Solution Structure of a Conserved DNA Sequence from the HIV-1 Genome: Restrained Molecular Dynamics Simulation with Distance and Torsion Angle Restraints Derived from Two-Dimensional NMR Spectra[†,‡]

Anwer Mujeeb, Sean M. Kerwin,[§] George L. Kenyon, and Thomas L. James*

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

Received June 4, 1993; Revised Manuscript Received September 30, 1993[⊗]

ABSTRACT: The three-dimensional solution structure of a trisdecamer DNA duplex sequence, d(AGCT-TGCCTTGAG)·d(CTCAAGGCAAGCT), from a conserved region of HIV-1 genome's long terminal repeat, has been investigated using NMR spectroscopy and restrained molecular dynamics calculations. Interproton distances derived from two-dimensional nuclear Overhauser enhancement (2D NOE) experiments, using the iterative complete relaxation matrix algorithm MARDIGRAS, and torsion angles for sugar rings, estimated from simulated fitting of double-quantum-filtered correlation (2QF-COSY) spectra, were obtained [Mujeeb, A., Kerwin, S. M., Egan, W. M., Kenyon, G. L., & James, T. L. (1992) *Biochemistry 31,* 9325–38]. These structural restraints have now been employed as the basis for structure refinement using restrained molecular dynamics (rMD) to search conformational space for structures consistent with the experimental restraints. Specifically, upper and lower bounds on the restraints were incorporated into the AMBER (version 4.0) total potential energy function of the system, the bounds being used to define the width of a flat-well penalty term in the AMBER force field. Confidence in the time-averaged structure obtained is engendered by convergence to essentially the same structure (root-mean-square deviation ~0.9 Å) when two quite different DNA models, A-DNA and B-DNA (RMSD ~6.5 Å), were employed as starting structures and when various initial trajectories were used for the rMD runs. The derived structure is further supported by the total energy calculated, the restraint violation energy, the restraint deviations, and the fit with experimental data. For the latter, the sixth-root residual index indicated a good fit of the determined structure with experimental 2D NOE spectral intensities ($R_1^x < 0.07$), and the RMS difference between vicinal proton coupling constants calculated for the derived structure and experimental coupling constants were also in reasonable agreement (J_{RMS} = 0.9 Hz). While the structure of the trisdecamer is basically in the B-DNA family, some structural parameters manifest interesting local variations. The helix parameters indicate that, compared with classical B-DNA, the structure is longitudinally more compressed. Local structural variations at the two TG steps in particular together create bending into the major groove of the duplex. Comparison of the two –CTTG– tetrads in the duplex reveals that they have similar structures, with the TT moieties being almost identical; however, the –CTTG–pur sequence has a larger roll and slide for the –TG– step than for the –CTTG–pyr sequence, in accord with published X-ray crystallographic conclusions.

We recently conducted a successful search for a highly conserved region in the genome of a human immunodeficiency virus (HIV-1)[1] that could lead to a potential target for attack (Mujeeb *et al.*, 1992). A 21-nucleotide sequence, immediately downstream of TAR in the long terminal repeat, was found to be essentially 100% conserved in the 148 HIV-1 sequences searched. A 13-bp double-helical segment of this conserved sequence was selected and synthesized, and structural restraints were determined via NMR (Mujeeb *et al.*, 1992). The present paper describes our determination of the three-dimensional structure of the duplex d(AGCTTGCCTTGAG)·d(CTCA-AGGCAAGCT) using restrained molecular dynamics (rMD) calculations.

In the past decade, understanding of the molecular basis of physiological processes has been considerably enhanced following studies which have deduced the three-dimensional solution structures of biomolecules. NMR, in conjunction with appropriate computational algorithms, has become the method of choice for determination of the high-resolution structure of proteins, nucleic acids, and complexes. Multi-dimensional NMR has the capability of yieldinig internuclear distances and bond torsion angles as experimental structural restraints (James & Basus, 1991; Oppenheimer & James, 1989; Wagner *et al.*, 1992). In the case of duplex DNA, we are interested in fairly subtle sequence-dependent structural details which can influence recognition and stability. These subtle variations require that the structure be determined to high resolution. A higher resolution structure can be achieved

[†] This work was supported by National Institutes of Health Grants GM39247 and GM41639. This research was also supported in part by Pittsburgh Supercomputing Center Grant 1 P41 RR06009 from the NIH National Center for Research Resources.

[‡] Coordinates of the final restrained molecular dynamics structures and a list of NOE and torsion angle restraints have been deposited in the Protein Data Bank, Brookhaven National Laboratory, Upton, NY 11973, under accession number 142D.

* Author to whom correspondence should be addressed (telephone, (415) 476-1569; Fax, (415) 476-0688).

[§] Present address: Division of Medicinal Chemistry, The University of Texas, Austin, TX 78712-1074.

[⊗] Abstract published in *Advance ACS Abstracts,* November 15, 1993.

[1] Abbreviations: HIV, human immunodeficiency virus; LTR, long terminal repeat; TAR, transactivating response element; bp, base pairs; 1D NMR, one-dimensional NMR; 2D NMR, two-dimensional NMR; 2D NOE, two-dimensional nuclear Overhauser effect; COSY, correlation spectroscopy; 2QF-COSY, double-quantum-filtered COSY; rMD, restrained molecular dynamics; r-MIN, restrained energy minimization; E_{restr}, restraint violation energy; Δd_{av}, average distance deviation; $\Delta\nu_{av}$, average torsion angle deviation; RMSD, root-mean-square deviation.

0006-2960/93/0432-13419$04.00/0

with more structural restraints and with more accurate structural restraints. Bond torsion angles can be determined using various correlation spectroscopic techniques, *e.g.*, E.COSY, PCOSY, and double-quantum-filtered COSY (2QF-COSY) (Griesinger *et al.*, 1985; Marion & Bax, 1988; Piantini *et al.*, 1982). Broad lines have prevented direct analysis of all coupling constants in DNA oligomers greater than ~8 bp in length, but fitting of simulated to experimental cross-peaks enables extraction of vicinal coupling constants and, subsequently, torsion angle restraints (Celda *et al.*, 1989; Macaya *et al.*, 1992; Schmitz *et al.*, 1990; Weisz *et al.*, 1992; Widmer & Wüthrich, 1987); the torsion angles for the deoxyribose rings are determined using a parametrization of the relationship between torison angles and coupling constants established by Rinkel and Altona (1987). The major difficulty with this approach is in establishing the correct line width to be employed (Schmitz *et al.*, 1990). However, the choice of line width can usually be constrained such that limits (bounds) on the torsion angles describing sugar pucker can be made.

We have found that complete relaxation matrix analysis of proton homonuclear 2D NOE spectra enables numerous accurate interproton distances to be calculated (Borgias & James, 1988, 1990; James, 1991; Keepers & James, 1984; Kumar *et al.*, 1992; Liu *et al.*, 1992). To that end, the most efficient techniques for calculating accurate distances entail an iterative approach (Boelens *et al.*, 1988; Borgias & James, 1988, 1989, 1990; Madrid *et al.*, 1991; Post *et al.*, 1990). In particular, the MARDIGRAS algorithm permits the determination of a large number of accurate distance restraints and aids in individually setting bounds for those distances. For the present studies, we have benefited from recent improvements in MARDIGRAS accounting for the presence of molecular motions (Liu *et al.*, 1992) and incorporation of exchange with bulk water for exchangeable protons (Liu *et al.*, 1993).

The resulting distance and torsion angle restraints can be used with restrained molecular dynamics calculations (Lane *et al.*, 1991; Nikonowicz & Gorenstein, 1992; van Gunsteren *et al.*, 1983). We incorporate both distance and torsion angle restraints as a flat-well penalty function into the AMBER force field for rMD calculations (Gochin & James, 1990). That capability is available on the latest version (4.0) of AMBER (Pearlman *et al.*, 1991). The established bounds set the width of the flat well for each individual restraint. We have found the procedure to yield reliable representations of time-averaged DNA duplex structure manifesting good agreement with experimental 2D NOE data via a sixth-root *R* factor, good agreement with experimental coupling constant data, and low energies (Kerwood *et al.*, 1991; Schmitz *et al.*, 1991, 1992b; Stolarski *et al.*, 1992). Quite recently, we developed a restrained Monte Carlo (rMC) technique, based on the Metropolis algorithm, which used solely distance restraints incorporated into the DNAminiCarlo force field to generate DNA duplex structure (Ulyanov *et al.*, 1993). The rMC structures obtained for d(GTATAATG)·d(CATATTAC) were quite similar to that refined by the rMD approach (RMSD < 0.5 Å) despite the different force fields used and despite the fact that rMD refinement was conducted with additional restraints imposed on the endocyclic torsion angles of deoxyriboses. This enables us to utilize the rMD calculations with an even greater degree of confidence in the resulting structures.

We have used our previously reported distance and torsion angle restraints with the AMBER 4.0 force field to determine the structure of d(AGCTTGCCTTGAG)·d(CTCAAGGCAAGCT) via rMD (Mujeeb *et al.*, 1992); the results are reported in the present paper. Figure 1 shows the sequence and numbering used for the trisdecamer. The approach described here utilizes a "standard rMD" protocol, where the time-averaged nature of the NMR restraints is implicit. As we have reported (Schmitz *et al.*, 1992a, 1993), molecular dynamics with exponentially weighted time-averaged restraints (MD-tar) may yield additional insight into the dynamic nature of the DNA backbone. However, MD-tar is not very efficient in propelling a B-DNA or A-DNA starting structure into the correct region of conformational space; consequently, it is more efficient to employ standard rMD utilizing different starting structures and initial trajectories to find the region of conformational space which satisfies the experimental restraints and the force field prior to any MD-tar simulations. Starting from two very different initial structures, A- and B-DNA (RMSD ≈ 6.4 Å), and using several different rMD protocols, conformational space was searched via rMD for the conformations best defined by the experimental restraints. Two families of converged structures (RMSD < 1.0 Å) were generated. The final structure was obtained by averaging these structures. A detailed analysis of the final structure is presented in terms of structural features and helical parameters.

	1	2	3	4	5	6	7	8	9	10	11	12	13	
5'-	A	G	C	T	T	G	C	C	T	T	G	A	G	-3'
3'-	T	C	G	A	A	C	G	G	A	A	C	T	C	-5'
	26	25	24	23	22	21	20	19	18	17	16	15	14	

FIGURE 1: Sequence of the HIV-1 trisdecamer duplex with base numbering. Two –CTTG– tetrads have been boxed.

METHODOLOGY

Generation of Experimental Restraints. Extraction of experimental restraints was described in detail in our earlier paper on d(AGCTTGCCTTGAG)·d(CTCAAGGCAAGCT) (Mujeeb *et al.*, 1992). Interproton distances entailing solely nonexchangeable protons were calculated from the 2D NOE cross-peak intensities at three mixing times (80, 120, and 200 ms) using the iterative complete relaxation matrix algorithm MARDIGRAS (Borgias & James 1989, 1990) which also enables accurate distances to be calculated for fast rotating methyl protons (Liu *et al.*, 1992). Distances to exchangeable imino protons were determined using 2D NOE cross-peak intensities obtained from the trisdeamer in H_2O at mixing times of 120 and 200 ms with MARDIGRAS, taking into account effects of exchange kinetics with bulk solvent on the intensities (Liu *et al.*, 1993). An isotropic tumbling motion was assumed for the molecule, with an overall correlation time of 2.0 ns (Mujeeb *et al.*, 1992). This is reasonable for a trisdecamer for which the hydrated length-to-radius ratio is about 2; this degree of anisotropy has a negligible effect on NOE values. Upper and lower bounds were determined for individual proton pairs, taking into account (a) the fit of the converged MARDIGRAS matrix with the experimental cross-peak intensities, (b) signal to noise, (c) distances obtained from different mixing time spectra, (d) distances obtained using different starting structures in the MARDIGRAS calculation, (e) worst case geometries for distances involving the methyl rotor (Liu *et al.*, 1992), and (f) estimates on the upper limit of the rate of exchange with bulk water for exchangeable protons (Liu *et al.*, 1993).

Torsional angle restraints for sugar rings were estimated via fitting of experimental with simulated 2QF-COSY cross-

peaks for deoxyribose spin systems. The programs SPHINX and LINSHA were used to simulate a stick spectrum and to add calculated line shapes, respectively (Widmer & Wüthrich, 1987). As discussed earlier, an examination of all the experimental data revealed that they were incompatible with a single conformation. However, this approach allowed us to use a two-state model for cross-peak simulations, which reflected the dynamics of sugar rings (Mujeeb *et al.*, 1992; Schmitz *et al.*, 1990). The coupling constants required for fitting the simulated spectra to the calculted spectra enabled an assessment of the pseudorotation phase angles P (Rinkel & Altona, 1987; van Wijk *et al.*, 1992); consequently, endocyclic torsion angles ν_0, ν_1, ν_2, ν_3, and ν_4 could be calculated for each deoxyribose ring in the trisdecamer (Saenger, 1984). This analysis yielded five torsion angle restraints per sugar residue for the subsequent rMD structure refinement.

Initial Model Building. All energy minimization and molecular dynamics calculations were carried out using the program package AMBER, version 4.0 (Pearlman *et al.*, 1991). The calculations were performed on a Cray Y-MP computer at the Pittsburgh Supercomputing Center and locally on a cluster of HP-730 computers or a SUN SPARCstation 2.

Initial structures for canonical A- and B-DNA conformations were generated using the NUCGEN module of AMBER 4.0; hydrogen atoms with standard geometries were added with the EDIT module. The backbone charges were reduced from −1.0 to −0.2 to implicitly incorporate counterion effects. A small number of calculations employed a −1.0 charge on phosphate but included explicit hydrated sodium ions. In comparison, the two different methods of accounting for counterions gave essentially the same results. The possible extra charges on 5′- and 3′-terminal nucleotides were neutralized. A distance-dependent dielectric constant ($\epsilon = r$) was used during *in vacuo* rMD calculations to mimic the presence of a high dielectric solvent. All molecular structures were displayed using the MIDASplus graphics package on a Silicon Graphics Iris workstation (Gallo *et al.*, 1989).

Restrained Molecular Dynamics. The SANDER (simulated annealing with NMR driven energy restraints) module in AMBER 4.0 allows one to perform energy refinements using NMR-derived distance and torsion angle restraints. The potential energy can incorporate experimental structural restraints within pseudoenergy terms E_{NOE} and $E_{J\text{-coupling}}$ with other terms common to the conventional force field:

$$E_{total} = E_{bond\ lengths} + E_{bond\ angles} + E_{dihedrals} + E_{electrostatics} + E_{NOE} + E_{J\text{-coupling}} \quad (1)$$

The precise nature of the individual terms in the conventional force field has been described previously (Pearlman *et al.*, 1991; Singh *et al.*, 1986).

The pseudoenergy terms added to the standard force field to enforce the experimental restraints take the form of a flat-well potential with quadratic boundaries. Restraint energy violations are calculated for each distance restraint, where the lower (r_l) and the upper (r_u) bounds derived from MARDIGRAS (*vide supra*) define the size of the flat-well width according to

$$E_{restr} = \sum_{all\ NOEs} \begin{cases} k_2(r - r_l)^2 & \text{when } r < r_l \\ 0 & \text{when } r_l \le r \le r_u \\ k_3(r - r_u)^2 & \text{when } r_u < r \\ 4k_2(r - r_u) & \text{when } r > r_u + 2.0 \end{cases} \quad (2)$$

Force constant values k_1 and k_2 can also incorporate confidence in distance restraints as well, *e.g.*, due to uncertainties from peak overlap in 2D NOE spectra. Similarly to eq 2, $E_{J\text{-coupling}}$ terms can be added for each torsion angle restraint. Force constant values for distance and torsion angle restraints ultimately utilized in the rMD structure refinement calculations (*vide infra*) were optimized in some preliminary simulations as described previously (Schmitz *et al.*, 1991).

To maintain right-handedness and prevent structural artifacts during the high-temperature and high force constant phase of simulated annealing (rMD) (Baleja *et al.*, 1990), backbone torsion angle restraints, permitting a broad range of torsion space, were employed with a force constant of 5.0 kcal/(mol·rad²): $-90° < \alpha < -30°$, $135° < \beta < 215°$, $30° < \gamma < 90°$, $140° < \epsilon < 300°$, $150° < \zeta < 315°$. Outside the range specified, the penalty increases quadratically for 10° and linearly beyond that. Backbone angle δ was not restrained, as its value is also reflected by endocyclic sugar ring torsion angle ν_3. It should be noted that the use of backbone torsion angle restraints hinders the search of unusual regions of conformational space, so it is not advisable to use such constraints in rMD simulations of unusual DNA structures such as loops and junctions.

The base-pair Watson–Crick hydrogen-bonding scheme was also maintained during rMD runs by applying hydrogen bond distance and flat angle restraints. A total of 66 hydrogen-bonding restraints were included. The lower and upper bounds defining the flat well for a G·C base pair were set to 2.81–3.01 Å (G-O6 to C-N4), 2.85–3.05 Å (G-N1 to C-N3), and 2.76–2.96 Å (G-N2 to C-O2); for A·T, the bounds were set to 2.72–2.92 Å (A-N1 to T-N3) and 2.85–3.05 Å (A-N6 to T-O4) (Saenger, 1984; Schmitz *et al.*, 1992b). Hydrogen bond angles were restrained to 170–190° with force constants of 5.0 kcal/(mol·rad²).

A typical rMD run consisted of 30 000 1-fs steps, *i.e.*, 30 ps. A cutoff value of 30 Å was chosen to calculate all nonbonded interactions in the system, and the nonbonded pair list was updated every 25 steps. The SHAKE algorithm was used to constrain all bonds (Ryckaert *et al.*, 1977); SHAKE removes bond-stretching freedom, which is the fastest motion, and consequently allows a larger time step (McCammon & Harvey, 1987). Translational and rotational motions were removed every 100 fs. Relative geometrical tolerance for coordinate resetting in SHAKE was 0.0005 Å. Initial structures for rMD were the coordinates of A- and B-DNA geometries, energy-minimized using steepest descent and conjugate gradient methods. A Maxwellian distribution for initial atomic velocities was assigned at 0.4 K. A one-step restrained energy minimization revealed experimental restraint violations for the initial structures, which were used as reference for further refinements. The system was slowly heated to 1000 K during the first 3 ps and then maintained at this high temperature for the next 10 ps to overcome high-energy barriers. The temperature then was gradually lowered to 300 K over the next 5 ps and maintained at 300 K to the end of the rMD simulation. The whole system was loosely coupled to a heat bath at 300 K. The strength of coupling was determined by a time constant of 0.2. Along with temperature, force constant weightings were also modulated from 0.2 to 250 kcal/(mol·Å²), kept constant during the high-temperature period, and then gradually reduced to 25 kcal/(mol·Å²) during cooling and maintained until the end of the run. With A-DNA as the initial structure, relative weights of all force-field energy terms were reduced to 25% during the high-temperature period, so the trajectory was mainly guided by experimental restraints. As no significant drift in either potential energy or restraint

deviations was observed during the final equilibration period, it can be concluded that the system reached a minimum energy conformation. As the coordinates were collected every 100 steps, 40 snapshots from the last 4 ps of the rMD trajectory were averaged and energy-minimized with restraints to remove averaging artifacts.

During all restrained energy minimizations, force constants were the same as those employed during the final rMD equilibration period. These final energy minimizations were carried out without using any backbone torsion angle restraints. The structures were energy-minimized by 200 steps of steepest descent, followed by conjugate gradient minimization until the energy gradient norm difference between successive minimization steps was <0.0001 kcal/(mol·Å^2).

RESULTS AND DISCUSSION

Distance and Torsion Angle Restraints. The flat-well form of the experimental restraint penalty functions used in the rMD procedure necessitates a choice for the size of the flat wells. While this could be set the same for all distances, for example, this practice results in a fairly arbitrary choice of the flat-well size. As we have demonstrated previously, assuming that distances are more accurately determined than is experimentally warranted can lead to inaccurate structures (Thomas *et al.*, 1991). Use of the iterative complete relaxation matrix MARDIGRAS, however, permits us to choose the size of the flat well individually for each proton pair in a reasonably logical fashion. As noted in the Methodology section, we have several considerations which enable us to calculate the upper and lower bounds for each restraint; these can subsequently be used to establish the size of the flat well. Although MARDIGRAS provides us with fairly accurate interproton distances, there are individual variations in the accuracy of these values. In particular, distances corresponding to overlapped cross-peaks or cross-peaks with low signal-to-noise ratio and distances entailing exchangeable protons are less accurately determined.

In the present case, a total of 267 experimental distance restraints were obtained. There were four cross-strand restraints between AH2–TH3 protons, derived from 2D NOE spectra acquired in H_2O, which were incorporated in the data set. There were 105 interresidue distances and 158 intraresidue distances. The average well width for the 267 distance restraints was found to be 0.28 Å.

The well-established relationship between proton–proton vicinal coupling constants and dihedral angles, in a modified Karplus relationship (Rinkel & Altona, 1987; van Wijk *et al.*, 1992), permits fairly accurate determination of deoxyribose conformation. Simulated fitting of double-quantum-filtered COSY cross-peaks provided values of pseudorotation angle P and the pucker amplitude ϕ_m for the minor N and major S conformers in dynamic equilibrium along with their fractional populations (Mujeeb *et al.*, 1992). The five individual endocyclic torsion angles ν_i, $i = 0$–4, in a particular sugar conformation can be related to the pseudorotation angle and the amplitude of pucker, which describe the deoxyribose conformation (Altona & Sundaralingam, 1972):

$$\nu_i^{\text{total}} = f_S\phi_m \cos(P_S + 2(i-2)(360°)/5) + f_N\phi_m \cos(P_N + 2(i-2)(360°)/5) \quad (3)$$

where f_S and f_N are the fractional populations of S and N states, respectively, and it is assumed the pucker amplitude is the same in both states. For each endocyclic torsion angle value, a flat-well width was calculated on the basis of the estimated uncertainty in the 2QF-COSY simulation. The average value of assigned flat-well widths was 9.88°. A total of 130 (five per residue) endocyclic torsion angle restraints were calculated.

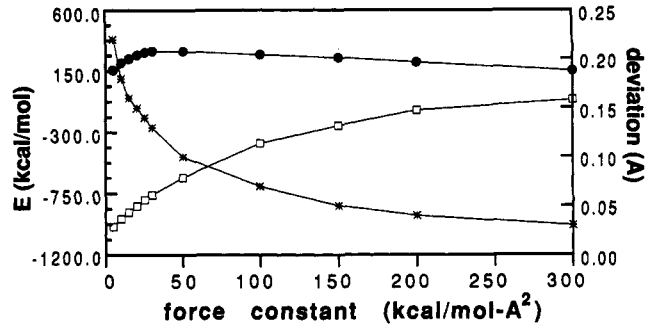

FIGURE 2: Total potential energy (□), restraint violation energy (●), and average distance restraint deviation (*) for restrained energy minimization of the B-DNA form of the HIV-1 trisdecamer duplex as a function of restraint force constant values.

An effort was made to find an optimum value to use for the force constants for distance and torsion angle restraints, separately. In a series of restrained energy minimizations, force constants were varied from 5 to 300 kcal/(mol·Å^2) for distance restraints. As discussed previously (Schmitz *et al.*, 1991), increasing the force constant for restraints should provide a better fit with experimental data. Figure 2 shows our results. Increasing force constant values certainly decreases average distance deviations (to 0.04 Å) for restraints, but the effect diminishes as the force constant values increase. On the other hand, increasing force constants increase the total potential energy of the system, which reaches a maximum of –62.9 kcal/mol at 300 kcal/(mol·Å^2). However, as the initial part of these curves exhibits, the system pays less of an energy penalty and still achieves a reasonably low distance deviation of 0.13 Å with a force constant of <30 kcal/(mol·Å^2). This suggests that using a force constant of ≤30 kcal/(mol·Å^2) would give good accord with experimental data (smaller deviations) along with a reasonable potential energy value. With this rationale, a value of 25 kcal/(mol·Å^2) was chosen as the force constant for all distance restraints during the final equilibration period. In a similar strategy, 60 kcal/(mol·rad^2) was assigned to all torsion angle restraint force constants. Hydrogen-bonding and backbone torsion angle restraints were 25% and 50%, respectively, of the experimental restraint force constants.

Restrained Molecular Dynamics. Structures determined via rMD have two major determinants: the experimental restraints and the force field applied. The role of the force field is to provide an energetically feasible structural framework for the molecule as the rMD protocol is utilized to search conformational space for a structure most consistent with the experimental restraints, *i.e.*, to an energy minimum balancing the restraint violation terms and the conventional potential energy terms. It is necessary to utilize an appropriate protocol to achieve that minimum via rMD simulations. This strategy is, more or less, case-dependent and is best optimized for each particular data set (Schmitz *et al.*, 1991).

To calibrate our perception of the final target structure, we ran one-step restrained energy minimizations on the starting structures. A-DNA exhibited a very high restraint violation energy of 2772.2 kcal/mol (see Table I), indicatinig that this starting structure is quite distant from the target structure manifest in the restraint data set. However, the B-DNA starting structure exhibited a restraint violation energy of 638.5 kcal/mol, much lower than that of the A-DNA initial structure. Since the A-DNA initial structure was very far

Table I: Energy Terms (kcal/mol), Average Distance Deviations Δd_{av} (Å), and Torsion Angle Deviations $\Delta\nu_{av}$(deg) for Initial and rMD Structures

structure	E_{total}	E_{bond}	E_{angle}	E_{vdW}	$E_{1-4elect}$	E_{Hbond}	E_{elect}	E_{NOE}	$E_{J\text{-coupling}}$	Δd_{av}	$\Delta\nu_{av}$
A-DNA	−980.0	9.7	128.7	−441.1	−1753.9	−7.0	720.7	2772.2	401.4	0.68	20.0
B-DNA	−1047.4	8.3	92.5	−436.5	−1728.5	−4.4	638.5	684.5	81.8	0.27	5.0
rMD-A	−762.0	27.6	223.8	−384.9	−1732.3	−6.2	661.8	212.3	18.1	0.24	1.2
rMD-B	−845.3	22.9	171.9	−385.6	−1731.2	−4.2	649.8	175.8	13.3	0.24	0.9
rMD-final	−851.3	22.2	173.2	−389.3	−1729.6	−5.0	649.8	181.7	14.1	0.24	0.9

from the target structure, it was anticipated that refinement from A-DNA would require a much more laborious rMD strategy than from B-DNA. It was found that, for energy-minimized B-DNA as the initial structure, convergence to a final structure could be readily achieved even at 300 K. In a 30-ps rMD run, the system was slowly heated from 100 to 300 K in 5 ps and maintained there for the remainder of the run. Simultaneously, weighting of all restraints, via force constant values, was also increased to maximum. Monitoring changes in total potential energy, restraint violation energy, average distance deviations (Δd_{av}), and average torsion angle deviations ($\Delta\nu_{av}$), it was evident that gradual application of restraints successfully guided the system to an energy minimum. Distance deviations reached a minimum of 0.12 Å from an initial average deviation of 0.27 Å. Torsion angle deviations also decreased to 1.1° from an initial $\Delta\nu_{av}$ of 5.0° for the B-DNA starting structure. As every restraint violation pays a penalty for deviating from the flat part of the well, the total penalty is reflected in the restraint violations energy. The lower value of the restraint violations energy, $E_{restr} = E_{NOE} + E_{J\text{-coupling}}$, indicates a better fit of restraints, within the limits of their bounds. In practice, due to the time-average nature of NMR restraints, the system always pays some penalty in terms of restraint energy. However, this value, along with Δd_{av} and $\Delta\nu_{av}$, does reflect the fitting. For refinement from B-DNA, E_{restr} gradually decreases to 150 kcal/mol (Table I) and subsequently fluctuates within 30–50 kcal/mol of this value; equilibration of the system for the next 25 ps does not reveal any major changes in energy or restraint deviations. This relatively mild rMD protocol using B-DNA as the starting structure is consistent with the final structure lying within the B-family of DNA structures.

With energy-minimized A-DNA as the starting structure, the rMD protocol required to achieve convergence was rather arduous. It was necessary to take the system to high temperature, utilizing high force constants, in order to pass over energy barriers prior to settling into a converged structure with reasonable energy. The potential energy profile and trajectory in this case were much more complicated as the target structure manifest by the experimental structural restraints was relatively far from the initial structure. As noted in the Methodology section, it was necessary to employ a simulated annealing protocol entailing heating to 1000 K in order to achieve successful convergence starting with A-DNA. The converged structures, resulting from different initial trajectories, had Δd_{av} values of ~0.13 Å and $\Delta\nu_{av}$ values of ~1.8°. These values are quite similar to those of the converged structures resulting from rMD simulations using B-DNA as the starting structure, also suggesting that convergence is achieved.

Following initial optimization of the rMD protocols and parameters, rMD simulations resulting in the final reported structure were carried out. Two sets of five 30-ps rMD runs, each set starting with either energy-minimized A-DNA or B-DNA, were performed using randomly assigned initial atomic velocities for each starting structure. For all ten rMD simulations, coordinates for the last 4 ps of the rMD trajectory were averaged, and the average was subsequently subjected to restrained energy minimization. The pairwise RMSD, calculated for all ten structures resulting from each run, was <1.1 Å for the two most divergent structures. The five structures in each of the two sets were further averaged and restrained-minimized to extract the average features of each set. This gave us two subfinal structures, rMD-A and rMD-B. These two structures exhibited an average RMSD of 0.9 Å. The coordinate sets of all ten structures from the two sets were superimposed using least-squares fitting, averaged, and restrained-minimized to achieve a single coordinate set expected to reflect the averaged conformational characteristics of all ten structures. This coordinate set was subsequently subjected to a final 20-ps rMD run, with coordinates from the last 4 ps being averaged and restrained-minimized to obtain our final structure, rMD-final. This ultimate rMD run was to ensure that observed conformational features of rMD-final were a direct reflection of experimental restraints and not of averaging effects. During the 20-ps run, however, no major structural change was observed, and the molecule maintained essentially the same energy. The pairwise RMSD between rMD-A, rMD-B, and rMD-final was 0.9 ± 0.2 Å.

General Validation of Converged Structures. An assessment of the refined structures in terms of energetics, atomic root-mean-square deviations per residue, restraint violations, and residual indices provides us with some perspective on the quality of the converged structures. The energy terms and average restraint deviations for the energy-minimized initial structures and resulting rMD structures are listed in Table I. For each energy term, variations among the three rMD structures are reasonably small. The total potential energy of the final structure, rMD-final, is −851.3 kcal/mol, which is ~190 kcal/mol higher than the energy of the initial B-DNA structure (−1047 kcal/mol) and ~140 kcal/mol higher than that of the initial A-DNA model (−980 kcal/mol). The restraint energy value, which represents violations of both experimental distance and torsion angle data, exhibits much lower values for the three rMD structures in comparison to our starting structures—rMD-A has a restraint violation energy of 212.3 kcal/mol, strikingly lower than the value of 2772 kcal/mol for the initial A-DNA structure. In rMD-B, the restraint violation energy is 175.8 kcal/mol, which is slightly better than for the rMD-A structure. Careful examination reveals that this difference arises from the terminal part of the molecule. In the rMD-A structure, fraying and destacking effects for terminal base pairs are more pronounced than in rMD-B. This behavior was observed previously (Schmitz *et al.*, 1992b; Stolarski *et al.*, 1992).

As well as energies, the average restraint deviation, Δd_{av} and $\Delta\nu_{av}$, also show large improvements for rMD structures relative to the initial structures. Average restraint deviations are defined in terms of the difference between the distance (or torsion angle) for any particular structure and the closer of the experimental upper or lower bound, *i.e.*, either r_2 or r_3. The average distance deviation is $\Delta d_{av} = (1/N)\sum\Delta r$. The

Table II: Atomic Root-Mean-Square Deviations (Å) between Initial Structures and Restrained Molecular Dynamics Structures

structure	B-DNA	rMD-A	rMD-B	rMD-final
A-DNA	6.43	3.20	2.72	3.05
B-DNA		3.65	4.03	3.99
rMD-A			0.94	1.13
rMD-B				0.89

summation is taken over all restrained distances including the terminal base pairs, N is the number of restraints, and

$$\Delta r = \begin{cases} 0 & \text{when } r_l \le r \le r_u \\ r_l - r & \text{when } r < r_l \\ r - r_u & \text{when } r > r_u \end{cases}$$

The average torsion angle deviation is defined similarly.

The average distance deviation, Δd_{av}, decreases from a value of $\sim 0.5 \pm 0.2$ Å for initial structures to $\sim 0.1 \pm 0.02$ Å for the three rMD structures. Among the three rMD structures, Δd_{av} values are comparable. This further corroborates convergence from starting A- and B-DNA models. Although rMD-final arises from a final 20-ps restrained molecular dynamics run, it manifests essentially the same Δd_{av} value as rMD-A and rMD-B. Torsion angle deviations, $\Delta \nu_{av}$, demonstrate the same behavior as Δd_{av}, reflecting general concordance of distance and torsion angle restraints.

Interproton distance restraint violations can be compared for different moieties. The intra- and interresidue distance restraint violations were calculated for the rMD-final structure. For most distances, average deviations were very low, ~0.05 Å. Intraresidue base–H1′, base–H3′, and H1′–H2″ distance deviations were 0.07 ± 0.02 Å. Cross-strand distances all were completely satisfied in the rMD-final structure. In the case of interresidue distances, deviations for base–H3′ and base–H2″ were surprisingly higher (~0.2 Å) than for other interresidue distances (~0.05 Å). These higher distance deviations are also reflected in the restraint violation energies. As the average flat-well width for rMD-final was 0.28 Å, a distance restraint exacted an energetic penalty, on average, if the deviation was greater than ~0.14 Å from the center of the flat well.

Atomic root-mean-square deviations among all rMD structures provide further insight. Table II reports RMSD values, comparing the various rMD structures and initial A- and B-DNA structures. The calculated RMSD between the two initial structures is 6.4 Å. The RMSD between any one of the three rMD structures and either of the starting structures is also quite large, but the RMSD between the rMD structures is much lower, ~1.0 Å. This is generally acknowledged as an indication that convergence has been achieved (Baleja *et al.*, 1990; Clore *et al.*, 1988; Gochin & James, 1990; Huang & Eisenberg, 1992; Kerwood *et al.*, 1991; Kim & Reid, 1992; Nikonowicz & Gorenstein, 1992; Schmitz *et al.*, 1991, 1992b; Stolarski *et al.*, 1992). Some additional illumination may be gained by comparing the RMSD at individual residues (Schmitz *et al.*, 1991), as shown in Figure 3. Clearly, rMD-A, rMD-B, and rMD-final structures converge to similar structures with low RMSD, which can be contrasted with the RMSD between starting structures. With small increases at the terminal base pairs, the others do not vary much. This may be considered a reflection of the good distribution of restraints across the whole molecule. Resulting structures from rMD calculations carried out on a restraint set deficient in experimental restraints for particular moieties exhibit a greater RMSD among rMD structures in the less restrained region (Schmitz *et al.*, 1991).

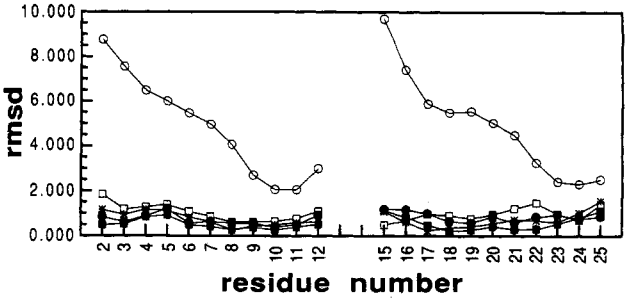

FIGURE 3: Atomic root-mean-square deviations (Å) calculated for individual residues of d(AGCTTGCCTTGAG)·d(CTCAAGGCAAGCT): canonical structures A-DNA vs B-DNA (○), rMD-A vs rMD-B (□), rMD-A vs rMD-final (*), and rMD-B vs rMD-final (●) and the average of all structures, rMD-av, vs rMD-final (■). Only the internal 11 base pairs are shown.

Another vital criterion for evaluation of structure quality emanates from direct comparison of theoretical 2D NOE spectra calculated for the structure with the experimental 2D NOE spectra. Various numerical indices could be used as a figure of merit to evaluate the overall fit of experimental and calculated intensities and consequently the consistency of that structure with the experimental data (Gochin & James, 1990; Suzuki *et al.*, 1986; Thomas *et al.*, 1991). Most commonly a residual index R, analogous to a crystallographic R factor, has been used. However, this R factor is dominated by evaluation of the fit of cross-peaks corresponding to very short (<2.5 Å) distances. As the 2D NOE cross-peak intensiities depend on distances essentially with a sixth-root dependence, we have found that a more sensitive monitor of fitting is the sixth-root residual index (James, 1991; Kerwood *et al.*, 1991; Thomas *et al.*, 1991):

$$R_1^x = \frac{\sum_i |I_o^{1/6}(i) - I_c^{1/6}(i)|}{\sum_i I_o^{1/6}(i)} \qquad (4)$$

where $I_o(i)$ is the experimental intensity of cross-peak i and $I_c(i)$ is the corresponding cross-peak intensity calculated for a particular structure. The summation is generally carried out for all observed cross-peaks. However, indices illuminating discrepancies between experimental and calculated intensities for smaller parts of the molecule or highlighting certain features of the molecule can also be defined. In an intraresidue index, both protons of the ith pair must belong to the same residue (nucleotide) to be included in the summation. And in an interresidue index, only those intensities are summed which arise from proton pairs belonging to different residues.

The program CORMA is capable of providing us with a set of theoretically calculated 2D NOE intensities for a given structure and calculates several quality factors including R_1^x (Borgias & James, 1988; James, 1991). Calculated values of the "crystallographic" R factor and R_1^x for the d(AGCTTGCCTTGAG)·d(CTCAAGGCAAGCT) structures are listed in Table III. The calculated intensities for each of the rMD structures were compared with the three 2D NOE data sets obtained at 80-, 120-, and 200-ms mixing times. For comparison, R and R_1^x values for the starting structures are also listed. It is evident that the rMD structures give a much better fit to the experimental intensities. In general, it appears that the calculated crystallographic R factor and R_1^x exhibit similar trends.

We can also make a limited comparison with our other type of experimental restraint—torsion angles. Deoxyribose proton

Table III: Comparison of Sixth-Root Residual Indices R_1^x and Standard R Factors for Initial and rMD Structures[a]

mixing time (ms)	structure	$R_{intraresidue}$	$R_{interresidue}$	R_{total}	$R_1^x{}_{intraresidue}$	$R_1^x{}_{interresidue}$	$R_1^x{}_{total}$
200	$B_{initial}$	0.46	1.31	0.58	0.09	0.25	0.12
	rMD-A	0.31	0.29	0.31	0.05	0.06	0.05
	rMD-B	0.31	0.25	0.31	0.05	0.05	0.05
	rMD-final	0.31	0.26	0.31	0.05	0.05	0.05
120	$B_{initial}$	0.75	1.15	0.80	0.13	0.30	0.17
	rMD-A	0.35	0.45	0.36	0.06	0.08	0.07
	rMD-B	0.36	0.43	0.36	0.06	0.08	0.06
	rMD-final	0.36	0.43	0.36	0.06	0.07	0.06
80	$B_{initial}$	0.63	1.56	0.77	0.13	0.26	0.16
	rMD-A	0.48	0.45	0.48	0.09	0.09	0.09
	rMD-B	0.47	0.43	0.47	0.08	0.08	0.08
	rMD-final	0.47	0.44	0.47	0.08	0.08	0.08

[a] Calculated for 2D NOE spectra obtained at mixing times of 80, 120, and 200 ms.

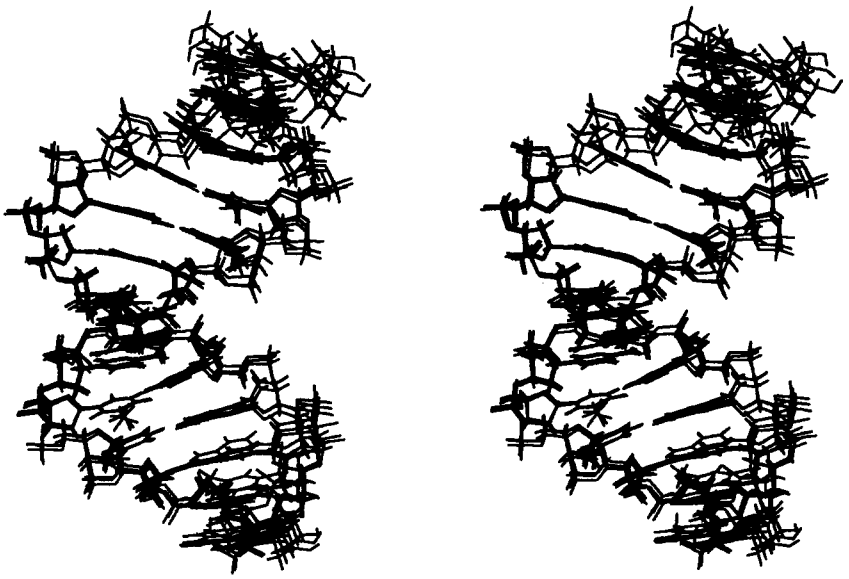

FIGURE 4: Stereoview of the three rMD structures superimposed: substructures rMD-A and rMD-B as well as the final structure rMD-final.

vicinal coupling constants calculated for the final structure, rMD-final, were compared with coupling constants derived from analysis of the 2QF-COSY spectra of d(AGCTTGCCTTGAG)·d(CTCAAGGCAAGCT). The coupling constants were calculated for rMD-final using the Karplus equation as modified by Altona and co-workers (Rinkel & Altona, 1987; van Wijk *et al.*, 1992). The pseudorotation angle P and pucker amplitude ϕ_m used in the calculation were determined using heavy-atom geometries of the sugars in the rMD-final structure. The room-mean-square deviation between experimental and theoretical coupling constants was calculated as $J_{RMS} = (1/N)[\sum(J_{exp}-J_{theor})^2]^{1/2}$, the summation being carried out over the coupling constants $J_{1'2'}$, $J_{1'2''}$, $J_{3'2'}$, and $J_{3'2''}$. J_{RMS} was calculated to be 0.9 Hz, including all coupling constants, and did not deviate much from this value if calculated for most individual residues; exceptions were G6, C16, C25, and T26, where J_{RMS} was slightly higher with values of 1.2 ± 0.1 Hz. A value of 0.5 Hz is an estimate of the precision of calculating vicinal coupling constants (Ulyanov *et al.*, 1993; van Wijk *et al.*, 1992). Although the J_{RMS} values are reasonably good, finding values above 0.5 Hz is likely due to the dynamic nature of the deoxyribose rings. Similar values for J_{RMS} were found for single time-averaged structures of d(GTATAATG)·d(CATATTAC), but values of $J_{RMS} \leq 0.5$ Hz were calculated for an 80%/20% mixture of B- and A-DNA forms, respectively (Ulyanov *et al.*, 1993). Likewise, we note that our previous analysis of the 2QF-COSY spectrum of the trisdecamer duplex was not consistent with any single sugar pucker (Mujeeb *et al.*, 1992). Rather, conformational mixtures were required to satisfactorily fit the 2QF-COSY data.

The general consensus of the above methods of analysis is that the rMD structures are consistent with the experimental data and have converged to within about 0.9 Å, which is comparable to the RMS fluctuation of atoms about their average positions. The extent of convergence can also be assessed by comparison of backbone torsion angles and helical parameters (*vide infra*) for RMD-A, rMD-B, and rMD-final.

Conformational Features of the rMD Structures. The three rMD structures, rMD-A, rMD-B, and rMD-final, are shown superimposed in Figure 4. To acquire additional insight regarding the structure, all helical parameters, backbone torsion angles, and sugar conformations of the resulting rMD structures were thoroughly analyzed with the program CURVES (Lavery & Sklenar, 1990), whose graphic module MACB enables display of the global helix axis. The CURVES algorithm provides a rigorous way to obtain local structural parameters and overall helical axis locus for irregular nucleic acid structures (Lavery & Sklenar, 1988). The output from CURVES includes "global helical parameters" defined relative to a global helix axis and "local helical parameters" defined relative to local helix axes at each base pair. Unless specified otherwise, we refer to the global helical parameters as calculated by CURVES. Plots of various helix parametes for d(AGCTTGCCTTGAG)·d(CTCAAGGCAAGCT) as a function of position in the duplex are shown in Figure 5. The structural parameters are shown for two classical structures, A-DNA and B-DNA, the rMD substructures rMD-A and

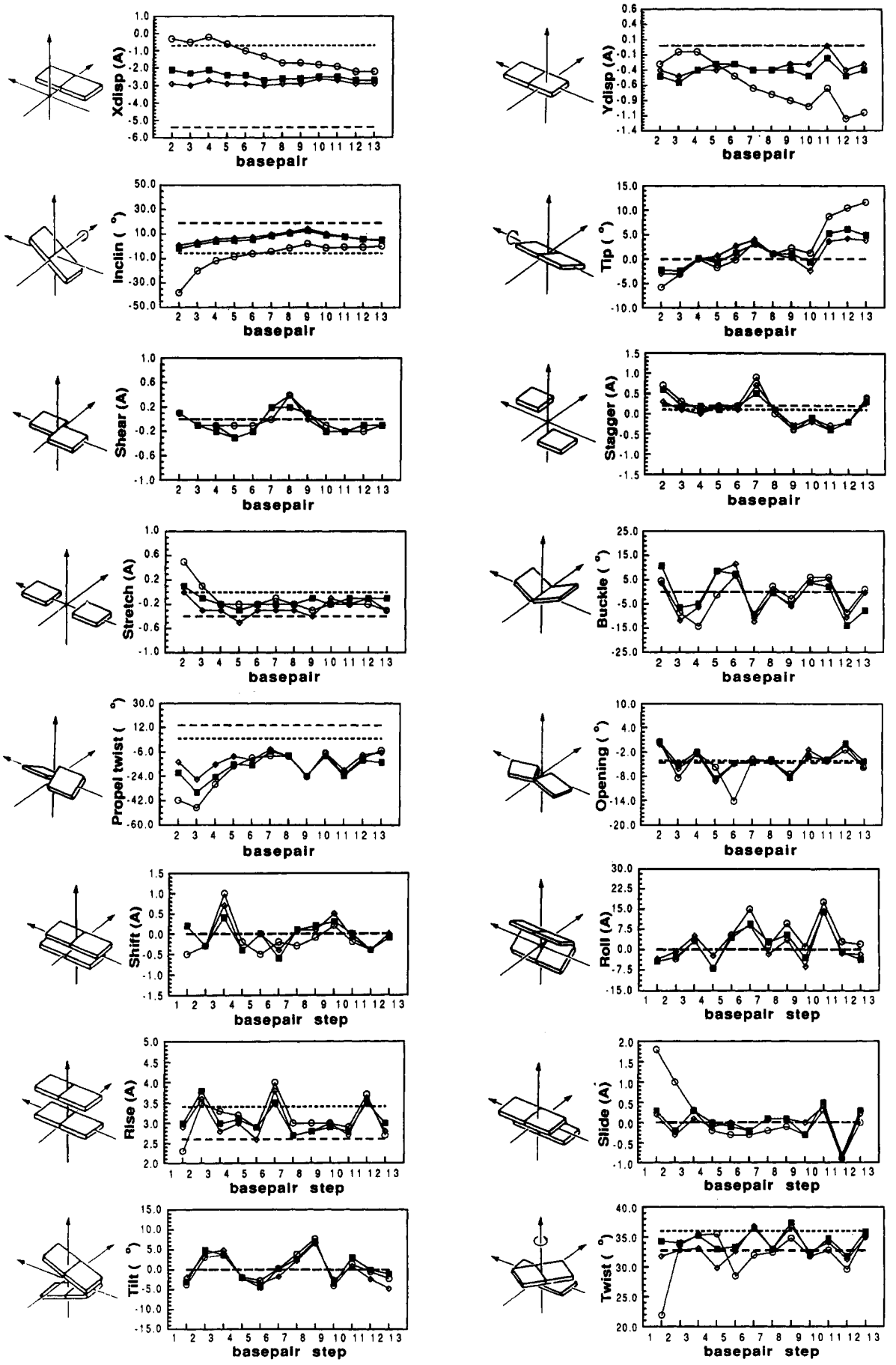

FIGURE 5: Helical parameters for d(AGCTTGCCTTGAG)·d(CTCAAGGCAAGCT) calculated for canonical A-DNA (– –) and B-DNA (- - -), the rMD substructures rMD-A (○) and rMD-B (◇), and the final structure rMD-final (■).

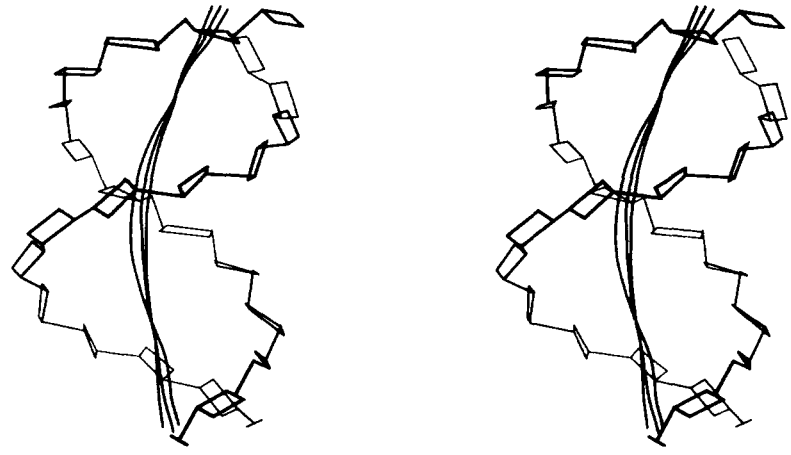

FIGURE 6: Stereoview of the global helix axis in rMD-A, rMD-B, and rMD-final.

rMD-B, and the final structure rMD-final. It should be appreciated that the experimental NMR parameters, being short range in nature, largely establish local structure; consequently, many helical parameters are only secondarily established to be consistent with the local structure and with the AMBER force field. So some structural parameters are better defined by the experimental data than others.

Basically, we can distinguish three categories of helical parameters: axis–base pair, intrabase pair, and interbase pair or base pair–step parameters according to the definitions of an EMBO workshop (Dickerson *et al.*, 1989). Many of these parameters can be defined both in a global and in a local sense. It is readily evident from Figure 5 that variations in helical parameters for the three rMD structures are fairly similar. When deviations occur, usually rMD-A differs from rMD-B and rMD-final. The following discussion will emphasize the characteristics of rMD-final. Among the *axis–base pair parameters*, *X*-axis displacements are ~–2.0 to –2.5 Å, larger than for canonical B-DNA. Unlike standard geometries, all three rMD structures also exhibit nonzero values for *Y*-axis displacements. A less negative value of *Y*-axis displacement for base pair 11 (G11–C16) is evident in all three rMD structures. Base pair inclination is approximately –6° for regular B-DNA and +19° for regular A-DNA; the rMD structures display inclination values midway between those of regular A- and B-DNA. Tip angles generally fluctuate within ±5°, similar to the canonical A- and B-DNA forms.

Among *intrabase pair parameters*, stretch and shear do not vary much from their ideal values. However, a slightly larger stretch (up to 0.24 Å for C8–G19) is found in the center of the duplex. This larger value of stretch for the C8–G19 base pair is manifest in all three rMD structures. Associated with it is a comparably larger stagger for adjacent base pair 7 (C7–G20), which reflects a relative shift in the *Z* direction. Buckle seems to experience large local variations but is almost the same for all three rMD structures. It is possible that the more negative (–10°) value of buckle for base pair 7 (C7–G20) is related to the stagger mentioned above for the same base pair. The following base pairs 8 (C8–G19), 9 (T9–A18), 10 (T10–A17), and 11 (G11–C16) exhibit minimal buckling. Propeller twist for all base pairs is negative, larger than for classical B-DNA. This is typical of other oligonucleotide whose structures in solution have been determined by NMR (Baleja *et al.*, 1990; Kerwood *et al.*, 1991; Kim & Reid, 1992; Schmitz *et al.*, 1991, 1992b; Stolarski *et al.*, 1992; Weisz *et al.*, 1993). For the first four base pairs, propeller twist variations among the three rMD structures are larger. Base pair opening demonstrates an alternating behavior around ideal values; this alternating behavior is preserved in all three rMD structures except base pair 6 (G6–C21) in rMD-A.

In regular A-DNA and B-DNA geometries, global values of the *interbase pair parameters*, roll, slide, shift, and twist, are essentially zero. As seen in Figure 5, there are small base pair shifts in d(AGCTTGCCTTGAG)·d(CTCA-AGGCAAGCT), mostly negative. An interesting facet exhibited by all three rMD structures though is the positive shift observed for base pair steps 3–4 and 9–10. Most slide values differ little from the classical forms; however, the value of *ca.* –0.9 Å at base pair step 11–12 is intriguing and corresponds to a few other structural variations, *e.g.*, rise and twist.

Values for rise are generally only about 3.0 Å, but the higher values (>3.5 Å) for the penultimate steps and for the 6–7 step are notable. The generally smaller values for rise are likely associated with other parameters such as *X* displacement. Of course, the effect is that d(AGCTTGCCTTGAG)·d(CTCA-AGGCAAGCT) is shorter than one would anticipate on the basis of a classical B-DNA structure; all three structures were found to have a helix ~35 Å long, *i.e.*, ~5 Å shorter than standard B-DNA. The roll parameter exhibits an alternating pattern along the duplex. Positive roll opens the angle between base pairs toward the minor groove; as a result, a wider minor groove and bending toward the major groove occurs. This alternation is manifest in distinct positive rolls at base pair steps G6–C7, C8–T9, and T10–G11, which together cause a global bending of the helix axis toward the major groove. In fact, the largest roll value of 14° is observed for the base pair step T10–G11. In a recent study of *in vitro* evolution of intrinsically bent DNA, TG/CA steps have been correlated with DNA bending (Beutel & Gold, 1992). A very recent study of an octamer motif in our laboratory reveals similar behavior for a TG step (Weisz *et al.*, 1993). In d(AGCT-TGCCTTGAG)·d(CTCAAGGCAAGCT), we have two TG steps; the T5–G6 base pair step has a lower (but still positive) value of roll (5°) than the T10–G11 step (14°). However, the next base pair step, G6–C7, has a roll angle of 10° in rMD-final. Due to the small local bendings correlated with these roll values, global helical curvature is effected. The global helix axis for the three rMD structures is illustrated in Figure 6. The helix axis curvature for these structures is ~30–40°; rMD-final has a helix axis curvature of 40.4°.

As noted above, CURVES reports both global and local helical parameters. One might anticipate some correlation between them (Lavery & Sklenar, 1989). Reduced global *X*-axis and *Y*-axis displacements did lead to smaller values of

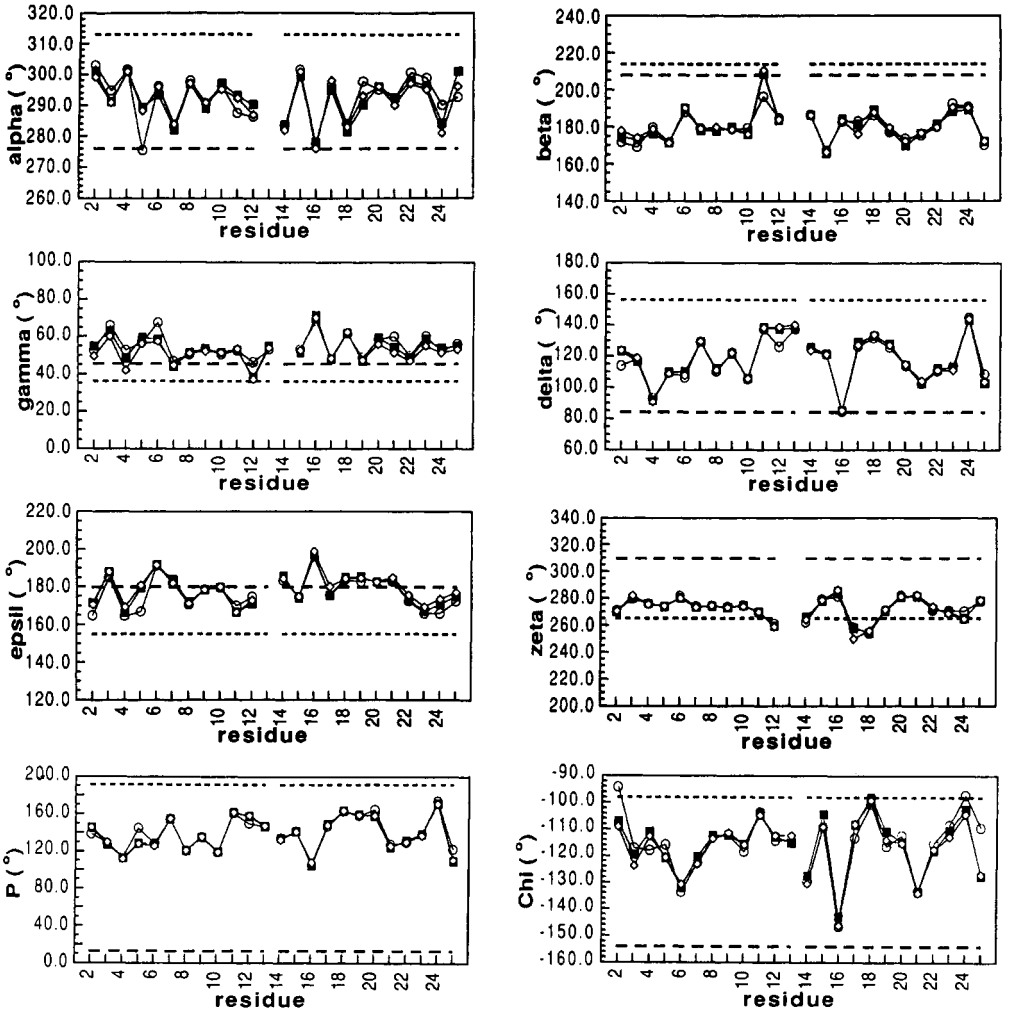

FIGURE 7: Backbone torsion angles for d(AGCTTGCCTTGAG)·d(CTCAAGGCAAGCT) calculated for canonical A-DNA (– –) and B-DNA (- - -), the rMD substructures rMD-A (○) and rMD-B (◇), and the final structure rMD-final (■).

local slide and shift parameters. Small positive inclination values in our rMD structures did result in anticipated changes in local slide (+) and roll (+) parameters. A positive inclination was also associated with reduced local rise and twist and with shortening of the DNA helix (*vide supra*).

Values of torsion angles α through ζ and χ for all three rMD structures as well as canonical A- and B-DNA are presented in Figure 7. Variations in torsion angle values for all rMD structures are strikingly similar. Maximum deviation for any of these parameters from an average value is 25°. With NOE distance restraints used, no backbone torsion angles are defined directly, except δ, which is intrinsically defined by endocyclic torsion angle ν_3 (*vide supra*). The convergence of these parameters may be attributed to the right-handedness restraints used in the present study. However, a low force constant value of 5.0 kcal/(mol·rad^2) was used for all backbone torsion angle constraints, which is only one fifth that used with experimental restraints. The presence of lightly weighted backbone torsion angle restraints helps to achieve a smoother and better converged backbone (Schmitz *et al.*, 1992b), but which is also conformationally compatible with other structural features. That is, the structure is largely being established by the experimental restraints. The nomenclature used here to define the ranges and conformations of various backbone torsion angles is as proposed by Klyne and Prelog and recommended by IUPAC–IUB (Saenger, 1984).

As seen in Figure 7, values of torsion angle α vary within the range of values characteristic for standard geometries, being *–syn clinal* (*–sc*), but with an alternating $\leq\pm10°$ variation evident in all three structures. Angle β represents rotation around the C5′–O5′ ester bond and, in the present case, differs from that of standard A- and B-DNA, *–anti periplanar* (*–ap*), to a less positive value in the *+ap* range. Residue G11 provides an exception in that β is more in the canonical range for all three rMD structures. Backbone torsion γ is only slightly more positive than that of the canonical structures for almost all residues in all structures, being in the *+sc* range. Backbone torsion angle δ, as noted already, reflects the deoxyribose puckering. Values for δ observed here are in the *+anti clinal* (*+ac*) range. This conformation is between the values characteristic of A- and B-DNA. Torsion angle ϵ, which represents rotation about the C3′–O3′ bond, is in the *+ap* range. Values of torsion angle ζ are quite similar to that of B-DNA (*–sc*).

Our previous 2QF-COSY analysis of d(AGCTTGCCTTGAG)·d(CTCAAGGCAAGCT) indicated that the deoxyribose rings were in dynamic equilibrium between at least two conformers but that one conformer dominated (Mujeeb *et al.*, 1992); the dominant conformer possessed a pseudorotation angle P in the S range, but generally not quite in the 2′-*endo* conformation characteristic of classical B-DNA. The time-averaged structure resulting from our rMD simulations is essentially consistent with that major conformer having a pseudorotation phase angle in the range 118–162°, with all three rMD structures quite consistent (Figure 7). As one might anticipate, there is a correlation between the

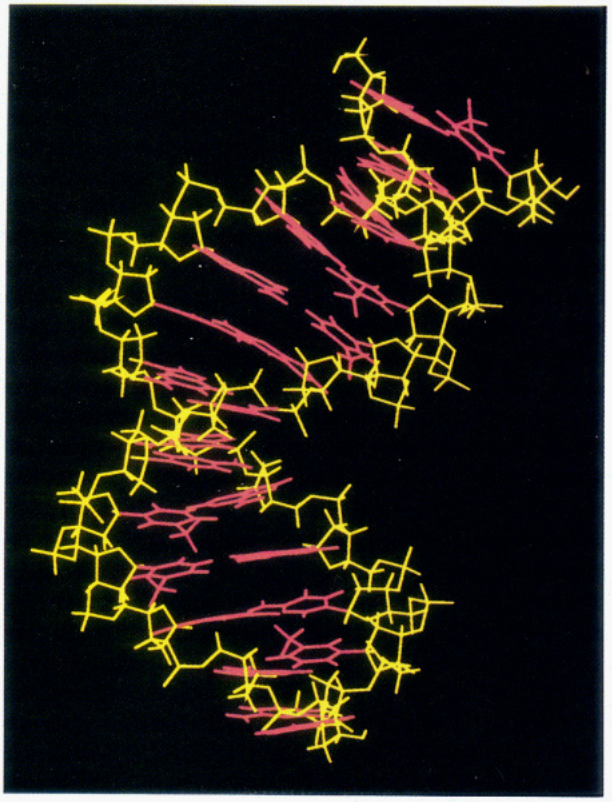

FIGURE 8: Final structure of d(AGCTTGCCTTGAG)·d(CTCA-AGGCAAGCT) determined via restrained molecular dynamics calculations: rMD-final.

pseudorotation angle and the glycosidic torsion angle χ of the individual residues.

The minor groove width in rMD-final (structure shown in Figure 8) was measured as the shortest interstrand phosphorus–phosphorus distance across the groove (minus the combined radii of the phosphorus atoms, 5.8 Å). The measured width varies within the range 5.7–9.5 Å. Around the T10–G11 step, with its larger roll, the minor groove widens to 9.5 Å. In many other places, it is closer to that of canonical B-DNA, which is ~6 Å. An alternate description of minor groove width, the shortest cross-strand C1′–C1′ distance, has recently been proposed (Bhattacharya & Bansal, 1992); since the C1′–C1′ vector lies in the base plane, it is less affected by minor backbone deformations. In the middle of the trisdecamer, accounting for atomic radii, the two definitions gave the same value of minor groove width but varied by up to 1.0 Å at the termini.

Our trisdecamer DNA sequence possesses a degree of symmetry, with two –CTTG– tetrads around the central C·G base pair (Figure 1). It has been suggested that at least a tetrad of four base pairs is necessary to define the local helix conformation (Yanagi *et al.*, 1991). While the flanking sequences differ, it is nevertheless interesting to compare the degree of preservation of local conformation in these two –CTTG– tetrads. Figure 9 shows a side view (a) and a view down the helix axis (b) for the two –CTTG– tetrads overlapped. The two tetrads have an RMSD of 0.6 Å, which indicates almost the same geometry for the two. It is evident from Figure 9 that the conformation of the middle TT step is much

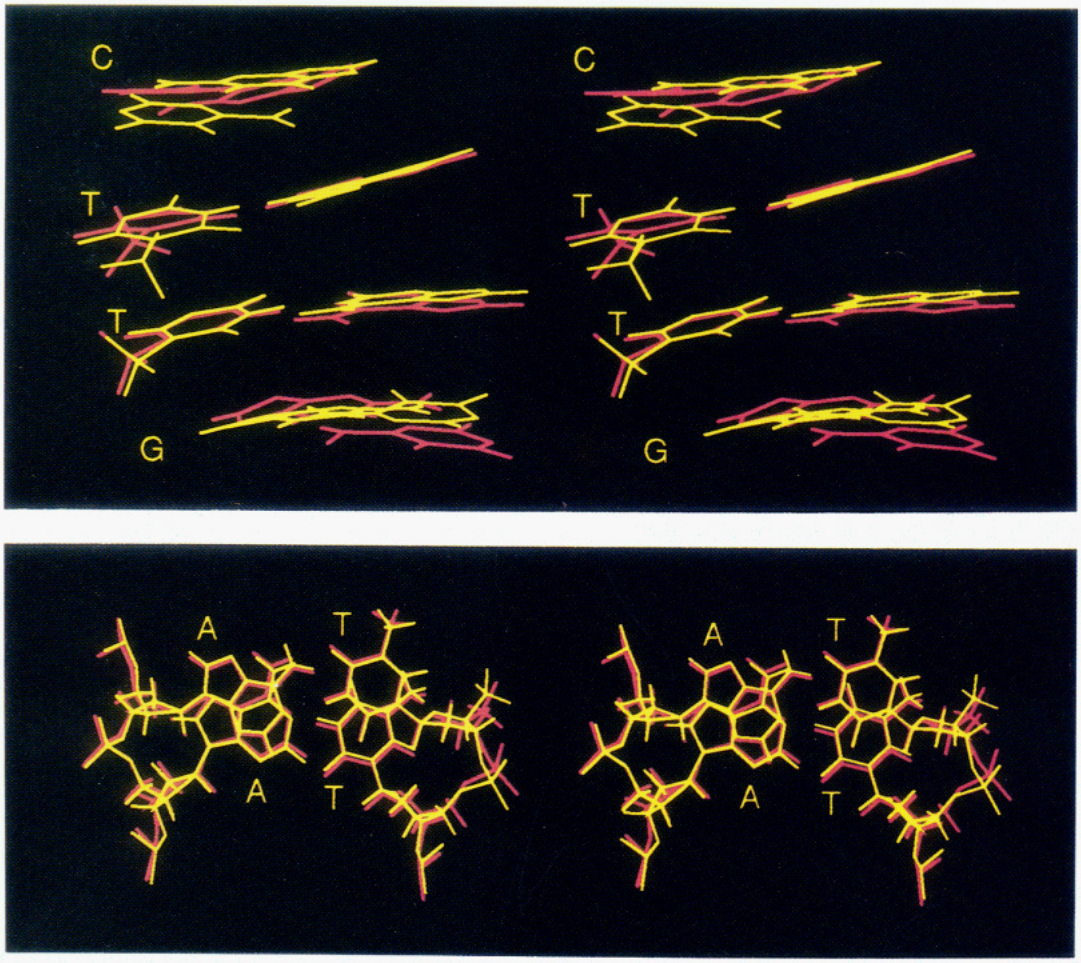

FIGURE 9: (a, top) Stereoview of superimposed base pairs of the two –CTTG– tetrads: –C3T4T5G6– (red) and –C8T9T10G11– (yellow). (b, bottom) Stereoview of the –TT– step of the two –CTTG– tetrads: –T4T5– (red) and –T9T10– (yellow).

better preserved. This is reflected in local helical parameters for the two TT steps. While the outer base pairs of the tetrads exhibit similar structure, there are larger variations in local helical parameters as well as RMSD; the second –CTTG– tetrad has a larger roll and local slide at the TG step (*vide supra*). We note, however, that the TG step is followed by C in the first tetrad and by A in the second tetrad. According to X-ray crystallographic studies, pyr–TG–pur sequences are found with larger roll and slide values for the TG step than pyr–TG–pyr sequences (Heinemann & Alings, 1989; Lipanov *et al.*, 1993; Prive *et al.*, 1987). This is consistent with our observations in d(AGCTTGCCTTGAG)·d(CTCAAGGCAAGCT).

CONCLUSIONS

The present study demonstrates the capability of determining the high-resolution, sequence-dependent structure of DNA duplexes using molecular dynamics calculations to search conformational space, restrained by experimental structure data: accurate MARDIGRAS-derived interproton distances from 2D NOE spectra and sugar conformation information deduced from 2QF-COSY simulations. The method employed was able to establish a good time-averaged representation of the solution conformation of d(AGCTTGCCTTGAG)·d(CTCAAGGCAAGCT), a highly conserved trisdecamer sequence from the long terminal repeat region of HIV-1.

The experimental restraints, interproton distances, and deoxyribose ring torsion angles were internally consistent, although they were derived independently from different 2D NMR experiments and are subject to different time-averaging effects. An increased number of experimental restraints help in establishing a better defined structure. However, due to the time-averaged nature of the NMR data and the conformational flexibility of most molecules in solution, it is not to be expected that all restraints will be perfectly satisfied simultaneously. Cross-strand restraints in a DNA duplex are helpful in defining interstrand geometry; however, most cross-strand restraints involve exchangeable protons which, if analyzed properly, entail extra care in analysis of interproton distances and bounds. This is now possible with the current version of the iterative relaxation matrix algorithm MARDIGRAS (Liu *et al.*, 1993).

Confidence in the time-averaged structure obtained is engendered by convergence to essentially the same structure (RMSD ~0.9 Å) when two quite different DNA models, A-DNA and B-DNA (RMSD ~6.5 Å), were employed as starting structures and when various initial trajectories were used for the rMD runs. Such convergence may be the most convincing argument for the validity of the resulting structure. This was further supported by the total energy, the restraint violation energy, the restraint deviations, and the fit with experimental data. In the case of the last, the sixth-root residual index indicated a good fit of the derived structure with the experimental 2D NOE spectral intensities, and the RMS differences between vicinal proton coupling constants calculated for the derived structure and the experimental coupling constants extracted from 2QF-COSY cross-peaks were also in reasonable agreement (J_{RMS} = 0.9 Hz). Consistent with the relatively even distribution of experimental restraints, none of the above parameters analyzed on a residue-by-residue basis suggested that any part of the molecule (save perhaps the terminal base pairs) was ill-defined.

Overall, the structure of d(AGCTTGCCTTGAG)·d(CTCAAGGCAAGCT) deviates somewhat from canonical B-DNA or A-DNA, with some helical parameters and backbone torsion angles manifesting interesting local variations, which might be important structural features. The helix parameters indicate that, compared with classical B-DNA, the structure is longitudinally more compressed; a pronounced decrease in rise and distinctive *X* displacement is seen at most base pairs. Many of the base pairs also exhibit significant propeller twist. Local structural variations at the two TG steps in particular together create bending into the major groove of the duplex. Comparison of the two –CTTG– tetrads reveals that they have similar structures, with the TT moieties being almost identical; in accord with X-ray crystallographic studies (Lipanov *et al.*, 1993; Prive *et al.*, 1987), however, we find that the –CTTG–pur sequence has a larger roll and slide for the TG step than does the –CTTG–pyr sequence.

ACKNOWLEDGMENT

We acknowledge Drs. Klaus Weisz and Nikolai Ulyanov for helpful discussions. We also acknowledge the use of the University of California, San Francisco, Computer Graphics Laboratory (supported by NIH Grant RR01081).

REFERENCES

Altona, C., & Sundaralingam, M. (1972) *J. Am. Chem. Soc. 94*, 8205–12.

Baleja, J. D., Pon, R. T., & Sykes, B. D. (1990) *Biochemistry 29*, 4828–39.

Beutel, A. B., & Gold, L. (1992) *J. Mol. Biol. 228*, 803–12.

Bhattacharya, D., & Bansal, M. (1992) *J. Biomol. Struct. Dyn. 10*, 213–26.

Boelens, R., Koning, T. M. G., & Kaptein, R. (1988) *J. Mol. Struct. 173*, 299–311.

Borgias, B. A., & James, T. L. (1988) *J. Magn. Reson. 79*, 493–512.

Borgias, B. A., & James, T. L. (1989) in *Methods in Enzymology, Nuclear Magnetic Resonance, Part A: Spectral Techniques and Dynamics* (Oppenheimer, N. J., & James, T. L., Eds.) Vol. 176, pp 169–83, Academic Press, New York.

Borgias, B. A., & James, T. L. (1990) *J. Magn. Reson. 87*, 475–87.

Celda, B., Widmer, H., Leupin, W., Chazin, W. J., Denny, W. A., & Wüthrich, K. (1989) *Biochemistry 28*, 1462–70.

Clore, G. M., Oschkinat, H., McLaughlin, L. W., Benseler, F., Happ, C. S., Happ, E., & Gronenborn, A. M. (1988) *Biochemistry 27*, 4185–97.

Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O., Lavery, R., Nelson, H. J. C., Saenger, W., Shakked, Z., Sklenar, H., Soumpasis, D. M., von Kitzing, E., Wang, A.-H.-J., & Zhurkin, V. B. (1989) *EMBO J. 8*, 1–4.

Gallo, K., Huang, C., Ferrin, T. E., & Langridge, R. (1989) *Molecular Interactive Display and Simulation (MIDASplus)*, Univesity of California, San Francisco.

Gochin, M., & James, T. L. (1990) *Biochemistry 29*, 11172–80.

Griesinger, C., Sørensen, O. W., & Ernst, R. R. (1985) *J. Am. Chem. Soc. 107*, 6394–6.

Heinemann, U., & Alings, C. (1989) *J. Mol. Biol. 210*, 369–81.

Huang, P., & Eisenberg, M. (1992) *Biochemistry 31*, 6518–32.

James, T. L. (1991) *Curr. Opin. Struct. Biol. 1*, 1042–53.

James, T. L., & Basus, V. J. (1991) *Annu. Rev. Phys. Chem. 42*, 501–42.

Keepers, J. W., & James, T. L. (1984) *J. Magn. Reson. 57*, 404–26.

Kerwood, D. J., Zon, G., & James, T. L. (1991) *Eur. J. Biochem. 197*, 583–95.

Kim, S.-G., & Reid, B. R. (1992) *Biochemistry 31*, 12103–16.

Kumar, A., James, T. L., & Levy, G. C. (1992) *Isr. J. Chem. 32*, 257–61.

Lane, A. N., Jenkins, T. C., Brown, D. J. S., & Brown, T. (1991) *Biochem. J. 279*, 269–81.
Lavery, R., & Sklenar, H. (1988) *J. Biomol. Struct. Dyn. 6*, 63–91.
Lavery, R., & Sklenar, H. (1989) *J. Biomol. Struct. Dyn. 6*, 655–67.
Lavery, R., & Sklenar, H. (1990) *CURVES 3.0. Helical Analysis of Irregular Nucleic Acids*, Laboratory for Theoretical Biology, CNRS, Paris.
Lipanov, A., Kopka, M. L., Kaczor-Grzeskowiak, M., Quintana, J., & Dickerson, R. E. (1993) *Biochemistry 32*, 1373–89.
Liu, H., Thomas, P. D., & James, T. L. (1992) *J. Magn. Reson. 98*, 163–75.
Liu, H., Kumar, A., Weisz, K., Schmitz, U., Bishop, K. D., & James, T. L. (1993) *J. Am. Chem. Soc. 115*, 1590–1.
Macaya, R. F., Schultze, P., & Feigon, J. (1992) *J. Am. Chem. Soc. 114*, 781–3.
Madrid, M., Llinas, E., & Llinas, M. (1991) *J. Magn. Reson. 93*, 329–46.
Marion, D., & Bax, A. (1988) *J. Magn. Reson. 80*, 528–33.
McCammon, J. A., & Harvey, S. C. (1987) *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge.
Mujeeb, A., Kerwin, S. M., Egan, W., Kenyon, G. L., & James, T. L. (1992) *Biochemistry 31*, 9325–38.
Nikonowicz, E., & Gorenstein, D. G. (1992) *J. Am. Chem. Soc. 114*, 7494–7503.
Oppenheimer, N. J., & James, T. L. (1989) *Methods in Enzymology, Nuclear Magnetic Resonance, Part A: Spectral Techniques and Dynamics*, Vol. 176, Academic Press, New York.
Pearlman, D. A., Case, D. A., Caldwell, J., Seibel, G. L., Singh, U. C., Weiner, P. K., & Kollman, P. A. (1991) *AMBER 4.0 (UCSF)*, University of California, San Francisco.
Piantini, U., Sørensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800–1.
Post, C. B., Meadows, R. P., & Gorenstein, D. G. (1990) *J. Am. Chem. Soc. 112*, 6796–6803.
Prive, G. G., Heinemann, U., Chandrasegaran, S., Kan, L.-S., Kopka, M. L., & Dickerson, R. E. (1987) *Science 238*, 498–504.
Rinkel, L. J., & Altona, C. (1987) *J. Biomol. Struct. Dyn. 4*, 621–49.
Ryckaert, J. P., Cicotti, G., & Berendsen, H. J. C. (1977) *J. Comput. Phys. 23*, 327–41.
Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer, New York.
Schmitz, U., Zon, G., & James, T. L. (1990) *Biochemistry 29*, 2357–68.
Schmitz, U., Pearlman, D. A., & James, T. L. (1991) *J. Mol. Biol. 221*, 271–92.
Schmitz, U., Kumar, A., & James, T. L. (1992a) *J. Am. Chem. Soc. 114*, 10654–6.
Schmitz, U., Sethson, I., Egan, W., & James, T. L. (1992b) *J. Mol. Biol. 227*, 510–31.
Schmitz, U., Ulyanov, N. B., Kumar, A., & James, T. L. (1993) *J. Mol. Biol.* (in press).
Singh, U. C., Weiner, P. K., Caldwell, J., & Kollman, P. A. (1986) *AMBER 3.0*, University of California, San Francisco.
Stolarski, R., Egan, W., & James, T. L. (1992) *Biochemistry 31*, 7027–42.
Suzuki, E.-I., Pattabiraman, N., Zon, G., & James, T. L. (1986) *Biochemistry 25*, 6854–65.
Thomas, P. D., Basus, V. J., & James, T. L. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 1237–41.
Ulyanov, N. B., Schmitz, U., & James, T. L. (1993) *J. Biomol. NMR 3*, 547–568.
van Gunsteren, W. F., Boelens, R., Kaptein, R., & Zuiderweg, E. R. P. (1983) in *Nucleic Acid Conformation and Dynamics, Report of NATO/CECAM Workship* (Olson, W. K., Ed.) pp 79–82, Orsay, France.
van Wijk, J., Huckriede, B. D., Ippel, J. H., & Altona, C. (1992) *Methods Enzymol. 211*, 286–306.
Wagner, G., Hyberts, S. G., & Havel, T. F. (1992) *Annu. Rev. Biophys. Biomol. Struct. 21*, 167–98.
Weisz, K., Shafer, R. H., Egan, W., & James, T. L. (1992) *Biochemistry 31*, 7477–87.
Weisz, K., Shafer, R. H., Egan, W., & James, T. L. (1993) *Biochemistry* (in press).
Widmer, H., & Wüthrich, K. (1987) *J. Magn. Reson. 74*, 316–36.
Yanagi, K., Prive, G. G., & Dickerson, R. E. (1991) *J. Mol. Biol. 217*, 201–14.