J. Mol. Biol. (1986) **188**, 455–475

Structure Refinement of Oligonucleotides by Molecular Dynamics with Nuclear Overhauser Effect Interproton Distance Restraints: Application to 5′ d(C-G-T-A-C-G)$_2$

Lennart Nilsson[1,3], G. Marius Clore[2], Angela M. Gronenborn[2] Axel T. Brünger[1,2] and Martin Karplus[1]

[1] *Department of Chemistry, Harvard University Cambridge MA 02138, U.S.A.*

[2] *Max-Planck-Institute für Biochemie D-8033 Martinsried bei München, F.R.G.*

[3] *Department of Medical Biophysics Karolinska Institute S-10401 Stockholm, Sweden*

(Received 25 November 1985)

The solution structure of the self-complementary DNA hexamer 5′ d(C-G-T-A-C-G)$_2$ is refined by restrained molecular dynamics in which 192 interproton distances, determined from pre-steady-state nuclear Overhauser enhancement measurements, are incorporated into the total energy of the system in the form of effective potentials. First the method is tested by applying an idealized set of distance restraints taken from classical *B*-DNA to a simulation starting off from *A*-DNA and *vice versa*. It is shown that in both cases the expected transition between *A*- and *B*-DNA occurs. Second, a set of restrained molecular dynamics calculations is carried out starting from both *A*- and *B*-DNA with the experimental interproton distances for 5′ d(C-G-T-A-C-G)$_2$ as restraints. Convergence to the same *B*-type structure is achieved with the interproton distances equal to the measured values within experimental error. The root-mean-square atomic difference between the two average restrained dynamics structures (<1 Å) is approximately the same as the root-mean-square fluctuations of the atoms.

1. Introduction

Nuclear magnetic resonance spectroscopy is being used increasingly for obtaining information concerning the solution structures of macromolecules of biological interest (Dwek, 1976; Jardetzky & Roberts, 1981). Although n.m.r.† has long been used for structure determinations of peptides and a wide range of organic molecules, the availability of high-field n.m.r. spectrometers and the advent of two-dimensional n.m.r. spectroscopy has led to significant progress in its application of macromolecules (Aue *et al.*, 1976; Jeener *et al.*, 1976; Wider *et al.*, 1984). With the development of sequential resonance assignment strategies based on the delineation of through-bond and through-space (<5 Å) connectivities, it has become possible to obtain complete or virtually complete and unambiguous resonance assignments for small proteins and oligonucleotides (Wagner & Wüthrich, 1982; Wüthrich *et al.*, 1982; Billeter *et al.*, 1982; Strop *et al.*, 1983; Zuiderweg *et al.*, 1983; Clore & Gronenborn, 1983, 1985*a*; Weiss *et al.*, 1984; Gronenborn & Clore, 1985) and to derive a large set of approximate interproton distances by means of pre-steady-state nuclear Overhauser enhancement measurements (Wagner & Wüthrich, 1979; Dobson *et al.*, 1982; Clore & Gronenborn, 1985*b*). These interproton distance estimates comprise the data for three-dimensional structure determination. Because of the limitation in the number and the range of the available NOE distances (generally <5 Å), the construction and refinement of a protein or oligonucleotide is not straightforward.

The most direct approach is to regard the distance estimates as elements of a distance matrix and to generate the structure by use of distance geometry algorithms coupled with the approxi-

† Abbreviations used: n.m.r., nuclear magnetic resonance; NOE, nuclear Overhauser effect; r.m.s., root-mean-square.

0022–2836/86/070455–21 $03.00/0

mately known covalent structure based on bond length and bond angle values (Crippen & Havel, 1978; Kuntz *et al.*, 1979; Havel & Wüthrich, 1984). To date, this approach has been applied to glucagon (Braun *et al.*, 1983), insectotoxin I_5A (Arseniev *et al.*, 1984) and bull seminal plasma inhibitor (Williamson *et al.*, 1985). Of particular importance is a test of the method with model data derived from the crystal structure of the bovine pancreatic trypsin inhibitor (Havel & Wüthrich, 1985). The results are encouraging since the overall shape, size and folding of the polypeptide chain were reasonably well reproduced. However, the structures obtained tended to be slightly expanded relative to the X-ray structure and the local backbone conformation was poorly determined. To overcome this difficulty, it was suggested (Havel & Wüthrich, 1985; Williamson *et al.*, 1985) that energy minimization be used for refinement of the structures obtained from the distance geometry program.

An alternative approach is based on the use of energy or related restraints as part of the entire model-building process. Ideally one would begin with a structure obtained on the basis of a qualitative interpretation of the NOE data, followed by structure refinement. A possibility is to do a least-squares refinement with a penalty function including the NOE and other restraints based on the local geometry (bond lengths, bond angles, planarity of groups) as well as van der Waals' repulsion terms. This has been applied with some degree of success to three oligonucleotides (Clore & Gronenborn, 1985*c*; Clore *et al.*, 1985*a*,*b*). However, the convergence properties of the method appear to be limited to a relatively small region of conformational space. Thus, whereas convergence to very similar *B*-DNA structures was achieved starting from two somewhat different *B*-DNA structures, the method was incapable of converging to such a *B*-DNA structure starting from *A*-DNA. For this reason it is desirable to develop a refinement method with a greater convergence range. A method that satisfies the criterion and introduces energetic considerations at all stages of the refinement process is restrained molecular dynamics (Clore *et al.*, 1985*c*). This involves the simulation of the system by the solution of the classical equations of motion for all atoms for a suitable time period at a convenient temperature (McCammon *et al.*, 1977, 1979; Karplus & McCammon, 1983) with the experimental interproton distances incorporated into the total energy function of the system in the form of effective potentials (Clore *et al.*, 1985*c*; Kaptein *et al.*, 1985). Such a restrained molecular dynamics simulation has recently been applied to a 17-mer peptide comprising the DNA binding helix of the cAMP receptor protein. It was shown that, within the limits of the rather imprecise distance restraints used three different initial structures (α-helix, β-strand or 3–10 helix) all converged to very similar α-helical structures (Clore *et al.*, 1985*c*).

In the present paper we extend the previous restrained molecular dynamics method (Clore *et al.*, 1985*c*) to oligonucleotides. This is particularly important because it is known from X-ray structure determinations (Dickerson & Drew, 1981; Shakked *et al.*, 1983; Dickerson *et al.*, 1983; Wang *et al.*, 1983) that the structures of DNA oligomers in crystals differ from the idealized regular helical structures based on fiber diffraction studies (Arnott & Hukins, 1972*a*,*b*; Arnott *et al.*, 1983). Because the structural changes involved tend to be relatively small and localized, appear to be sequence-specific and may be influenced by ligands (e.g. DNA binding drugs and proteins), it is necessary to have available a method for determining the solution structure of oligonucleotides with rather higher accuracy.

In this paper we describe the restrained molecular dynamics method for oligonucleotides and apply it to the self-complementary DNA hexamer 5′ $d(C\text{-}G\text{-}T\text{-}A\text{-}C\text{-}G)_2$ for which the ^1H n.m.r. spectrum is completely assigned and 192 interproton distances, determined from pre-steady-state NOE measurements, are available (Clore & Gronenborn, 1983; Gronenborn *et al.*, 1984). This DNA oligonucleotide is presumed to be of the *B*-type in solution as judged both by circular dichroism spectroscopy and by a qualitative interpretation of the NOE data. First, we test the method with an idealized set of 192 distance restraints taken from classical *B*-DNA and *A*-DNA by starting with *A*-DNA and *B*-DNA, respectively. In both cases, the desired transition between *A*- or *B*-DNA occurs. Second, restrained molecular dynamics calculations with the experimental interproton distance restraints are carried out starting from classical *A*- and *B*-DNA. In both cases covergence to a *B*-type structure is achieved and the r.m.s. difference between the two average restrained dynamics structures is less than 1 Å.

2. Methodology

The present approach to structure refinement requires a method for energy minimization and molecular dynamics, a method for interpreting NOE data in terms of distance restraints, and finally a strategy for combining the two. Each of these elements of the refinement methodology is described briefly in this section.

(a) *Energy calculations*

All energy minimizations and molecular dynamics calculations were carried out using the program CHARMM (Brooks *et al.*, 1983) with an empirical energy function developed for nucleic acids in which all hydrogen atoms are treated explicitly (Nilsson & Karplus, 1986). The potential energy term, representing the interproton distance restraints, was added to the total energy function of the system in the form of a skewed biharmonic effective potential described below (Clore *et al.*, 1985*c*). Solvent molecules were not explicitly included in the calculations, but the effect of solvent was approximated by multiplying the electrostatic energy term by a $(1/r)$ screening function (Gelin & Karplus, 1977;

Brooks *et al.*, 1983) and by reducing the net charge on the phosphate group to −0·32e (Tidor *et al.*, 1983). The non-bonded interactions were switched off, using a cubic switching function, between 9·5 and 10·5 Å, with pairs up to 11·5 Å apart included in the non-bonded list. Integration of the equations of motion was performed by use of a Verlet integration algorithm (Verlet, 1967) with initial velocities assigned from Maxwellian distribution at the appropriate temperature. The time step of the integrator was 1 fs and the non-bonded interaction lists were updated every 20 fs. Bond lengths involving hydrogen atoms were kept fixed with the SHAKE algorithm (Ryckaert *et al.*, 1977).

(b) *Interproton distance determination*

The structure of the 6-mer was restrained by a total of 192 interproton distances determined from pre-steady-state NOE measurements (Gronenborn *et al.*, 1984). This distance set is illustrated in Fig. 1 on a *B*-DNA framework and their various categories are summarized in Table 1. As can be seen in Fig. 1, the distance restraints provide an extensive network of information that can be used to determine the structure of the 6-mer. There is, of course, a certain redundancy in this interproton distance data set, particularly as regards the intranucleotide sugar–sugar distances. Also, there are relatively few of the important inter-residue and inter-chain distances and no long-range data for determining the overall structure (e.g. bending) of the molecule. Nevertheless, the number of restraints is of the same order of magnitude as the number of degrees of freedom that are allowed by the 78 backbone (α to ξ) and glycosidic bond (χ) torsion angles and the 24 amplitude and phase parameters defining the sugar puckers. Moreover, the glycosidic (χ), C-4′–C-3′ (δ) and C-5′–C-4′ (γ) bond torsion angles can be directly related to specific interproton distances (Clore & Gronenborn, 1983, 1985*a*; Gronenborn & Clore, 1985).

The $(\langle r^{-6}\rangle)^{-1/6}$ mean interproton distances obtained by Gronenborn *et al.* (1984) were found from the equation:

$$(\langle r_{ij}^{-6}\rangle)^{-1/6} = (\sigma_{kl}/\sigma_{ij})^{1/6} \times r_{kl} \sim [N_{kl}(t)/N_{ij}(t)]^{1/6} \times r_{kl}, \quad (1)$$

where r_{ij} is the unknown distance, r_{kl} is the appropriate fixed internal reference distance, σ_{ij} and σ_{kl} are the corresponding cross-relaxation rates, and $N_{ij}(t)$ and $N_{kl}(t)$ the corresponding NOE values at a short irradiation time or mixing time t. (For a detailed discussion of the appropriate choice of reference distances for different classes of interproton distances see Gronenborn *et al.* (1984), Gronenborn & Clore (1985) and Clore & Gronenborn (1985*a*).) It is particularly important from the experimental view point that the approximation in eqn (1) remains valid for values of t up to 3 to 4 times longer than the initial rate approximation $N_{ij}(t) \sim \sigma_{ij} \times t$ (Clore & Gronenborn, 1985*b*). Nevertheless, in addition to the random experimental errors that are expected to be Gaussian in distribution, there is a small systematic error involved in the determination of interproton distances using eqn (1), due to the presence of a small amount of unavoidable spin-diffusion. This effect, as well as dynamical averaging corrections should in principle be included in the distance estimates by a full multiple-spin treatment (Olejniczak *et al.*, 1984). For practical purposes, however, a 3-spin system approximation is sufficient (Wagner & Wüthrich, 1979; Dobson *et al.*, 1982; Clore & Gronenborn, 1985*b*). Here we make use of the analysis of the systematic error introduced by spin-diffusion made by Clore & Gronenborn (1985*b*). They showed that the apparent value of the unknown distance will become closer to that of the reference distance as the extent of spin-diffusion increases. In other words, if the

Table 1
Number and type of interproton distance restraints

Type	Number	Examples
Intranucleotide		
sugar–sugar	108	H-1′(i)-H-2′/H-2″/H-4′(i)
base–sugar	34	H-1′/H-2′/H-3′(i)-H-8/H-6(i)
Internucleotide/intrastrand		
sugar–base	34	H-1′/H-2′/H-2″(i)-H-8/H-6($i+1$)
base–base	4	H-8/H-6(i)-H-5/CH_3($i+1$)
Internucleotide/interstrand		
Interbase pair	6	A(H-2)-T(H-3), G(H-1)-C(NH_2)
interbase pair	6	T(H-3)/A(H-2)-G(H-1)
Total	192	

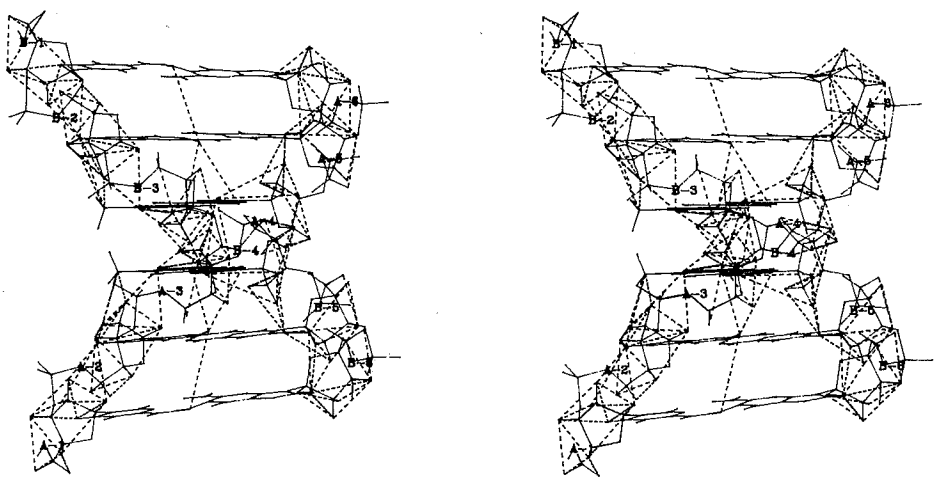


Figure 1. Stereo view of the interproton distance restraints as broken lines on a classical *B*-DNA framework; the nucleotides are labeled A_1 to A_6 on one strand and B_1 to B_6 on the other.

unknown distance is *smaller* than the reference distance, its values will be systematically *overestimated*; in contrast, if the unknown distance is *larger* than the reference distance, its value will be systematically *underestimated*; in the case of the data on the 6-mer, the latter condition always applies. Thus, the errors in the measured value of an unknown distance are skewed with the size of the systematic error depending upon the value of the unknown distance. The smaller the value of the unknown distance, the smaller will be the contribution of spin-diffusion to the observed NOE, and consequently the smaller the error. Based on the calculations of Clore & Gronenborn (1985*b*), we have estimated the errors as follows: for $r < 2{\cdot}4$ Å, the error is $-0{\cdot}1$ Å/$+0{\cdot}2$ Å; for $2{\cdot}4$ Å $< r < 3{\cdot}4$ Å it is $-0{\cdot}15$ Å/$+0{\cdot}3$ Å; and for $3{\cdot}4$ Å $< r < 4{\cdot}4$ Å it is $-0{\cdot}2$ Å/$+0{\cdot}4$ Å.

As a result of these considerations, the NOE effective potential terms introduced as restraints in the energy minimization and molecular dynamics calculations are chosen to have a skewed biharmonic effective potential form (Clore *et al.*, 1985*c*); that is:

$$E_{\mathrm{NOE}}(r_{ij}) = \begin{cases} c_1(r_{ij}-r_{ij}^0)^2, & \text{if } r_{ij} > r_{ij}^0, \\ c_2(r_{ij}-r_{ij}^0)^2, & \text{if } r_{ij} < r_{ij}^0, \end{cases} \qquad (2)$$

where r_{ij}^0 and r_{ij} are the target and calculated interproton distances, respectively, and c_1 and c_2 are force constants given by:

$$c_1 = \frac{kTS}{2(\Delta_{ij}^+)^2}, \quad c_2 = \frac{kTS}{2(\Delta_{ij}^-)^2}, \qquad (3)$$

where k is the Boltzman constant, T is the absolute temperature, S is a scale factor, and Δ_{ij}^+ and Δ_{ij}^- are the positive and negative error estimates of r_{ij}, respectively. In the molecular dynamics simulations, the scale factor S in eqn (3) was set to 3 so that error estimates of 0·1, 0·15, 0·2, 0·3 and 0·4 Å in the distances correspond to force constants of 119, 53, 29·8, 13·2 and 7·5 kcal mol^{-1} Å^{-2}, respectively (1 kcal = 4·184 kJ).

(c) *Calculational strategy*

The strategy employed in the calculations was to start with the classical *B*- or *A*-DNA structure for the 6-mer (Arnott & Hukins, 1972*a*,*b*; Arnott *et al.*, 1973). The difference between these two structures is sufficiently large to provide a good test of the convergence properties of the method. The two initial structures were subjected to 800 cycles of adapted-basis-set Newton–Raphson (Brooks *et al.*, 1983) restrained energy minimization with soft harmonic constraints applied to the starting positions (Bruccoleri & Karplus, 1985) and with the scale factor S in the NOE effective potential term set to 1. In the molecular dynamics simulations, the first minimization was followed by a 1-ps heating period during which time the system was heated from 50 K to 400 K with S set to 2, then 9 ps of dynamics was calculated without adjusting the temperature of the system with S set to 3. Averaged dynamics structures were obtained by averaging the co-ordinates over the last 5 ps of the trajectories (i.e. from 5 to 10 ps). In the restrained energy minimizations, the first minimization was followed by a further 3200 cycles of refinement with S set to 3. In order to obtain a set of energies for the average restrained dynamics structures that could be directly compared to the restrained energy minimized structures, the average restrained dynamics structures were subjected to an additional 1000 cycles of restrained minimization with S also set to 3. In addition, a corresponding set of energy minimization and free dynamics calculations were carried out with S set to 0 (i.e. with no interproton distance restraints).

3. Results and Discussion

In this section we describe first the results with idealized NOE restraints and then consider the calculations with the experimental NOE restraints.

(a) *Molecular dynamics with idealized* A- *and* B-*DNA distance restraints*

To establish the feasibility of the method, a series of molecular dynamics calculations were carried out without restraints (i.e. free dynamics) and with a set of distance restraints taken from classical *B*- and *A*-DNA comprising the same 192 proton pairs as in the experimental set: the interproton target distances (r_{ij}^0) were set equal to the values calculated from the ideal *B*-DNA and *A*-DNA structures and the effective force constants (c_1 and c_2 in eqn (3)) were evaluated in the same way as for the experimental NOE values. Two restrained dynamics runs were performed to determine whether it was possible to convert *A*-DNA to *B*-DNA and *vice versa*. In the first run the initial structure was classical *A*-DNA with the idealized *B*-DNA restraints: this yielded the average dynamics structure designated "ABR". In the second run the initial structure was classical *B*-DNA with the idealized *A*-DNA restraints: this yielded the average dynamics structure designated "BAR".

The r.m.s. difference between the 192 interproton distances in ideal *B*-DNA and *A*-DNA is 0·93 Å, a value which is reduced to 0·12 Å in going from *B*-DNA to *A*-DNA (structure BAR) and to 0·07 Å in going from *A*-DNA to *B*-DNA (structure ABR) as indicated in Table 2. As is clear from the Table, the intranucleotide distances and the interstrand distances (involving the hydrogen bonding protons) converge to values very close ($<0{\cdot}1$ Å) to the restraints; the r.m.s. differences in interproton distances for the intrastrand inter-residue distances, involving both sugar–base and base–base interactions (see Table 1), tend to be a little larger (0·1 to 0·2 Å).

Stereo views of the initial structures and the average free restrained dynamics structures are shown in Figure 2. Best fit superpositions of pairwise combinations of classical *B*-DNA, the *B*-DNA average free dynamics structure BFD and the average restrained *A* to *B* dynamics structure ABR on the one hand, and of classical *A*-DNA, the *A*-DNA average free dynamics structure AFD, and the average restrained *B* to *A* dynamics structure BAR on the other hand are shown in Figure 3. In addition, the r.m.s. atomic differences between various combinations of structures for all the backbone and the base atoms as a function of residue number are shown in Figure 4. From these Figures, as well as from the r.m.s. differences in

Table 2
r.m.s. differences between observed and idealized A- *and* B-*DNA interproton distance restraints*

Structure	Distance set	r.m.s. difference (Å)			
			Intrastrand		
		All (192)	Intranucleotide (142)	Internucleotide (38)	Interstrand (12)
ADNA	BDNA	0·93	0·78	1·43	0·26
BDNA	ADNA	0·93	0·78	1·43	0·26
ABR	BDNA	0·07	0·05	0·12	0·02
ABR	ADNA	0·94	0·79	1·46	0·26
BAR	BDNA	0·94	0·79	1·44	0·30
BAR	ADNA	0·12	0·07	0·22	0·08
BFD	BDNA	0·39	0·32	0·58	0·33
BFD	ADNA	0·90	0·77	1·34	0·48
AFD	BDNA	1·07	0·87	1·61	1·06
AFD	ADNA	0·55	0·31	0·94	0·96

r.m.s. differences between the observed and idealized *A*- and *B*-DNA interproton distance restraints for the initial (BDNA, ADNA), the average free dynamics (BFD, AFD) and the average restrained dynamics structures obtained starting off from *A*-DNA with the idealized BDNA restraints (ABR) and from *B*-DNA with the idealized *A*-DNA restraints. The idealized restraints comprise the same distance set as the experimental ones.

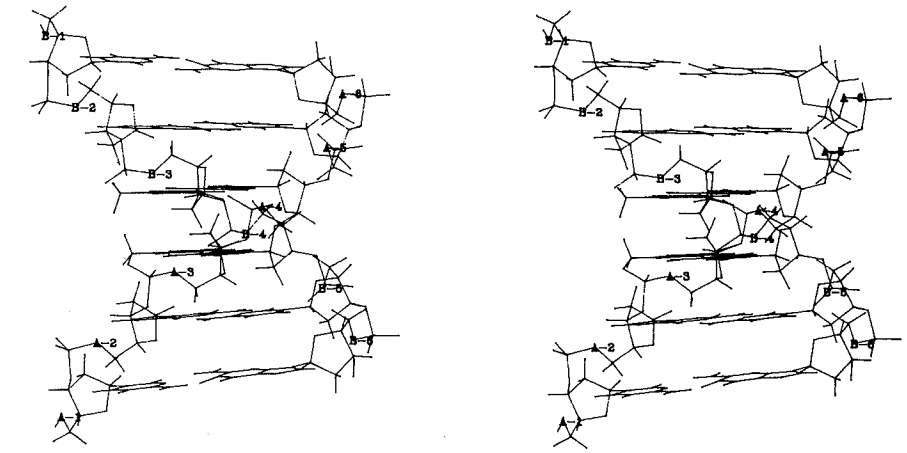
interproton distances given in Table 2, and the r.m.s. atomic differences between the various structures in Table 3, it is clear that the restrained dynamics has converted the classical *A*-DNA structure into a *B*-DNA structure, and the classical *B*-DNA structure into an *A*-DNA structure. The r.m.s. values in Table 3 for the free dynamics structures confirm that they remain similar to the starting structure (i.e. AFD like *A*-DNA and BFD like *B*-DNA), but that the dynamics averages deviate significantly from the idealized structures, as they should. Indeed, both the values and ranges of the various structural parameters describing the two average free dynamics structures are closer to those of the crystal structures of *A*- and *B*-DNA (Dickerson *et al.*, 1983; Shakked *et al.*, 1983) than to the idealized structures, suggesting that the potential energy function has well-defined local minima in the neighbourhood of the crystal structures.

The atomic r.m.s. differences are reduced from 2·89 Å, the difference between ideal *B*- and *A*-DNA structures, to 1 Å for the conversion from *A*-DNA to *B*-DNA and 2·1 Å for the conversion from *B*-DNA to *A*-DNA. The average restrained dynamics structure ABR is equally close to both classical *B*-DNA and the average free dynamics structure BFD (r.m.s. difference ~1 Å); the average restrained dynamics structure BAR, on the other hand, is closer to the average-free dynamics structure AFD (r.m.s. difference 1·51 Å) than to classical *A*-DNA (r.m.s. difference 2·09 Å). The

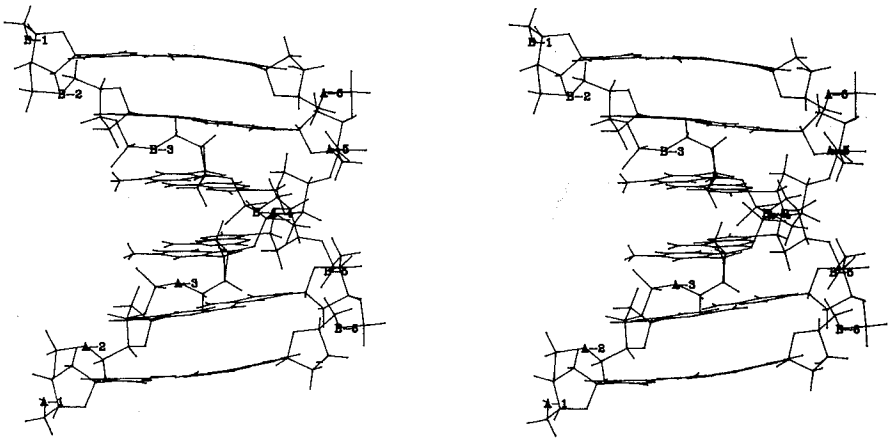
Table 3
r.m.s. differences between the B- *and* A-*DNA structures*

	Overall r.m.s. difference (Å)					
	BDNA	ADNA	BFD	AFD	ABR	BAR
BDNA		2·89	1·43	3·94	1·01	4·36
ADNA	1·75		3·52	2·00	3·08	2·09
BFD	1·37	2·38		4·40	1·00	5·00
AFD	2·32	1·33	2·90		3·96	1·51
ABR	0·83	1·99	0·93	2·52		4·51
BAR	2·55	1·32	3·34	1·18		
	r.m.s. difference for the central 4 base-pairs (Å)					

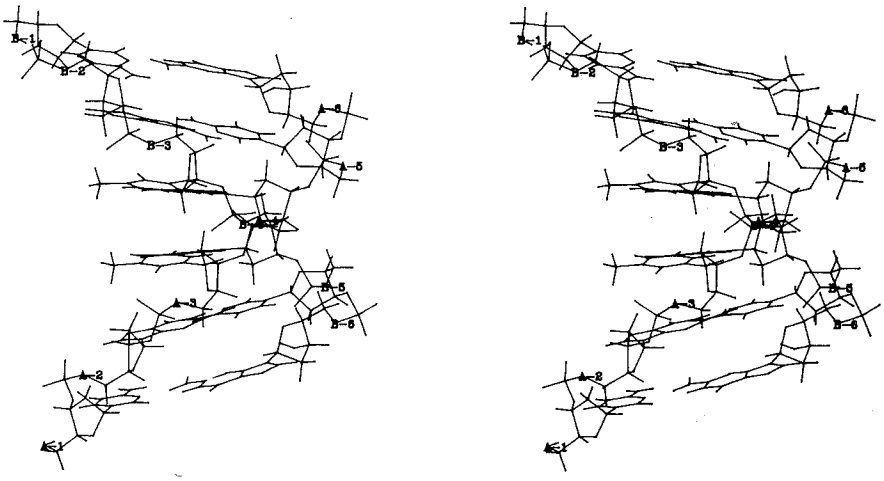
r.m.s. differences between the initial classical *B*- and *A*-DNA structures, the average free dynamics structures starting from *B*-DNA (BFD) and *A*-DNA (AFD), and the average restrained dynamics structures obtained starting off from *A*-DNA with idealized *B*-DNA restraints (ABR) and from *B*-DNA with idealized *A*-DNA restraints (BAR). The r.m.s. values above the main diagonal are obtained including all 6 base-pairs, and those below the main diagonal including only the central 4 base-pairs.



B-DNA



ABR



BFD

(a)

Fig. 2.

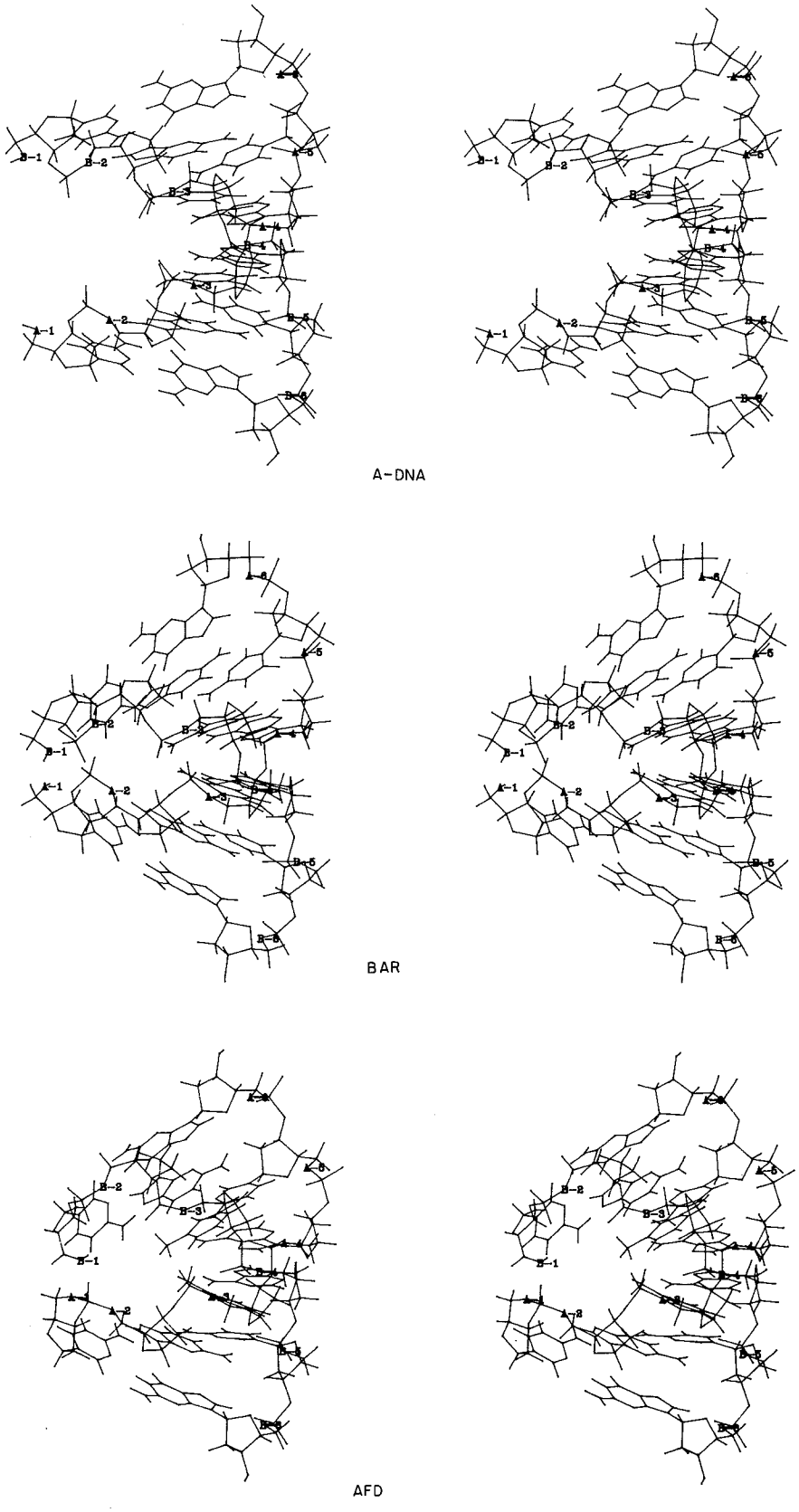
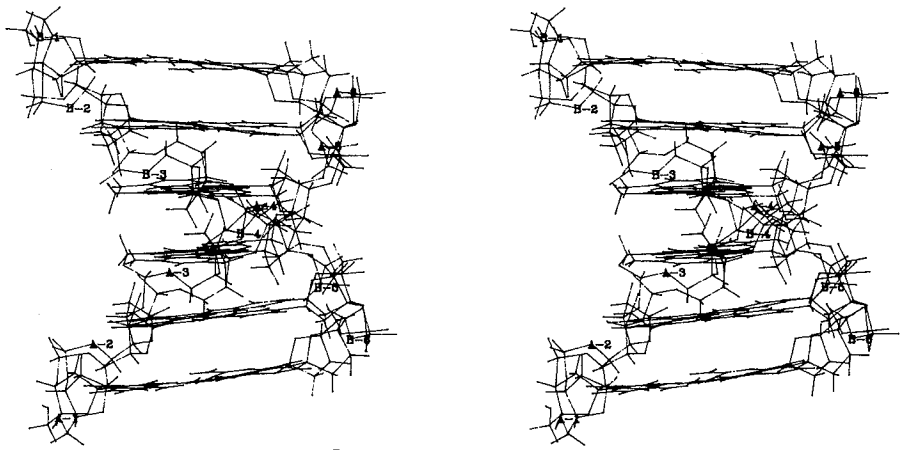
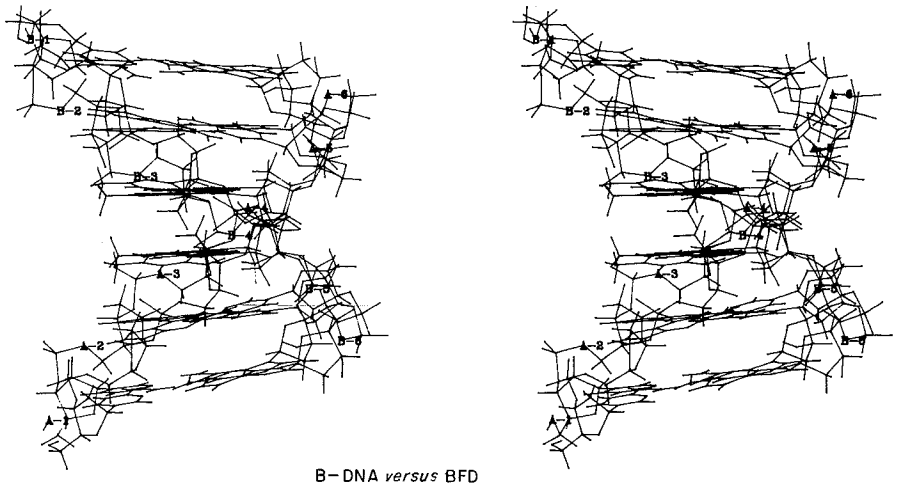


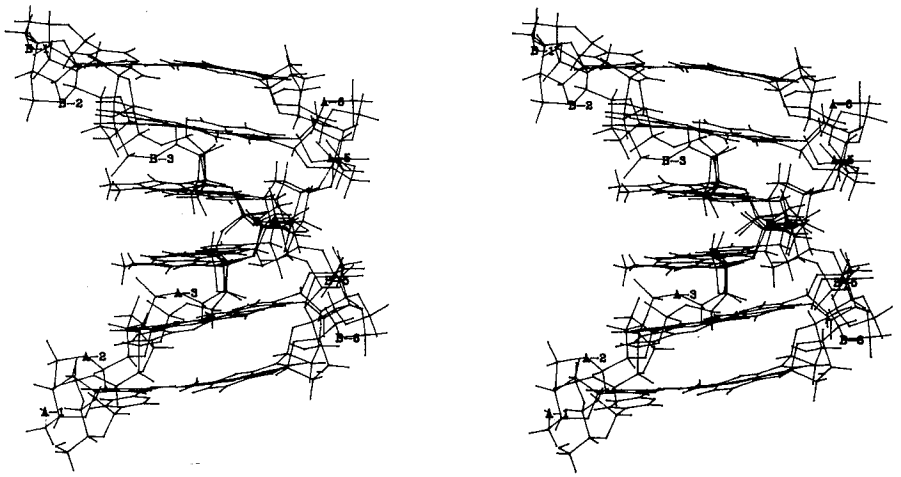
Figure 2. Stereo views along the helix axis of classical *B*- and *A*-DNA, the average restrained dynamics structures obtained using idealized distances for *B*- and *A*-DNA (ABR and BAR), and the average free dynamics structures (BFD and AFD). ABR and AFD were obtained starting off from classical *A*-DNA, and BAR and BFD from classical *B*-DNA (see the text). (a) Comparison of *B*-DNA, ABR and BFD; (b) comparison of *A*-DNA, BAR and AFD.



B-DNA *versus* ABR



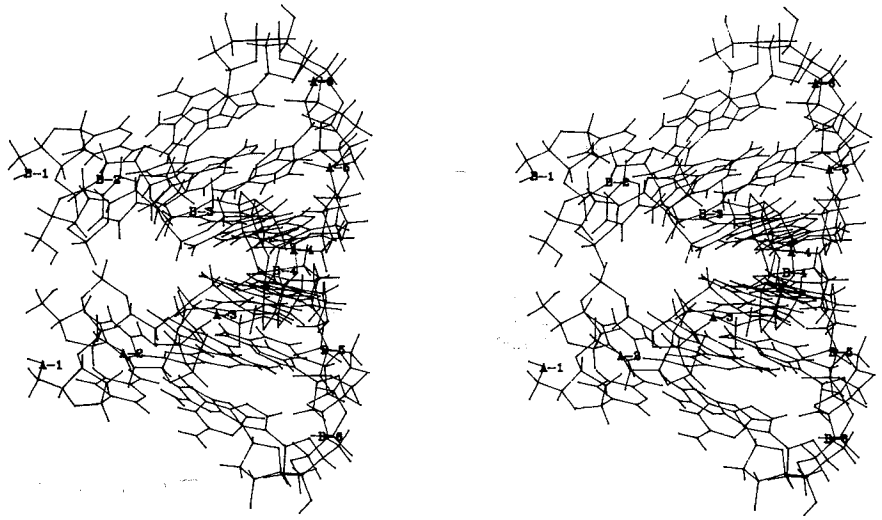
B-DNA *versus* BFD



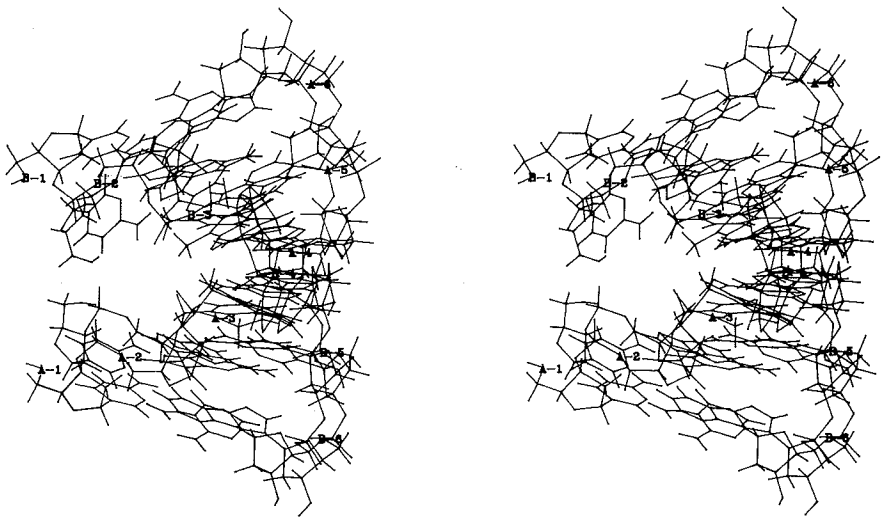
ABR *versus* BFD

(a)

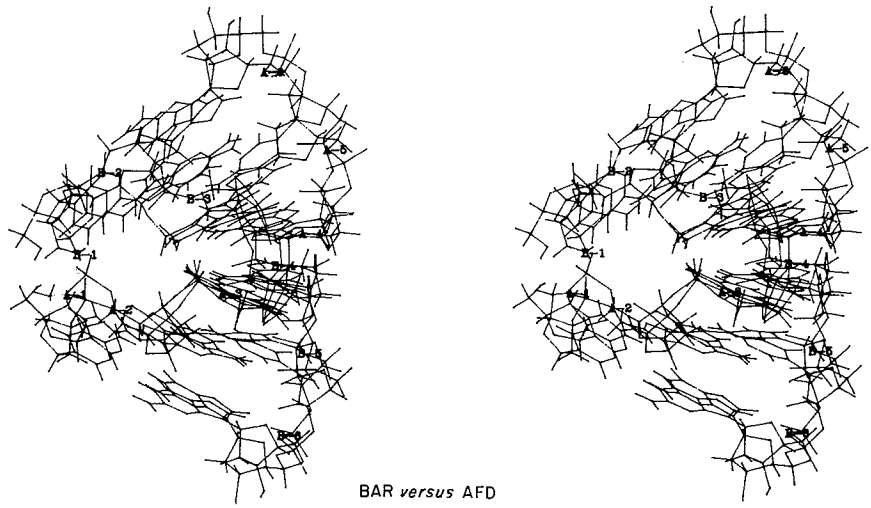
Fig. 3.



A-DNA *versus* BAR



A-DNA *versus* AFD



BAR *versus* AFD

(b)

Figure 3. Best fit superposition minimum r.m.s. of different structures in pairs: (a) classical *B*-DNA, ABR and BFD; and (b) classical *A*-DNA, BAR and AFD.

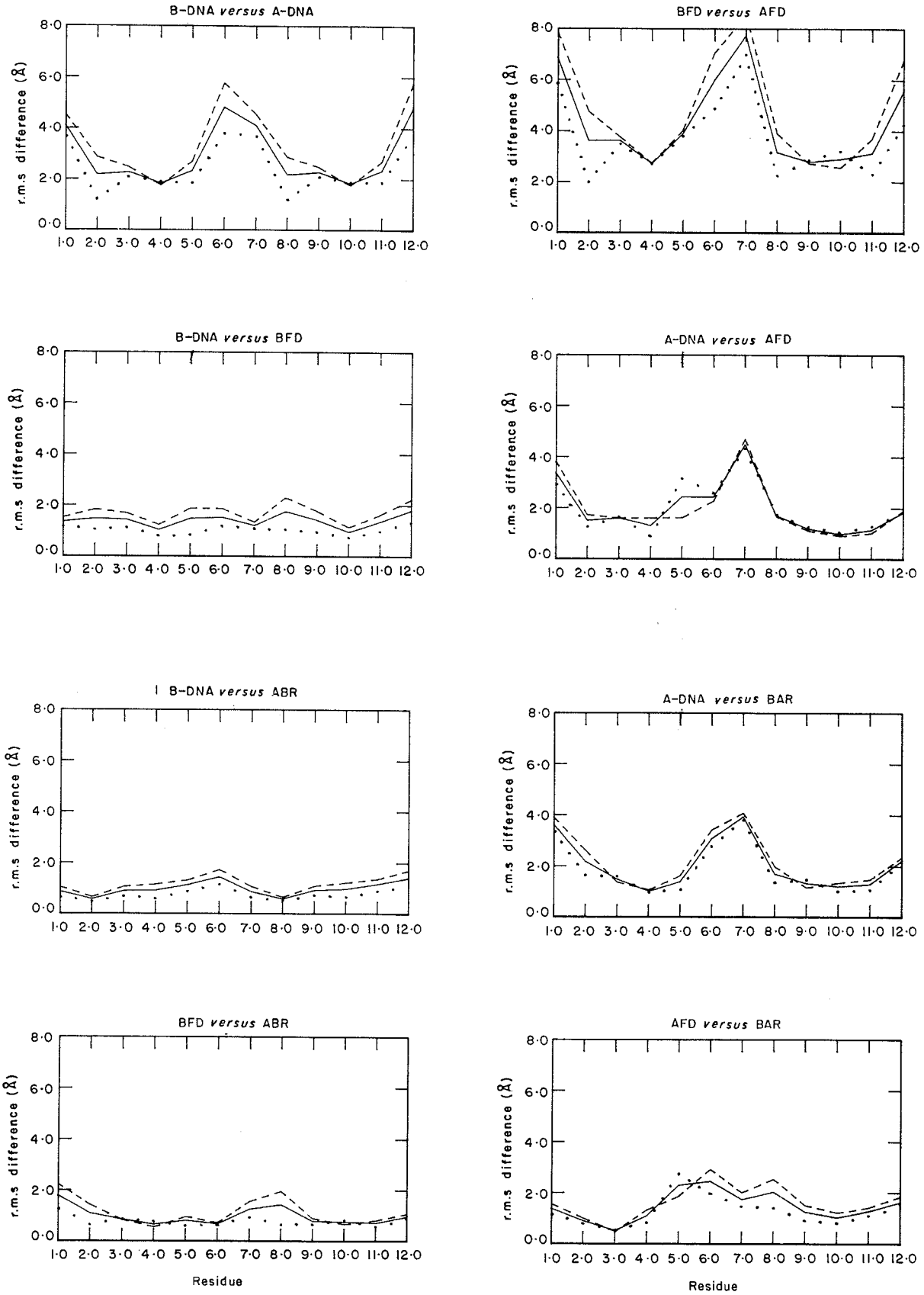


Figure 4. r.m.s. differences for all (——), the sugar–phosphate backbone (-----), and the base (.....) atoms as a function of residue number for various pairs of structures involving the initial (*A*-DNA, *B*-DNA), the average free dynamics (AFD, BFD) and the average restrained dynamics structures obtained using the idealized distance set for *A*- and *B*-DNA (BAR, ABR).

Table 4

Individual energy terms for the initial (BDNA, ADNA), the free (BFM, AFM) and restrained (BRM, ARM) minimized, and the average free (BFD, AFD) and restrained (BRD, ARD) dynamics structures

	Energy (kcal · mol^{-1}) (number of terms											r.m.s. difference of interproton distances (Å)
Structure	Total[a]	Potential[a]	Kinetic	Bond (406)	Angle (732)	Torsion (344)	Improper (172)	Electrostatic	van der Waals'	Hydrogen bonding	Restraints[b] (192)	
Initial												
BDNA	1188	150	[c]	18	149	262	4	−208	−38	−36	1038	0·56
ADNA	2804	403	[c]	86	163	201	0	−197	186	−36	2041	0·68
Free energy minimized												
BFM	[757]	−215	[c]	4	101	150	4	−251	−178	−45	[971][d]	0·55
AFM	[2395]	−226	[c]	5	101	138	2	−252	−181	−40	[2622][d]	0·75
Restrained energy minimized												
BRM	61	−115	[c]	18	119	211	6	−257	−167	−43	176	0·25
ARM	75	−75	[c]	19	133	225	5	−257	−155	−42	147	0·22
Free dynamics average[f]												
BFD	[2218(1)]	138(13)	403(13)	111(8)	292(12)	161(7)	46(5)	−252(6)	−160(5)	−39(3)	[1656(165)][d]	0·60
AFD	[4206(1)]	158(13)	403(13)	110(9)	287(13)	163(7)	47(7)	−262(6)	−166(6)	−41(4)	[3665(380)][d]	0·88
Restrained dynamics average [f]												
BRD	998(1)	332(10)	484(16)	140(11)	316(15)	249(8)	50(6)	−242(7)	−145(6)	−36(3)	182(11)	0·20
ARD	1013(1)	342(10)	488(16)	134(11)	320(15)	254(8)	50(6)	−236(7)	−144(7)	−36(3)	183(11)	0·18
Restrained minimization of restrained dynamics average[e]												
RMBRD	26	−97	[c]	16	127	228	5	−262	−169	−44	123	0·20
RMARD	28	−94	[c]	16	122	231	3	−254	−169	−42	123	0·20

The symbols A and B refer to the structures starting off from *A*- and *B*-DNA, respectively. The average dynamics structures are the averages over 5 to 10 ps.

[a] The total energy includes the restraints energy, whereas the potential energy does not.

[b] The scale factor S in eqn (3) used for the restraints was 3.

[c] The initial and the energy minimized structures have zero kinetic energy.

[d] The restrained energies for the free energy minimized and average free dynamics structure (in square brackets) were not included in the minimization or dynamics: i.e. they are the values calculated for the corresponding final (minimized) or average (dynamics) structures.

[e] The structures RMBRD and RMARD were obtained by restrained energy minimization (1000 cycles) of the average restrained dynamics structures BRD and ARD, respectively. The r.m.s. atomic shifts between these restrained minimized structures and their parent structures are less than 0·2 Å.

[f] The energies for the dynamics structures are obtained by averaging the energies over the structures from 5 to 10 ps. The r.m.s. fluctuations in the energy terms are given in parentheses.

Table 5

Overall r.m.s. differences (Å) between the restrained structures obtained using the experimental interproton distances and the unrestrained structures

	Overall r.m.s. difference (Å)			
	Restrained energy minimized		Average restrained dynamics	
	BRM	ARM	BRD	ARD
BDNA	0·73	2·65	1·62	1·49
ADNA	2·78	1·31	2·35	2·43
BFD	1·56	3·18	2·03	1·82
AFD	3·76	2·11	3·41	3·37
BRM		2·46	1·36	1·20
ARM			1·88	1·84
BRD				0·94

The notation is as follows: BDNA and ADNA, the initial classical *B*- and *A*-DNA structures, respectively; BFD and AFD, the average free dynamics structures obtained starting off from *B*- and *A*-DNA, respectively; BRM and ARM, the restrained energy minimized structures obtained starting off from *B*- and *A*-DNA, respectively; BRD and ARD, the average restrained dynamics structures obtained starting off from *B*- and *A*-DNA, respectively.

Table 6

r.m.s. differences between the experimental and calculated interproton distances for the initial, free dynamics average, restrained energy minimized and restrained dynamics average structures

	r.m.s. difference in interproton distances(Å)			
		Intrastrand		
	All (192)	Intraresidue (142)	Inter-residue (38)	Intrastrand (12)
Initial				
BDNA	0·56	0·55	0·66	0·30
ADNA	0·68	0·61	0·92	0·51
Free dynamics average				
BFD	0·60	0·52	0·86	0·37
AFD	0·89	0·68	1·41	1·19
Restrained energy minimized				
BRM	0·25	0·24	0·31	0·08
ARM	0·22	0·22	0·26	0·08
Restrained dynamics average				
BRD	0·20	0·20	0·22	0·07
ARD	0·18	0·18	0·22	0·07
Restrained minimization of restrained dynamics average				
RMBRD	0·20	0·21	0·21	0·05
RMARD	0·20	0·21	0·22	0·06

behavior of the *A*-type DNA structures appears to be due to the fact that the bases are more tilted with respect to the helix axis, particularly at the ends of the helix, in both the average free and restrained dynamics structures relative to classical *A*-DNA. This structural feature, which could be an artefact of the vacuum simulation, results in increased base stacking with a consequent improvement in the van der Waals' energy term.

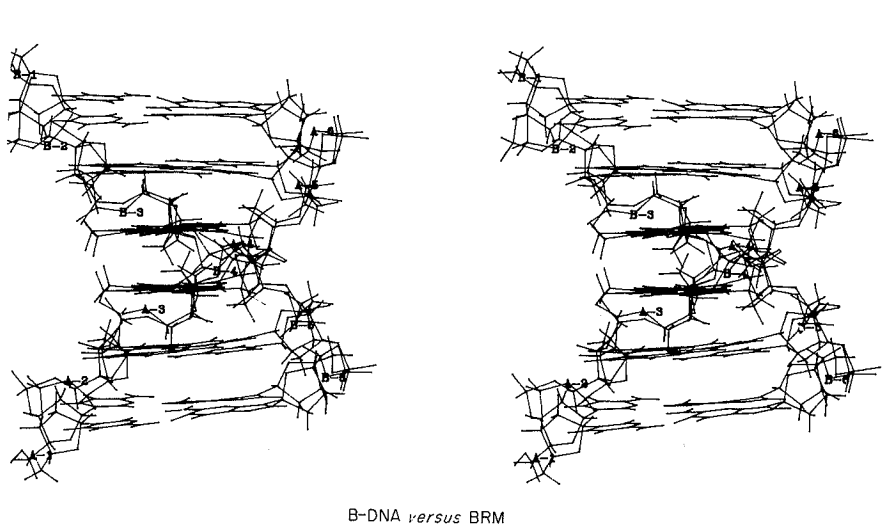
(b) *Molecular dynamics with the experimental interproton distance restraints*

Evaluation of the molecular dynamics simulation method requires comparison of several different DNA structures: they are the idealized *B*- and *A*-DNA structures (*B*-DNA, *A*-DNA) the free (BFM, AFM) and restrained (BRM, ARM) minimized structures obtained from the ideal *B*- and *A*-DNA geometries, and the free (BFD, AFD) and restrained (BRD, ARD) average molecular dynamics structures obtained from the ideal *B*- and *A*-DNA geometries. The restraints used for the various cases were the 192 experimental interproton distances for the presumed *B*-type oligomer. The energies including the energy components of the various structures are given in Table 4 and the r.m.s. differences between the structures are given in Table 5. Table 6 lists the r.m.s. differences between calculated and experimental interproton distances. Stereo views of the restrained energy minimized and average molecular dynamics structures are shown in Figures 5 and 6. The r.m.s. differences between all atoms, the sugar–phosphate backbone atoms and the base atoms as a function of residue number are illustrated in Figure 7 for some combinations of structures.

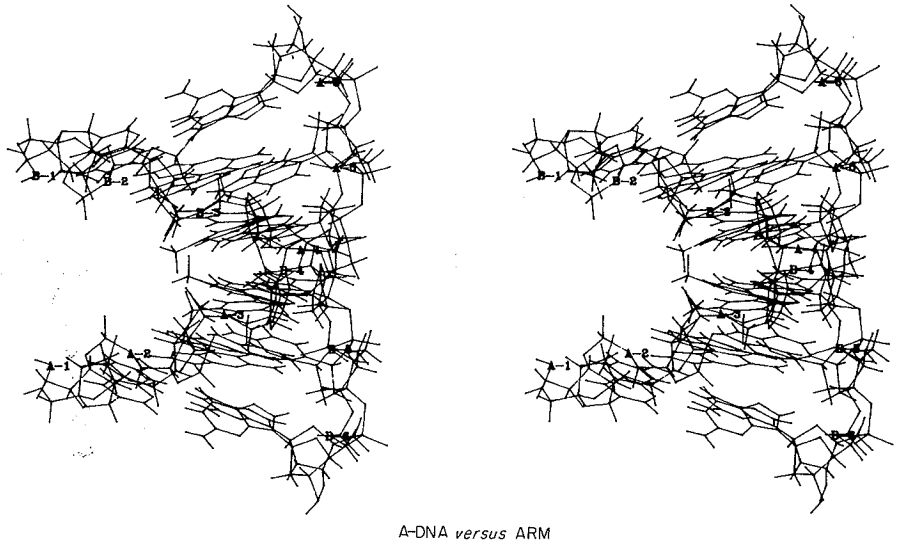
The r.m.s. difference between the experimental and calculated interproton distances (Tables 4 and 6) is greater than 0·5 Å for both classical *B*- and *A*-DNA, as well as the free energy minimized (BFM and AFM) and average free molecular dynamics (BFD and AFD) structures. These differences are significantly larger than the experimental errors. They indicate that with the present potentials, a free vacuum simulation (i.e. without NOE restraints) is not sufficiently accurate to reproduce the experimental structures. It is of interest, however, that the unrestrained *B*-DNA set of structures (classical *B*-DNA, BFM and BFD) all have lower r.m.s. deviations for the interproton distances than the unrestrained *A*-DNA set of structures (classical *A*-DNA, AFM and AFD). Also, the average-free molecular dynamics structures have slightly larger deviations for the interproton distances than their corresponding parent idealized DNA structures.

The restrained energy minimization and

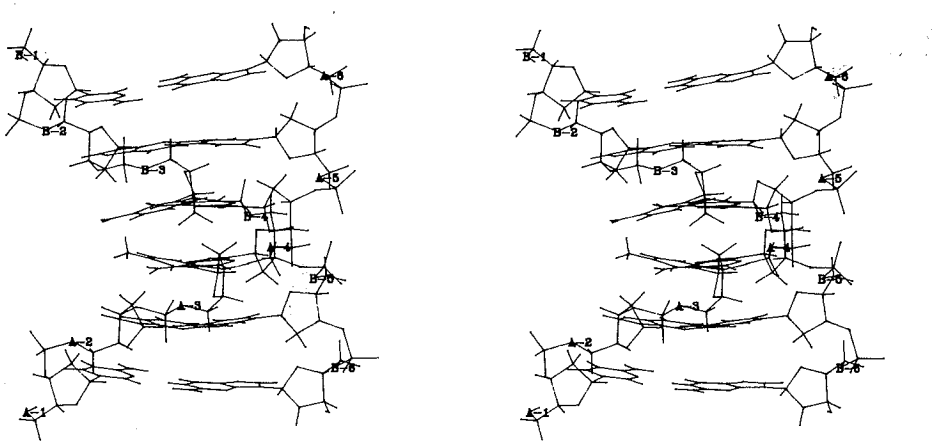
Figure 5. Stereo views of classical *B*-DNA superimposed with the restrained minimized structure BRM, classical *A*-DNA superimposed with the restrained minimized structure ARM, and the average restrained dynamics structures BRD and ARD. The experimental interproton distances are used as restraints. BRM and BRD are obtained starting from classical *B*-DNA, and ARM and ARD from classical *A*-DNA.



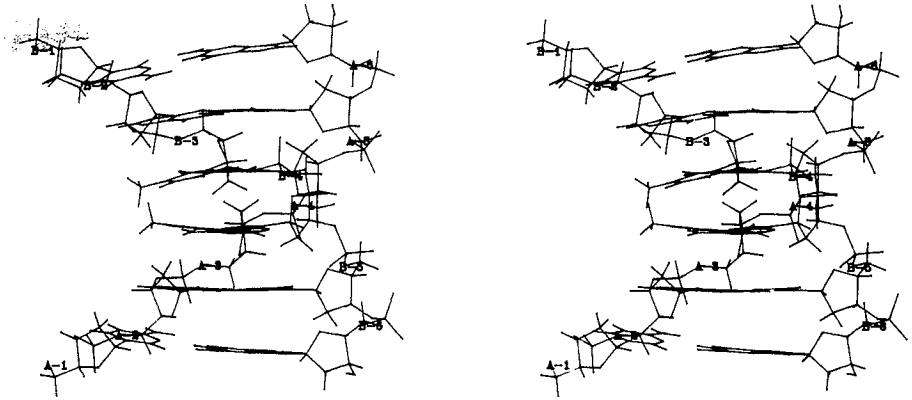
B-DNA *versus* BRM



A-DNA *versus* ARM



BRD



ARD

Fig. 5.

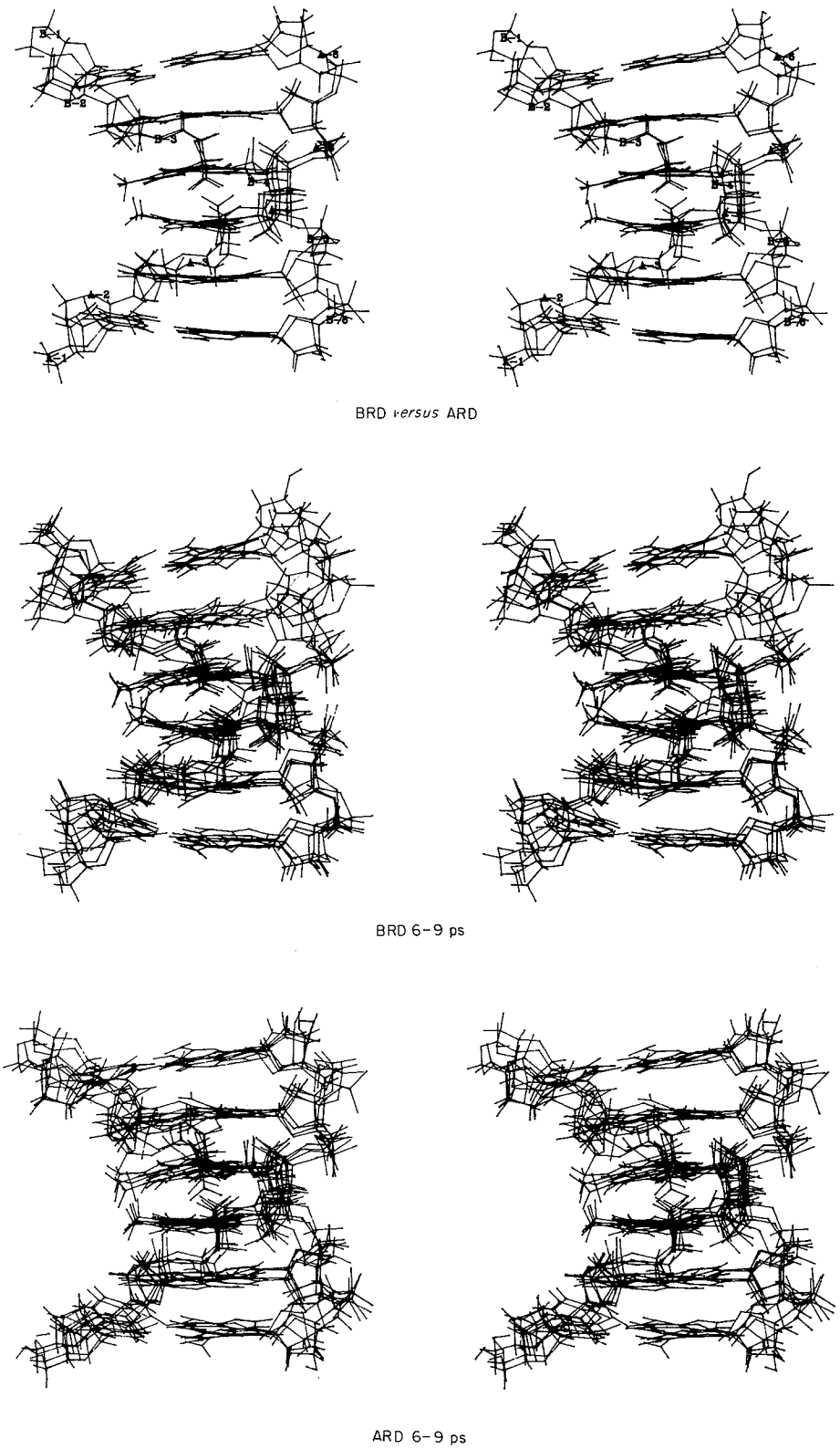


Figure 6. Best fit superposition of (1) the average restrained dynamic structures BRD and ARD and (2) the structures at 6, 7, 8 and 9 ps for the restrained dynamics structures BRD and ARD.

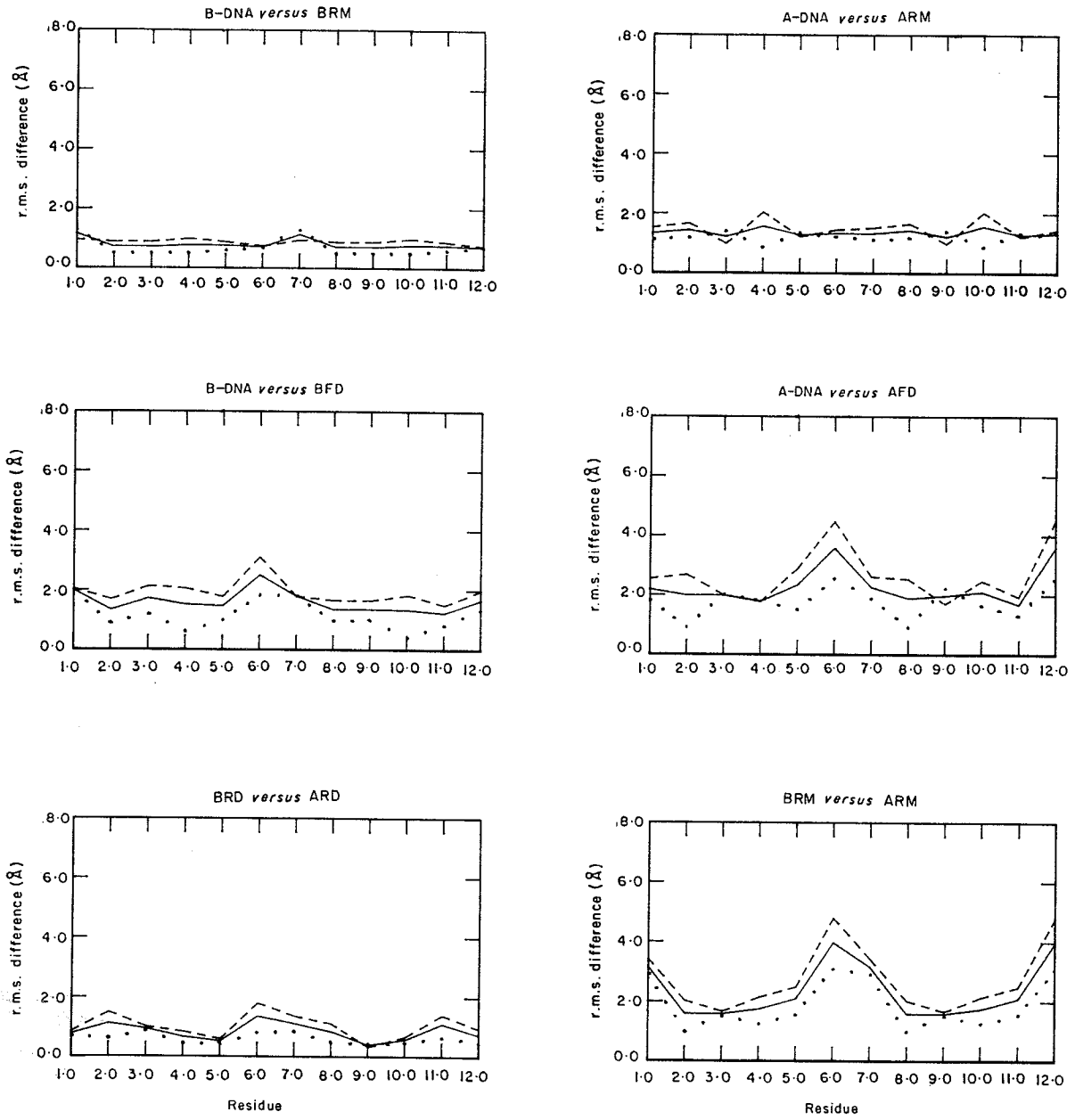


Figure 7. r.m.s. differences (Å) for all (——), the sugar–phosphate backbone (-----), and the base (.....) atoms as a function of residue number for various pairs of structures involving the initial (*A*-DNA, *B*-DNA), the restrained minimized (ARM, BRM), and average restrained dynamics structures (ARD, BRD).

molecular dynamics calculations starting from either *A*- or *B*-DNA reduce the r.m.s. difference in the interproton distances to 0·25 Å or less, with the average restrained molecular dynamics structure ARD obtained starting off from *A*-DNA having the lowest value of 0·18 Å. Of the three interproton distance groups, the interstrand distances involving the hydrogen-bonding protons show the closest agreement between experimental and calculated values. Minimization of the restrained dynamics average structures has only a small effect on the results. In all cases the individual energy terms, with the obvious exception of the restraints energy, have similar values to those in the corresponding free simulations, indicating that no sterically bad local structures have been introduced by the restraining procedure.

The stereo plots of best fit superpositions of the initial and restrained energy minimized structures (Fig. 5) show that while restrained energy minimization changes the local structure quite significantly, it leaves the global structure essentially unchanged. This is evident from the comparison of the initial classical *A*-DNA structure with the restrained energy minimized structure ARM, which has *B*-type glycosidic bond and sugar pucker conformations but globally is still an *A*-type structure. A comparison of the r.m.s difference

between the ideal *B*-DNA and *A*-DNA structures (2·89 Å) with the difference between the restrained minimized structure ARM and *B*-DNA (2·65 Å) and the difference between ARM and *A*-DNA (1·31 Å) confirms this conclusion. Thus, the available interproton distances can be satisfied within experimental error by an *A*-type structure. This suggests that structural refinement with NOE data must be supplemented by other information to obtain meaningful results.

Restrained molecular dynamics starting with idealized *A*-DNA or *B*-DNA results in structural convergence to *B*-type structures (see Fig. 5, and Table 5). The r.m.s. difference between the two average restrained molecular dynamics structures BRD and ARD is ~0·9 Å and both are much closer to the idealized *B*-DNA than to *A*-DNA. Just how similar these two structures are can be seen from the best fit superposition shown in Figure 6. In fact, the structural differences between the two average restrained dynamics structures are approximately of the same magnitude as the atomic r.m.s. fluctuations. This can be ascertained visually by a comparison of the best fit superposition of the two average structures ARD and BRD with the best fit superpositions of snapshots taken at 6, 7, 8 and 9 ps (Fig. 6), and, on a more quantitative basis, by comparing the r.m.s. differences and fluctuations for

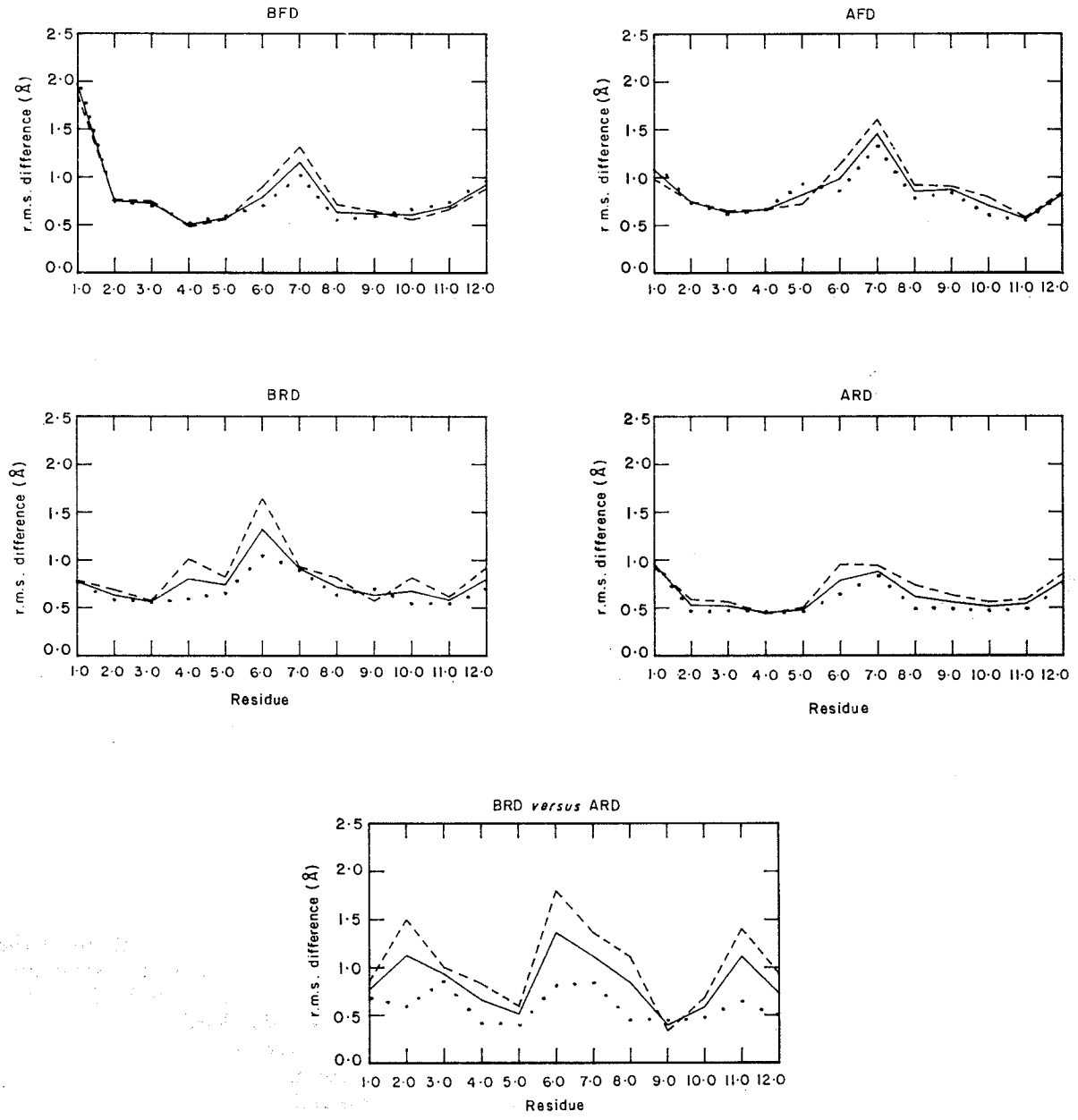
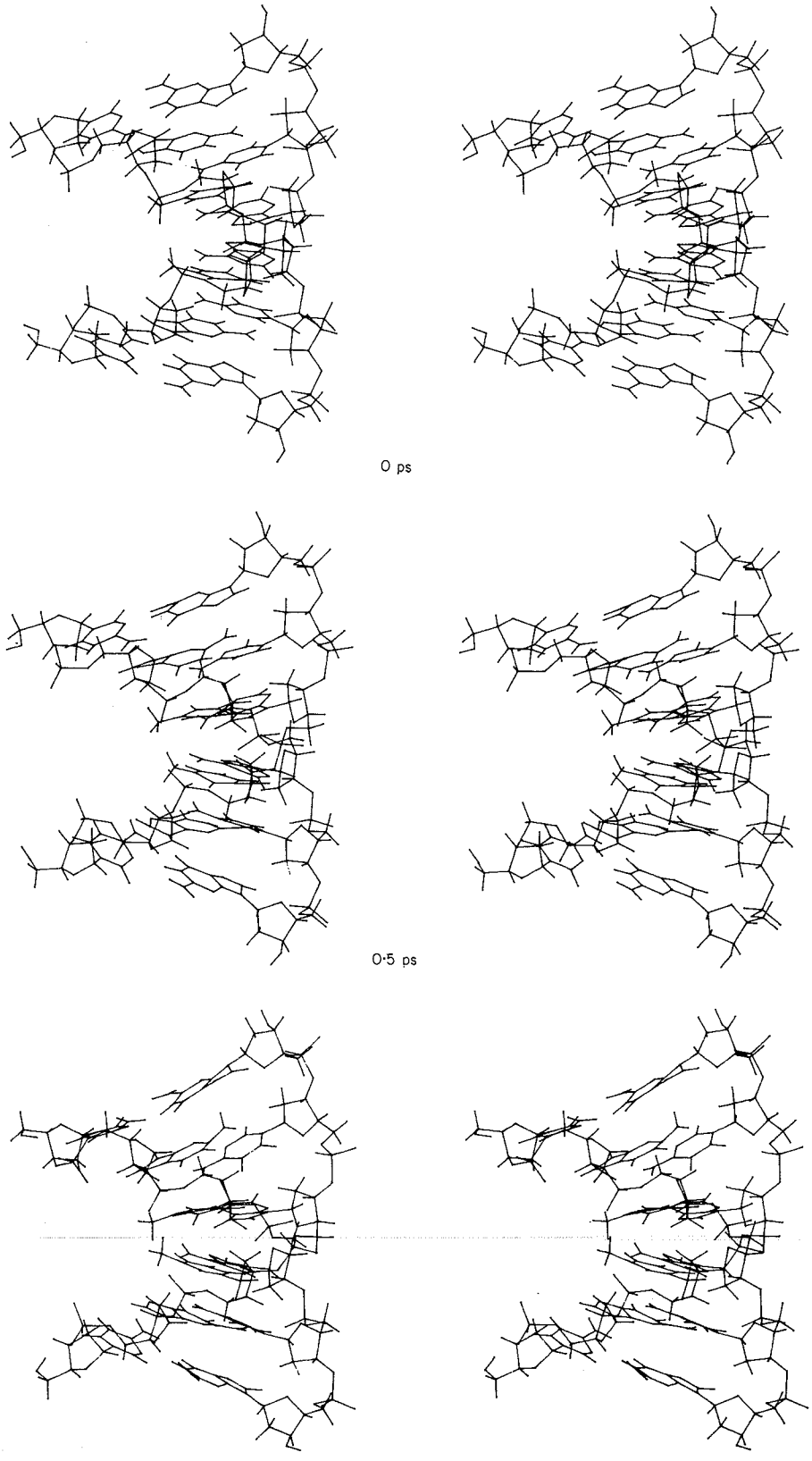
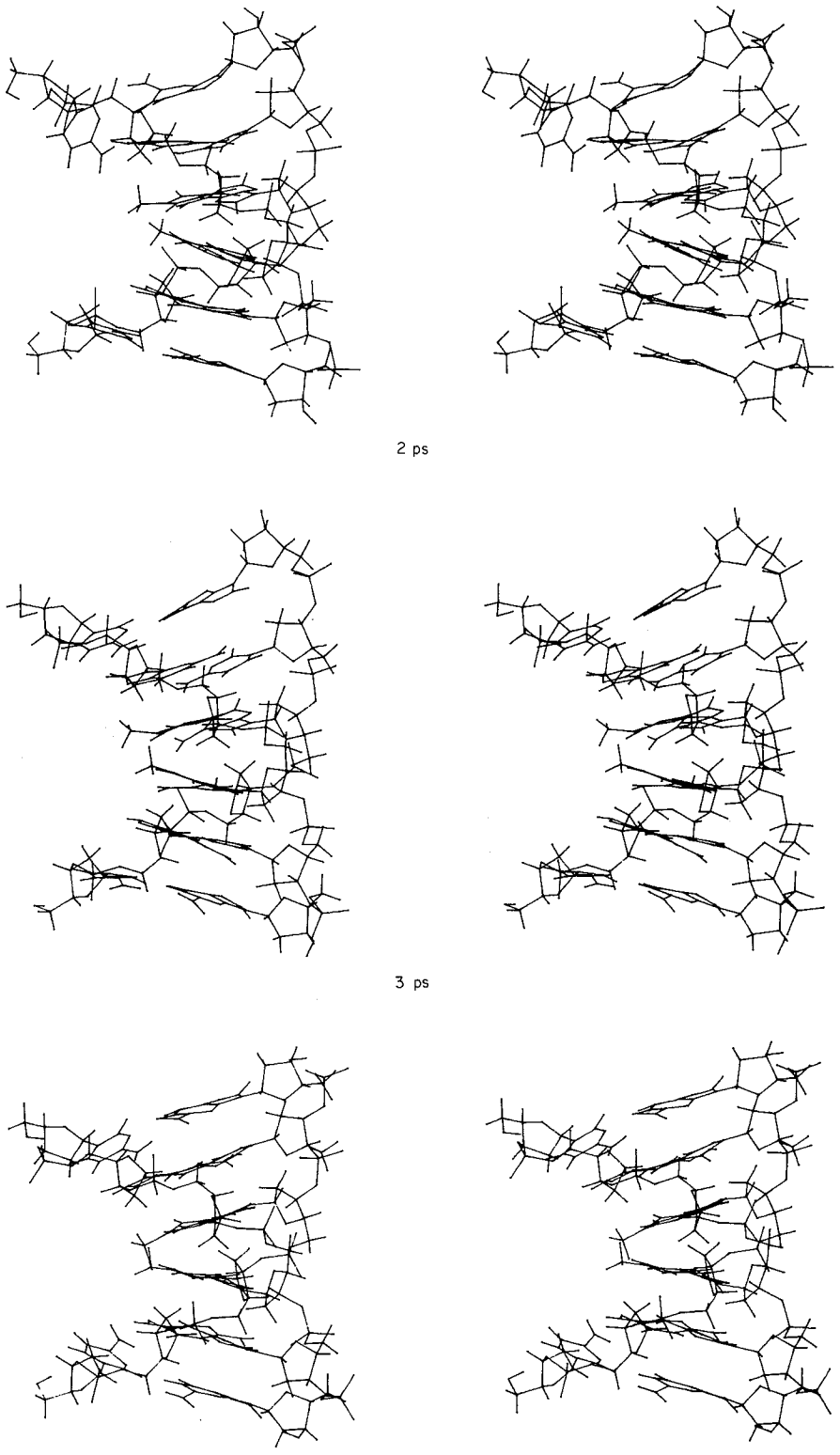


Figure 8. r.m.s. fluctuations (Å) for all (——), the sugar–phosphate backbone (-----), and base (.....) atoms as a function of residue number for the average free (AFD, BFD) and restrained (ARD, BRD) dynamics structures. The r.m.s. atomic difference between the average restrained dynamics structures (ARD and BRD) are shown on the same scale for comparison.



(a)

Fig. 9.



(b)

Fig. 9.

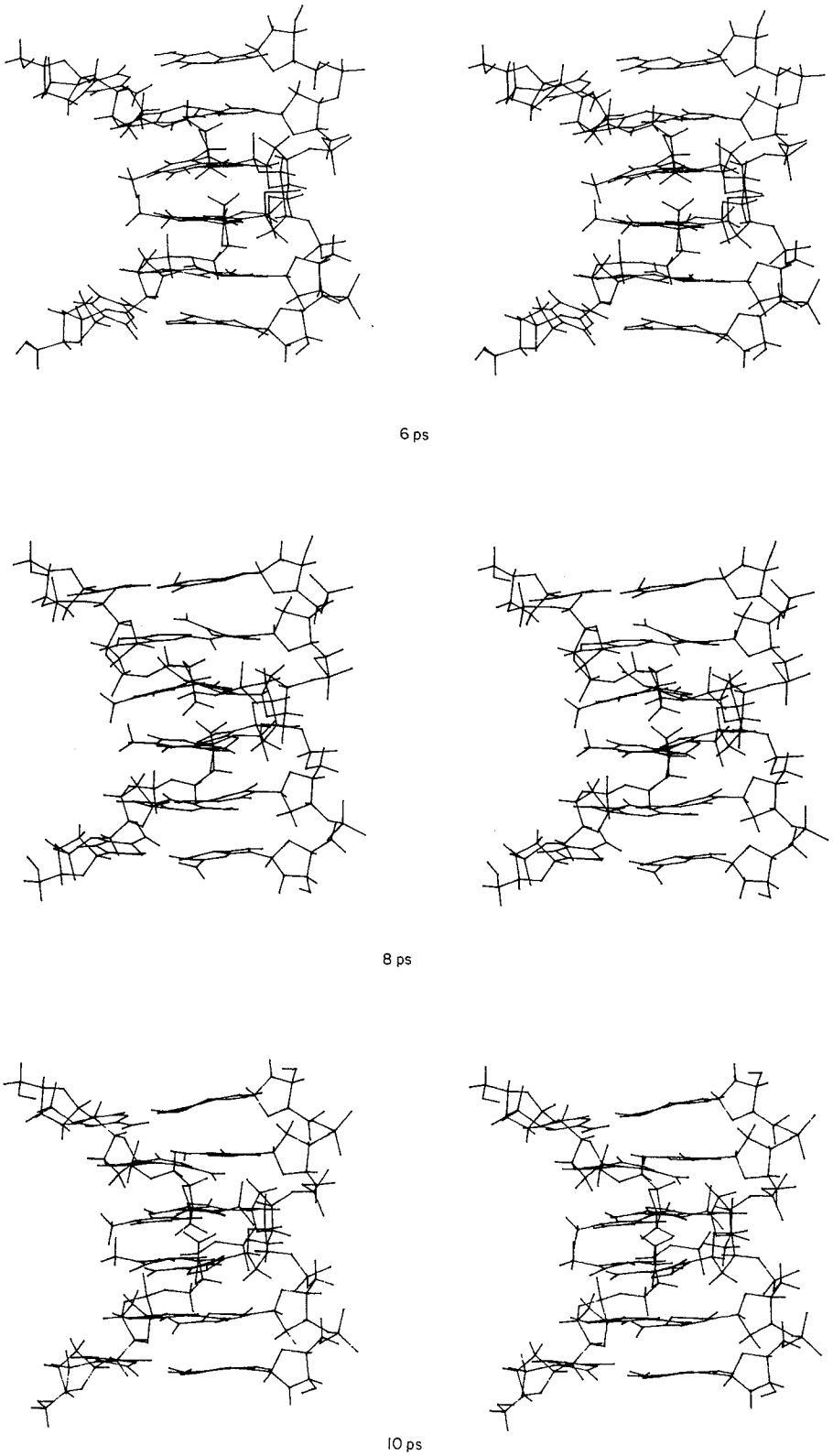


Figure 9. Snapshots of the trajectory of the restrained dynamics run showing the conversion of the *A*-type restrained energy minimized structure at 0 ps to the *B*-type restrained dynamics structure ARD.

the sugar–phosphate and base moieties as a function of residue number (Fig. 8). It is interesting that both the r.m.s. atomic differences and the mobility of the sugar–phosphate backbone are in general larger than that of the bases, and that the r.m.s. atomic differences and mobility for the terminal base-pairs are larger than those for the central ones. These results on the r.m.s. fluctuations are in accord with a free dynamics simulation of another *B*-type DNA hexamer (Tidor *et al.*, 1983). In addition, both the magnitude and pattern of the r.m.s. fluctuations are similar in both the free and restrained dynamics structures, although the fluctuations at the ends are slightly smaller in the restrained case.

(c) *Time course of the transition from A-DNA to B-type DNA*

The time course of the transition from *A*-DNA to the *B*-type structure that fits the experimental interproton distances is depicted in Figure 9 by a series of snapshots from the restrained molecular dynamics simulation starting with *A*-DNA. Between 4 ps and 6 ps the global *B*-DNA character of the structure becomes clear. Since the restraints energy is apparently just as well satisfied in the restrained minimized structure ARM as in the average restrained dynamics structure ARD, what is the reason for the conversion from the global *A* topology to the *B*-type structure? To investigate this we have plotted the time course of the restraints, van der Waals' and electrostatic energy components as a function of time during the course of the *A* to *B* transition (Fig. 10). From this plot it is apparent that the principal driving force arises from a decrease in the restraints energy term by $\sim$25 to 30 kcal mol^{-1} coupled with a smaller improvement ($\sim$15 kcal mol^{-1}) in the van der Waals' energy term. There is no significant change in the electrostatic energy component during the course of the transition. To examine this point we subjected the average restrained dynamics structures BRD and ARD to a further 1000 cycles of restrained energy minimization to obtain a set of energies that are directly comparable to those of the restrained energy minimized structures BRM and ARM. This process generated structures RMBRD and RMARD, which deviated by less than 0·2 Å from their respective parent structures BRD and ARD. The energies of RMBRD and RMARD are included in Table 4. A comparison of the energies for RMARD and ARD confirms the view that the restrained dynamics results in an improvement in both the restraints (by $\sim$25 kcal mol^{-1}) and van der Waals' (by $\sim$15 cal mol^{-1}) energy terms, leaving the electrostatic term essentially unchanged. Thus, although the NOE restraints can be satisfied by a restrained minimization in the neighborhood of *A*-DNA, the greater range of the restrained molecular dynamics simulation leads to a minimum in the neighborhood of *B*-DNA that not only satisfies the restraints but also has a lower van der Waals' energy. In the case of RMBRD and BRM, the only significant difference lies in the restraints energy term that is approximately 50 kcal mol^{-1} lower for RMBRD relative to BRM.

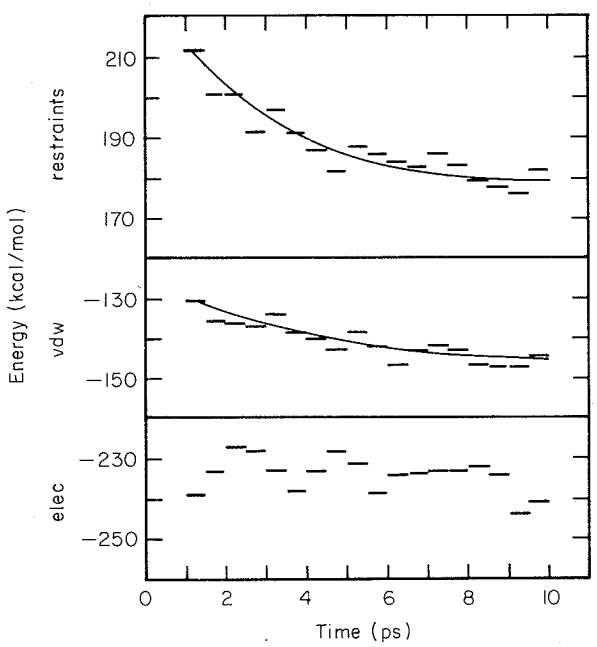


Figure 10. Time dependence of the restraints, van der Waals' (vdw) and electrostatic (elec) energy components during the course of the *A* to *B* transition for the ARD trajectory.

4. Concluding Remarks

We have demonstrated that restrained molecular dynamics has a satisfactory radius of convergence for structure refinement of oligonucleotides with interproton distance restraints. When experimental or idealized interproton distances are used as restraints, the molecular dynamics simulation converges to the final structure independent of the choice of initial structures: e.g. with the experimental NOE data, both an ideal *A*- and *B*-DNA starting structure yield a *B*-type DNA with the r.m.s. difference between the two average restrained dynamics structures comparable to the r.m.s. fluctuations of the atoms. This is in marked contrast to the results of energy minimization, which alters only the local structure and leaves the global structure unchanged.

It is important to notice that the above results are obtained in spite of the fact that the change in global conformation from an *A*- to a *B*-type structure during the course of the restrained molecular dynamics simulation does *not* arise solely from an improvement in the restraints energy terms themselves. This follows from the fact that the final r.m.s. deviations in the interproton distances are not very different for both the restrained energy minimized and average restrained dynamics structures. Thus, the present results are a consequence of the use of the empirical energy function

to guide the restrained dynamics. Comparison of the non-bonding energy terms (i.e. van der Waals', electrostatic and hydrogen bonding) shows a significant improvement in the converged dynamics values of the van der Waals' component relative to the energy minimized structure. The important role of the energy function in this work makes it clear that successful use of interproton NOE distance data for the refinement of DNA structures depends upon the accuracy of these functions, particularly that of the non-bonding energy terms. Since the latter are expected to be sensitive to solvent effects, it is important to extend the present results to DNA simulations in aqueous solution. Further, it suggests that some caution is required concerning NOE refinement approaches for oligonucleotides based solely on distance geometry algorithms, particularly if no independent information is available on the type of DNA (*A*-DNA, *B*-DNA, etc . . .) being studied.

This work was supported by the Max-Planck-Gesellschaft (G.M.C., A.M.G. and A.T.B.) and the National Institutes of Health (A.T.B., M.K. and L.N.). L.N. held a Swedish National Research Council Fellowship while at Harvard.

References

Arnott, S. & Hukins, D. W. L. (1972*a*). *Biochem Biophys. Res. Commun.* **47**, 1504–507.

Arnott, S. & Hukins, D. W. L. (1972*b*). *J. Mol. Biol.* **81**, 93–105.

Arnott, S., Chandrasekharan, R., Hall, I. H., Puijganer, L. C., Walker, J. K. & Wang, M. (1983). *Cold Spring Harbor Symp. Quant. Biol.* **47**, 53–66.

Arseniev, S. A., Kondakov, V. I., Maiorov, V. N. & Bystrov, V. F. (1984). *FEBS Letters,* **165**, 57–62.

Aue, W. P., Bartholdi, E. & Ernst, R. R. (1976). *J. Chem. Phys.* **64**, 2229–2246.

Billeter, M., Braun, W. & Wüthrich, K. (1982). *J. Mol. Biol.* **155**, 321–346.

Braun, W., Wider, G., Lee, K. H. & Wüthrich, K. (1983). *J. Mol. Biol.* **169**, 921–948.

Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983). *J. Comput. Chem.* **4**, 187–217.

Bruccoleri, R. E. & Karplus, M. (1986). *J. Comp. Chem.* In the press.

Clore, G. M. & Gronenborn, A. M. (1983). *EMBO J.* **2**, 2109–2113.

Clore, G. M. & Gronenborn, A. M. (1985*a*). *FEBS Letters,* **179**, 187–198.

Clore, G. M. & Gronenborn, A. M. (1985*b*). *J. Magn. Reson.* **61**, 158–164.

Clore, G. M. & Gronenborn, A. M. (1985*c*). *EMBO J.* **4**, 829–835.

Clore, G. M., Gronenborn, A. M., Moss, D. S. & Tickle, I. J. (1985*a*). *J. Mol. Biol.* **185**, 219–226.

Clore, G. M., Gronenborn, A. M. & McLaughlin, L. W. (1985*b*). *Eur. J. Biochem.* **151**, 153–165.

Clore, G. M., Gronenborn, A. M., Brünger, A. T. & Karplus, M. (1985*c*). *J. Mol. Biol.* **186**, 435–455.

Crippen, G. M. & Havel, T. F. (1978). *Acta Crystallogr. sect. A,* **34**, 282–284.

Dickerson, R. E. & Drew, H. R. (1981). *J. Mol. Biol.* **149**, 761–785.

Dickerson, R. E., Drew, H. R., Conner, B. N., Kopka, M. L. & Pjura, P. E. (1983). *Cold Spring Harbor Symp. Quant. Biol.* **47**, 13–24.

Dobson, C. M., Olejniczak, E. T., Poulsen, F. M. & Ratcliffe, R. G. (1982). *J. Magn. Reson.* **48**, 87–110.

Dwek, R. A. (1976). *Nuclear Magnetic Resonance in Biochemistry,* Oxford University Press, Oxford.

Gelin, B. & Karplus, M. (1977). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 801–805.

Gronenborn, A. M. & Clore, G. M. (1985). *Progr. Nucl. Magn. Reson. Spect.* **17**, 1–33.

Gronenborn, A. M., Clore, G. M. & Kimber, B. J. (1984). *Biochem. J.* **221**, 723–736.

Havel, T. F. & Wüthrich, K. (1984). *Bull. Math. Biol.* **46**, 673–698.

Harvel, T. F. & Wüthrich, K. (1985). *J. Mol. Biol.* **182**, 281–294.

Jardetzky, O. & Roberts, G. C. K. (1981). *NMR in Molecular Biology,* Academic Press, New York.

Jeener, J., Meier, B. H., Bachmann, P. & Ernst, R. R. (1976). *J. Chem. Phys.* **71**, 4546–4553.

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R. & van Gunsteren, W. F. (1985). *J. Mol. Biol.* **182**, 179–182.

Karplus, M. & McCammon, J. (1983). *Annu. Rev. Biochem.* **52**, 263–300.

Kuntz, I. D., Crippen, G. M. & Kollman, P. A. (1979). *Biopolymers,* **18**, 939–957.

McCammon, J. A., Gelin, B. R. & Karplus, M. (1977). *Nature (London),* **267**, 585–590.

McCammon, J. A., Wolynes, P. G. & Karplus, M. (1979), *Biochemistry,* **18**, 927–942.

Nilsson, L. & Karplus, M. (1986). *J. Comput. Chem.* In the press.

Olejniczak, E. T., Dobson, C. M., Karplus, M. & Levy, R. M. (1984). *J. Amer. Chem. Soc.* **106**, 1923–1930.

Ryckaert, J. P., Cicotti, G. & Berendsen, H. J. C. (1977). *J. Comput. Phys.* **23**, 327–337.

Shakked, Z., Robinovick, D., Kennard, O., Cruse, W. B. T., Salisbury, S. A. & Viswanitra, M. A. (1983), *J. Mol. Biol.* **166**, 183–201.

Strop, P., Wider, G. & Wüthrich, K. (1983). *J. Mol. Biol.* **166**, 641–667.

Tidor, B., Irikura, K. K., Brooks, B. R. & Karplus, M. (1983). *J. Biomol. Struct. Dynam.* **1**, 231–252.

Verlet, L. (1967). *Phys. Rev.* **159**, 98–105.

Wagner, G. & Wüthrich, K. (1979). *J. Magn. Reson.* **33**, 675–680.

Wagner, G. & Wüthrich, K. (1982). *J. Mol. Biol.* **160**, 343–361.

Wang, A. H. J., Fujii, S., van Boom, J. H. & Rich, A. (1983). *Cold Spring Harbor Symp. Quant. Biol.* **47**, 33–44.

Weiss, M. A., Patel, D. J., Sauer, R. T. & Karplus, M. (1984). *Proc. Nat. Acad. Sci., U.S.A.* **81**, 130–134.

Wider, G., Macura, S., Kumar, A., Ernst, R. R. & Wüthrich, K. (1984). *J. Magn. Reson.* **56**, 207–234.

Williamson, M. P., Havel, T. F. & Wüthrich, K. (1985). *J. Mol. Biol.* **182**, 295–315.

Wüthrich, K., Wider, G., Wagner, G. & Braun, W. (1982). *J. Mol. Biol.* **155**, 311–319.

Zuiderweg, E. R. P., Kaptein, R. & Wüthrich, K. (1983). *Eur. J. Biochem.* **137**, 279–282.

Edited by G. A. Gilbert