, 1989). The final structures have $R$ values of 0.17, and their convergency improves slightly to 0.64 Å (rmsd), although relatively large structural shifts ($\sim$1 Å) occurred in the back-calculation refinement.

## Materials and Methods

### *2D NOE Intensities and Interproton Distances*

All NMR experimental data were from the previous NMR studies of the X·A 9-mer at pH 5.8 (Kouchakdjian et al., 1989). In these studies, two-dimensional phase-sensitive NOESY spectra of the X·A 9-mer in $D_2O$ buffer were recorded at two mixing times: 50 and 250 ms. The volume integrals of individual NOE cross peaks were measured from the NOESY contour plots. A total of 115 (50 ms) and 173 (250 ms) observed volume integrals with a range of ±5% for the bounds were used in our NOE-based back-calculation refinements. The range for the bounds of volume integrals reflects the integration error due to noise and spectral overlap.

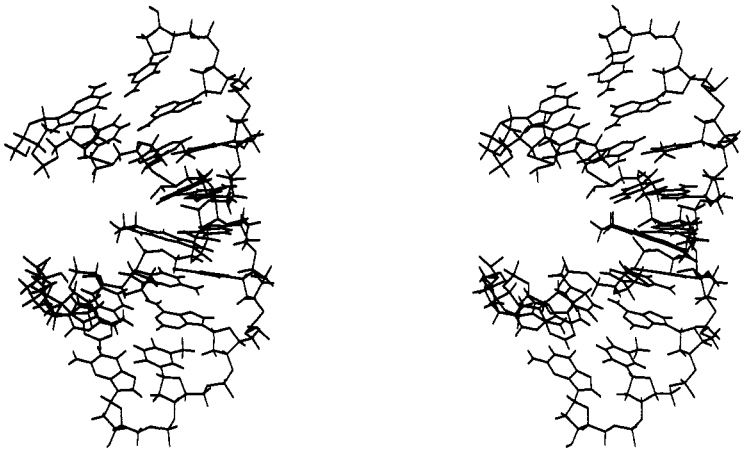
The interproton distances were estimated from the NOESY spectra on the basis of the relationship

$$r_{ij}^6/r_{kl}^6 = \sigma_{kl}/\sigma_{ij} \quad (4)$$

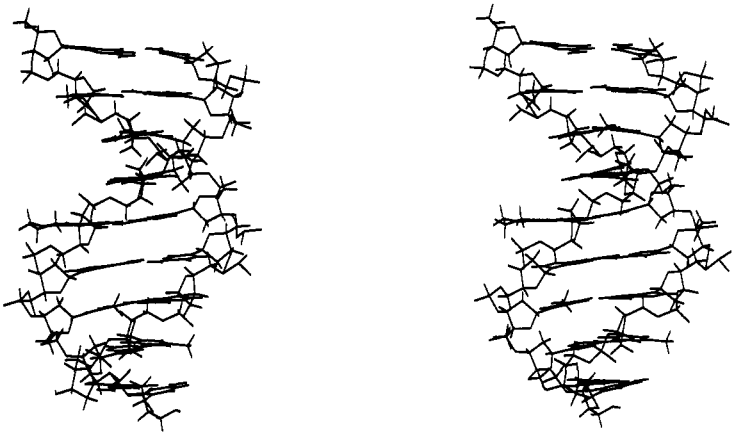
where $r_{ij}$ is the distance between protons $H_i$ and $H_j$, $r_{kl}$ is the fixed internal reference distance (cytidine H5–H6 distance of 2.45 Å), and $\sigma_{ij}$ and $\sigma_{kl}$ are the corresponding cross-relaxation rates. The ratio $\sigma_{kl}/\sigma_{ij}$ can be calculated by measuring the volume integrals of the cross peaks corresponding to interactions $k$–$l$ and $i$–$j$. The distances obtained from NOE experiments were defined by broad lower ($\Delta_-$ = 0.5–1.2 Å) and upper ($\Delta_+$ = 0.8–1.2 Å) bounds since they are dependent on the mixing time value within the initial buildup region and estimation of this range is somewhat imprecise when based on one short (50-ms) and one long (250-ms) mixing time, particularly since spin diffusion contributes at the longer mixing time. A total of 188 experimentally determined distances (91 interresidue and 97 intraresidue) were included in our restrained molecular dynamics simulation. The complete tables of volume integrals, interproton distances, and error estimates are included in the Supplementary Material.

### *Starting Models*

The initial model building was carried out on an Evans and Sutherland PS390 graphics workstation driven by an Alliant FX-1 minisupercomputer using the BioGraf program (BioDesign, Inc.). Two different initial structures, classical A and B DNA 9-mers (Arnott & Hukin, 1972, 1973), were constructed with the 1,$N^2$-propanodeoxyguanosine (**1**) which was generated from the deoxyguanosine base at position 5 incorporating the propano exocyclic bridge. Guided by the results of previous NMR studies (Kouchakdjian et al., 1989), the two initial structures were built in a way such that X5 adopts a syn orientation while A14 adopts an anti orientation in the center of the X·A 9-mer. Furthermore, the N1 position



Ini-A



Ini-B

FIGURE 1: Stereoviews of the initial structures. Ini-A and Ini-B are classical A and B DNA, respectively.

in A14 was protonated as shown in (3). These two starting structures were then energy minimized using BioGraf to relieve strain in the systems such that excessive forces and accelerations are not present. The resultant structures (Figure 1) are our starting models, labeled Ini-A (for A DNA) and Ini-B (for B DNA).

*Restrained Molecular Dynamics*

All subsequent energy minimization and molecular dynamics calculations were performed on a Silicon Graphics IRIS 4D/220 GTX workstation with the X-PLOR program (Brünger et al., 1990) in which all hydrogen atoms are treated explicitly. The calculations are based on an energy function approach: the total energy of the molecule is given by the combination of an empirical and an effective energy term.

$$E_{\text{total}} = E_{\text{empirical}} + E_{\text{effective}} \tag{5}$$

The empirical function we used (Nilsson & Karplus, 1986) consists of the usual energy terms for bonds, bond angles, dihedral torsions, planar or tetrahedral geometries, explicit hydrogen bonding, and nonbonded interactions (van der Waals and electrostatic). The van der Waals energy is approximated by the Lennard-Jones potential energy function. The electrostatic potential energy term uses Coulomb function and is based on a fully charged set of partial charges (−1/residue, −2 for A14) using a dielectric constant of 80. A cutoff radius of 12.0 Å was applied for nonbonded interactions, and the nonbonded pair list was updated if any atom had moved more than 0.5 Å. Solvent molecules and counterions were not explicitly included in our calculations, but their effects are implicitly reflected by the NOE restraints which are a direct measure of the actual solution structure under the experimental conditions employed.

The effective energy function is composed of two terms representing distance and dihedral restraints ($E_{\text{dist}}$ and $E_{\text{dihe}}$). We applied square-well potential (Clore et al., 1986b) for $E_{\text{dist}}$, given by

$$\begin{aligned} E_{\text{dist}} &= k_{\text{dist}}\sum(r_{ij}-r_{ij}^{+})^2 \qquad \text{if } r_{ij} > r_{ij}^{+} \\ &= 0 \qquad \text{if } r_{ij}^{-} \le r_{ij} \le r_{ij}^{+} \\ &= k_{\text{dist}}\sum(r_{ij}-r_{ij}^{-})^2 \qquad \text{if } r_{ij} < r_{ij}^{-} \end{aligned} \tag{6}$$

where $r_{ij}^{+}$ and $r_{ij}^{-}$ are the upper and lower bounds of the distance, respectively, $r_{ij}$ is the calculated distance from NMR data, and $k_{\text{dist}}$ is a scale factor in kcal·mol$^{-1}$·Å$^{-2}$. $E_{\text{dihe}}$ is

$$\begin{aligned} E_{\text{dihe}} &= k_{\text{dihe}}\sum C(\phi_k - \phi_k^{+})^2 \qquad \text{if } \phi_k > \phi_k^{+} \\ &= 0 \qquad \text{if } \phi_k^{-} \le \phi_k \le \phi_k^{+} \\ &= k_{\text{dihe}}\sum C(\phi_k - \phi_k^{-})^2 \qquad \text{if } \phi_k < \phi_k^{-} \end{aligned} \tag{7}$$

where $\phi_k^{+}$ and $\phi_k^{-}$ are the upper and lower bounds of dihedral angles, respectively, $\phi_k$ is the calculated dihedral angle. The value of C was 20.0 kcal·mol$^{-1}$·rad$^{-2}$, and $k_{\text{dihe}}$ is an overall weight factor. Bond lengths involving hydrogen atoms were

kept fixed with SHAKE algorithm (Ryckaert et al., 1977) throughout our calculations.

*NOE-Based Back-Calculation Refinement*

Interproton distances estimated by assuming inverse proportionality to the sixth root of the NOE cross-peak intensities (eq 4) are only approximate since cross-peak intensities due to direct cross-relaxations between spins $i$ and $j$ are modified by additional cross-relaxations with any spin $k$, especially if spin $k$ is such that $r_{ik} < r_{ij}$ or $r_{jk} < r_{ij}$. NOE cross-peak intensities may be predicted, however, via a full relaxation matrix approach (Keepers & James, 1984) or by numerical integration of the coupled differential equations which determine the relaxation (Marion et al., 1987; Lefevre et al., 1987) from the structures produced by dynamics calculations, and compared directly to the observed intensities, eliminating approximate distance calculation.

Our structural refinements were carried out by using an extended version (Nilges et al., 1991) of the X-PLOR program (Brünger et al., 1990) for back-calculations. The basis of the refinement is the calculation of the volume of a cross peak between spins $i$ and $j$ from their atomic coordinates by means of the relaxation matrix (Keepers & James, 1984). The structure is then refined by minimizing directly the difference between the observed 2D NOE intensities and those calculated by the full relaxation matrix approach. To achieve this, a term ($E_{relax}$) proportional to this difference was added to the potential energy function to replace the experimental distance restraint potential, and a technique, in which derivatives with respect to atomic coordinates are calculated analytically (Yip & Case, 1989), was used during the back-calculation refinement.

The relaxation potential, $E_{relax}$, is given by

$$\begin{aligned} E_{relax} &= k_{relax}\sum w_i(I_i - k_s I_i^+)^2 \quad \text{if } I_i > k_s I_i^+ \\ &= 0 \quad \text{if } k_s I_i^- \le I_i \le k_s I_i^+ \\ &= k_{relax}\sum w_i(I_i - k_s I_i^-)^2 \quad \text{if } I_i < k_s I_i^- \end{aligned} \tag{8}$$

where $k_{relax}$ is the energy constant for the relaxation term, $I_i$ is the calculated intensity, and $I_i^+$ and $I_i^-$ are the upper and lower bounds of intensities, respectively. $w_i$ is a weight factor, and $k_s$ is a calibration factor for each spectrum and is evaluated as

$$k_s = \sum I_i / \sum I_i^o \tag{9}$$

where $I_i$ and $I_i^o$ are the calculated and observed intensities, respectively. For a peak to be included in the calculation of the calibration factor, it has to be at least three times as big as the error estimate.

To measure the fit of a refined structure to the NOE data, we use an NOE $R$ value which corresponds to the $R$ factor commonly used in crystallography (Stout & Jensen, 1989):

$$R = \sum|I_i - k_s I_i^o| / \sum k_s I_i^o \tag{10}$$

But we also report a "weighted" $R$ value (Nilges et al., 1991) which is used in X-PLOR and is given by

$$R_w = \sum(w_i \cdot \text{well}) / \sum(w_i k_s I_i^o) \tag{11}$$

where

$$\begin{aligned} \text{well} &= |I_i - k_s I_i^+| \quad \text{if } I_i > k_s I_i^+ \\ &= 0 \quad \text{if } k_s I_i^- \le I_i \le k_s I_i^+ \\ &= |I_i - k_s I_i^-| \quad \text{if } I_i < k_s I_i^- \end{aligned} \tag{12}$$

Note that no error offsets are used in the "crystallographic" $R$ value (see eq 10) and this $R$ value is defined in analogy to the way crystallographers define the agreement index $R$ for a crystal structure. Error offsets, however, are applied in the "weighted" $R$ value, which is used in X-PLOR and closely relates to the relaxation energy term (eq 8). Based on the different approaches, the numerical values of these two $R$ values should not be expected to be comparable.

*Analysis of Helical Parameters*

Analyses of helical parameters were performed with the program Dials-and-Windows (kindly provided by Dr. G. Ravishanker, Wesleyan University) in which comprehensive visualization of results and pattern recognition is greatly facilitated (Ravishanker et al., 1989). With this program, the complete time evolution of the conformational and helicoidal parameters of a DNA double helix can be depicted, and threrefore any conformational changes, such as unexpected motions of the nucleic acid backbone, in the underlying molecular dynamics are revealed. This is informative and valuable in characterizing and validating the computational protocols and methodology. The detailed analyses provide a comprehension of the consequences of various necessary assumptions made in setting up a calculation, such as choices for temperature, restraints and their weighting factors, time needed in a dynamics process, treatment of solvent and electrostatics, etc.

*Refinement Protocol*

(*I*) *Restraints.* In addition to the application of experimental interproton distance restraints in molecular dynamics (MD) and observed volume integrals in back-calculation refinement, some other restraints have been used in our calculations. Base pairs were kept Watson–Crick hydrogen bonded by distance restraints between bases. These were as follows for all base pairs. For A·T base pairs, $r_{A(N6)-T(O4)} = 2.80 \pm 0.20$ Å and $r_{A(N1)-T(N3)} = 2.95 \pm 0.10$ Å. For C·G base pairs, $r_{C(N4)-G(O6)} = 2.70 \pm 0.20$ Å, $r_{C(N3)-G(N1)} = 2.91 \pm 0.10$ Å, and $r_{C(O2)-G(N2)} = 3.01 \pm 0.20$ Å. These distance values were taken from X-ray diffraction analyses of DNA crystalline fibers (Arnott & Hukins, 1973). We chose distance bounds such that weaker restraints applied to the noncentral distances of each base pair to allow propeller twist motion. For the X·A pair, no distance restraints were applied in the MD runs, but restraints of $r_{X(N7)-A(N1)} = d_1 \pm 0.10$ Å and $r_{X(O6)-A(N6)} = d_2 \pm 0.20$ Å were used in the back-calculation refinement, where $d_1$ and $d_2$ were the distances from the structure obtained by MD.

To avoid collapse of the major groove during the high temperature dynamics simulation, we added four distance restraints between C1′ atoms on opposite sides of the major groove (the list of restraints is available in the Supplementary Material). Since these distances are 17.71–17.82 and 19.06–19.37 Å in A and B DNA, respectively, we restrained them to be greater than 16.0 Å during the calculations. For the use of distance restraints in the total energy term, scale factors of $k_{dist}$(exp), $k_{dist}$(hbond), and $k_{dist}$(grv) in kcal·mol$^{-1}$·Å$^{-2}$ have been used for experimental, hydrogen-bonded, and major groove distance restraints individually. Values of these scale factors are given in the calculational strategy.

In order to preserve the right-handed character of the DNA during our calculations, it was necessary to restrain backbone dihedral angles in an allowed region of torsional angle space. The allowed angles ($\alpha$ = 265–325°, $\beta$ = 160–220°, $\gamma$ = 27–57°, $\delta$ = 80–160°, $\epsilon$ = 145–205°, $\zeta$ = 260–320°, and $\chi$ = 200–270°) were derived from the canonical A- and B-form DNA defined by the analyses of fiber diffraction by Arnott and Hukins (1972, 1973). These dihedral restraints cover the ranges characteristic of both right-handed A and B DNA. For

the X5(*syn*)·A14(*anti*) pair, no restraint on torsion angle $\chi$ was used in order to allow different orientations of the bases. In the back-calculation refinement, the backbone restraints on the six residues (4–6, 13–15) positioned in the middle of X·A 9-mer duplex were removed to allow further reduction of the difference between experimental and calculated NOE intensities. In addition, for the purpose of avoiding serious stagger or buckle displacements which we had observed during MD, we applied four dihedral restraints on each base pair (except X·A pair) with an allowed range of ±10° in our calculations. These dihedral restraints are listed in a table which is available in the Supplementary Material. Weight factors of $k_{dihe}$(backbone) and $k_{dihe}$(plane) were used to apply the dihedral restraints of backbone and base-pair plane, respectively, in the energy calculation. Values of these weight factors are given in the calculational strategy.

*(II) Calculational Strategy.* Each of the initial structures (Ini-A and Ini-B) was subjected to the following refinement protocol with restraints of $k_{dihe}$(plane) = 30, $k_{dihe}$(backbone) = 50, and appropriate values of $k_{dist}$ and $k_{relax}$. Five hundred steps of minimization were performed, with $k_{dist}$(exp, hbond,grv) = 0.5, followed by a simulated annealing stage (restrained MD).

*(1) 7-ps High-Temperature (450 K) Dynamics.* $k_{dist}$(exp,hbond,grv) was increased from 0.5 to 500 in the first 1 ps by multiplying its value by 2 every 0.1 ps and then remained at 500 throughout the simulation. The initial velocities were assigned with a Maxwell distribution at 450 K. The temperature of the system was maintained at 450 K by rescaling the velocities of atoms every 0.2 ps. The time step of the integrator was set to 0.001 ps.

*(2) 15-ps Equilibration (300 K).* $k_{dist}$(exp,hbond,grv) was maintained at 500. The velocities of atoms were rescaled every 0.1 ps to keep the temperature at 300 K. The time step of the integrator was 0.001 ps. The coordinates of the last 5 ps were averaged and subjected to restrained energy minimization to generate the restrained dynamics structures, designated as MD-A and MD-B for those starting from Ini-A and Ini-B, respectively.

*Back-Calculation Refinement.* A uniform isotropic correlation time of 5.0 ns was used. With restraints of $k_{dist}$(exp) = 0, $k_{dist}$(hbond,grv) = 100, and appropriate $k_{relax}$ values, structures resulting from MD (MD-A and MD-B) were subjected to: (1) 25 steps of energy minimization, $k_{relax}$ = 100; (2) contiguous dynamics stages with time step = 0.5 fs, tolerance = 0.02 Å, cutoff = 4.5 Å, (a) 0.5 ps dynamics (300 K), $k_{relax}$ = 100, (b) 0.5 ps dynamics (300 K), $k_{relax}$ = 200, and (c) 1.0 ps dynamics (300 K), $k_{relax}$ = 400. The final structures R-A and R-B, refined from MD-A and MD-B, respectively, were obtained by averaging the coordinate trajectories over the last 0.5 ps followed by restrained energy minimization.

In the back-calculation refinement, tolerance denotes the maximum distance a hydrogen atom is allowed to move before the relaxation matrix and gradient are recalculated, and the cutoff value is the distance cutoff spheres around spins *i* and *j* assuming that only those spins within the spheres contribute to the cross-relaxation between spins *i* and *j* (Nilges et al., 1991). The calibration factors $k_s$ were recalculated after the initial minimization and at the end of each dynamics stage in the back-calculation refinement.

## Results and Discussion

*Structures before and after Refinements.* A survey of atomic rms differences among X·A 9-mer structures is given in Table I. The two starting structures Ini-A and Ini-B (A- and B-type DNA models, respectively) are depicted in Figure 1, and they have a rms atomic deviation of 5.18 Å relative to each other. After molecular dynamics and mechanics simulations as described under Materials and Methods, structures MD-A and MD-B are obtained. These structures are shown in the top of Figure 2, and they have a rms atomic deviation of 0.69 Å relative to each other. This is well within the rms fluctuations of the atoms about their average positions as can be ascertained from a comparison of the best fit superposition of the structures taken at 19, 20, 21, and 22 ps (i.e., the last 4 ps) of the dynamics run (center and bottom in Figure 2). This indicates that molecular dynamics has sampled sufficient conformational space and that both MD runs have converged essentially to identical structures. Further refinement of MD-A and MD-B by back-calculations, yields the final structures R-A and R-B, which are shown in Figure 3. The convergency of the resultant structures is slightly improved (from 0.69 to 0.64 Å rms) after the back-calculation refinement. A structural shift of ~1 Å was detected after both runs starting from MD-A and MD-B (see Table I). The final structures are B-type in character, having rms deviations approximately 1.9 Å from classical B-type (Arnott & Hukins, 1973) and 4.3 Å from classical A-type DNA (Arnott & Hukins, 1972).

Table I: Atomic rms Differences (Å) between X·A 9-mer Structures

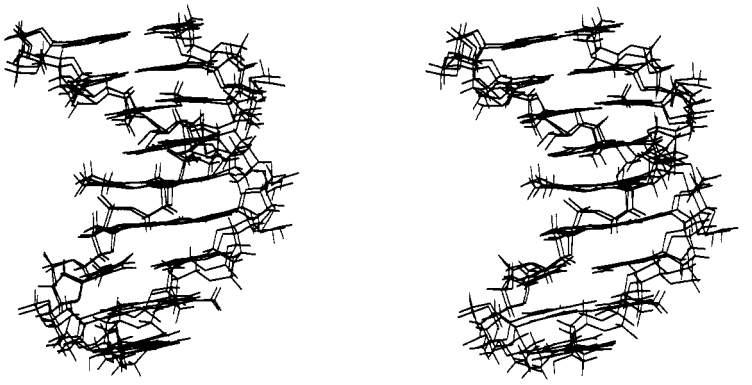
| | Ini-B | MD-A | MD-B | R-A | R-B |
|---|---|---|---|---|---|
| Ini-A | 5.18 | 3.80 | 4.16 | 4.10 | 4.41 |
| Ini-B | | 2.19 | 1.78 | 2.06 | 1.71 |
| MD-A | | | 0.69 | 1.03 | 1.25 |
| MD-B | | | | 1.11 | 1.01 |
| R-A | | | | | 0.64 |

Table II: rms Difference of Interproton Distances (Å)

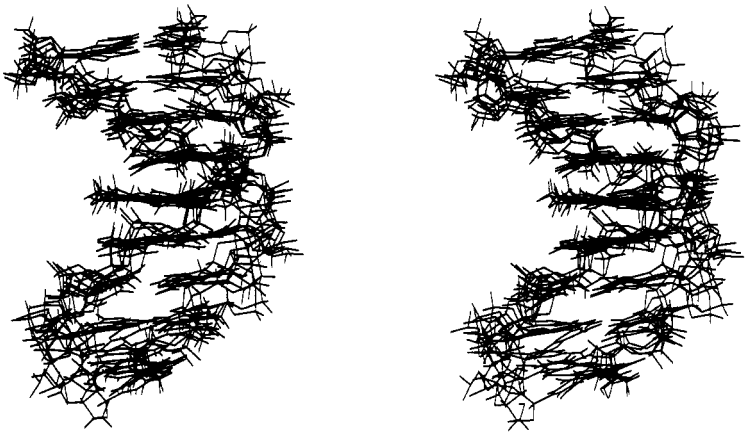
| | experimental distances (188) |
|---|---|
| Ini-A | 0.337 |
| Ini-B | 0.177 |
| MD-A | 0.001 |
| MD-B | 0.001 |

The rms differences between all atoms as a function of residue number are illustrated in Figure 4. This shows that the rms differences between R-A and R-B are only about 0.5–0.8 Å in most of the residues. These values are less than the values for any other pairs of structures in Figure 4. The rms differences between MD-A and MD-B are also quite small, only slightly larger than those in R-A vs R-B. This indicates that the convergence to essentially identical structures has been achieved starting from the initial structures Ini-A and Ini-B between which the rms differences are greater than 3 Å for all residues. The structural shifts which resulted from the back-calculation refinement are shown in MD-A vs R-A and MD-B vs R-B plots. In both cases, rms deviations of about 1 Å occur in most of the residues, and the results are quite similar for A and B. This indicates that changes made by back-calculations are essentially the same in both structures and are simply a consequence of the approximate nature of the distances data used in restrained molecular dynamics.

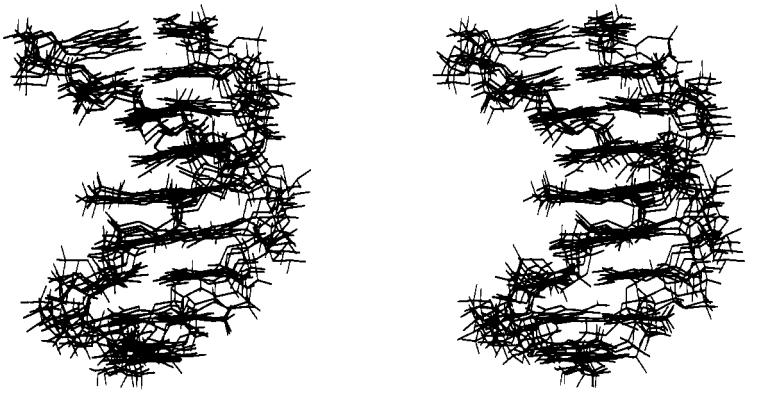
*Agreement with Experimental NMR Data.* The rms differences between the calculated and experimental interproton distances are listed in Table II. As can be seen, the initial structure Ini-B is more consistent with the experimental distance data than Ini-A. This suggests that the solution structure should be closer to B- than A-form DNA. After restrained molecular dynamics, the agreement with experimental distance data is improved significantly: both structures MD-A and MD-B satisfy the experimental distance restraints very well,



MD-A and MD-B



MD-A structures at 19, 20, 21, 22 ps



MD-B structures at 19, 20, 21, 22 ps

FIGURE 2: Stereoviews of the best fit superposition of (1) the restrained molecular dynamics structures MD-A and MD-B, (2) structures at 19, 20, 21, 22 ps of the dynamics run for MD-A, and (3) structures at 19, 20, 21, 22 ps of the dynamics run for MD-B.

with the rms deviations from experimental distances of only 0.001 Å in both cases. The comparison between the distances in our derived structures with the experimental interproton distances is illustrated in Figure 5. The experimentally derived distances' upper and lower bounds are represented by two lines, sorted in decreasing order by the upper bounds. The interproton distances in the Ini-A structure are widely distributed and display large discrepancies with distances represented by upper and lower bounds. For the structure Ini-B, in contrast, most of the distances lie well within the experimental distances. MD-A and MD-B obviously satisfy the experimental distances. However, the back-calculation causes some distance changes so that these distances are out of the distance bounds derived from experimental results. This indicates that some of the original distance restraints are incorrect, as we suspected from the approximate nature of the method. Further, we did not observe a systematic increase in distances after back-calculation refinements. This indicates that spin diffusion is not the most significant effect in this case, probably due to sufficiently large estimates of upper bounds which make up for the underestimation of distances derived from the isolated two-spin approximation.

R-A and R-B

FIGURE 3: Stereoview of the best fit superposition of back-calculation refined structures R-A and R-B.

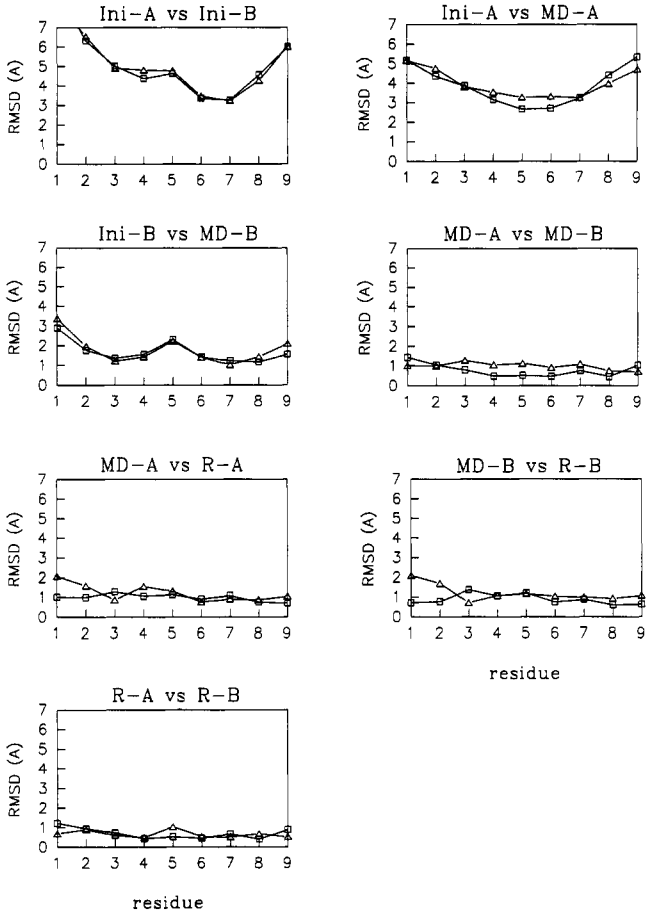



FIGURE 4: rms deviations (Å) for all atoms as a function of residue number. (□) Strand no. 1. (△) Strand no. 2 (note: residues 10–18 are shown as 1–9 of strand no. 2 in the figure).

The agreement with experimental 2D NOE intensities are represented by NOE *R* values. Table III lists the *R* values of the initial, restrained molecular dynamics, and back-calculation refined structures. These *R* values are calculated in two different ways: "X-ray" is unweighted and defined in eq 10. No error offsets are applied. It is defined analogously to the crystallographic residual index; "weighted" *R* value (for its definition, see eqs 11 and 12) is used by X-PLOR and is closely related to the target function (eq 8) used in the back-calculation refinement. Error offsets are included in the calculation of the "weighted" *R* value. Table III shows that

Table III: NOE *R* Values of the Structures

| structure | X-ray | weighted[a] |
|---|---|---|
| Ini-A | 0.544 | 0.823 |
| Ini-B | 0.358 | 0.481 |
| MD-A | 0.325 | 0.394 |
| MD-B | 0.330 | 0.408 |
| R-A | 0.167 | 0.200 |
| R-B | 0.170 | 0.203 |

[a] The weighted *R* value is defined in eqs 11 and 12.

the numerical values of each *R* value in MD-A and MD-B are similar, so are in R-A and R-B. Also, significantly large reductions of *R* values occurred after both back-calculation refinements. *R* values also decreased in restrained molecular dynamics simulations, especially starting from Ini-A. In fact, molecular dynamics simulation produced the largest reductions on *R* values from Ini-A to MD-A, but not much from Ini-B to MD-B. This again indicates that Ini-A is inconsistent with the experimental NOE data, while Ini-B is not.

Deviations between the calculated initial, restrained molecular dynamics, and back-calculation refined structures and observed NOE intensities are shown in Figure 6. The improvement from the initial to the final refined structures is evident. The divergence of the deviations is generally reduced by restrained molecular dynamics. Some of the large deviations, however, remain, and some even become larger. Back-calculation refinement further improves the structures. This step seems to be very efficient since the resultant structures (R-A and R-B) display *only* small deviations from observed intensities, including those deviations which were large after restrained dynamics.

Simulations by Borgias and James (1988) have shown that, for short DNA oligomers (such as the case in this study), the errors introduced by the assumption of a single correlation time are small. We chose to use an isotropic correlation time of 5 ns in the back-calculation refinements. This was derived from a comparison of the "weighted" *R* values of the MD structures subjected to refinements over a range of correlation times (1–9 ns). This comparison shows that the *R* values at the correlation times of 4 and 5 ns are smaller than at any other correlation times (data not shown). Thus, 5 ns was selected. Since the refined structures agree very well with the observed NOE intensities, no attempt was made to optimize this correlation time.

*Potential Energies and Deviations from Idealized Geometry.* Potential energies of the starting and resultant structures, in

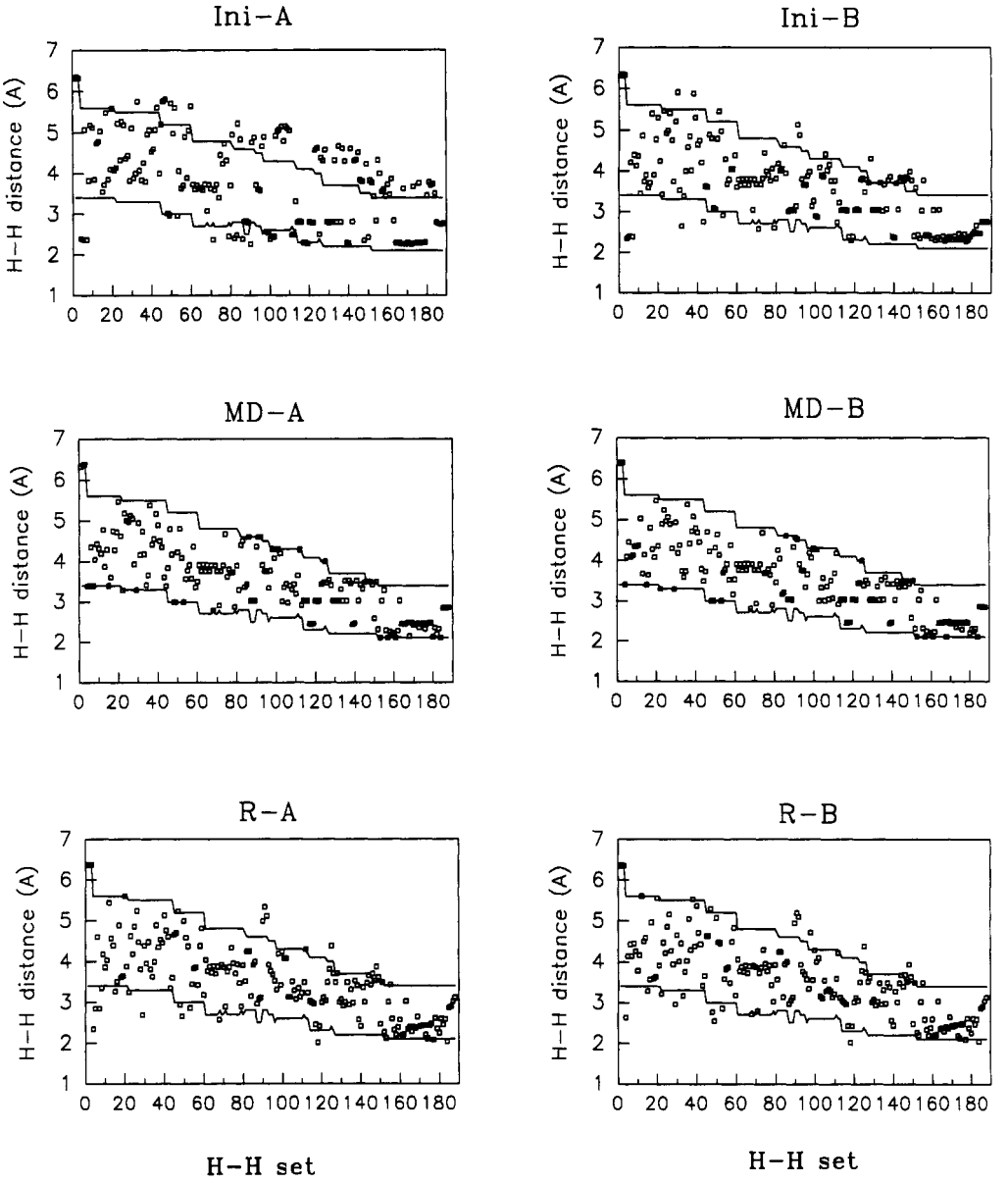



FIGURE 5: Agreement with the experimental interproton distances. The two lines in each graph define the upper and lower bounds of the distances. (□) The distance (Å) of a specific H-H set in a structure.

Table IV: Potential Energies (kcal/mol) for X·A 9-mer Structures

| structure[a] | total[b] | bond | angle | dihedral | improper | van der Waals | electrostatic | hydrogen bonding | dihedral restraint[c] | distance restraint[d] | NOE restraint[e] |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | Molecular Dynamics | | | | | | |
| Ini-$A_m$ | 10067 | 4.6 | 163.9 | 222.3 | 3.1 | −303.0 | −1.3 | −58.0 | 0.07 | 10035 | |
| Ini-$B_m$ | 1360 | 4.2 | 158.6 | 232.6 | 3.7 | −278.7 | −2.6 | −59.5 | 0.03 | 1302 | |
| MD-A | 35.6 | 4.8 | 161.7 | 215.7 | 7.6 | −297.7 | −1.9 | −54.7 | 0.02 | 0.11 | |
| MD-B | 28.8 | 4.3 | 159.9 | 216.2 | 7.0 | −299.4 | −2.0 | −57.3 | 0.02 | 0.04 | |
| | | | | | Back-Calculation Refinement | | | | | | |
| MD-A | 730.7 | 4.8 | 161.7 | 215.7 | 7.6 | −297.7 | −1.9 | −54.7 | 0.02 | 0.004 | 695.2 |
| MD-B | 774.8 | 4.3 | 159.9 | 216.2 | 7.0 | −299.4 | −2.0 | −57.3 | 0.02 | 0.001 | 746.0 |
| R-A | 209.9 | 8.4 | 197.7 | 231.2 | 11.2 | −282.4 | −2.4 | −58.4 | 0.08 | 0.07 | 104.5 |
| R-B | 211.3 | 8.0 | 196.9 | 232.9 | 12.0 | −279.9 | −2.4 | −61.8 | 0.09 | 0.02 | 105.4 |

[a] Ini-$A_m$ and Ini-$B_m$ are the energy-minimized structures of Ini-A and Ini-B, MD-A and MD-B are the restrained molecular dynamics structures, and R-A and R-B are the back-calculation refined structures. [b] The total energy includes the empirical and restraints energies. [c] The restraint scale factor is always 30 for the base-pair plane and 50 for backbone torsions. [d] The distance restraint scale factor is 500 in molecular dynamics runs and 100 in the back-calculation refinement. No experimental distance restraints are used in the back-calculation refinement. [e] The scale factor $k_{relax}$ is 400.

restrained molecular dynamics and back-calculation refinement, are listed in Table IV. The energy comparisons start from energy-minimized structures Ini-$A_m$ and Ini-$B_m$, instead of Ini-A and Ini-B, so as to exclude any possible bad contacts or excess strain and forces in the system. Ini-$A_m$ and Ini-$B_m$ are obtained by subjecting Ini-A and Ini-B to 500 steps of energy minimization (see the section on calculational strategy). This procedure results in atomic rms shifts of about 0.5 Å. As shown in the table, the values of the empirical energy terms including the bond, angle, dihedral, improper, van der Waals,

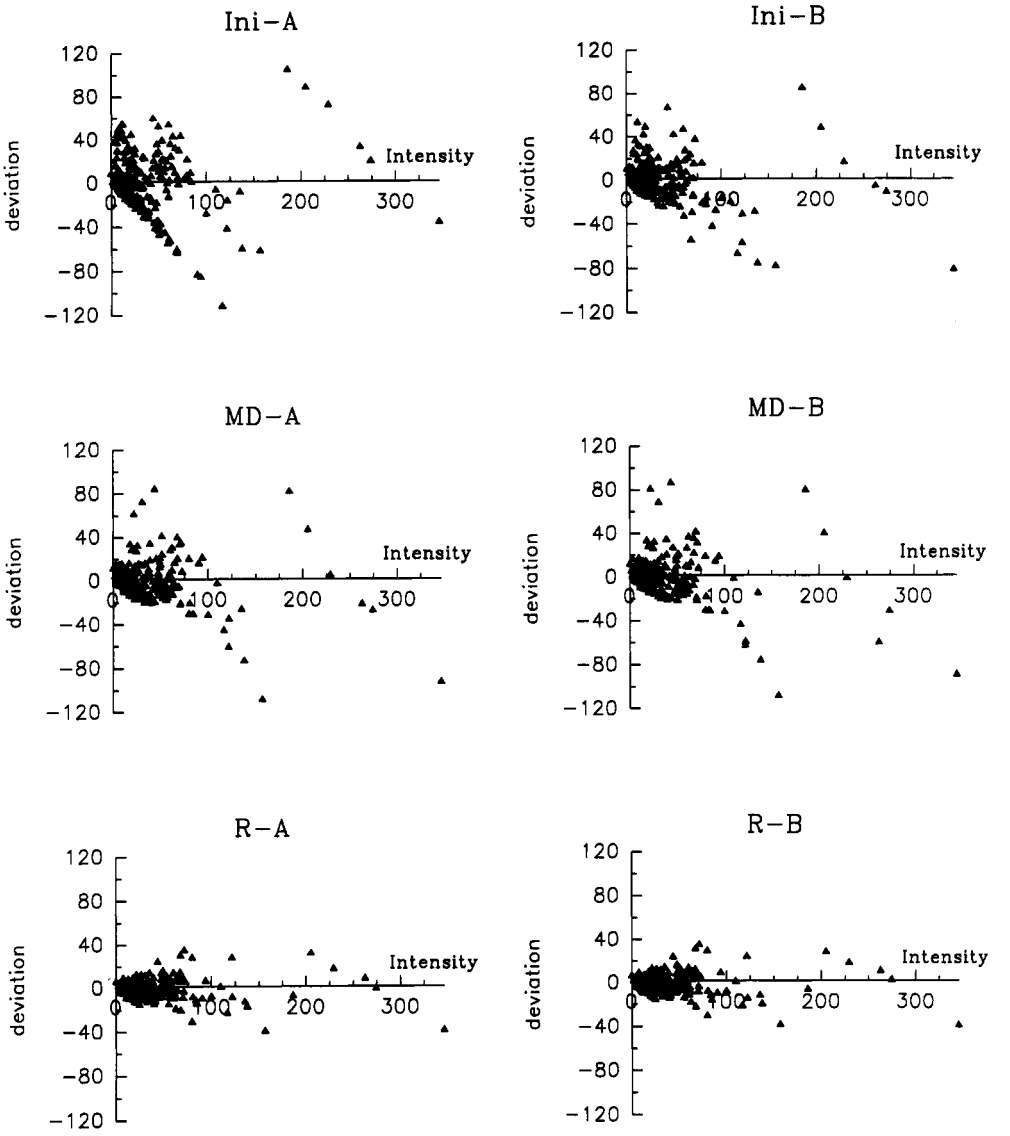



FIGURE 6: Deviations of calculated from observed NOE intensities.

electrostatic, and hydrogen-bonding energies are similar in various structures. The dihedral restraints energies are essentially unchanged. The main energy differences in these structures come from the distance restraints energy in molecular dynamics and the NOE restraints energy in back-calculation refinement. This suggests that the principal driving force for the structural changes occurring during the course of restrained molecular dynamics and the back-calculation refinement arises from the interproton distance restraints and NOE intensity restraints, respectively.

Deviations from idealized geometry are compiled in Table V. The small values for these deviations indicate that the restraining dynamics calculations have not resulted in any distortion of the covalent structure. The empirical energies maintain good covalent geometry and nonbonded contacts, thus all the structures exhibit little deviations from idealized geometry (Table V) and have good nonbonded contacts, as evidenced by the similar and large negative values of Lennard-Jones van der Waals energy (Table IV). Dihedral restraints, moreover, do *not* provide the driving force for the structural changes, since the dihedral restraints energies are small and essentially identical for all structures. The dihedral restraints are only used to preserve the right-handed character of DNA during the calculations. In addition, a comparison of the values in Tables IV and V for R-A with those for R-B confirms the

Table V: Deviations from Idealized Geometry

| structure[a] | bonds (Å) | angles (deg) | dihedrals (deg) | impropers (deg)[b] |
|---|---|---|---|---|
| Ini-$A_m$ | 0.005 | 3.448 | 31.805 | 1.991 |
| Ini-$B_m$ | 0.005 | 3.391 | 27.580 | 2.099 |
| MD-A | 0.005 | 3.422 | 26.816 | 3.545 |
| MD-B | 0.005 | 3.404 | 26.925 | 3.335 |
| R-A | 0.007 | 3.875 | 27.486 | 4.143 |
| R-B | 0.007 | 3.872 | 27.589 | 4.326 |

[a] Ini-$A_m$ and Ini-$B_m$ are the energy-minimized structures of Ini-A and Ini-B, MD-A and MD-B are the restrained molecular dynamics structures, and R-A and R-B are the back-calculation refined structures. [b] The improper torsion terms serve to maintain planarity and chirality.

convergence of the refined structures. The same is true for MD-A and MD-B.

*Structural Features.* As suggested by the best fit superpositions, and the similar energies and *R* values, the convergence of the refined structures can also be seen by examining the details of the helical parameters. Comparisons of the helical parameters are shown in Figures 7, 8, and 9. In these figures, each of the conformational parameters is presented in the form of a dial and each of the helicoidal parameters is drawn on a window. These dials and windows can be compared directly with those for the canonical forms, shown at

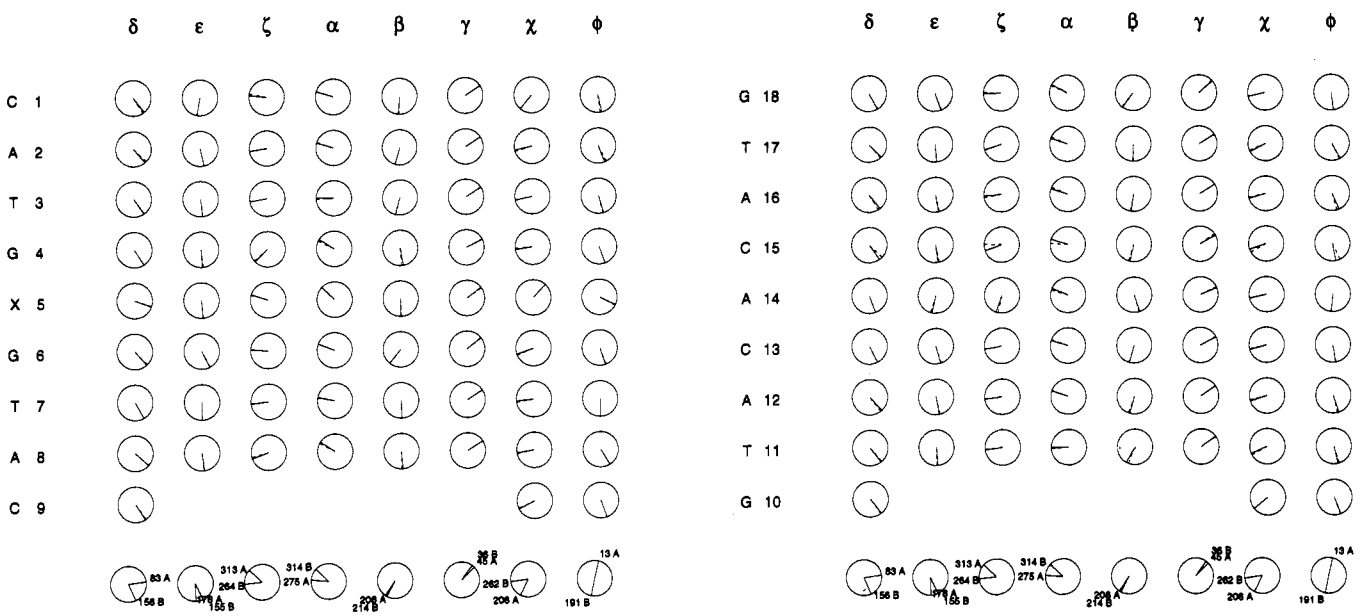



FIGURE 7: Conformational dials for analysis of the refined structures R-A and R-B. (---) R-A; (—) R-B. The dials for canonical A-form (Arnott & Hukins, 1972) and B-form (Arnott & Hukins, 1973) DNA are shown at the bottom with the actual values.

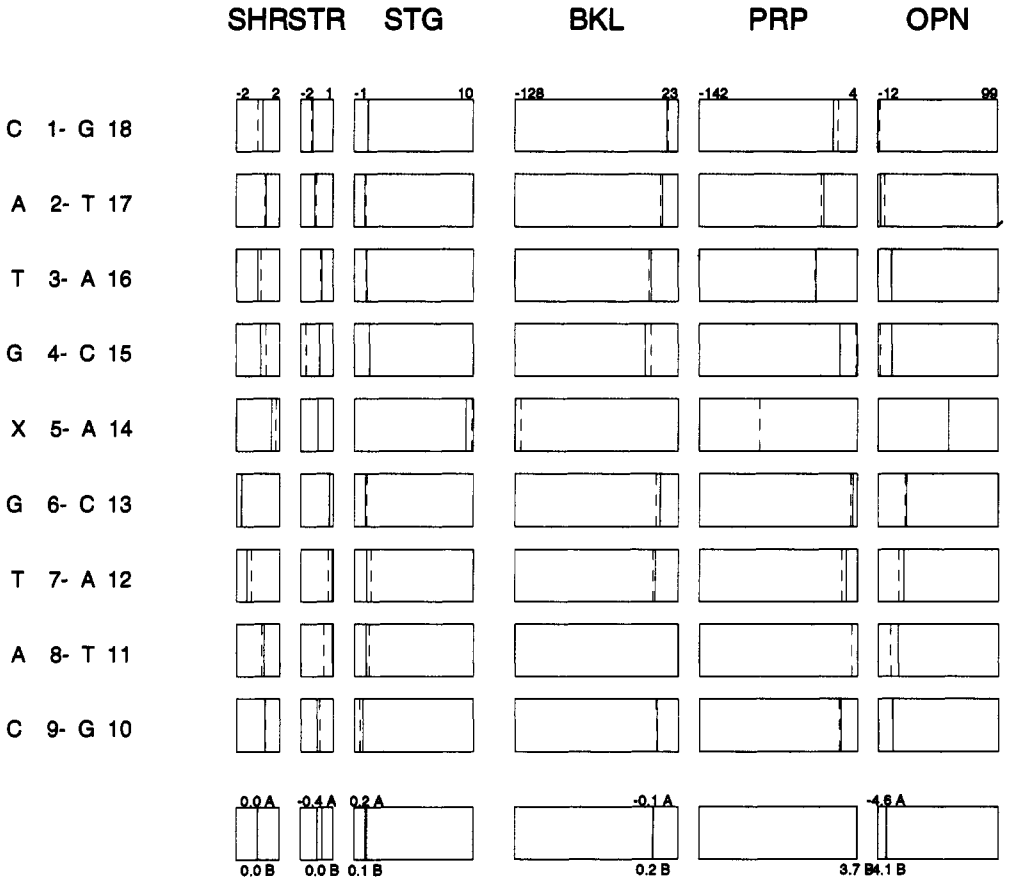



FIGURE 8: Helicoidal windows for analysis of *intra*-base pair parameters of the refined structures R-A and R-B. (---) R-A; (—) R-B. The windows for canonical A-form (Arnott & Hukins, 1972) and B-form (Arnott & Hukins, 1973) DNA are shown at the bottom with the actual values.

the bottom of each figure. The canonical forms considered here are from Arnott and co-workers for A-form (1972) and B-form (1973). Conventions adopted for the presentation of the dials and windows are fully described in the article by Ravishanker et al. (1989).

Figure 7 shows the conformational dials for analysis of the torsion angles in the refined structures R-A and R-B. In this figure, the 360° conformation wheels ("dials") for each of the backbone dihedral angles, $\alpha$, $\beta$, $\gamma$, $\delta$, $\epsilon$, and $\zeta$, the glycosidic bond sugar–base torsion $\chi$, and the sugar pucker $\phi$, are presented with reference to successive residues and with chain direction of top to bottom for 5′ to 3′ ends. The termini of the chain may have incomplete sets of dials, depending on which atoms are present. The vertical position, corresponding to "north" on a compass wheel, is taken as 0° in the figure. It is clear from the figure that for all residues of the refined structures (R-A and R-B) the backbone torsion angles lie in the conformational range expected for canonical B-form DNA.

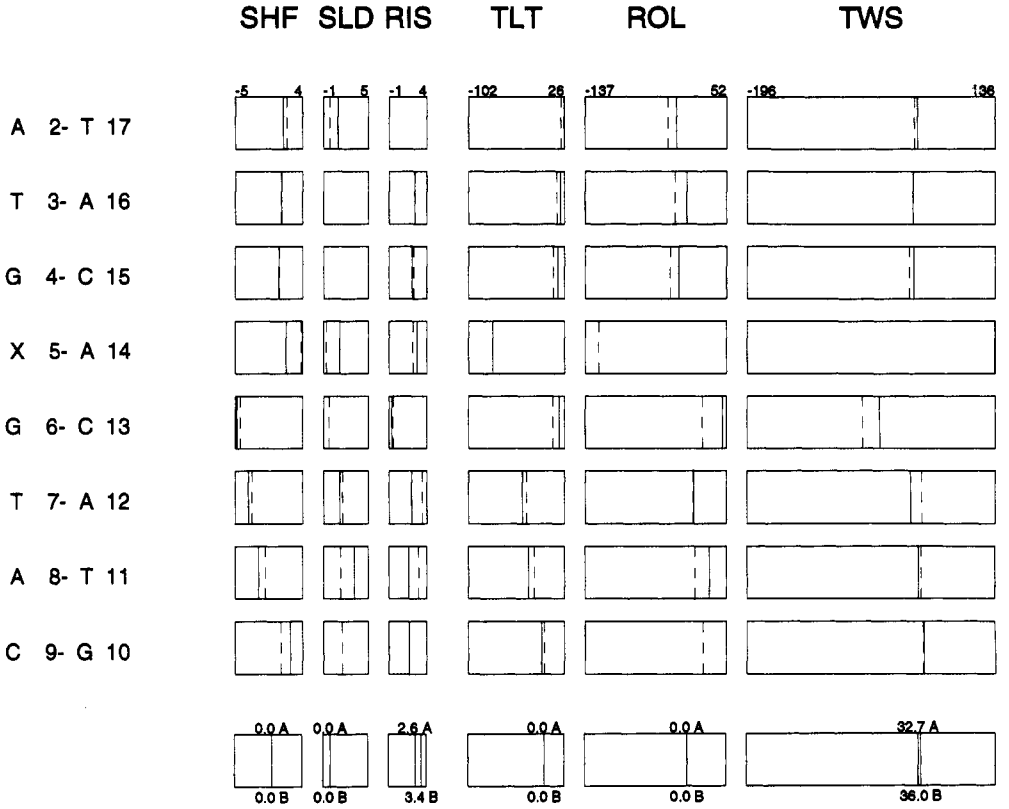



FIGURE 9: Helicoidal windows for analysis of *inter*-base pair parameters of the refined structures R-A and R-B. (---) R-A; (—) R-B. The windows for canonical A-form (Arnott & Hukins, 1972) and B-form (Arnott & Hukins, 1973) DNA are shown at the bottom with the actual values.

Another feature of interest in the figure is the excellent agreement in the values of torsion angles for R-A and R-B. These results demonstrate the convergence of conformational features in the refined structures to the B-type character. In fact, individual torsion angles of various residues align in general to the canonical B value, except in X5 which adopts a syn orientation and thus has a different $\chi$ value. All sugars in R-A and R-B are puckered in $C_2'$-*endo* with the $\phi$ values of 144–190°. Some small perturbations in and near X5 are also observed in the figure. These include the smaller $\delta$ and $\phi$ values in X5, larger $\beta$ in G6, and smaller $\beta$ and $\zeta$ in A14. These perturbations occur nearby X5 and are caused by the modified deoxyguanosine. In fact, the perturbed torsion angle $\zeta$ (C3′–O3′–P–O5′) for C13pA14 is consistent with the $^{31}P$ NMR results which indicated a phosphorus resonance downfield shift of the C13pA14 phosphodiester linkage at the lesion site (Kouchakdjian et al., 1989).

The *intra*- and *inter*-base pair parameters of the helicoidal set are depicted in Figures 8 and 9, respectively, where the parameters referenced to successive base pairs are laid out in the graphics as "windows", defined on a suitable range of values for each parameter. Precise definitions for these parameters are given by Ravishanker et al. (1989). In Figures 8 and 9, the displayed results for the X5·A14 pair can be simply ignored since X5 is in syn orientation, while the calculation of the helicoidal parameters is based on anti orientation.

The *intra*-base pair parameters, which describe the relative position of bases in a base pair, are shear (SHR), stretch (STR), stagger (STG), buckle (BKL), propeller twist (PRP), and opening (OPN). Here SHR and STR are linear in-plane displacements, and STG is out-of-plane linear displacement. All of these displacements are zero or very near zero for both canonical A- and B-form DNA. As shown in Figure 8, these parameters for the refined structures R-A and R-B are also very near zero (ignore the X5·A14 pair). BKL is an angular intra-base pair parameter, and its values in R-A and R-B seem small and similar. The propeller twist (PRP) is 13.7° in canonical A DNA and 3.7° in canonical B DNA (the PRP value for canonical A DNA is out of the region shown in the figure). As seen in the figure, the propeller twists in R-A and R-B are all similar and close to that for canonical B-form DNA. The base pair opening parameter (OPN) is −4.6° and −4.1° for canonical A and B DNA, respectively, with the negative value indicative of an opening toward the minor groove. R-A and R-B have like OPN values too.

The *inter*-base pair parameters, which describe the relative position of successive base pairs, are shift (SHF), slide (SLD), rise (RIS), tilt (TLT), roll (ROL) and twist (TWS). Here both SHF and SLD describe linear shearing displacements. A positive SHF corresponds to displacement of the "lower" base pair in the direction of the minor groove. SHF and SLD are both zero for canonical A and B DNA, and near zero in the refined structures R-A and R-B. Rise (RIS) is the vertical displacement of one base pair with respect to another. Its values for R-A and R-B are near those for canonical A and B DNA. The parameters TLT and ROL are both zero for canonical A and B DNA, and near zero (ignore the X5·A14 pair) throughout the structures R-A and R-B. The small displacements and small tilt and roll angles indicate that local twist axes are nearly the same and are coincident with the global twist axis. The winding angles (TWS) are similar in canonical A and B DNA and likewise in refined structures R-A and R-B. One exception is for the G6·C13 pair which can be ignored because this TWS is related to the X5·A14 pair which should not be under consideration.

*The Transition of A → B DNA.* The time-dependent structural transformation of A- to B-type DNA during dynamics has been of particular interest in our investigation. Analysis of the conformational parameters was carried out

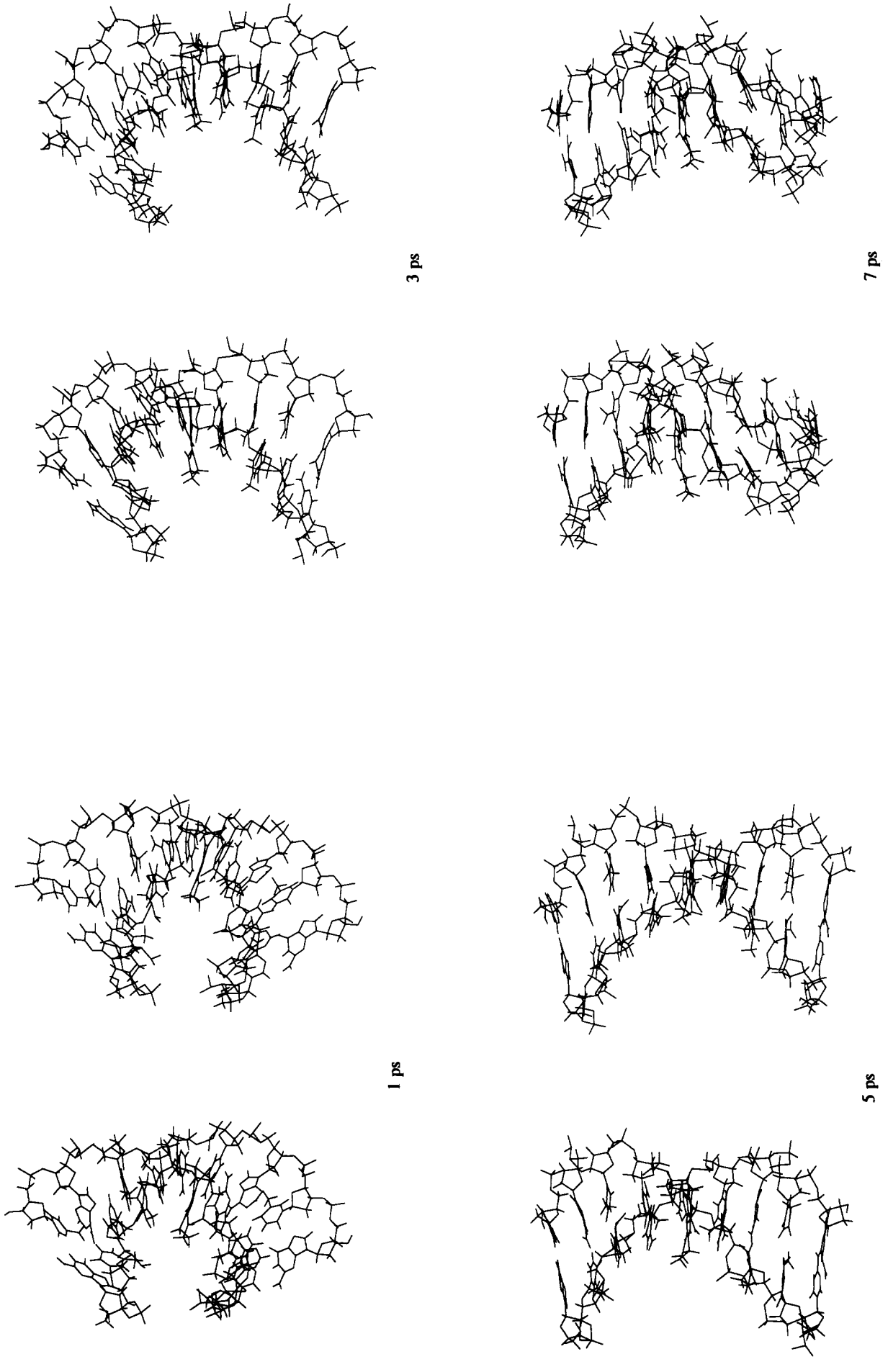



FIGURE 10: Snapshots of the trajectory of the restrained dynamics run showing the conversion of A DNA to the B-type structure.

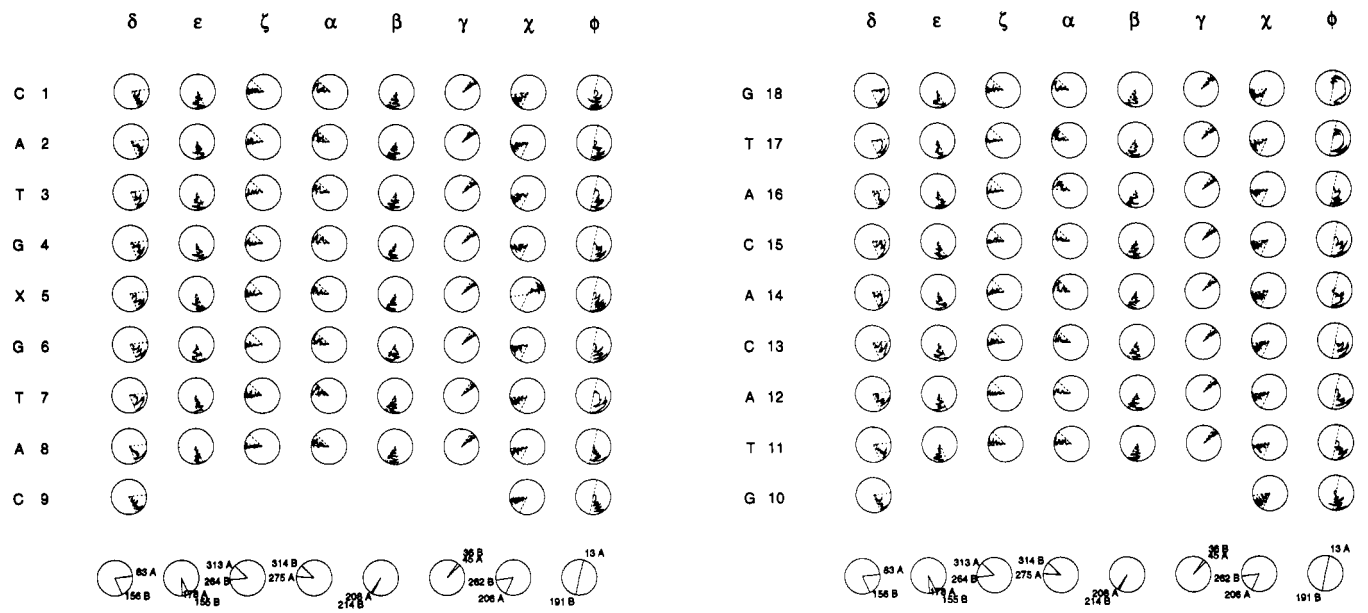



FIGURE 11: Conformational dials for analysis of the dynamics behavior of the transition from A- to B-type DNA in the 7-ps high-temperature (450 K) stage. The radial coordinate is the time axis, with $t = 0$ ps at the center and $t = 7$ ps at the circumference. The torsion angles for canonical A- and B-type DNA are also given in each dial to facilitate cross-reference.

with the program Dials-and-Windows for illustrative purposes over the full molecular dynamics simulation, beginning with the A-minimized structure and including the high-temperature (450 K) and equilibration at 300 K stages. A complete record of the time evolution of structural conformation during the course of the calculation is produced in the analysis. We observed that the transition from A- to B-type DNA is essentially complete during the high-temperature stage. This has been confirmed by the fact that the structure, captured at the end of the high-temperature dynamics stage starting from A-DNA, is found to be B-type in character. In fact, during the 300 K equilibration stage, no obvious change is observed. The conformational parameters only fluctuate randomly and remain in the B-type DNA region. The transition of A- to B-form DNA was significantly detected by the changes of sugar pucker ($\phi$) in nucleosides, which has values of 0–36° for $C_3'$-*endo* and 144–190° for $C_2'$-*endo*. We also noticed that transitions of A- to B-type DNA were incomplete for dynamics simulations at lower temperatures or shorter times (data not shown). This suggests that the high-temperature dynamics (450 K) is the conformation-determining stage in our calculational strategy. In principle, dynamics simulations at high temperature allow the structure to cross energy barriers.

The time course of the structural change from A- to B-type DNA structure is depicted in Figure 10 by a series of snapshots from the restrained molecular dynamics starting with A-DNA. At about 5 ps, the global B-DNA character of the structure becomes clear. To investigate this A → B DNA transition, the dynamics behavior of conformational parameters in the 7-ps high-temperature (450 K) dynamics was monitored as shown in Figure 11. In this figure, the radius of the conformational dials is the 7-ps time coordinate of the dynamics simulation, beginning from the center. Each dial, therefore, contains a record of the dynamics trajectory of the corresponding structural parameter in the course of the simulation. The dotted lines in each of the dials represent the values for canonical A- and B-form DNA, and they are given for reference. Dials for the canonical A- and B-form DNA are also provided at the bottom of the figure. It can be seen that distinct A → B transitions occur in the $\phi$ and $\delta$ parameters. Note that most of these transitions occur in the early stages of the dynamics and remain in the region of the B-type DNA throughout the simulation. But two late A → B transitions in both $\phi$ and $\delta$ parameters occur at T17 and G18. It is clear from the figure that by the end of the high-temperature stage all parameters (except $\chi$ for X5) are in the region of B-type DNA, and thus the overall A → B transition is finished. The only exception is X5, which adopts the syn orientation, and its $\chi$ value thus does not lie in the canonical B region.

In contrast with the dynamics behavior of A DNA, molecular dynamics starting from B DNA does not show any significant transitions. All parameters are stable and remain in the B-type DNA region during the high-temperature dynamics simulation (the figure is included in the Supplemental Material). A critical question is, What are the major factors which are responsible for the different dynamics behaviors between A and B DNA? In order to answer this, we performed a test by carrying out a molecular dynamics simulation starting from A-DNA (Ini-A) in the absence of experimental interproton distance restraints, while other restraints and the calculation strategy remained unchanged. The result of this dynamics calculation is shown in Figure 12. It is obvious that for all residues the conformational parameters exhibit stable and oscillatory behavior in the region of the canonical A-form DNA. The fact that no A → B DNA transition occurs in the absence of experimental interproton distance restraints demonstrates that *the driving force of the A → B DNA transition, consequently the convergence of refined structures, is the experimental interproton distances data itself.* The convergence to identical structures starting from A and B DNA is due to restrained molecular dynamics as a tool to overcome local energy barriers, guided by distance restraints, to the energy minimum region. This is consistent with the conclusions derived from energy analysis.

*Hydrogen Bonds in the X5(syn)·A14(anti) Pair.* The X5(syn)·A14(anti) pair in the average of the final structures R-A and R-B is shown in Figure 13. In addition to two Hoogsteen-type hydrogen bonds between O6 of X5 and N6H of A14 and between N7 of X5 and N1H of A14, an intrastrand hydrogen bond (N–H···O) between the imino group formed at N2 by the exocyclic adduction in X5 and the oxygen (O2P)

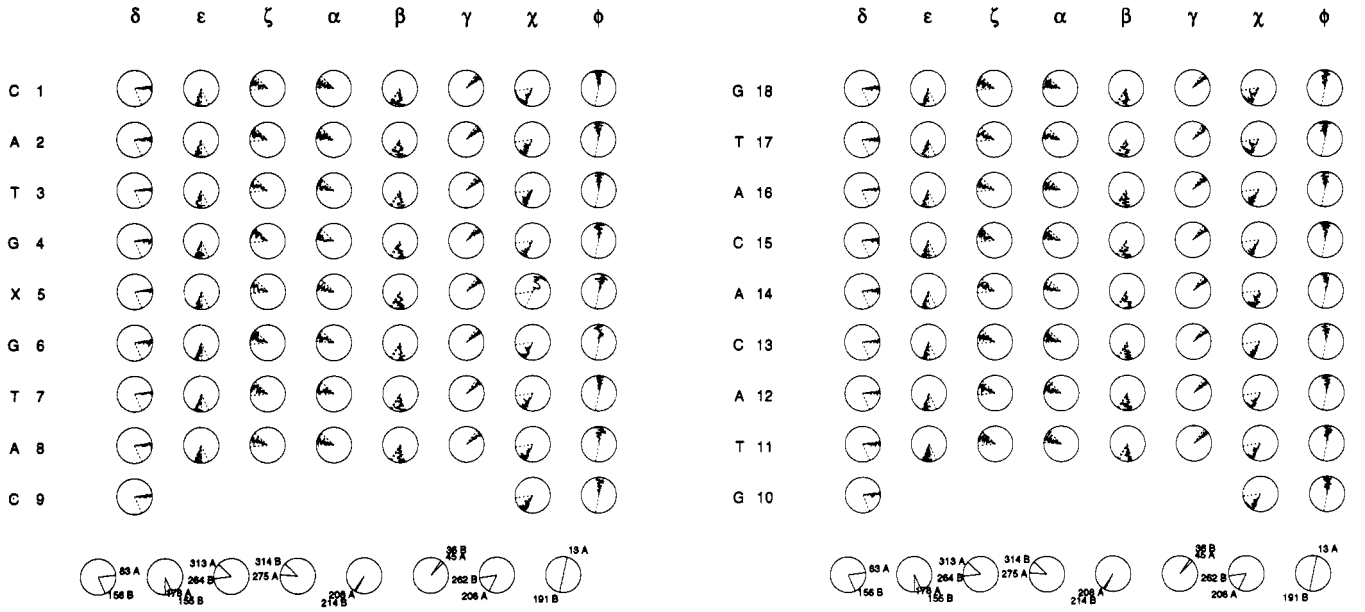



FIGURE 12: Conformational dials for analysis of the dynamics starting from A DNA in the absence of experimental interproton distance restraints. The radial coordinate is the time axis, with $t = 0$ ps at the center and $t = 7$ ps at the circumference. The torsion angles for canonical A- and B-type DNA are also given in each dial to facilitate cross-reference.

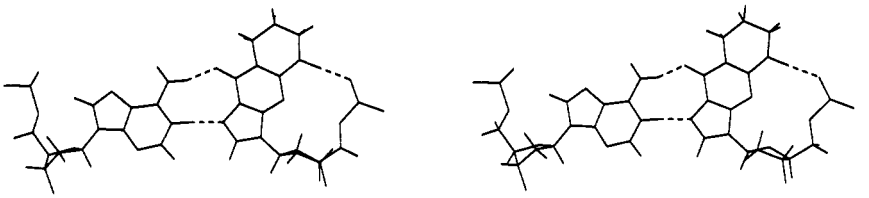


FIGURE 13: X5(syn)·A14(anti) pair in the average structure of R-A and R-B. Three hydrogen bonds are indicated by dotted lines. The distances are $d_{X5(O6)-A14(N6)} = 2.78$ Å, $d_{X5(N7)-A14(N1)} = 2.87$ Å, and $d_{X5(N2)-X5(O2P)} = 3.03$ Å.

in its own backbone phosphate is also predicted because the distance between N2 and O2P is 3.0 Å. This hydrogen bonding is, however, expected to be weaker than the other two because of the partial $\pi$ character in the P–O bond. As a result, X5 is in syn orientation through formation of three hydrogen bonds in the X5(syn)·A14(anti) pair at pH 5.8.

## Conclusions

In this paper, we have used restrained molecular dynamics guided by experimental interproton distances, followed by a NOE-based back-calculation refinement which circumvents the inaccuracies in distances, to study the three-dimensional structure of the X·A 9-mer duplex in solution at pH 5.8. Separate simulations start from classical A- and B-form DNA and converge to essentially identical structures. This indicates that the calculations have sampled sufficient conformational space, and the converged refined structures represent a reasonable approximation of the actual structures in solution. The observation of the time dependence of conformational parameters during dynamics simulations has demonstrated that the structural features that emerge are entirely a result of the incorporation of the NOE experimental data in the calculations. The final refined structures are B-type DNA and satisfy the NOE intensity restraints very well, with the "crystallographic" $R$ values of only 0.17.

It seems safe to use distance restraints derived from isolated two-spin approximation, even with relatively large bounds, for initial structure calculation, provided that the initial calculations are followed by a back-calculation refinement which directly improves the agreement of calculated to observed NOE intensities. The fact that in our case convergence to a unique structural set that satisfies the experimentally observed NOE intensities occurs gives us confidence that a realistic picture of the solution structure has been obtained.

Fully validated protocols for nucleic acids structural analyses have not yet been unequivocally established. One of our motivations for carrying out these calculations was to search for a workable strategy and thus provide a means to study the three-dimensional structures of nucleic acids in general. We are currently applying these findings to elucidate the dynamics structure of the X·A 9-mer at high pH and the nature of the syn–anti transition at X5.

## Acknowledgments

We thank Dr. D. J. Patel for providing the NMR experimental data which were used in our calculations, Professor Axel T. Brünger for the X-PLOR program, Dr. M. Nilges for helpful discussions, and Professor G. Ravishanker for the helix analysis program Dials-and-Windows.

## Supplementary Material Available

Tables of volume integrals for X·A 9-mer at the mixing times of 50 and 250 ms, interproton distances and error estimates, lists of major groove distance restraints and dihedral restraints for base-pair planes, and a figure of conformational dials for the dynamics behavior starting from B DNA (13 pages). Ordering information is given on any current masthead page.

## References

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

Arnott, S., & Hukins, D. W. (1973) *J. Mol. Biol. 81*, 93–105.

Baleja, J. D., Pon, R. T., & Sykes, B. D. (1990) *Biochemistry 29*, 4828-4839.
Borgias, B. A., & James, T. L. (1988) *J. Magn. Reson. 79*, 493-512.
Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983) *J. Comput. Chem. 4*, 187-217.
Brünger, A. T. (1990) X-PLOR *version 2.1*, user manual, Yale University, New Haven, CT.
Brünger, A. T., Clore, G. M., Gronenborn, A. M., & Karplus, M. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3801-3805.
Clore, G. M., Brünger, A. T., Karplus, M., & Gronenborn, A. M. (1986a) *J. Mol. Biol. 191*, 523-551.
Clore, G. M., Nilges, M., Sukumaran, D. K., Brünger, A. T., Karplus, M., & Gronenborn, A. M. (1986b) *EMBO J. 5*, 2729-2735.
Grollman, A. P. (1989) *Proc. Am. Assoc. Cancer Res. 30*, 682.
Gronenborn, A. M., & Clore, G. M. (1989) *Biochemistry 28*, 5978-5984.
Kalk, A., & Berendsen, J. C. (1976) *J. Magn. Reson. 24*, 343-366.
Keepers, J. W., & James, T. L. (1984) *J. Magn. Reson. 57*, 404-426.
Kouchakdjian, M., Marinelli, E., Gao, X., Johnson, F., Grollman, A., & Patel, D. J. (1989) *Biochemistry 28*, 5647-5657.
Kouchakdjian, M., Eisenberg, M., Live, D., Marinelli, E., Grollman, A., & Patel, D. J. (1990) *Biochemistry 29*, 4456-4465.
Lefevre, J. F., Lane, A. N., & Jardetzky, O. (1987) *Biochemistry 26*, 5076-5090.
Marion, D., Genest, M., & Ptak, M. (1987) *Biophys. Chem. 28*, 235-244.
Nilges, M., Clore, G. M., & Gronenborn, A. M. (1987a) *Biochemistry 26*, 3718-3733.
Nilges, M., Clore, G. M., & Gronenborn, A. M. (1987b) *Biochemistry 26*, 3734-3744.
Nilges, M., Habazettl, J., Brünger, A. T., & Holak, T. A. (1991) *J. Mol. Biol. 219*, 499-510.
Nilsson, L., & Karplus, M. (1986) *J. Comput. Chem. 7*, 591-616.
Nilsson, L., Clore, G. M., Gronenborn, A. M., Brünger, A. T., & Karplus, M. (1986) *J. Mol. Biol. 188*, 455-476.
Powell, M. J. D. (1977) *Math. Program. 12*, 241-254.
Ravishanker, R., Swaminathan, S., Beveridge, R. L., & Sklenar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 669-699.
Ryckaert, J. P., Cicotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327-337.
Singer, B., & Grunberger, D. (1983) in *Molecular Biology of Mutagens and Carcinogenesis*, Plenum Press, New York.
Singer, B., & Bartsch, H. (1986) *The Role of Cyclic Nucleic Acid Adducts in Carcinogenesis and Mutagenesis*, IARC Scientific Publications 70, International Agency for Research on Cancer, Lyon, France.
Stout, G. H., & Jensen, L. H. (1989) *X-ray Structure Determination*, Wiley, New York.
Verlet, L. (1967) *Phys. Rev. 159*, 98-105.
Yip, P., & Case, D. A. (1989) *J. Magn. Reson. 83*, 643-648.

# Secondary Structure of Streptokinase in Aqueous Solution: A Fourier Transform Infrared Spectroscopic Study

Heinz Fabian,*,‡ Dieter Naumann,§ Rolf Misselwitz,‡ Otto Ristau,‡ Dieter Gerlach,‖ and Heinz Welfle‡

*Max Delbrück Center for Molecular Medicine, Robert-Rössle-Strasse 10, D-1115 Berlin-Buch, FRG, Robert Koch Institute of the Federal Health Office of Germany, Nordufer 20, D-1000 Berlin 65, FRG, and Institute of Experimental Microbiology, Beutenbergstrasse 11, D-6900 Jena, FRG*

*Received January 10, 1992; Revised Manuscript Received April 6, 1992*

ABSTRACT: The secondary structure of streptokinase (Sk) in aqueous solution was quantitatively examined by using Fourier transform infrared (FT-IR) spectroscopy. Resolution enhancement techniques, including Fourier deconvolution and derivative spectroscopy, were combined with band curve-fitting procedures to quantitate the spectral information from the amide I bands. Nine component bands were found under the broad, nearly featureless amide I bands which reflect the presence of various substructures. The relative areas of these component bands indicate an amount of $\beta$-sheet between 30 and 37% and an $\alpha$-helix content of only 12-13% in Sk. Further conformational substructures are assigned to turns (25-26%) and to "random" structures (15-16%). Additionally, the correlation of a pronounced component band near 1640 $cm^{-1}$ (10-16% fractional area) with the possible presence of $3_{10}$-helices is discussed.

The bacterial protein streptokinase (Sk)[1] can activate human plasminogen, and it is in clinical use as a thrombolytic agent (Collen & Gold, 1990). For understanding its structure–function relationships and to improve further its therapeutic properties by techniques of genetic engineering, knowledge on the secondary structure of Sk in solution is desirable.

Sk has a molar mass of about 47 000 $g \cdot mol^{-1}$, and its primary structure was estimated by protein sequencing (Jackson & Tang, 1982) and from the nucleotide sequence of the cloned Sk gene (Malke et al., 1985). From CD studies, a content of 17% $\alpha$-helices, 28% $\beta$-sheet, 21% turns, and 34% disordered structures was derived, in agreement with results obtained by secondary structure prediction methods (Radek & Castellino,

* Author to whom correspondence should be addressed.
‡ Max Delbrück Center for Molecular Medicine.
§ Robert Koch Institute of the Federal Health Office of Germany.
‖ Institute of Experimental Microbiology.

[1] Abbreviations: CD, circular dichroism; FT-IR, Fourier transform infrared; IR, infrared; Sk, streptokinase.

0006-2960/92/0431-6532$03.00/0 