Journal of Biomolecular NMR, 3 (1993) 547–568
ESCOM

J-Bio NMR 132

Metropolis Monte Carlo calculations of DNA structure using internal coordinates and NMR distance restraints: An alternative method for generating a high-resolution solution structure

Nikolai B. Ulyanov*, Uli Schmitz and Thomas L. James**

Department of Pharmaceutical Chemistry, University of California, San Francisco, CA 94143-0446, U.S.A.

Received 5 March 1993
Accepted 26 April 1993

Keywords: Monte Carlo simulations; Molecular dynamics simulations; Two-dimensional nuclear Overhauser effect (2D NOE); Complete relaxation matrix analysis; DNA duplex; Pribnow consensus sequence

SUMMARY

A new method, a restrained Monte Carlo (rMC) calculation, is demonstrated for generating high- resolution structures of DNA oligonucleotides in solution from interproton distance restraints and bounds derived from complete relaxation matrix analysis of two-dimensional nuclear Overhauser effect (NOE) spectral peak intensities. As in the case of restrained molecular dynamics (rMD) refinement of structures, the experimental distance restraints and bounds are incorporated as a pseudo-energy term (or penalty function) into the mathematical expression for the molecular energy. However, the use of generalized helical parameters, rather than Cartesian coordinates, to define DNA conformation increases efficiency by decreasing by an order of magnitude the number of parameters needed to describe a conformation and by simplifying the potential energy profile. The Metropolis Monte Carlo method is employed to simulate an annealing process. The rMC method was applied to experimental 2D NOE data from the octamer duplex d(GTA-TAATG) · d(CATTATAC). Using starting structures from different locations in conformational space (e.g. A-DNA and B-DNA), the rMC calculations readily converged, with a root-mean-square deviation (RMSD) of < 0.3 Å between structures generated using different protocols and starting structures. Theoretical 2D NOE peak intensities were calculated for the rMC-generated structures using the complete relaxation matrix program CORMA, enabling a comparison with experimental intensities via residual indices. Simulation of the vicinal proton coupling constants was carried out for the structures generated, enabling a comparison with the experimental deoxyribose ring coupling constants, which were not utilized in the structure determination in the case of the rMC simulations. Agreement with experimental 2D NOE and scalar coupling data was good in all cases. The rMC structures are quite similar to that refined by a traditional restrained MD approach (RMSD < 0.5 Å) despite the different force fields used and despite the fact that MD refinement was conducted with additional restraints imposed on the endocyclic torsion angles of deoxyriboses. The computational time required for the rMC and rMD calculations is about the same. A comparison of structural parameters is made and some limitations of both methods are discussed with regard to the average nature of the experimental restraints used in the refinement.

*On leave from the Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, 117984 Moscow B-334.
**To whom correspondence should be addressed.

0925-2738/$ 10.00 © 1993 ESCOM Science Publishers B.V.

INTRODUCTION

Experimental structural restraints obtained from NMR experiments, especially interproton distances extracted from two-dimensional nuclear Overhauser effect (2D NOE) spectra and sometimes torsion angles from multidimensional scalar coupling-based NMR spectra, can lead to high-resolution molecular structures only when used in conjunction with appropriate structure algorithms. Various computational methods, e.g. a distance geometry (DG) algorithm (Havel et al., 1983; Braun, 1987; Oshiro et al., 1991), energy minimization (EM) (Pearlman et al., 1991), and molecular dynamics (MD) (van Gunsteren et al., 1983; McCammon and Harvey, 1987; Pearlman et al., 1991; Gorenstein, 1992) have been adapted and widely used to yield molecular structures. In principle, these computational procedures provide a means of globally searching conformational space to find a set of molecular structures consistent with the experimentally determined restraints. In current practice, protein structures are commonly determined using distance restraints via DG calculations; further refinement is then carried out using the resultant DG structures as starting structures in restrained MD calculations. (Torsion-angle data may or may not be included in the rMD calculations.) In the case of DNA duplexes, obtaining starting structures via DG is unnecessary as it can be readily ascertained if the antiparallel duplex is a left- or right-handed helix. A right-handed helix suggests that canonical A-form and B-form DNA (and possibly other models) can be used as starting structures for rMD calculations.

Generally, we have only a limited number of structural restraints, which are also of limited accuracy. We can construct an objective function which measures the deviation of the individual distances from their individual target values; an increasing deviation from the target value results in a larger penalty. An example would be to incorporate a pseudo-energy term in the molecular force field. The pseudo-energy term might typically be of the form of a flat-well potential such that the estimated experimental error can even be incorporated as the size of the flat well (Baleja et al., 1990; Kerwood et al., 1991; Stolarski et al., 1992). The problem of structure refinement is to find the global minimum in potential energy when the potential energy surface has multiple minima; this multiple-minima problem can be especially severe when the starting structure is far from the true global minimum. Methods entailing systematic searches of accessible conformational space have been advocated, but they are computationally rather expensive at the present time for molecules the size of proteins or nucleic acids. This problem may be successfully overcome by utilizing a simulated annealing procedure, implemented with MD simulation algorithms (Nilsson et al., 1986; Nilges et al., 1987; Brünger and Karplus, 1991). The annealing protocol entails a heating period which enables the simulated molecule to pass over potential barriers due to its elevated kinetic energy; a slow cooling from this elevated temperature will presumably enable the lowest energy to be achieved. Confidence is enhanced when different starting structures and different initial trajectories lead to the same global minimum.

All existing program packages for MD simulation use atomic Cartesian coordinates as independent variables. For a double-stranded DNA octamer, this corresponds to over 1500 degrees of freedom, most of which are not 'conformationally interesting' in the sense that they correspond

---

Abbreviations: MC, Monte Carlo; rMC, restrained Monte Carlo; MD, molecular dynamics; rMD, restrained molecular dynamics; DG, distance geometry; EM, energy minimization; 2D NOE, two-dimensional nuclear Overhauser effect; DQF-COSY, double-quantum-filtered correlation spectroscopy; RMSD, root-mean-square deviation.

to high-frequency bond stretching and bond-angle fluctuations. Parallel to the coordinate-based approach to DNA conformation calculations, computational methods utilizing internal coordinates have been developed (Olson, 1977; Zhurkin et al., 1978; Gupta et al., 1980; Chuprina et al., 1981). These methods use idealized geometries of bases, treating them as rigid bodies; bond lengths and often bond angles are fixed at their idealized values. In recent years, these approaches have been extended to include all 'essential' degrees of freedom inherent to the DNA double helix (Lavery, 1988; Ulyanov et al., 1989), as defined by the Cambridge convention (Dickerson et al., 1989). An obvious advantage of using a method based on internal coordinates such as helical parameters is the large reduction in the number of independent variables in the system. Typically, one needs 120–270 parameters (depending on the model of deoxyribose used) to define the double-helical structure of a DNA octamer, which is an order of magnitude less than in the coordinate-based approach. So far no annealing algorithm has been implemented with the helical parameters-based methods, although other methods such as scanning procedures or adiabatic mapping have been employed to overcome the multiple-minima problem (Mauffret et al., 1992; Ulyanov et al., 1992).

Although rMD calculations are commonly used, there are other possible methods of generating structures using NMR experimental data. For example, a Monte Carlo search using NOE distance restraints has been developed for peptides (Levy et al., 1989; Ripoll and Ni, 1992).

A Monte Carlo simulation is amenable to the annealing process. In the present paper, we describe the use of a restrained Monte Carlo (rMC) procedure. The Metropolis Monte Carlo algorithm was used to generate the Boltzmann distribution of DNA structures at a series of different temperatures (Metropolis et al., 1953). Generalized helical parameters were used to define each DNA structure. Interproton distances were obtained via complete relaxation-matrix analysis of 2D NOE spectra with the MARDIGRAS algorithm (Borgias and James, 1989, 1990; Kumar et al., 1992; Liu et al., 1992). The following DNA octamer was used:

5′ - G1 T2 A3 T4 A5 A6 T7 G8 - 3′
3′ - C16 A15 T14 A13 T12 T11 A10 C9 - 5′

The solution structure of this octamer, which contains the Pribnow consensus sequence, has recently been determined in this laboratory by conventional rMD methods (Schmitz et al., 1992b), utilizing the AMBER program suite (Pearlman et al., 1991). The results from the rMC calculations are compared to those from the rMD calculations.

METHODS

DNA structure generation

All conformational calculations, including energy minimizations and Monte Carlo simulations, were performed with the program developed by Zhurkin and co-workers (Ulyanov et al., 1989; Gorin et al., 1990), which has recently been named DNAminiCarlo. Calculations were carried out on the Cray Y-MP at the Pittsburgh Supercomputing Center as well as on a Sun SPARCstation 2 in our laboratory. The program is capable of both energy minimization and Monte Carlo simulation of the Boltzmann distribution of DNA structures. DNAminiCarlo uses generalized helical parameters rather than atomic Cartesian coordinates as independent varia-

bles. The program generates the DNA structure in three steps. First, the position of bases in space is calculated using the complete set of helical parameters; the definitions of parameters are in accord with the Cambridge convention (Dickerson et al., 1989). All bases are treated as rigid bodies with idealized geometry. Next, sugar rings are attached to the N1 or N9 atoms of the bases. The orientation of the sugar relative to the attached base is defined by the glycosidic torsion angle χ, which is another independent variable. Two different models were utilized to calculate the conformation of the flexible sugar ring — a four-parameter model and a one-parameter model. Generally, a five-membered ring has four independent internal degrees of freedom, provided that all bond lengths are fixed. The four-parameter sugar model, when implemented in DNAminiCarlo, employs as parameters the dihedral angles C2′-C3′-C4′-O1′ and C3′-C4′-O1′-C1′ and bond angles C3′-C4′-O1′ and C4′-O1′-C1′. The one-parameter model utilizes the concept of pseudorotation (Altona and Sundaralingam, 1972). In the one-parameter sugar model, four internal degrees of freedom are tabulated as a function of a single variable, the pseudorotation phase angle P. The tabulated values correspond to the minimum energy conformation for the deoxyribose sugar for a particular value of the pseudorotation angle P (V.B. Zhurkin, private communication). As the pseudorotation angles varied in the course of refinement, the look-up table was employed to select those low-energy tabulated values for the four parameters characterizing the deoxyribose conformation. Usually, exocyclic bond angles associated with the sugar ring are omitted from the list of independent variables; they vary as a function of pseudorotation angle P similar to the four internal sugar parameters. In the present study however, a number of calculations were made with independently varied exocyclic bond angles associated with sugar atoms C1′, C3′ and C4′, each exocyclic bond having two degrees of freedom. In total, including exocyclic bond angles, each deoxyribose had one to ten independent variables, depending on the precise refinement procedure employed.

As the last step in the generation of the double helix, the conformation of the backbone is calculated using a special chain-closure algorithm (Zhurkin et al., 1978; Gorin et al., 1990). Consequently, the backbone torsion angles are not treated as independent variables in DNAminiCarlo. Rather, they depend on the relative position of deoxyriboses in the strands.

Calculation of conformational energy

The conformational energy E_{conf} of a DNA structure is calculated using additive empirical atom–atom potential functions (Zhurkin et al., 1981; Poltev and Shulyupina, 1986), which include '6–12' van der Waals terms, '10–12' hydrogen bond terms, electrostatic interactions, torsion-angle and bond-angle distortion terms. All calculations were performed in vacuo, with the effect of solvent being modeled implicitly by a distance-dependent dielectric constant $\varepsilon(r) = r$. The effect of counterions was modeled by neutralizing the effective charges on the phosphodiesters.

In contrast to previous calculations with the DNAminiCarlo program which utilized few distance restraints and a simple harmonic penalty (Zhurkin et al., 1991), here we use a flat-well potential:

$$E_{restr} = \sum_{NOE} \begin{cases} k_{NOE}\,(r - r_l)^2 & \text{when } r < r_l \\ 0 & \text{when } r_l \le r \le r_u \\ k_{NOE}\,(r - r_u)^2 & \text{when } r_u < r \end{cases} \qquad (1)$$

where r is an interproton distance and r_l and r_u are the corresponding lower and upper distance bounds estimated from experimental 2D NOE intensities (vide infra). k_{NOE} is a pseudo-force constant selected according to the annealing protocol. The total potential energy of a system, $E = E_{conf} + E_{restr}$, is used in minimization and MC routines of the DNAminiCarlo program.

Distance restraints and comparison with the experimental 2D NOE data

Distance restraints were calculated for d(GTATAATG) · d(CATTATAC) from two 500 MHz homonuclear proton 2D NOE data sets acquired with mixing times of 100 and 150 ms, employing iterative complete relaxation matrix analysis using the program MARDIGRAS (Borgias and James, 1989, 1990; Liu et al., 1992). Acquisition of these restraints has been described in detail previously (Schmitz et al., 1992b), and they are available from the Brookhaven Protein Data Bank (accession number 1D70). However, we have recently improved the quality of the distance restraints by using structures obtained from restrained MD calculations, rather than canonical A- or B-DNA, as starting models for MARDIGRAS (Schmitz et al., 1993). In addition, the error bounds were evaluated in a different manner to better reflect inaccuracies in peak intensities, especially for weak peaks. Overall, this led to a small decrease in flat-well widths for short distances and an increase for larger distances compared to the restraint set published. Flat-well widths for the present study range from 0.05 to 1.70 Å; the average flat-well width of the new set is 0.31 Å, compared to 0.25 Å for the published set. In the present study, for the sake of compatibility with the original MD refinement of the octamer, we omit all distance restraints involving H3′ protons (Schmitz et al., 1992b). However, distances involving protons of the four terminal residues were treated in a different way compared to the published rMD refinement. Here, we used the MARDIGRAS-generated restraints for these distances, rather than theoretical values taken from B- and wrinkled D-forms of DNA. The reason for this is the ambiguity in selection of 'B-type restraints' due to the conformational flexibility of the B-form of DNA. In addition to experimentally determined interproton distance restraints, we used pseudo-restraints (r_l = 1.8 Å, r_u = 2.0 Å) for all protons involved in hydrogen bonds. Although these restraints are usually satisfied in unrestrained MC simulations at room temperature, they are important to keep the base pairs intact at elevated temperatures during simulated annealing.

Progress in refinement, as well as the results of the various refinement protocols, was monitored by comparing the simulated and experimental 2D NOE intensities. For model structures generated during the MC refinement procedure, 2D NOE intensities were calculated taking into account all dipole–dipole multispin effects (i.e. 'spin-diffusion') with the full relaxation-matrix method (Keepers and James, 1984), using the program CORMA (Borgias and James, 1988; Liu et al., 1992). A single effective correlation time of 2.8 ns was utilized for the calculations (Schmitz et al., 1992b). Two residual indices were used to quantify the level of agreement between experimental and simulated spectra: an analog of the crystallographic residual index or R-factor

$$R = \frac{\sum_i | a_o(i) - a_c(i) |}{\sum_i a_o(i)} \qquad (2)$$

and the sixth-root residual index (James, 1991; James et al., 1991; Thomas et al., 1991):

$$R^x = \frac{\sum_i | a_o^{1/6}(i) - a_c^{1/6}(i) |}{\sum_i a_o^{1/6}(i)} \quad (3)$$

where a_o is the observed experimental 2D NOE peak intensity and a_c is the corresponding theoretical intensity; the summation is carried out over all measured nonzero experimental cross peaks for nonterminal residues (including the H3′ protons). Both indices were calculated individually for the two mixing times used. R^x attempts to express a figure of merit which relates all NOE intensities to the coordinate space of the structure, assuming here an approximate distance dependence of r^{-6}. Because of the extreme distance dependence of NOE intensities, errors in the shortest distances tend to dominate in a 'crystallographic' R-factor. The sixth-root scaling allows longer-range NOE interactions (e.g. ~ 4–5 Å) to be considered as well.

We also calculate the average deviation of interproton distances from the closest of the two bounds, either r_u or r_l, as $R_{dev} = (1/N)\Sigma\Delta r$. The summation is taken over all restrained distances including the terminal base pairs. N is the number of restraints (151 in the case of the octamer), and

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

Torsion-angle restraints were not employed during structure refinement. However, deoxyribose proton vicinal coupling constants calculated for the simulated structures were compared with coupling constants derived from analysis of the DQF-COSY spectra of d(GTATAATG) · d(CATTATAC) (Schmitz et al., 1992b). The coupling constants were calculated according to the Karplus equation as modified by Altona and co-workers (Haasnoot et al., 1980; De Leeuw et al., 1983); our computations were carried out using the B set of parameters (Haasnoot et al., 1980). The RMSD between experimental and theoretical coupling constants was calculated as $J_{rms} = (1/N)\sqrt{\Sigma(J_{exp} - J_{theor})^2}$, the summation being carried out over the coupling constants $J_{1'2'}$, $J_{1'2''}$, $J_{2''3'}$, and $J_{2''3'}$ in nonterminal residues.

In addition to these figures of merit, the convergence of refinement was assessed by calculating the atomic RMSD between superimposed final structures. Only the heavy atoms of nonterminal residues of d(GTATAATG) · d(CATTATAC) were used to calculate the superposition and RMSD.

Monte Carlo simulations

The Metropolis Monte Carlo simulations at a particular temperature were performed as previously described (Ulyanov and Zhurkin, 1982; Zhurkin et al., 1991). All independent helical variables were subdivided into subsets of correlated variables, each consisting of six to ten variables. For example, for calculations with a single sugar pucker parameter (vide supra), there were seven subsets corresponding to the seven steps between the eight base pairs of the octamer with six parameters describing each of the seven steps; in addition, there were ten parameters

characterizing each base pair (six for base-pair geometry, two pseudorotation angles P, and two glycosidic torsion angles χ). Obviously, there were 15 subsets of variables in this case. With calculations entailing the four-parameter sugar model, there were approximately twice as many subsets, as the base-pair parameter subsets were further subdivided.

All variables from a particular subset were altered randomly within specified limits, while parameters in the other subsets were maintained. The structure was then generated with the new parameters, and its energy was compared with the energy of the previous structure. The new structure is accepted with the probability $\exp(-\Delta E/RT)$, where ΔE is the difference between old and new energies (Metropolis et al., 1953). We use the term 'iteration' for an event such that each of the variable subsets has been sequentially (randomly) changed, with the new structure being accepted or rejected according to the probability calculation. Thus, each iteration consisted of 15 to 30 trials of new structures, depending on the deoxyribose model used. After completion of each iteration, all independent variables and a number of other parameters were summed for subsequent calculation of mean values. The subsets of variables and their maximum increments are specified as input for the DNAminiCarlo program, so it is possible to keep some of the variables

TABLE 1
REFINEMENT PARAMETERS ASSOCIATED WITH THE DIFFERENT SIMULATED-ANNEALING PROTOCOLS EMPLOYED

	Deoxyribose variables[a]	Terminal base-pair conformation	Starting structure	Annealing, no. of iterations	Averaging, no. of iterations
B-series					
B1	1	fixed	B[b]	31 000	10 000
B2	4	fixed	B1(min)[c]	21 000	10 000
B3	8	fixed	B2(av)[d]	22 000	10 000
B4	8	flexible	B3(av)[d]	22 000	10 000
B5	10	flexible	B4[e]	40 000	20 000
A1-series					
A1.1	1	fixed	A[f]	29 000	10 000
A1.2	1	flexible	A1.1[e]	29 000	10 000
A1.3	10	flexible	A1.2(av)[d]	40 000	20 000
A2-series					
A2.1	1	flexible	A[f]	22 000	10 000
A2.2	10	A2.1(av)[d]	A2.1(av)[d]	40 000	20 000

[a] Number of variables associated with each deoxyribose. 1: one-parameter model of sugar ring; internal sugar parameters and exocyclic bond angles vary as a function of a single parameter, i.e. pseudorotation angle P. 4: four-parameter model of sugar; exocyclic bond angles vary as a function of pseudorotation angle P. 8: four-parameter sugar model; exocyclic bond angles associated with atoms C4′ and C3′ vary independently, while those associated with C1′ are fixed. 10: four-parameter sugar model; all exocyclic bond angles vary independently.

[b] Energy-minimized B-conformation.

[c] The average structure calculated in the previous annealing was further restrained-minimized.

[d] Structure was averaged along the last fragment of the Markov chain of the previous annealing.

[e] Structure calculated at the last iteration of the previous annealing.

[f] Energy-minimized A-conformation.

'frozen' during the simulation. For example, the terminal base-pair parameters were kept fixed in some cases during the initial stages of refinement (see Table 1).

To achieve the fastest convergence of the Markov chain, the maximum increments in the variables must be chosen to provide an acceptance-to-rejection ratio as close to 1.0 as possible. If this ratio is too low, one needs a longer Markov chain to generate a sufficient number of new conformations. On the other hand, if the ratio is too high, the increments are very small and, again, one needs a longer Markov chain to deviate significantly from the starting structure. In practice, as the ratio of rejected structures depends both on temperature and constant k_{NOE} (Eq. 1), this ratio varied in the range of 0.6–1.0 at different stages of the annealing procedure.

Annealing procedure

We have performed three separate refinements (series of annealings) starting from energy-minimized (without distance restraints) B- or A-forms of the octamer d(GTATAATG) · d(CATTATAC). Each series included several sequential annealings with an increasing number of 'unfrozen' variables associated with the deoxyriboses and the terminal residues. The parameters for the different annealing protocols are listed in Table 1. The longest series (B-series) consisted of five sequential annealings, the shortest one (A2-series) consisted of two annealings. Each annealing comprised several (10–13) Markov chain fragments with heating and cooling periods, each fragment being generated at a constant temperature and with a nonvarying force constant k_{NOE} (Eq. 1). Within each annealing, the last structure of each individual fragment of Markov chain served as starting structure for the next fragment. The annealing started at a low temperature (100 K) and low force constant of 1–4 kcal/mol · Å^2. During the heating period, the temperature was increased in small increments to 600 K with a simultaneous increase in k_{NOE}, typically to 64 kcal/mol · Å^2. The fragment of Markov chain with these extreme values of temperature and force constant was typically 2 000–4 000 iterations long. During the cooling period, the temperature and force constant were gradually decreased to 300 K and 20 kcal/mol · Å^2, respectively. The exact temperature and force constant profiles had little effect on results, with the exception of the total length of annealing (see Results). These profiles are shown in Fig. 1a for the simplest series of annealings (A2).

The structures from the last 10 000–20 000 iterations of each annealing were used to calculate the 'average' structure (vide infra), which was subsequently restrained-minimized with k_{NOE} = 20 kcal/mol · Å^2. The first annealing in each series started with energy-minimized B-DNA (B-series) or A-DNA (A1- and A2-series). Each subsequent annealing was started with one of the following: (a) the last structure of the previous annealing, (b) the structure which was averaged at the end of the previous annealing, or (c) the result of restrained minimization of the previous annealing's final averaged structure. Several test simulations were carried out to check that the choice of one of these three options did not significantly affect the annealing results.

Calculation of the average structure

A simple procedure was used to calculate the average structure for individual Markov chain fragments. Mean values of all independent variables for that fragment, automatically calculated by DNAminiCarlo, served as input for DNAminiCarlo to calculate an 'average' structure. This procedure, first used in DNA hydration studies (V.B. Zhurkin, private communication), replaces the coordinate averaging procedure which is routinely used in MD simulations.

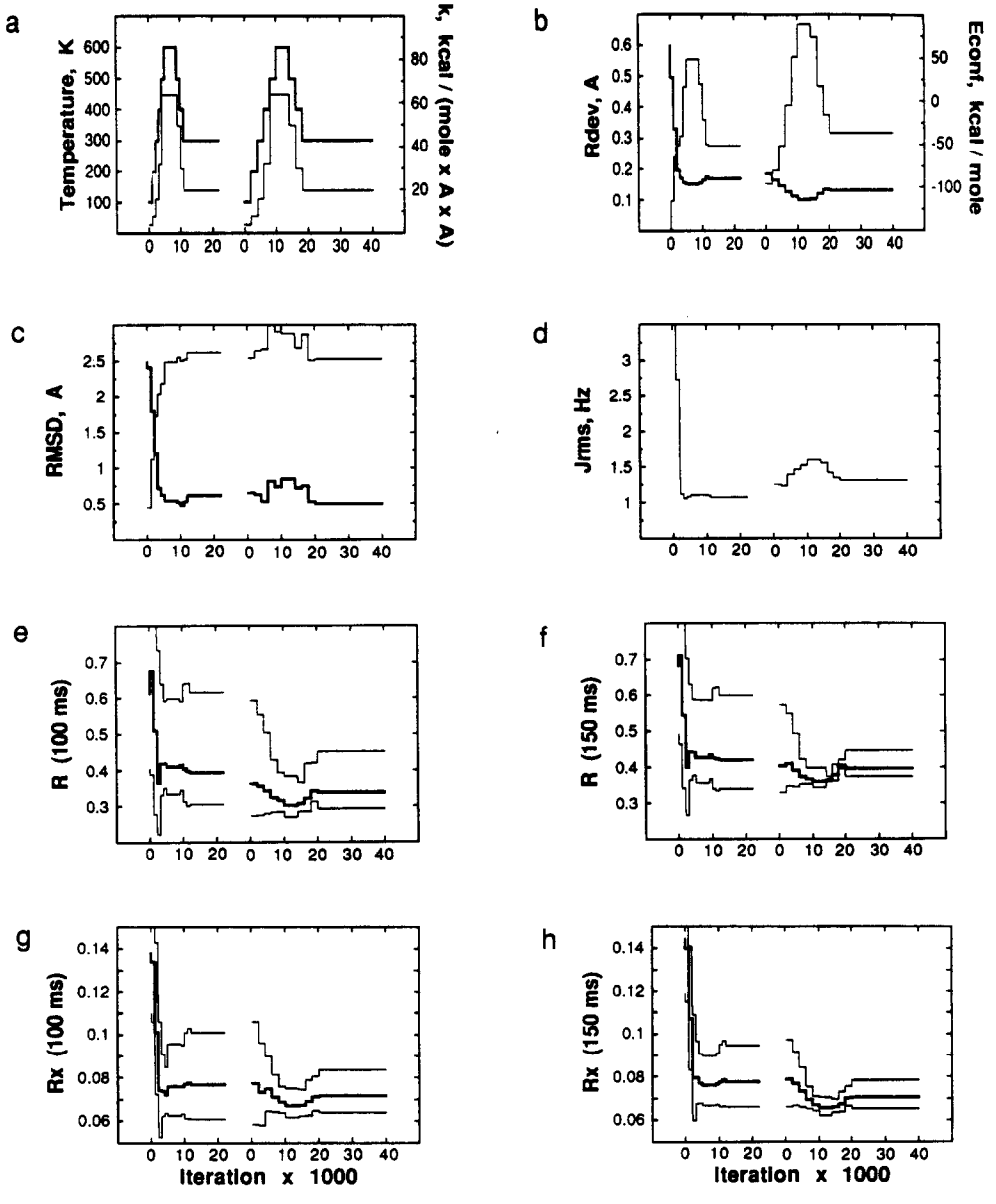

Fig. 1. Parameters of refinement (a) and indices of agreement (b–h) for the A2 series of rMC annealings. The indices were calculated as described in the Methods section; each fragment of the Markov chain, simulated at constant temperature and force constant, is represented by a single structure which was averaged along that fragment of the chain for purposes of calculating the indices. In all cases, the horizontal axis is the iteration number; note that the iteration number is marked separately for annealing A2.1 (iterations 1 through 22 000) and subsequent annealing A2.2 (1 through 40 000). Different lines in each panel are distinguished as follows: (a) temperature (thick line) and force constant k_{NOE} (thin line); (b) R_{dev} (thick line) and conformational energy E_{conf} (thin line); (c) RMSD from the starting A conformation (thin line) and from the rMD structure (Schmitz et al., 1992b), MD_{fin} (thick line); (d) J_{rms}; (e–h) 2D NOE residual indices R (e,f) and R^x (g,h), which were calculated for the experimental 2D NOE intensities acquired at mixing times of 100 ms (e,g) and 150 ms (f,h); the middle (thick) line always refers to the total index, while the upper and lower (thin) lines always refer to the inter- and intraresidue components of the indices, respectively.

TABLE 2
INDICES OF AGREEMENT FOR THE STARTING STRUCTURES AND STRUCTURES RESULTING FROM rMC AND rMD CALCULATIONS[a]

	R_{dev} (Å)	R (100 ms)			R^x (100 ms)			R (150 ms)			R^x (150 ms)			J_{rms} (Hz)
		total	intra	*inter*	**total**	intra	*inter*	**total**	intra	*inter*	**total**	intra	*inter*	
Initial structures														
B	0.275	**0.493**	0.417	*0.691*	**0.0874**	0.0658	*0.1210*	**0.532**	0.473	*0.667*	**0.0949**	0.0731	*0.1270*	1.08
A	0.597	**0.614**	0.404	*1.16*	**0.1380**	0.1100	*0.1810*	**0.683**	0.494	*1.13*	**0.1440**	0.1190	*0.1820*	5.20
B-series rMC protocol														
B1	0.185	**0.415**	0.342	*0.601*	**0.0733**	0.0633	*0.0882*	**0.439**	0.375	*0.587*	**0.0782**	0.0681	*0.0931*	1.08
B2	0.162	**0.384**	0.312	*0.562*	**0.0748**	0.0612	*0.0953*	**0.411**	0.348	*0.557*	**0.0762**	0.0665	*0.0907*	1.42
B3	0.154	**0.394**	0.344	*0.520*	**0.0726**	0.0612	*0.0897*	**0.420**	0.377	*0.519*	**0.0731**	0.0653	*0.0846*	1.31
B4	0.148	**0.393**	0.341	*0.528*	**0.0730**	0.0618	*0.0900*	**0.417**	0.370	*0.525*	**0.0732**	0.0657	*0.0844*	1.35
A1-series rMC protocol														
A1.1	0.238	**0.419**	0.329	*0.647*	**0.0810**	0.0644	*0.1060*	**0.444**	0.362	*0.633*	**0.0822**	0.0698	*0.1010*	1.08
A1.2	0.170	**0.419**	0.351	*0.592*	**0.0730**	0.0642	*0.0862*	**0.445**	0.387	*0.580*	**0.0782**	0.0696	*0.0909*	1.08
A2-series rMC protocol														
A2.1	0.171	**0.397**	0.315	*0.608*	**0.0769**	0.0616	*0.0998*	**0.423**	0.349	*0.593*	**0.0778**	0.0669	*0.0939*	1.07
Final structures														
B_{fin}	0.137	**0.350**	0.305	*0.466*	**0.0738**	0.0653	*0.0864*	**0.402**	0.380	*0.453*	**0.0715**	0.0664	*0.0792*	1.36
$A1_{fin}$	0.134	**0.337**	0.296	*0.442*	**0.0719**	0.0643	*0.0834*	**0.392**	0.372	*0.437*	**0.0705**	0.0656	*0.0779*	1.36
AB_{fin}	0.137	**0.340**	0.297	*0.451*	**0.0725**	0.0644	*0.0847*	**0.396**	0.377	*0.441*	**0.0710**	0.0660	*0.0783*	1.35
$A2_{fin}$	0.133	**0.343**	0.303	*0.446*	**0.0678**	0.0569	*0.0841*	**0.400**	0.383	*0.438*	**0.0706**	0.0657	*0.0777*	1.32
MD-refined structures[b]														
MD_{fin}	0.176	**0.295**	0.236	*0.445*	**0.0610**	0.0494	*0.0784*	**0.330**	0.262	*0.487*	**0.0685**	0.0535	*0.0907*	1.04
MD_{subst}	0.194	**0.310**	0.243	*0.482*	**0.0700**	0.0591	*0.0864*	**0.367**	0.296	*0.529*	**0.0779**	0.0642	*0.0982*	1.04

[a] Indices R_{dev}, R, R^x and J_{rms} are defined in the Methods section. For the R and R^x indices, the total value of the index and its intra- and interresidue components are printed with the use of bold, ordinary and italic fonts, respectively.

[b] The MD_{fin} structure was refined against a slightly different set of constraints for the internal residues (Schmitz et al., 1992b) and completely different constraints for the termini (see Methods), which explains the increased value of R_{dev} for this structure.

Final structures

Each of the three rMC refinement protocols (B-, A1- and A2-series) yielded a 'final' structure (B_{fin}, $A1_{fin}$ and $A2_{fin}$, respectively). Final structures were obtained by restrained minimization of structures averaged (vide supra) during the last stage of the last annealing in each series. A fourth 'final' structure, AB_{fin}, was obtained by averaging the average structures obtained in the B- and A1-series of refinements, with subsequent energy minimization.

RESULTS AND DISCUSSION

Three series of rMC annealings yielded four 'final' structures as described above. All rMC

simulations showed a high degree of convergence with the RMSD between different final structures being within 0.3 Å (vide infra). The indices of agreement (the R factors, R_{dev}, J_{rms}) are also quite similar for these final structures (Table 2), including the structures refined from both B- and A-DNA. Consequently, we will discuss in detail results from the simplest protocol of refinement (the A2-series, Fig. 1) and only briefly mention the other series.

The indices of agreement exhibit similar profiles over the course of annealing for each of the annealing protocols, with the exception of J_{rms} (Fig. 1d) which will be discussed later. R_{dev}, R and R^x decrease during the heating period, reach their minimum at the maximum temperature and force constant, and then decrease somewhat during the cooling period (Figs. 1b,e–h). R_{dev} is closely related to the restraint energy E_{restr} (Eq. 1), so it is trivial that an increase in force constant k_{NOE} causes a decrease in R_{dev}. In contrast, there is no direct relationship between E_{restr}, which drives the refinement, and 2D NOE residual indices R and R^x. In this respect, we wish to note that a very strong correlation between R_{dev} and the R factors (Fig. 1 and Table 2) demonstrates that the MARDIGRAS-derived set of distances adequately represents the experimental 2D NOE intensities. Both R-factors, R and R^x, change almost synchronously during annealing, the most significant difference between them being in the contribution of intra- and interresidue components (compare Figs. 1e and f with 1g and h). It is beyond the scope of the current work to discuss the details of the differences between R and R^x (James, 1991).

The most dramatic improvement in all indices occurs during the heating period of the first annealing. In the case of refinements starting with energy-minimized A-DNA (A1- and A2-series), the octamer is generally converted into the B-type conformation by the end of the first heating period. This transition was monitored by inspection of helical parameters and sugar pseudorotation angle; it is also reflected in a very sharp drop in J_{rms} (Fig. 1d). A rapid increase in RMSD relative to the starting A-DNA conformation and simultaneous decrease in RMSD relative to the rMD structure MD_{fin} also confirm the A to B transition at a very early stage of refinement (Fig. 1c).

All subsequent annealings lead to smaller improvements in the agreement indices (Table 2, Fig. 1); presumably, these improvements are achieved by introduction of additional degrees of freedom (Table 1). The anticorrelated behavior of these indices with the conformational energy (Fig. 1b) suggests that, for a particular set of conditions, the simulated molecule achieves an optimal compromise between low conformational energy and low restraints violation energy. In fact, this compromise hardly depends on the particular protocol used as indicated by the similarity of all final structures. Furthermore, consider the pair of averaged structures A1.2 and A2.1, which were refined with the one-parameter sugar model and flexible terminal base pairs (Table 1). Both structures manifest a value for R_{dev} of about 0.17 Å and similar indices R and R^x (Table 2). Structure B1, which was also refined with the one-parameter sugar model but with fixed terminal base pairs (Table 1) has a somewhat larger R_{dev} of 0.185 Å. However, the value of R_{dev} can be decreased easily to 0.174 Å after the restrained minimization of B1 with flexible terminal base pairs; the indices R and R^x change almost imperceptibly during this minimization (data not shown), because they are calculated for the central six base pairs only. Also, several test simulations showed that repeated annealings with the same number of degrees of freedom (same sugar model with fixed or flexible termini) do not improve the quality of refinement (data not shown).

The only parameter in the protocol which significantly affected the results of refinement was the total length of annealings. The optimal annealing length apparently depends on the number

of variables in the system, despite the fact that each iteration includes incrementation of all variables; for example, instead of annealing protocol A2.2 (Table 1), if we performed the annealing with the same parameters but with a total length of 22 000 iterations (similar to annealing protocol A2.1, Table 1), a rather different structure resulted. Even though R_{dev} and the other indices of the structure thus obtained were similar to those of the final structures, the conformational energy was quite high, mainly because the torsion angle γ of the A5–A6 dinucleotide moiety was trapped in a high-energy minimum, characterized as *gauche*$^-$. The subsequent restrained minimizations could not significantly improve the structure. It is possible that the cooling period of the annealing is responsible, the length of which may be critical for location of the true minimum. However, we did not explore this hypothesis with more calculations.

Finally, it is important to acknowledge that the annealing procedure itself is necessary for correct refinement, at least, if the starting structure is far from the true minimum. We performed 20 000 iterations of rMC simulation at a temperature of 300 K and $k_{NOE} = 20$ kcal/mol · Å^2 starting from the A-conformation. All other parameters of refinement were similar to the A2.1 annealing protocol (i.e. one-parameter sugar model and flexible termini). The structure converts into the B-type conformation during simulation. However, the energy of the structure was unrealistically high, and the agreement indices were higher than those of structures refined with the same degrees of freedom (A1.2 and A2.1).

Dependence of R_{dev} on averaging procedure — insights into parameter accuracy

The agreement indices shown in Fig. 1 were calculated for 'average' structures, i.e. for structures constructed from the helical parameters which were averaged for any particular fragment of the Markov chain (see Methods). However, there are other ways to calculate average values of the indices. We have employed four different definitions to compute the average distance deviation for each fragment of the Markov chain. First, R_{dev} was calculated for each structure of the chain individually, and the results were averaged, yielding $R_{dev}^{(1)}$. Second, instead of averaging the individual R_{dev} values, we averaged the restrained interproton distances along the Markov chain; the distance deviation calculated for these averaged distances gives $R_{dev}^{(2)}$. Index $R_{dev}^{(3)}$ was calculated similar to $R_{dev}^{(2)}$, but with $<r^{-6}>$ averaging of distances. And finally, the structure itself was averaged along each fragment of the Markov chain (at constant temperature and force constant), as described in the Methods section, and $R_{dev}^{(4)}$ was calculated as the distance deviation for this single average structure. Deviation $R_{dev}^{(4)}$ is shown in Fig. 1b for the A2 refinement series. The distance deviations reported in Table 2 are similar to $R_{dev}^{(4)}$, but they are calculated for the structures which were further restrained-minimized after the averaging procedure.

We found small but reproducible differences between these four distance deviation parameters. The following inequality holds for averaging of all fragments of the Markov chain:

$$R_{dev}^{(3)} \leq R_{dev}^{(4)} \leq R_{dev}^{(2)} < R_{dev}^{(1)} \tag{4}$$

We find this inequality not quite trivial and think that it deserves some consideration. A typical example of these four indices is: $R_{dev}^{(3)} = 0.130$, $R_{dev}^{(4)} = 0.131$, $R_{dev}^{(2)} = 0.133$ and $R_{dev}^{(1)} = 0.144$; these numbers were computed for the last 20 000 iterations of annealing protocol A2.2 (Table 1).

We can consider some individual aspects of Inequality 4. One of the most peculiar relationships is the one between $R_{dev}^{(4)}$, calculated for the average structure, and $R_{dev}^{(1)}$, which was averaged for all

structures of the simulated ensemble. Indeed it is not obvious why the average structure has reasonable values of E_{conf} and E_{restr} at all. Before we first applied the averaging procedure, we were sceptical about it, because the DNAminiCarlo program requires the input of helical parameters of sufficient quality to close the backbone chains. However, the average structures appeared to be not only stereochemically reasonable, they always had a better energy (both E_{conf} and E_{restr}) than the mean values of energy calculated for the ensemble; this was true even without any additional minimization of the average structure. In the last fragment of the A2.2 annealing protocol mentioned above, the energy of the average structure was 64.8 kcal/mol lower than the mean value of energy along the Markov chain (E_{conf} and E_{restr} improved by 48.5 and 16.3 kcal/mol, respectively). Subsequent restrained minimization of the average structure decreased E_{conf} and E_{restr} by an additional 12.4 and 1.9 kcal/mol, respectively; this particular minimization yielded the final structure $A2_{fin}$. However, R_{dev} increased slightly from 0.131 to 0.133 Å (Table 2).

To explain these observations, we must assume that along each individual fragment of the Markov chain, for a particular temperature and force constant: (1) all simulated structures remain within the 'sphere of gravity' of the same potential minimum, and (2) within this minimum potential well, the energy surface can be approximated by some effective harmonic potential with respect to independent variables used. The first assumption is quite reasonable, taking into account the relatively large force constants k_{NOE} employed. The second assumption was shown to be valid in energy calculations of bending rigidities even with a more loosely defined minimum of B-family DNA conformations (Ulyanov and Zhurkin, 1984). With these assumptions, averaging of helical parameters should give a structure which is close to the equilibrium point (potential minimum), and this is exactly what we found in our rMC simulations. We can see, however, that the average structure, while being close to the minimum, does not precisely coincide with it, because the energy was further improved by restrained minimization. There are two possible explanations for this: insufficient length of the fragment being averaged (20 000 iterations in this case) or asymmetry in the effective potential with respect to some of the helical parameters. While we cannot completely exclude the first explanation, the second one is more likely; for example, such asymmetry has been found in the bending energy of B-DNA (Ulyanov and Zhurkin, 1984). It was calculated that DNA bending rigidities differ when DNA is bent into the minor or major groove (and depend also on nucleotide sequence). This difference in rigidities, even though small, led to a difference between average bending angle and that of the energy-minimum conformation. Such an asymmetry in potential energy with respect to bending angle (and possibly to other parameters as well) accounted for the improvement observed after minimization of the averaged structures.

As another consequence of the harmonicity of the effective potential, we can estimate the difference between mean energy of the Boltzmann ensemble and the minimum energy. This difference depends only on the temperature and number of degrees of freedom in the system: $\Delta E = nRT/2$. For the case of the A2.2 annealing protocol with ten degrees of freedom per sugar ring (Table 1), $n = 266$; this gives $\Delta E = 79.8$ kcal/mol at 300 K. Remarkably, that value corresponds quite closely to the figure obtained in the simulation: the difference between the mean total energy of this fragment of the Markov chain and the energy of $A2_{fin}$ is 78.9 kcal/mol. This coincidence shows that 20 000 iterations are probably sufficient for correct averaging in this particular system. (Although, of course, 'correct' averaging is not required for our refinement procedure.)

Now we proceed to the other part of Inequality 4: $R_{dev}^{(3)} \leq R_{dev}^{(2)}$. The difference between these two was very small but was nevertheless reproduced in absolutely all fragments of the Markov chain. The $<r^{-6}>$ averaging used in the definition of $R_{dev}^{(3)}$ gives smaller distances compared to the simple arithmetic averaging used for $R_{dev}^{(2)}$. The fact that the averaging procedure, which is biased towards smaller distances, *always* gives rise to a smaller deviation from the experimental restraints, implies that the latter are somewhat underestimated. This is not surprising, because the experimental 2D NOE intensities reflect the $<r^{-6}>$ averaging over the conformational ensemble and, therefore, the NOE-derived distances must be underestimated due to thermal fluctuations of DNA structure. In our rMC simulations, a very small magnitude of the improvement in $R_{dev}^{(3)}$ relative to $R_{dev}^{(2)}$ may be explained by small amplitudes of thermal fluctuations due to the relatively high force constants used in the refinement.

The relationship between $R_{dev}^{(4)}$ and two other indices, $R_{dev}^{(2)}$ and $R_{dev}^{(3)}$, are understood least of all. Inequality 4 shows that interproton distances in the average structure are somewhat larger than the $<r^{-6}>$ average but smaller than $<r>$ (again, because the decreased distances are in better agreement with the experimental restraints). If we denote the translational and angular parts of the independent variables by $\mathbf{t}$ and Θ, respectively, then the difference between $R_{dev}^{(2)}$ and $R_{dev}^{(4)}$ will be the result of differences in $<r(\mathbf{t},\Theta)>$ and $r(<\mathbf{t}>,<\Theta>)$ averaging. The translational variables are unlikely to affect the two averagings differently. However, in the case of symmetrically distributed angular variables Θ, it is easy to show that the sinusoidal dependence of r on Θ leads to values of $<r(\Theta)>$ which are underestimated compared to $r(<\Theta>)$, contrary to what was observed in our rMC simulations. This discrepancy suggests that the distribution of angular variables is not symmetric, which agrees with the observation that the average structure does not coincide exactly with the minimum energy structure (vide supra). However, the differences between $R_{dev}^{(4)}$, $R_{dev}^{(2)}$ and $R_{dev}^{(3)}$ are very small, which precludes drawing any definite conclusion.

Characteristics of final rMC structures and comparison with the rMD structure

Table 3 compares atomic RMSD values for structures refined via rMC and rMD. The MC_{interm} structures are specifically those refined using protocols B1 through B4, A1.1, A1.2 and A2.1 (see Tables 1 and 2); each of these utilized a reduced set of independent variables. The MC_{fin} structures

TABLE 3
ATOMIC ROOT-MEAN-SQUARE DEVIATIONS (Å) BETWEEN STARTING MODELS AND STRUCTURES OBTAINED FROM RESTRAINED MONTE CARLO AND RESTRAINED MOLECULAR DYNAMICS CALCULATIONS[a]

	A	MC_{interm}[b]	MC_{fin}	MD_{fin}
B	2.0	0.9–1.3	0.9–1.0	1.0
A		2.5–2.9	2.5–2.6	2.5
MC_{interm}		0.3–0.5	0.3–0.7	0.5–0.7
MC_{fin}			0.2–0.3	0.5

[a] The structures were superimposed and the RMSD values were calculated for the heavy atoms of the internal residues. In the case of groups of structures, the range of RMSD values is given.

[b] MC_{interm} structures are those refined with a reduced set of independent variables, i.e. structures resulting from protocols B1 through B4, A1.1, A1.2 and A2.1 (see Tables 1 and 2).

result from protocols B5, A1.3 and A2.2, i.e. from subsequent refinement of the MC_{interm} structures using more parameters (Tables 1 and 2).

As already mentioned, our method of rMC refinement manifested a remarkable convergence: pair-wise RMSD values between the four different final structures are within 0.3 Å (Table 3), structures refined from A- and B-DNA conformations being essentially the same. We attribute this success of the rMC refinement to the use of internal coordinates to define the DNA structure. First, use of internal coordinates reduces the number of degrees of freedom in the system at the expense of 'conformationally uninteresting' variables (see introductory comments). Second, we greatly simplify the potential energy profile, which is often pseudoharmonic for helical parameters (vide supra) while it can presumably be very complicated for each individual Cartesian coordinate. Finally, it is important to consider the nature of the internal coordinates employed. The generalized helical parameters define the *relative* position of neighboring base pairs (see Methods). That means, if one needs to change, e.g., helical rise in the middle of a DNA fragment, it can be done by changing a single variable, theoretically in a single step of Markov chain. In contrast, Cartesian coordinates used in rMD protocols are absolute variables. Thus, the same procedure of changing helical rise requires moving all atoms in half of the molecule; this could be especially aggravating when dealing with longer oligonucleotides. In other words, using the *relative* internal coordinates eliminates inertia from the system.

Remarkably, the final structures from rMC converged to within 0.5 Å RMSD of MD_{fin}, the 'final' structure obtained by rMD refinement, despite the different force fields used and despite the fact that MD refinement was conducted with additional restraints imposed on the endocyclic torsion angles of deoxyriboses derived from DQF-COSY analysis; in contrast, only the distance restraints were used in the rMC refinement (see Methods). Nevertheless, the final rMC structures differ slightly more from MD_{fin} than between themselves (Table 3), which may be attributed to the absence of torsion-angle restraints in the rMC refinement. The superimposed rMC structures

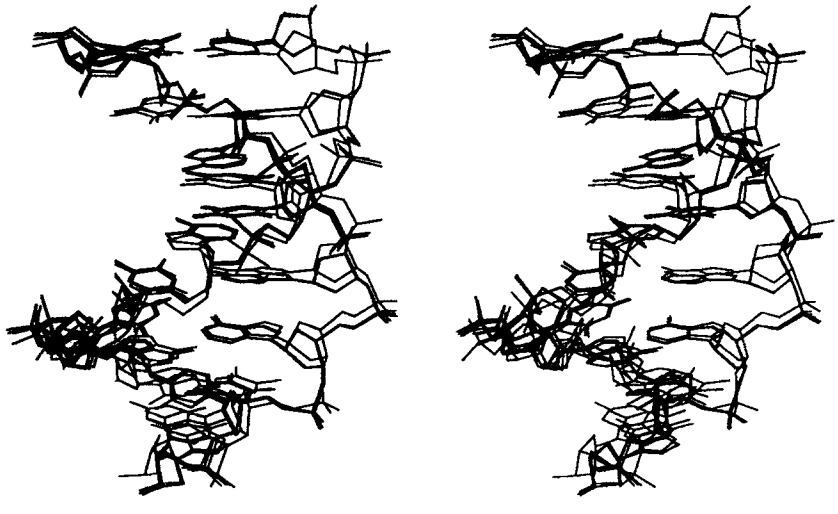

Fig. 2. Superposition of three structures of d(GTATAATG) · d(CATTATAC) resulting from restrained MC calculations, B_{fin}, $A1_{fin}$ and AB_{fin}, and the rMD structure MD_{fin}. The G1 residue is located at the upper left corner. All four structures are drawn with the same line width; the most deviant structure is MD_{fin}.

look essentially like a single structure drawn by a thick line (Fig. 2), while some differences are seen between them and MD_{fin}. These differences are more significant in the sugar-phosphate backbone (see also Table 4) and, of course, in the terminal residues, which had different restraints in the rMC and rMD refinements (see Methods).

Another conclusion which can be drawn from inspection of atomic coordinate RMSD values between the different structures is that those refined with fewer variables (i.e. MC_{interm} structures in Table 3) deviate more, but still < 0.7 Å, from both MD_{fin} and the final rMC structures (see also Fig. 1c). Moreover, convergence among the MC_{interm} structures, 0.3 to 0.5 Å, is slightly worse than among the MC_{fin} structures (Table 3). This observation is in accord with the differences in helical parameters between the different groups of structures (data not shown). More importantly, introduction of a more flexible model for sugar ring and exocyclic bond angles leads to an improved agreement with experimental NOE intensities, which is especially noticeable in the case of interresidue cross peaks (Figs. 1e–h and Table 2). On first sight, it is a rather logical result, because it is easier to fit a molecule to the experimental data with more variables. However, further inspection of the agreement indices in Table 2 shows that the MD_{fin} structure exhibits lower R and R^x indices — the intraresidue indices especially — even compared to the MC_{fin} structures which were refined with the most flexible sugar model. To rationalize this observation, we must consider what additional degrees of freedom a DNA duplex possesses in MD simulations compared to our MC refinement. Even though bond lengths and angles can be constrained during calculations of MD trajectories by the SHAKE routine (Ryckaert et al., 1977), this procedure is not used in the restrained energy minimization, which is the last step of a typical MD-based refinement (Schmitz et al., 1992b). In many cases, variation of bond lengths and angles must be considered as non-essential; nevertheless, it does improve the apparent agreement with experimental restraints. To assess the impact of X-Y-H bond-angle distortion, we substituted the protons of the MD_{fin} structure using idealized values for bond angles and bond lengths. The parameters of the resulting structure, MD_{subst}, are listed in Table 2. The indices R and R^x increased to some extent for MD_{subst} compared to MD_{fin}; the sixth-root weighted indices R^x are somewhat higher and indices R somewhat lower for MD_{subst} relative to the final structures from rMC.

Further analysis of the MD_{fin} structure revealed deviations from planarity for some bases, which also contributed to the improvement in apparent agreement with the experimental data. One of the most distorted bases in MD_{fin} is A5 (Fig. 3); the reason for this distortion is elucidated by analysis of experimental distance restraints imposed on A5. The restraints H8A5–H8A6

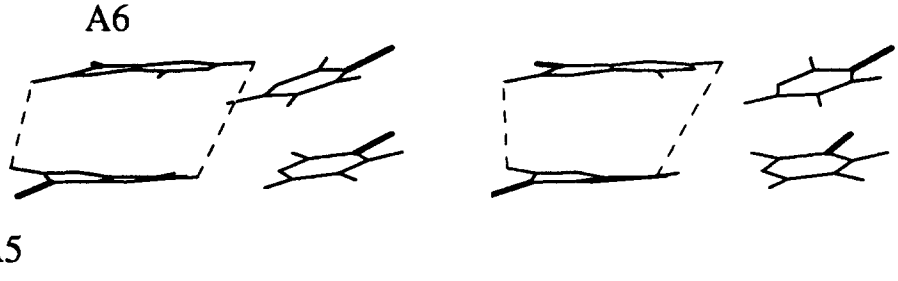

Fig. 3. Base-pair step A5–A6 · T11–T12 of the rMD structure, MD_{fin}, viewed from the minor groove. Glycosidic bonds are emphasized; dashed lines show the contacts H8A5–H8A6 and H2A5–H2A6.

TABLE 4
COMPARISON OF LOCAL HELICAL PARAMETERS FOR THE INNER SIX BASE PAIRS OF THE rMD AND rMC STRUCTURES OF d(GTATAATG)·d(CATTATAC)[a]

Step (wedge) local parameters

Step	Twist	Tilt	Roll	D_x	D_y	D_z
T2–A3	37.5 (1.5)	−0.2 (1.4)	0.8 (2.5)	−0.73 (0.41)	−0.40 (0.22)	2.69 (0.29)
A3–T4	30.4 (0.2)	−0.8 (1.2)	−5.2 (6.7)	0.06 (0.25)	−0.58 (0.22)	2.99 (0.09)
T4–A5	40.0 (7.5)	−2.9 (4.5)	4.6 (2.8)	0.28 (0.14)	−0.18 (0.56)	2.88 (0.31)
A5–A6	32.4 (2.7)	2.9 (1.2)	7.1 (2.9)	0.18 (0.10)	−0.66 (0.14)	2.90 (0.13)
A6–T7	31.2 (2.4)	−0.3 (2.2)	−5.9 (1.6)	0.76 (0.46)	−0.85 (0.13)	3.14 (0.13)
Max. dev.	7.5	4.5	6.7	0.46	0.56	0.31

Base pair parameters

Pair	Propeller	Buckle	Opening	S_x	S_y	S_z
T2·A15	−4.5 (3.4)	−5.9 (2.5)	−0.3 (1.3)	0.05 (0.07)	−0.08 (0.09)	−0.27 (0.24)
A3·T14	−1.4 (3.0)	−0.7 (2.9)	1.5 (0.6)	0.13 (0.11)	−0.06 (0.07)	−0.15 (0.12)
T4·A13	−3.0 (1.6)	2.9 (4.0)	1.3 (1.3)	−0.04 (0.13)	−0.08 (0.09)	−0.14 (0.12)
A5·T12	−12.0 (4.3)	3.6 (3.9)	0.7 (4.0)	−0.01 (0.06)	−0.11 (0.09)	−0.21 (0.28)
A6·T11	−4.0 (1.5)	1.3 (2.1)	1.7 (1.0)	−0.01 (0.07)	−0.13 (0.09)	−0.31 (0.10)
T7·A10	−0.8 (2.7)	1.9 (1.8)	−2.4 (1.6)	0.01 (0.07)	−0.10 (0.04)	−0.28 (0.13)
Max. dev.	4.3	4.0	4.0	0.13	0.09	0.28

Backbone parameters

Residue	P	τ_{max}	χ	γ	β	α	ε	ζ
T2	146 (11)	35 (4)	241 (6)	56 (3)	192 (1)	295 (13)	259 (4)	182 (3)
A3	147 (1)	34 (4)	257 (3)	55 (2)	182 (1)	294 (9)	259 (2)	181 (7)
T4	133 (24)	33 (13)	257 (5)	53 (7)	183 (10)	294 (4)	262 (8)	174 (1)
A5	149 (39)	36 (4)	264 (8)	53 (2)	188 (7)	299 (2)	270 (43)	165 (19)
A6	143 (27)	36 (3)	258 (4)	45 (3)	184 (7)	293 (4)	271 (6)	175 (7)
T7	129 (32)	34 (4)	242 (14)	56 (3)	176 (24)	296 (4)	263 (23)	173 (33)
A10	152 (15)	33 (4)	253 (5)	52 (8)	191 (35)	293 (25)	256 (28)	182 (22)
T11	128 (19)	32 (3)	244 (3)	61 (6)	178 (9)	298 (2)	260 (9)	171 (12)
T12	133 (16)	38 (4)	246 (8)	54 (1)	174 (2)	294 (3)	269 (8)	179 (4)
A13	147 (12)	34 (3)	256 (2)	53 (8)	184 (11)	292 (8)	259 (28)	179 (18)
T14	134 (33)	34 (2)	254 (14)	52 (1)	182 (14)	291 (3)	260 (6)	176 (14)
A15	145 (0)	38 (8)	241 (20)	64 (4)	171 (5)	291 (3)	264 (16)	174 (11)
Max. dev.	39	13	20	8	24	9	43	33

[a] Translations are given in Å, rotations are given in degrees. Helical parameters for the MD_{fin} structure were calculated by the Fitparam program (N.B. Ulyanov, unpublished); those for the four rMC structures were calculated as described earlier (Ulyanov et al., 1992). Positive signs of base-pair parameters correspond to the conformations depicted in the report of the Cambridge convention (Dickerson et al., 1989); these signs are in accord with definitions used in the CURVES program (see Mauffret et al., 1992). Parameters for the nonterminal residues of MD_{fin} are shown; numbers in parentheses refer to the maximum deviations from MD_{fin} among the four rMC structures. Pseudorotation angle P and pucker amplitude τ_{max} are calculated according to Altona and Sundaralingam (1972). Glycosidic angles χ are defined as torsions O1′-C1′-N9-C4 for purines and O1′-C1′-N1-C2 for pyrimidines. Torsion angles for each residue refer to the backbone which connects this residue with the 5′-neighboring one.

(2.69–2.99 Å) and H2A5–H2A6 (3.16–3.39 Å) can hardly be satisfied simultaneously by any single conformation. In rMC structures with idealized base geometries, these two distances are 3.61–3.64 and 3.98–3.99 Å, respectively; but they are only 3.52 and 3.75 Å in MD_{fin}, largely due to distortion of A5. The latter values are the result of a compromise between repulsive van der Waals forces and the pseudoforces which are trying to satisfy the imposed experimental restraints. In our view, the internally inconsistent experimental restraints are an indication of internal dynamics in the octamer. In this particular case, it was shown that H2A5 is involved in a dynamic process as experimentally manifested in $T_{1\rho}$ measurements (Schmitz et al., 1992b). Similar indications of internal motion in the TpA junction were recently reported for another oligonucleotide (Kim and Reid, 1992). Our theoretical investigation of dynamics in the octamer d(GTATAATG) · d(CATTATAC) via MD simulations with time-averaged restraints (Torda et al., 1990; Pearlman and Kollman, 1991; Schmitz et al., 1992a) showed that, indeed, the H8A5–H8A6 and H2A5–H2A6 distances decreased after $<r^{-3}>$ averaging (data not shown). Another example of internally inconsistent sets of distances are the H6/H8–H2′/H2″ and H6/H8-H3′ intraresidue distances. The H6/H8–H2′/H2″ distances (and 2D NOE intensities) are generally satisfied in all refined structures. Significant H6/H8–H2′/H2″ interactions are strongly associated with a major S-type deoxyribose conformer. At the same time, the H6/H8–H3′ intraresidue distances are overestimated by the resulting theoretical models vis-à-vis the experimental data; this apparent inconsistency may emanate from the north-south equilibrium of deoxyribose ring conformations (Ulyanov et al., 1992). Indeed, both sets of distances have been found to be better satisfied on a time-average basis in MD simulations of the octamer with weighted, time-averaged restraints which exhibit repuckering of the sugar rings (Schmitz et al., 1993). It should be recalled that both rMC and rMD refinements of the octamer were made without restraints involving H3′ protons; such H6/H8–H3′ intraresidue interactions would be strongly associated with a major S-type deoxyribose conformer.

We note that the main differences in helical parameters between the rMC final structures and MD_{fin} are concentrated around residue A5, most notably in the T4–A5:T12–A13 step (Table 4). However, these differences may be an artifact of ambiguities connected with extraction of local helical parameters for structures with non-flat bases. Indeed, superposition of this step for two structures, MD_{fin} and $A1_{fin}$, as illustrated in Fig. 4, shows that the two structures are more similar than might be suggested by numerical values of their helical parameters. A close similarity of these structures indicates that introduction of a more flexible backbone model (four-parameter sugar model and unfrozen exocyclic bond angles, see Methods) enables the rMC refinement to reproduce the positions of the bases obtained by rMD. However, experience with the dynamic behavior of A5 teaches us to be cautious with interpretation of these results, because the NOE-derived distance restraints may represent quantities which are $<r^{-3}>$-averaged over the conformational ensemble (or worse, $<r^{-6}>$-averaged if conformational fluctuations are sufficiently slow). If we retain structures with somewhat higher distance deviations, such as those refined with the one-parameter sugar model, then the resulting structures may vary as much as those shown in Fig. 5. Again, it is interesting that the most variable part of the octamer among these structures is in the region adjacent to the A5 residue.

Finally, we consider J_{rms}, the root-mean-square deviation between experimental and theoretical vicinal coupling constants. Figure 1d shows that J_{rms} stops improving after the early steps of the distance-driven refinement, i.e. after the deoxyriboses have converted into S-type conformations.

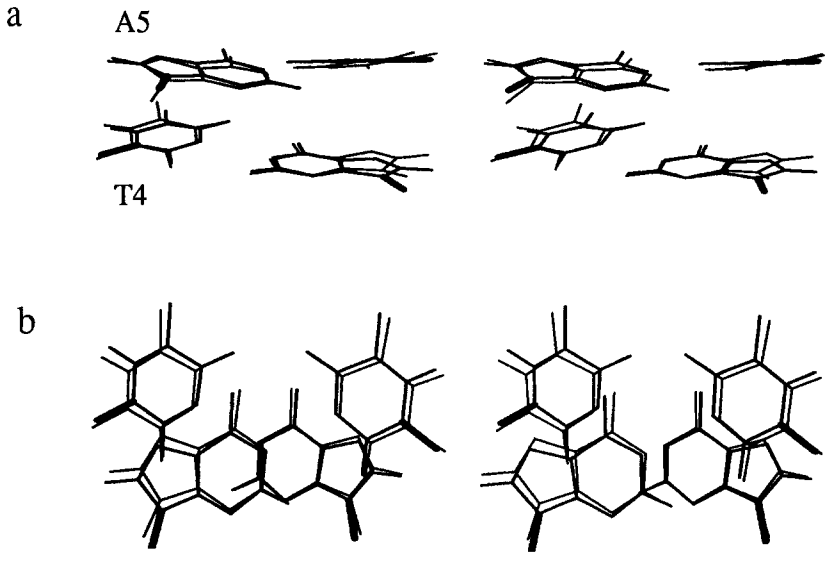

Fig. 4. Superposition of base-pair steps T4–A5 · T12–A13 in the structures $A1_{fin}$ (thick line) and MD_{fin} (thin line). Glycosidic bonds are emphasized. (a) View from the minor groove; (b) Top view.

More unexpected, however, is the observation that this index increases after additional variables are introduced into the deoxyriboses and further structure refinement is carried out (Fig. 1d). This increase in J_{rms} may indicate that, in fact, these additional degrees of freedom permit distortion of the conformation of the five-membered sugar rings, which is not necessarily consonant with the existing experimental data. Furthermore, no single structure, including MD_{fin} which was refined with torsion-angle restraints derived from the experimental coupling constants (Schmitz et al., 1992b), has a J_{rms} value below 0.5 Hz (see Table 2), which is an estimate of the precision of calculating vicinal coupling constants (Haasnoot et al., 1980). In contrast, we calculate that J_{rms} is less than 0.5 Hz for an 80/20% mixture of d(GTATAATG) · d(CATTATAC) in B- and A-DNA forms, respectively. (Details of this calculation are omitted here.) This strongly suggests that one should account for the dynamic S-N equilibrium of deoxyriboses in order to accommodate the experimentally observed coupling constants for the octamer.

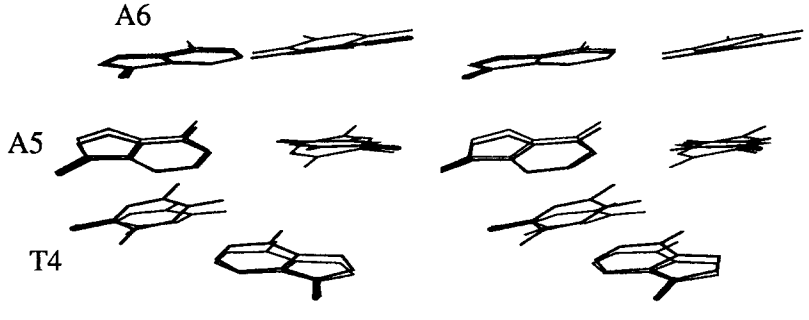

Fig. 5. Superposition of base pairs T4 · A13, A5 · T12 and A6 · T11 in the structures $A1_{fin}$ (thick line) and A1.2 (thin line) viewed from the minor groove; glycosidic bonds are emphasized.

CONCLUSIONS

A method for the refinement of duplex DNA solution structures is described in this paper. This method has the following features: (1) interproton distances and bounds obtained via complete relaxation matrix analysis of 2D NOE spectra using the MARDIGRAS algorithm comprise the experimental restraints; (2) DNA conformations are generated by a helical-parameter-based algorithm; (3) an annealing process is simulated by generating Boltzmann ensembles of structures via the Metropolis Monte Carlo procedure. It has been previously demonstrated that interproton distances derived from 2D NOE intensities by the MARDIGRAS algorithm constitute a set of experimental distance restraints (including bounds) with the requisite accuracy for determination of high-resolution solution structures (Liu et al., 1992; Mujeeb et al., 1992; Weisz et al., 1992). This conclusion is confirmed in the present study on the octamer d(GTATAATG) · d(CATTATAC).

Simulations and energy minimizations were performed by the DNAminiCarlo program, which uses generalized helical parameters to calculate the structure of the DNA double helix. This approach enforces the proper stereochemistry of the DNA molecule — the planarity of bases in particular, which may pose a problem in Cartesian-coordinate-based algorithms, due to the relatively high contribution of pseudoenergy from the experimental restraints. The annealing procedure permits the simulated DNA molecule to convert easily from a starting structure of A-DNA to a structure more like the canonical B-DNA conformation, which is more closely related to the final structure; this is observed to occur rather early in the simulation. This easy transition is in distinct contrast to our experience with restrained MD calculations, where the conversion is difficult to effect in the absence of explicit experimental sugar torsion-angle restraints.

Using different starting structures and annealing protocols, the rMC method of refinement has led to a number of structures which are quite similar to that refined by a traditional rMD approach, despite the different force fields used in the two methods. This similarity to the rMD structure is quite good when a simple one-parameter sugar model is employed (RMSD < 0.7 Å) but is especially striking when a more flexible model of deoxyriboses is utilized in the rMC refinements (RMSD < 0.5 Å). However, additional investigation is required to clarify the question of what sugar model should generally be used for solution structure determinations via rMC. (Although it would appear that any model described herein would certainly be more than adequate.) Proper description of the deoxyribose conformation in solution may be compromised by internal motion; this can manifest itself in comparisons between experimental 2D NOE intensities or vicinal coupling constants and theoretical simulations based on refined structures or ensembles of structures. The presence of such internal motions is suggested by the existence of some mutually inconsistent groups of experimental restraints for the octamer. An analysis of mutually contradicting average interproton distances may clarify some aspects of the dynamic equilibrium of different DNA conformers; such a study is in progress.

The rMC simulated-annealing protocol requires about 45 min of CPU time on a Cray Y-MP supercomputer when using the one-parameter sugar model; the protocol with the most flexible model of deoxyribose requires about 4 CPU h. These figures are comparable to the computational time required for an rMD simulation of the octamer. In summary, we consider the method proposed here to be a worthy alternative to the traditional methods of DNA structural refinement.

ACKNOWLEDGEMENTS

This work was supported by National Institutes of Health Grants GM41639, GM39247 and RR01695. We are grateful to Drs. Klaus Weisz and Victor Zhurkin for useful discussions. We gratefully acknowledge use of the Computer Graphics Laboratory at the University of California, San Francisco (supported by NIH Grant RR01081). We also acknowledge use of the Cray Y-MP supercomputer, supported by a grant from the Pittsburgh Supercomputing Center through the NIH Division of Research Resources cooperative agreement U41RR04154 and a grant from the National Science Foundation Cooperative Agreement ASC-8500650.

REFERENCES

Altona, C. and Sundaralingam, M. (1972) *J. Am. Chem. Soc.*, **94**, 8205–8212.
Baleja, J.D., Pon, R.T. and Sykes, B.D. (1990) *Biochemistry*, **29**, 4828–4839.
Borgias, B.A. and James, T.L. (1988) *J. Magn. Reson.*, **79**, 493–512.
Borgias, B.A. and James, T.L. (1989) *Methods Enzymol.*, **176**, 169–183.
Borgias, B.A. and James, T.L. (1990) *J. Magn. Reson.*,, **87**, 475–487.
Braun, W. (1987) *Q. Rev. Biophys.*, **19**, 115–157.
Brünger, A.T. and Karplus, M. (1991) *Acc. Chem. Res.*, **24**, 54–61.
Chuprina, V.P., Khutorskii, V.E. and Poltev, V.I. (1981) *Stud. Biophys.*, **85**, 81–88.
De Leeuw, F.A.A.M., Van Beuzekom, A. and Altona, C. (1983) *J. Comp. Chem.*, **4**, 438–448.
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. and Zhurkin, V.B. (1989) *EMBO J.*, **8**, 1–4.
Gorin, A.A., Ulyanov, N.B. and Zhurkin, V.B. (1990) *Molek. Biol. (Eng. transl.)*, **24**, 1036–1047.
Gupta, G., Bansal, M. and Sasisekharan, V. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 6486–6490.
Haasnoot, C.A.G., De Leeuw, F.A.A.M. and Altona, C. (1980) *Tetrahedron*, **36**, 2783–2792.
Havel, T.F., Kuntz, I.D. and Crippen, G.M. (1983) *Bull. Math. Biol.*, **45**, 665–720.
James, T.L. (1991) *Curr. Opin. Struct. Biol.*, **1**, 1042–1053.
James, T.L., Gochin, M., Kerwood, D.J., Pearlman, D.A., Schmitz, U. and Thomas, P.D. (1991) In *Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy* (Eds, Hoch, J.C., Poulsen, F.M. and Redfield, C.) Plenum Press, New York, pp. 331–347.
Keepers, J.W. and James, T.L. (1984) *J. Magn. Reson.*, **57**, 404–426.
Kerwood, D.J., Zon, G. and James, T.L. (1991) *Eur. J. Biochem.*, **197**, 583–595.
Kim, S.-G. and Reid, B.R. (1992) *Biochemistry*, **31**, 12103–12116.
Kumar, A., James, T.L. and Levy, G.C. (1992) *Isr. J. Chem.*, **32**, 257–261.
Lavery, R. (1988) In *DNA Bending and Curvature* (Eds, Olson, W.K., Sarma, M.H., Sarma, R.H. and Sundaralingam, M.) Vol. 3, Adenine Press, New York, pp. 191–211.
Levy, R.M., Bassolino, D.A., Kitchen, D.B. and Pardi, A. (1989) *Biochemistry*, **28**, 9361–9372.
Liu, H., Thomas, P.D. and James, T.L. (1992) *J. Magn. Reson.*, **98**, 163–175.
Mauffret, O., Hartmann, B., Convert, O., Lavery, R. and Fermandjian, S. (1992) *J. Mol. Biol.*, **227**, 852–875.
McCammon, J.A. and Harvey, S.C. (1987) *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, Cambridge.
Metropolis, N.A., Rosenbluth, A.M., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953) *J. Chem. Phys.*, **21**, 1087–1092.
Mujeeb, A., Kerwin, S.M., Egan, W., Kenyon, G.L. and James, T.L. (1992) *Biochemistry*, **31**, 9325–9338.
Nikonowicz, E. and Gorenstein, D.G. (1992) *J. Am. Chem. Soc.*, **114**, 7494–7503.
Nilges, M., Clore, G.M., Gronenborn, A., Piel, N. and McLaughlin, L.W. (1987) *Biochemistry*, **26**, 3734–3744.
Nilsson, L., Clore, G.M., Gronenborn, A., Brünger, A.T. and Karplus, M. (1986) *J. Mol. Biol.*, **188**, 455–475.
Olson, W.K. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 1775–1779.
Oshiro, C.M., Thomason, J.F. and Kuntz, I.D. (1991) *Biopolymers*, **31**, 1049–1064.

Pearlman, D.A. and Kollman, P.A. (1991) *J. Mol. Biol.*, **220**, 457–479.
Pearlman, D.A., Case, D.A., Caldwell, J., Seibel, G.L., Singh, U.C., Weiner, P.K. and Kollman, P.A. (1991) *AMBER 4.0*, University of California, San Francisco.
Poltev, V.I. and Shulyupina, N.V. (1986) *J. Biomol. Struct. Dyn.*, **3**, 739–765.
Ripoll, D.R. and Ni, F. (1992) *Biopolymers*, **32**, 359–365.
Ryckaert, J.P., Cicotti, G. and Berendsen, H.J.C. (1977) *J. Comp. Phys.*, **23**, 327–341.
Schmitz, U., Kumar, A. and James, T.L. (1992a) *J. Am. Chem. Soc.*, **114**, 10654–10656.
Schmitz, U., Sethson, I., Egan, W. and James, T.L. (1992b) *J. Mol. Biol.*, **227**, 510–531.
Schmitz, U., Ulyanov, N.B., Kumar, A. and James, T.L. (1993) *J. Mol. Biol.*, in press.
Stolarski, R., Egan, W. and James, T.L. (1992) *Biochemistry*, **31**, 7027–7042.
Thomas, P.D., Basus, V.J. and James, T.L. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 1237–1241.
Torda, A.E., Scheek, R.M. and Van Gunsteren, W.F. (1990) *J. Mol. Biol.*, **214**, 223–235.
Ulyanov, N.B. and Zhurkin, V.B. (1982) *Molek. Biol. (Eng. transl.)*, **16**, 857–867.
Ulyanov, N.B. and Zhurkin, V.B. (1984) *J. Biomol. Struct. Dyn.*, **2**, 361–385.
Ulyanov, N.B., Gorin, A.A. and Zhurkin, V.B. (1989) In *Proc. Int. conf. Supercomp. '89: Supercomputer Applications* (Eds, Kartashev, L.P. and Kartashev, S.I.) Int. Supercomputing Inst., Inc., St. Petersburg, Florida, pp. 368–370.
Ulyanov, N., Gorin, A.A., Zhurkin, V.B., Chen, B.-C., Sarma, M.H. and Sarma, R.H. (1992) *Biochemistry*, **31**, 3918–3930.
Van Gunsteren, W.F., Boelens, R., Kaptein, R. and Zuiderweg, E.R.P. (1983) In *Nucleic Acid Conformation and Dynamics, NATO/CECAM Workshop Report* (Ed. Olson, W.K.) Orsay, pp. 79–82.
Weisz, K., Shafer, R.H., Egan, W. and James, T.L. (1992) *Biochemistry*, **31**, 7477–7487.
Zhurkin, V.B., Lysov, Y.P. and Ivanov, V.I. (1978) *Biopolymers*, **17**, 377–412.
Zhurkin, V.B., Poltev, V.I. and Florentiev, V.L. (1981) *Molek. Biol. (Eng. transl.)*, **14**, 882–895.
Zhurkin, V.B., Ulyanov, N.B., Gorin, A.A. and Jernigan, R.L. (1991) *Proc. Natl. Acad. Sci. USA*, **88**, 7046–7050.