Nucleic Acids Research, 1999, Vol. 27, No. 24 **4759–4767**

An NMR and molecular modelling analysis of d(CTACTGCTTTAG)·d(CTAAAGCAGTAG) reveals that the particular behaviour of TpA steps is related to edge-to-edge contacts of their base-pairs in the major groove

S. Leporc, O. Mauffret, G. Tevanian, E. Lescot, M. Monnot and S. Fermandjian*

Département de Biologie Structurale, UMR 8532 CNRS, Institut Gustave Roussy, 39–53 rue Camille Desmoulins, 94805 Villejuif, France

Received August 18, 1999; Revised and Accepted November 3, 1999

ABSTRACT

In a previous NMR study we detected the presence of particular motions and hydration properties within the DNA fragment d(CTACTGCTTTAG)·d(CTAAAGCAGTAG). Now, we report on an NMR and molecular modelling analysis of this sequence focusing our attention on the biologically important TpA steps. NOe and coupling constant restraints were introduced in three different modelling protocols: X-PLOR and JUMNA used with *Flex* and *AMBER94* as force-fields. Despite their differences the protocols produce similar mean B-DNA structures (r.m.s.d. <1 Å). The new information confirms our previous experimental results on the narrowing of the minor groove along the T8T9T10/A17A16A15 run and the sudden widening at the T10pA11 step ending this run. It is further shown that this step displays a large positive roll with its T10:A15 and A11:T14 base-pairs likely stabilised by amino–amino and amino–carbonyl interactions in the major groove. A relationship between roll values and amino–amino and amino–carbonyl distances strongly suggests that electrostatics or bifurcated hydrogen-bonds could be responsible for induction of positive rolls in TpA steps. Such edge-to-edge interactions could explain the slower motions shown by the adenine A15. The influence of these interactions on the stabilisation of particular DNA conformers is discussed using our data and those provided by the recent literature.

INTRODUCTION

Conformational transitions occurring at TpA steps in DNA have been reported in the past by several experimental groups (1–6). These studies have been made by NMR spectroscopy and visualised as excess line widths of aromatic proton resonances (2,7) and ^{13}C relaxation data (3). The time scale of these dynamic motions has been found to range between the microsecond and the millisecond (3,5,6).

In a recent NMR study (4) we identified strong sequence dependence of the dynamics in the d(CTACTGCTTTAG)·d(CTAAAGCAGTAG) duplex. While each strand of the oligonucleotide contains two TpA dinucleotides it is mainly the T14pA15 step included in the 5′-TAAA-3′ segment which shows slow motions. In addition, the adenines of the 5′-TAAA-3′ run display positive nOe cross-peaks indicating substantial residence times of water molecules (>0.5 ps) at their proximities (4,8). In contrast, the adenine of the facing 5′-TTTA-3′ segment did not show either this dynamical process or retention of water molecules. Results underlined the high sensitivity of the water-DNA nOes towards small differences between the TpA structures, induced by different context sequences. Similar to any pyrimidine–purine step, the TpA step is very malleable (1,2,7,9–13) as it presents little overlap between its two base-pairs. This allows significant bending through rolling in the major groove although the faculty of TpA to be bent depends also on its interactions with its environment (neighbouring residues, ions, proteins etc.). Notably, the TpA step has been identified in protein–DNA complexes as a site of major bending (14–16).

The main purpose of this NMR and molecular modelling work on d(CTACTGCTTTAG)·d(CTAAAGCAGTAG) is to improve our structural and dynamical information on the TpA steps. Yet, the structure determination of double-stranded nucleic acids by NMR combined to molecular modelling still remains a hazardous task and recent studies in this field have underlined a significant influence on results of the modelling method and of the way of using experimental restraints (17–20). To avoid these pitfalls the DNA structure was assessed with three different protocols, namely JUMNA with two different sets of force fields (*Flex* and *AMBER94*) (17,21), and molecular dynamics through X-PLOR (22). In both cases the use of relaxation matrix methods permitted a better account of experimental data (23). Present results describe the main properties of the d(CTACTGCTTTAG)·d(CTAAAGCAGTAG) structures provided by the three different modelling protocols. However, particular attention is paid to the peculiar behaviour of the T10pA11 step.

*To whom correspondence should be addressed. Tel: +33 1 42 11 49 85; Fax: +33 1 42 11 52 76; Email: sfermand@igr.fr

MATERIALS AND METHODS

DNA synthesis and purification

The dodecadeoxyribonucleotides d(CTACTGCTTTAG) and d(CTAAAGCAGTAG) were synthesised using the solid phase procedure on an Applied Biosystems 381A automated apparatus, according to the standard phosphoramidite chemistry and purified by HPLC column followed by dialysis. Nucleotides in the two strands were numbered in ascending order from the 5′ to the 3′ end, as shown below:

	1	2	3	4	5	6	7	8	9	10	11	12	
5'	C	T	A	C	T	G	C	T	T	T	A	G	3' (strand I)
3'	G	A	T	G	A	C	G	A	A	A	T	C	5' (strand II)
	24	23	22	21	20	19	18	17	16	15	14	13	

NMR spectroscopy

NMR experiments were performed at 500 MHz on a Bruker AMX-500 spectrometer. All data were processed on an X32 computer with the UXNMR software or a Silicon Graphics Workstation with the Felix 2.30 (M.S.I. Technologies, San Diego, CA).

The sample was dissolved in 0.4 ml of potassium dihydrogen phosphate-disodium hydrogen phosphate buffer containing 1 mM EDTA. The resulting solution was nearly 4 mM in duplex, ionic strength 0.1, and pH 6.9.

For assignments of non-exchangeable protons the sample was lyophilised several times from ^2H_2O and finally dissolved in 0.4 ml 99.996% ^2H_2O. NOESY spectra were collected at 35°C using the phase-sensitive mode with the TPPI method (24). For NOESY experiments, 4096 real points in t2 and 800 in t1 were collected with 2 s relaxation delays using a spectral width of 4504.5 Hz. The NOESY data sets were apodized with a skewed-sinebell-squared 70° phase-shifted function to 2048 points in t2 and 800 in t1. For P-COSY and TOCSY experiments at 35°C, 4096 real points in t2 and 800 in t1 were collected using a spectral width of 4032 Hz. TOCSY experiments were collected in the clean mode with a dipsi2 mixing sequence and mixing times were 40, 80 and 110 ms (25). The P-COSY and TOCSY data set were apodized with a sinebell-squared 45° phase-shifted function to 2048 points in t2 and 800 in t1.

For ^{31}P assignments, 2D heteroTOCSY experiments (26) were recorded at 35°C with a mixing time of 70 and 40 ms; 1024 points in t2 (proton dimension) and 100 in t1 (phosphorus dimension); spectral width of 2404 Hz in t2 and of 405 Hz in t1.

For assignments of exchangeable protons, the sample was lyophilised several times in H_2O and then dissolved in a mixture of 90% H_2O/10% ^2H_2O. 2D NOESY-Watergate (using a 32 kHz rf field for all pulses) (27) were recorded at 10°C with 4096 points in t2 and 800 in t1; a spectral width of 10 204 Hz. The NOESY data sets were apodized with a skewed-sinebell-squared 70° phase-shifted function to 1024 points in t2 and 800 in t1 and the baseline was corrected in F2 dimension with a five order polynomial function.

Simulation methodology

JUMNA. The 3D structures of the dodecamer were determined with the version 10.0 of the JUMNA (JUnction Minimization of Nucleic Acids) (28) program on a Silicon Graphics O2 R10000 workstation. JUMNA allows the use of *Flex* (29,30) and *AMBER94* (31) force fields. JUMNA is able to handle distance constraints, either with respect to a chosen fixed value or within specified upper and lower bounds (32,33). In both cases, violations of constraints are prevented by a simple quadratic penalty term using a force constant of either 6 or 12 kcal.mol^{-1}.Å^{-1}. It is also possible to constrain sugar puckering in terms of phase, amplitude or both these variables in combination (32,33). Differences between two torsion angles or between two distances can be also constrained; as these types of constraints are less stringent than constraints on a single angle or distance, they allow a lower energetic cost (32,33). The force constant used for the angles is 1000 kcal.mol^{-1}.rad^{-1}.

The structures obtained were analysed with the CURVES program (version 5.2) which provides a rigorous way to obtain local structural parameters and overall helical axis locus for irregular structures of nucleic acids (34). The output files from this program include two sets of helical parameters: the 'global parameters' defined relatively to a global helical axis and the 'local parameters' defined relatively to a local axis (34). Note that the structural analysis and r.m.s.d. calculations exclude the terminal base-pairs of oligomer, which are subject to fraying in solution.

Two starting structures were generated by constructing the d(CTACTGCTTTAG)·d(CTAAAGCAGTAG) duplex in canonical B-form, the first one using the Biopolymer module of Insight and the second one using JUMNA. The other starting conformations for the dodecamer were generated by a combinatorial search of low and high purine phase sugar puckers (17,35) and by minimising without constraints coordinates obtained previously after JUMNA minimisation with *Flex* and *AMBER94* force fields.

Then, these starting structures were angle constrained (sugar phases and the ε-ζ torsion angles). To avoid the influence of angle constraint definition on the obtained structures, each structure was minimised using two different sets of angle constraints. In the first set, the sugar phases were classified as low (120–170°) or high (150–200°) (35). In the second set, all the sugar phases were constrained between 120 and 200° while the phase differences were constrained between –10 and +10° for two residues displaying close phases and between –5 and +60° for residues having very different phases. So, for one given initial structure, we obtained two structures, minimised with the first and the second set of angle constraints. Finally, angles restrained structures were submitted to back-calculation iterative refinement of distances. The NOE_Simulate routine integrated into an Insight environment (M.S.I. Technologies) was used for the calculation of theoretical NOESY spectra (36,37). The full relaxation matrix used here takes into account all the spin diffusion pathways. A single correlation time of 2.5 ns was used for every interproton vector. This parameter and the Z-leakage factor were determined using the procedure described by Banks *et al.* (38). The refinement procedure was performed according to Lefebvre *et al.* (32,33). The theoretical NOESY spectra of a given structure were calculated at various mixing times, and the calculated cross peak volume $V^s ij$ compared to the experimental cross peak volume $V^e ij$ at the corresponding mixing time. To enable this comparison, a scaling factor was determined at each mixing time by comparing the non-terminal cytosine H5-H6 reference cross peak volumes in the calculated and experimental spectra.

The fit of the calculated nOe intensities ($V^s ij$) to the experimental data ($V^e ij$) was evaluated by calculating the nOe residual R factor: $R = (\Sigma|V^e ij - V^s ij|)/\Sigma V^e ij$ and the $R^{1/6}$ factor: $R^{1/6} = (\Sigma|V^e ij^{1/6} - V^s ij^{1/6}|)/\Sigma V^e ij^{1/6}$ where the calculations run over all mixing times and over all resolved cross peaks among the base (H6/H8 or H2) to sugar protons (H1′, H2′, H2″, H3′ or H4′) connectivities, intrasugar connectivities and internucleotide base to base connectivities. The distance constraints were then submitted to an automated refinement procedure (33).

A run with JUMNA was then performed with the new constraints set (angle restraints and new distance constraints). This refinement was repeated until the distance bound restraints and the structural parameters of the molecule showed no further changes.

X-PLOR. All energy minimisations and molecular dynamics calculations, using the program X-PLOR 3.1 (22), were performed on Silicon Graphics INDIGO2 R8000. All structures shown in this work were displayed using Insight97 (M.S.I. Technologies). Initial starting structures designated by Ini-A and Ini-B, respectively, were generated by constructing the d(CTACTGCTTTAG)·d(CTAAAGCAGTAG) duplex in canonical form A- or B-form using the Biopolymer module of Insight.

Determination of restraints

As described in Huang and Eisenberg (39), in addition to the application of experimental interproton distance restraints in molecular dynamics (MD) and observed volume integrals in back-calculation refinement, some other restraints (mentioned in the following) have been used in our calculations.

The interproton distances were calculated from the distance extrapolation method (40). The internal references for translating the nOe cross-peak intensities in ^2H_2O experiments were 2.95 Å (CH_3-H6 distance) for cross-peaks implying a methyl resonance and 2.45 Å (cytosine H5-H6 distance) for all the other cross-peaks.

NOe cross-peaks presenting overlaps making their integrations difficult were used as restraints by classification into strong (1.8–3 Å), medium (1.8–4 Å) and weak (1.8–6 Å) intensities. 383 distances were calculated and in general assigned the following boundaries: $d < 2.6$ Å (–0.2/+0.3 Å), 3.0 Å $< d <$ 4.0 Å (–0.3/+0.4 Å) and $d > 4$ Å (–0.4/+0.5 Å).

The NOESY data acquired in H_2O (two mixing times 120 and 180 ms) were quantified and translated into distances using the H2-NH3 cross-peaks as reference (2.9 Å).

Base-pairs were kept Watson–Crick hydrogen bonded by distance restraints between bases. For A-T base-pairs, A(N6) – T(O4) = 2.95 ± 0.1 Å and A(N1) – T(N3) = 2.82 ± 0.1 Å; for G-C base-pairs, C(N3) – G(N1) = 2.95 ± 0.1 Å, C(N4) – G(O6) = 2.91 ± 0.1 Å and C(O2) – G(N2) = 2.86 ± 0.1 Å. To avoid problems during molecular dynamic calculations, distances restraints were added between C1′ atoms on opposite sides of the major groove (these distances were restrained to be >16 Å) (39).

We also restrained the backbone dihedral angles to preserve the right-handed character of the DNA duplex (40,41). The values used were: $\alpha = 290 \pm 50°$, $\beta = 180 \pm 50°$, $\gamma = 30 \pm 35°$, $\varepsilon = 180 \pm 50°$ and $\zeta = 275 \pm 30°$.

δ-dihedral torsion angle restraints were included in the calculations on the basis of measured $J_{H1'-H2'}$ and $J_{H1'-H2''}$, and on qualitative estimations of $J_{H3'-H4'}$ and $J_{H2''-H3'}$ for all non-terminal residues.

Structure refinements

Nucleic acid parameters and force constants were obtained from the parameter files parallhdg.dna (22). This file was however modified for the valence angles in the sugar which are chosen to fit those given in Saenger (42). These modifications permit to obtain sugar conformations with correct puckering amplitudes (35–40°).

For the electrostatic component of the empirical energy function, the effect of solvent was approximated by a $\varepsilon = \varepsilon_0 \times r$ dielectric function and by reducing the net charge on the phosphate group to –0.32e. All bond lengths were kept fixed with the SHAKE algorithm (43) (tolerance 5.0×10^{-4}).

Calculations were performed in two steps. First, distance restrained molecular dynamics calculations were carried out. Second, the relaxation matrix refinement was applied, as described in other works (41,44,45). In the following, these steps will be denoted as distances-restrained molecular dynamics (d-MD) and relaxation matrix refinement molecular dynamics (RM-MD).

The starting structures (Ini-A and Ini-B) presented a r.m.s.d. of 5.69 Å (taken into account only the 10 central base-pairs). They were minimised with 1000 steps of conjugate gradients. After minimisation, the two structures (Ini-Amin and Ini-Bmin) present a r.m.s.d. of 5.68 Å.

Distance restrained molecular dynamics calculations with initial velocities assigned from a Maxwellian distribution at 1000 K were performed for a total of 38 ps. Three random number seeds were used for the initial velocity assignments in order to generate an ensemble of six structures.

The dihedral restraints were 100 $kcal.mol^{-1}.rad^{-2}$, the distance restraints were introduced gradually, the initial values for the experimental distance restraints were 0.4 $kcal.mol^{-1}.Å^{-2}$ and for the hydrogen bond restraints 0.5 $kcal.mol^{-1}.Å^{-2}$. These values were increased to 30 and 50 $kcal.mol^{-1}.Å^{-2}$, respectively, in 8 ps at 1000 K and were maintained at these values during the rest of the time course of d-MD. The system was equilibrated by an additional 8 ps at 1000 K, and cooled gradually to 300 K in 14 ps. An additional 8 ps of restrained-molecular dynamic at 300 K was performed. The coordinates were averaged over the last 2 ps and subjected to 1000 steps of conjugate gradient minimisation to produce $DR\text{-}A_{1,2,3}$ and $DR\text{-}B_{1,2,3}$ structures.

Relaxation matrix refinement

The six distance refined structures were refined against the nOe cross-peaks intensities with the relaxation matrix refinement routine of the X-PLOR program. The nOe intensities (total 899 on three NOESY data sets collected at mixing times of 70, 120 and 200 ms) were incorporated as penalty functions in the relax energy terms, in which exponent 1/6 was used. An isotropic correlation time of 3 ns derived from a systematic grid search (X-PLOR Manual) along with a cut-off distance of 5.5 Å was used during the relaxation matrix refinement calculations (46). The tolerance was set to 0.03 Å. The distance refined structures (DR) were submitted to 2 ps of molecular dynamics at 1000 K during which the weights for the non-exchangeable nOe intensities were increased from 0 to 400 $kcal.mol^{-1}.Å^{-2}$ while weights for distance restraints were decreased from 30 to 0.01 $kcal.mol^{-1}.Å^{-2}$. At the same time, the weights for non-exchangeable proton intensities, hydrogen-bond

and dihedral restraints were kept to 30, 50 and 100 kcal.mol^{-1}.Å^{-2}, respectively. Next, the system was gradually cooled from 1000 to 300 K in 5.6 ps, then an additional 4 ps dynamics was performed at 300 K. The coordinates during the last 1.0 ps of dynamics were averaged and these averaged coordinates were subjected to energy minimisation.

RESULTS

Distance and torsion angle restraints

Sugar puckers were estimated using the procedure described by Kim and Reid (36). This approach involves the quantitative measurements of $J_{H1'-H2'}$ and $J_{H1'-H2''}$ and the qualitative estimation of $J_{H2''-H3'}$ and $J_{H3'-H4'}$ in a P-COSY spectrum. In every residue, $J_{H1'-H2'}$ is found to be greater than $J_{H1'-H2''}$, thus limiting all the deoxyribose pseudorotation angles to 100–200° (36,47). In the H2′/H2″-H3′ region of the P-COSY spectrum, no residue, except terminal residues (C1, C13, G12 and G24), exhibits H2″-H3′ cross-peak, thus supporting the range of pseudorotation angles provided by $J_{H1'-H2'}$ and $J_{H1'-H2''}$.

The use of the H1′-H4′ nOe cross-peaks allows the reduction of the range of individual pseudorotation angles (36,47,48). This has been possible for all the residues except A11. The good agreement observed between the COSY and the NOESY experiments allowed the use of a larger range of pseudorotational angle restraints and the refinement of the H1′-H4′ distances to constrain sugar.

The phosphodiester backbone was constrained from ^{31}P NMR data (not shown). The ^{31}P chemical shift values were obtained from a heteroTOCSY experiment (26). We also recorded a heteroCOSY spectrum to obtain the proton–phosphorus couplings. However, these were not measurable due to strong signal overlaps, since the 22 phosphorus resonances are located between –4.5 and –4.1 p.p.m. Thus, we used the earlier reported relationship between the phosphorus chemical shifts and the ε-ζ angles (21,32,49).

JUMNA calculation results

The JUMNA procedure was applied using either the *Flex* force field or the *AMBER94* force field. With *Flex*, nine starting structures were generated exhibiting energies of nearly –590 kcal.mol^{-1}, with r.m.s.d. between 0.26 and 1.4 Å. With *AMBER94*, eight starting structures were generated displaying energies of nearly –400 kcal.mol^{-1}, with r.m.s.d. ranging from 0.24 to 1.95 Å. We note that the r.m.s.d. range is larger with *AMBER94*.

By minimisations with the two different sets of angle constraints, we obtained 18 structures with the *Flex* force field, and 16 structures with the *AMBER94* force field. For each force field, we selected the four structures (two for each set of angle constraints) presenting the highest r.m.s.d.

We applied to the above sets of structures the distance refinement protocol described in Materials and Methods. All the non-exchangeable protons were used, except the methyl protons which were ignored in the JUMNA program. The two force fields provided similar structures (r.m.s.d. 0.70 Å) although the convergence with *Flex* (r.m.s.d. 0.20 Å) was found better than with *AMBER94* (r.m.s.d. 0.38 Å). There was a total loss of 30 and 37 kcal.mol^{-1} for *Flex* and *AMBER94*, respectively, but structures were in good agreement with experimental data (*Flex*: $R^{1/6}$ = 0.056 and R = 0.27; *AMBER94*: $R^{1/6}$ = 0.057 and R = 0.26).

X-PLOR calculation results

The two starting structures Ini-A and Ini-B were generated as described in Materials and Methods. Three dynamics runs were performed on each starting structure by random assignment of initial velocities, producing six structures (Ini-Amin: DR-A1, DR-A2 and DR-A3, Ini-Bmin: DR-B1, DR-B2 and DR-B3). Details on the protocols are also outlined in Materials and Methods.

The convergence is always higher for structures arising from the same starting structure (DR-A: r.m.s.d. = 0.92 ± 0.07 Å, DR-B: r.m.s.d. = 0.95 ± 0.22 Å), compared to the convergence for structures arising from different starting structures (r.m.s.d. = 1.16 ± 0.16 Å). However, the application of a relaxation matrix refinement increases the convergence, a gain of 0.24 Å being observed on the r.m.s.d.

The energetic terms are listed in Table 1A. The relaxation matrix refinement procedure results in an ~35 kcal.mol^{-1} loss

Table 1. (**A**) Energy terms (kcal mol^{-1}) and (**B**) deviations from ideal stereochemistry and from experimental restraints of structures obtained at various stages of the X-PLOR modelling protocol

A

	Total	Bonds	Angles	Impropers	vdW	Elec	hbond
A-DNA	1017	90	712	150	32	–303	–73
B-DNA	778	97	628	125	–124	–339	–77
d-MD	–312.1	24.1	180.8	14.1	–354	–418	–74
RM-MD	–277.6	30.6	202.6	14.8	–346.8	–415.1	–74.8

B

	Deviations from ideal stereochemistry			Deviations from experimental restraints		
	Bonds (Å)	Angles (°)	Impropers (°)	cdih (kcal mol^{-1})	R factor	$R^{1/6}$ factor
A-DNA	0.0104	1.77	1.43	2715	1.47	0.166
B-DNA	0.0109	1.6	1.31	20	0.86	0.101
d-MD	0.0054	0.89	0.43	2.9	0.28	0.047
RM-MD	0.0061	0.94	0.44	0.81	0.14	0.030

A-DNA and B-DNA correspond to the starting structures, d-MD and RM-MD designate the structures obtained with d-MD and RM-MD, respectively.

in total energy. This rather small energetic cost arises mainly from an increase of the angle energy. The parameters related to the stereochemistry of the molecules (Table 1B) may be judged as good according to several criteria that have been also used in similar studies (46,50). The deviation on stereochemical parameters introduced by the refinement also appears minimal (Table 1B). The nOe R factors show clearly that the molecule adopts a structure near the B-form. The refinement procedure improves significantly the R factor: R decreases from 0.28 to 0.14 and $R^{1/6}$ from 0.0047 to 0.0030.

Comparison of the results provided by the three approaches

At the end, we obtained six molecules with the X-PLOR rMD (restrained Molecular Dynamics) strategy, eight structures with JUMNA-*Flex* and eight structures with JUMNA-*AMBER94*. An average structure was calculated in each case and the obtained three structures, called X-PLOR-f, *AMBER*-f and *Flex*-f, were compared as well as being compared to the A- and B-DNA canonical forms (Table 2).

Table 2. r.m.s.d. in Å between the structures of d(CTACTGCTTTAG)·d(CTAAAGCAGTAG) (for the 10 central base-pairs) provided by the three modelling methods (X-PLOR-f, Flex-f and AMBER-f) and their corresponding minimised canonical A- and B-forms

	A-DNA	B-DNA	X-PLOR-f	Flex-f	AMBER-f
A-DNA	0				
B-DNA	5.68	0			
X-PLOR-f	4.80	1.78	0		
Flex-f	4.77	1.79	0.80	0	
AMBER-f	4.18	2.27	1.00	0.70	0

The r.m.s.d. are <1 Å for the final DNA structures, indicating that these structures depend only a little on the method employed and are essentially defined by the experimental restraints. Note, for instance, that with X-PLOR we used the canonical A- and B-DNA forms as starting structures, while with JUMNA-*Flex* and JUMNA-*AMBER94* we used the energetical minima. Furthermore, the three force fields were also rather different, and the way the restraints were applied was totally different. Also, X-PLOR works through molecular dynamics with cartesian coordinates while JUMNA proceeds through molecular mechanics and internal coordinates. Therefore, it was reassuring that different approaches could provide results in so close agreement.

Structure analysis

In the first step, we compared the average helical parameters (calculated as the average of the individual values of each molecule) in the molecules provided by the three approaches. Among these, the X-disp and inclination parameters help to identify the nature of the DNA family. X-disp values (X-PLOR: –2.4 Å, AMBER: –2.2 Å and Flex: –1.6 Å) are nearer the B-form (0 Å) than the A-form (–5 Å), and are similar to the values found for other published structures (41,51,52). For comparison, the X-disp values are –1.0 (X-PLOR), –1.3 (AMBER) and –0.7 Å (Flex), for the energy minimised structures without restraints. Inclination values are slightly negative in both the X-PLOR and the *Flex* molecules (–2.2 and –3.5°, respectively), while being near 0° for the *AMBER94* molecules (0.7°). They also fit the B-family (–6 Å) better than the A-family (+19 Å), and values are globally in agreement with results provided by other NMR studies for B-DNA (41,53), although large positive inclinations have also been observed in some cases (20,52).

The values of the intra base-pair parameters: shear, stretch and stagger, are null, and those of the buckle are either slightly positive (*AMBER94*: 1.0°) or negative (X-PLOR: –2.0°, *Flex*: –1.4°). The propeller twists are always negative with values depending on the method used (X-PLOR: –3.9°, AMBER: –5.2° and Flex: –7.0°), but always in good agreement with the propeller twists found either in solutions (–5.4°) or in crystals (–6.9°) (53). The local slides are negative, ranging from –1.2 to –1.6 Å, while the global slides are near 0 Å (–0.1 Å), as found in other works (52,53). The rises are, whatever the method used, very close (3.2–3.3 Å) and also similar to those found by NMR and X-ray studies (53,54), while the mean twist values are identical to values generally found in solution (34–34.3°) (53,55,56). Globally, the values of helical parameters are in good accord with those reported in the literature.

In the following we concentrate on the local variations within the double helix. The analysed helical parameters, either at the base-pair level or at the base-pair step level, are given in Figure 1. Values of the tip angles fluctuate within ±5°, with the largest positive tip value for the A11:T14 base-pair (with the three methods). The latter further displays a large positive value for its buckle, although this is found to be largest at the central G6:C19 base-pair. Such a large positive buckle for the first G:C base-pair of a GpC step has been already reported by several authors (20,52). The propeller twist is negative at each base-pair with JUMNA (both *AMBER94* and *Flex*), while some positive values are obtained with X-PLOR. Indeed, the three protocols provide almost identical propeller twist profiles for all the steps, except T8pT9 which is poorly defined by NMR data, as we will see below. Globally, the propeller twist values appear strongly negative in the 5′-T2A3C4-3′ segment of the molecule, then near null or weak (positive or negative) in the 5′-C4T5G6C7-3′ segment (rich in G-C base-pairs) and rather negative in the last segment of the molecule (rich in A:T base-pairs). The A:T base-pairs generally prefer negative propeller twists as its has been shown by both NMR (18,41,51,57) and high resolution X-ray crystallography (54).

The values of rolls, twists and rises (local inter base-pair parameters) are also shown in Figure 1. Although a certain scattering among the twist values provided by the three protocols existed, the twist is always large at the G6pC7 and T2pA3 steps, while it appears similar to that found in regular B-DNA at the T10pA11 step. The discrepancy between some of the X-PLOR and the JUMNA data mentioned above for the T8pT9 step is also shown by the twist. Due to cross-peak overlaps, the number of nOes that are collected to define the conformation of T8 relative to T9 appears too weak to provide a correct distance refinement with JUMNA. The latter, in contrast to X-PLOR, cannot accommodate restraints implying the methyl groups. Although this represents a real handicap, we noted that the use of the available restraints produces a significant effect during the data refinement. For instance, the energetical

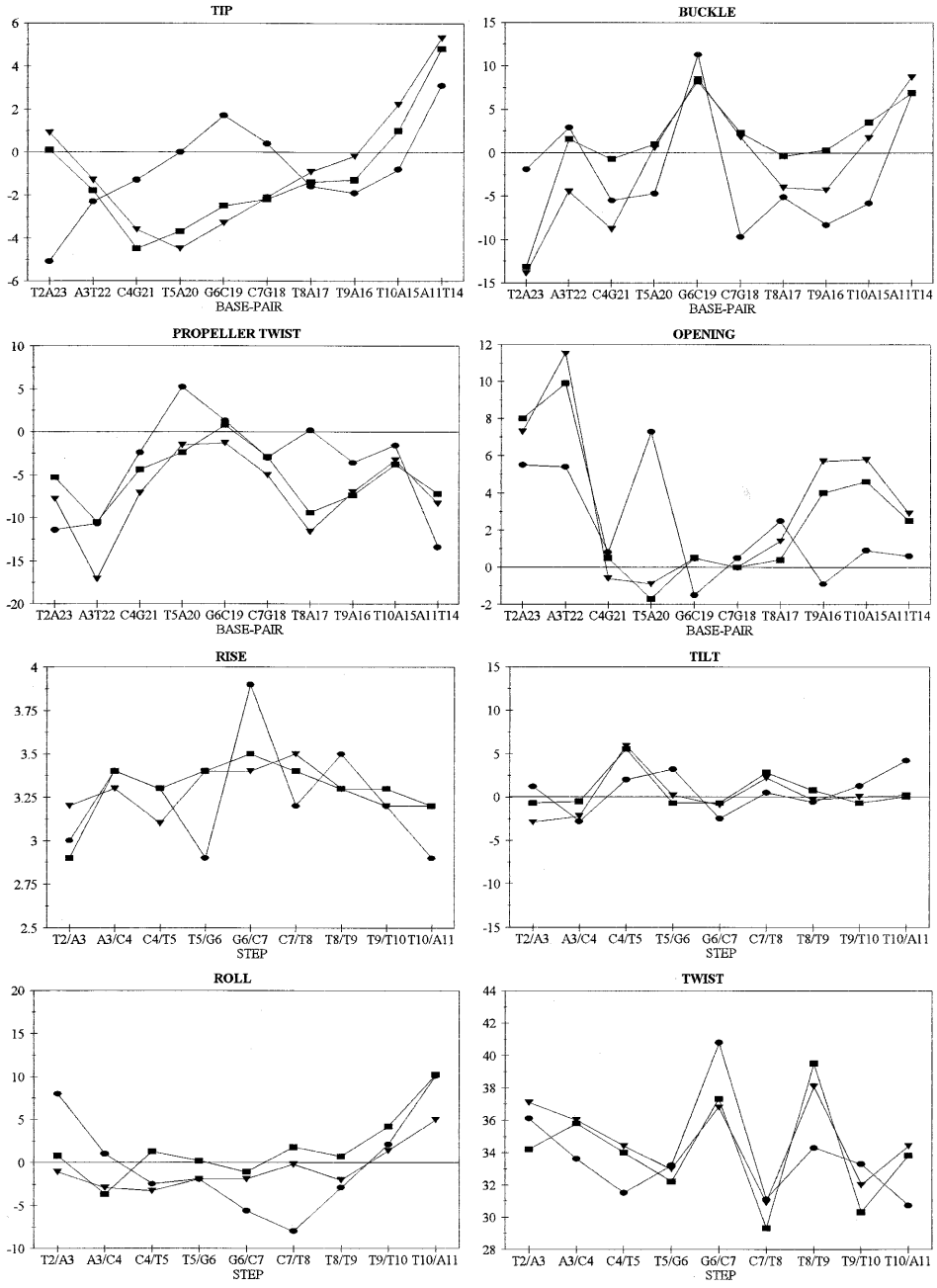

Figure 1. Mean helical parameters of the d(CTACTGCTTTAG)·d(CTAAAGCAGTAG) structures obtained with the three different approaches [(square) JUMNA-*AMBER94*, (triangle) JUMNA-*Flex* and (circle) X-PLOR]; calculated as the average of the individual values of each molecule.

minima provided by JUMNA (*Flex* and *AMBER94*) without restraints yield high twist values (~40°) for both T8pT9 and T9pT10. The introduction of restraint lowers the twist of T9pT10 to ~30–32° and remains ineffective on the T8pT9 twist for which nearly no NMR restraints are available. In contrast, in the restrained X-PLOR structures where the distances with methyl groups are taken into account, the twist value is found the same within T8pT9 and T9pT10 (~33°). Thus, a doubt might be cast on the twist value of T8pT9 provided by JUMNA, although this value enters the range of values reported for a TpT step: 32° (52), 41 and 35° (51), 34° (58), 35° (41), 32, 35 and 37° (59), 35° (53) for solution structures; and 33, 37° for X-ray crystallographic structures at high resolution (54). The large twist value (37–41°) found for the G6pC7 step is similar to the twist values often reported in the literature (18,51–53). Also noticeable is the difference observed on the twist values of the two TpA steps, this being rather high in T2pA3 and medium to low in T10pA11, illustrating once again the influence of the sequence context. Twist values in the other steps were rather in the average for B-DNA.

The sequence effects on the rise appear more accentuated with X-PLOR compared to others. The rise profiles illustrate the propensity of the TpA steps for the low rise values, and, in contrast, of the GpC step for the high rise values. For each

protocol the roll profile reflects a large positive value at the T10pA11 step, while for the T2pA3 step only X-PLOR leads to a significantly positive roll value. Such a positive roll for the TpA steps conforms to the recent NMR literature (41,51,52,54,58). The remaining pyrimidine–purine step of the sequence, T5pG6, displays a near null roll value, while in the literature positive values for this step (18,20) are more frequent than low values (52). Thus, here again, the roll angle depends greatly on the sequence context (51). The main characteristic of the oligonucleotide is the large positive roll at the T10pA11 step correlated to a low rise and to a widening of the minor groove as shown in Table 3. The refinement of distances between the adenine H2 proton and H1′ proton of a 3′-neighbouring residue on the complementary strand allows us to constraint minor groove. The absence of constraints for residues at the centre of sequence [including the A20(H2)-G6(H1′) distance that could not be measured because of severe overlap of resonances] can explain the differences observed among the structures obtained with each protocol. Although the three force fields give similar trends, there are noticeable differences where constraints are lacking, showing the importance of the latter for structure determination.

Table 3. Minor groove width (phosphorus–phosphorus separation minus 5.8 Å) of the average molecules obtained with the three different approaches

	X-PLOR (Å)	AMBER (Å)	Flex (Å)
T(5)-G(24)	5.5	6.4	5.8
G(6)-A(23)	5.5	5.8	5.6
C(7)-T(22)	5.1	6.3	5.8
T(8)-G(21)	4.0	5.9	5.5
T(9)-A(20)	3.9	5.6	4.9
T(10)-C(19)	4.3	5.9	4.9
A(11)-G(18)	6.3	6.9	5.5
G(12)-A(17)	7.2	7.5	6.8

The profile of rise, tilt, roll and twist values of each strand taken separately have been also examined (not shown). An important point is the agreement between the three protocols which are always found to be better for strand II than for strand I. This difference was assigned to the fact that strand I possesses more thymines than strand II and can therefore be better assessed by X-PLOR, which uses restraints on methyls, than by JUMNA. Moreover, the sequence effects described above for the double strand always appear larger for the strands taken separately. For instance, the local inter base-pair tilt can be very weak while the local inter base tilt can be large: either negative or positive. Examples are the TpA steps and also the central G6pC7 step.

DISCUSSION

The salient feature of the d(CTACTGCTTTAG)·d(CTAAAG-CAGTAG) double helix is its conformational heterogeneity, shown by the three minimisation protocols, X-PLOR, JUMNA-*AMBER94* and JUMNA-*Flex*. This can be attributed to the particular sequence of the oligonucleotide that contains a 5′-TTT-3′ run, several TpA steps with one of them ending the 5′-TTT-3′ run, and a GpC step at its centre. No curvature is detected as the distribution of roll angles (or tilt angles) in the double helix does not allow its induction. Several structural parameters deviate from the standard B-DNA values in the 5′-TTT-3′ run and also at other steps. Most of our present (calculated) and previously reported (experimental) parameter values (4) are found consistent with those observed in the recent refined structure of the Dickerson dodecamer (54). For instance, there are the same minor groove narrowing at the AT tracts and the same large twist at the GpC steps in both dodecamers. The large positive roll at the T10pA11 (T14pA15) step is shown by the three protocols X-PLOR, JUMNA-*AMBER94* and JUMNA-*Flex*. A large positive roll at the T10pA11 step, relative to the other TpA steps of the molecule, thus appears as a characteristic independent on the methodology used. Positive rolls at the pyrimidine–purine steps and opening of the minor groove in DNA duplexes have been identified by NMR and other methods (18,20,51,52,60). The bias of these steps for positive rolls has been also observed in protein–DNA crystal structures (14–16,61). Usually, large positive rolls are accompanied by extremely severe distortions at the base-pair or at the backbone (14,16,61–63).

Various edge-to-edge contacts can be detected in the major groove of our calculated structures. Note for instance the four distances from the amino protons N4H/N6H to the N4/N6 atom, and from the N4H/N6H protons to the O4/O6 atom between two adjacent base-pairs in Figure 2. Since all the base-pairs have one amino group and one carbonyl group oriented in the major groove, the above four contacts, either inter or intra strand, can occur whatever the step is. A rather good linear relationship (R = 0.8) between the average of these four distances and the rolls is obtained considering all the base-pairs of the dodecamer. They are the positive rolls that bring closer the amino and carbonyl functional groups in the major groove. For sufficiently large rolls, as it is the case for the T10pA11 step, the distance between the amino and the carbonyl groups and between an amino group and another amino group of adjacent base-pairs become compatible with the formation of electrostatic or even hydrogen bonds. Of the four interactive distances we note that it is the intra-strand distance between the 5′-carbonyl and the 3′-amino which is the shortest. Although, the T10:A15 base-pair is implied in non-concerted interactions with A11 for T10 and with T14 and A11 for A15, its planarity is not significantly altered and its propeller twist and its buckle can be considered to be small. They are, however, larger at the succeeding base-pair A11:T14, leading to an intra-strand roll between T14 and A15 (carbonyl to amino distance: 2.4 Å) larger the one between T10 and A11 (carbonyl to amino distance: 2.8 Å).

Our structures suggest further that, besides the amino to carbonyl interactions occurring between the edges of adjacent base-pairs (64,65), there are also interactions taking place between two or several amino groups. Such interactions have been found in biomolecules and particularly in nucleic acids and have been discussed in detail in recent papers (66–69). In nucleic acids they may help to maintain or induce positive rolls. In the inventoried DNA–protein complexes the edge-to-edge interactions taking place in the DNA major groove can be correlated to the observation of high positive rolls particularly at the pyrimidine–purine steps, TpA, TpG(CpA) and CpG. In both the TpA and the TpG(CpA) steps, the rolls seem to be

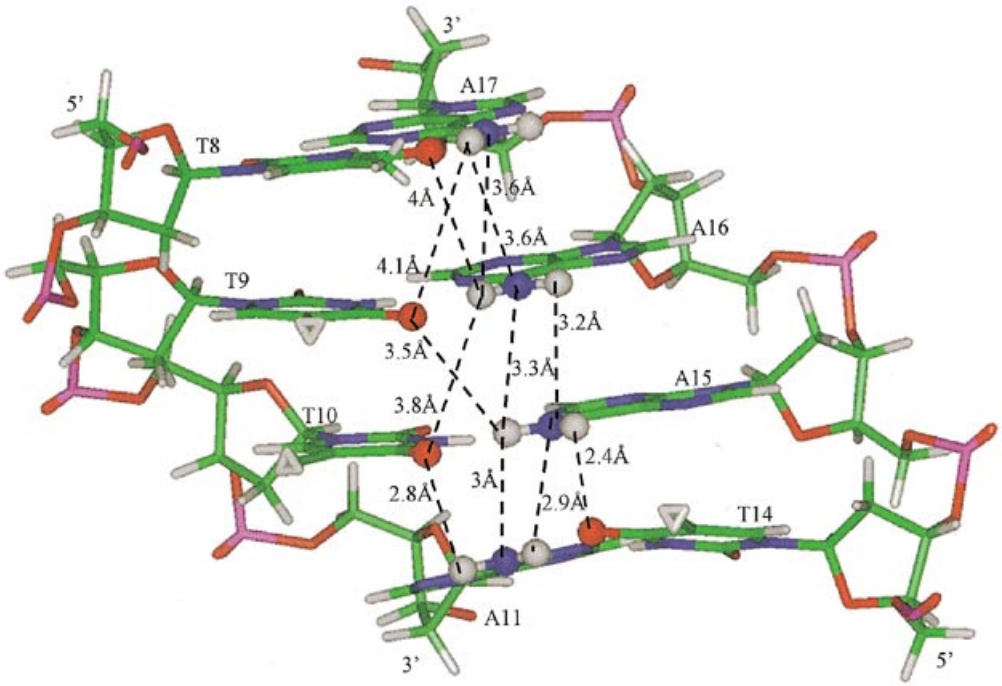

Figure 2. The edge-to-edge interactions implicating the amino (blue) and carbonyl (red) groups along the major groove of the TTTA/AAAT segment for the average structure obtained with X-PLOR. These represent the shortest distances, designated by dashed lines, between the amino protons H6a or H6b (in grey) and the N6 atom or the O4 atom of two contiguous base-pairs. The Watson–Crick interactions are not shown. Note that a positive roll brings closer the interacting atoms.

stabilised by bifurcated hydrogen bonds (15), as illustrated, for instance, in the crystal structure of the TATA box complexed to a yeast TBP (62). Note that interacting groups are not available in the minor groove of the TpA and TpG steps, presumably leading to the rarity of negative rolls at these steps. In contrast, both positive and negative rolls can be observed at the CpG steps (12,15,17) since these may contract amino–carbonyl and amino–amino interactions both in the minor groove and in the major groove. Finally, considering that only the pyrimidine–purine steps can adopt high roll values, their weak stacking properties have been suggested as the best explanation (12,15). It is tempting to assume that when embodied in pyrimidine–purine steps the bases possess more liberties to optimise their edge-to-edge interactions.

Another point was the slow motions exhibited by the adenine A15 (4). The signal broadening effects could be partly caused by the ring currents which are large in the TpAA context, leading in some cases to large changes in chemical shift for small changes in local conformation (5,6). Yet, it seems likely that most of the effects are due to the array of short edge-to-edge interactions slowing the motions of this residue. Figure 2 shows that these interactions are favoured by the propeller twist occurring in the A11:T14 base-pair and that brings closer the edges of the A15 and T14 bases in the major groove. Note also that m

ethylation of the adenine base in TpA steps leads to abolition of its motions (7), likely because the presence of two consecutive methyls on the same strand hinders the inclination of bases, and hence, the maximisation of the edge-to-edge interactions in the major groove. The same kind of reasoning helps us to understand why in straight A-runs only large propeller twists and no large rolls are observed (69), since, on the strand containing thymines, the successive methyls hinder close edge-to-edge contacts in the major groove.

In conclusion, the oligonucleotide d(CTACTGCTTTAG)·d(CTAAAGCAGTAG) duplex adopts a non-regular B-DNA helix conformation with unequal distortions and motions in the two strands. The three approaches X-PLOR, JUMNA-*Flex* and JUMNA-*AMBER94* lead to similar structures where, however, the influence of experimental restraints is the determining factor. Among the various steps, those TpA appear to play a particular role in the DNA structure and dynamics, although the extent of their effects is strongly sequence context dependent. Their possible role of hinge during protein–DNA complex formation has been often evoked (16,61,62,70), as well as their capability to disrupt B′-DNA (16,61,71,72). The recurrent presence of TpA steps in duplexes known for their opening properties is intriguing. Thus, a characteristic feature of the DNA sequences directing transcription initiation, replication initiation and site-specific recombination processes, is that they contain the TpA step either in isolation or as the tandem repeat TATA.

ACKNOWLEDGEMENT

We gratefully acknowledge support for this study from La Ligue Contre le Cancer, Comite du Val de Marne.

REFERENCES

1. Lefevre,J.-F., Lane,A. and Jardetzsky,O. (1985) *FEBS Lett.*, **190**, 37–40.
2. Kennedy,M.A., Nuutero,S.T., Davis,J.T., Drobny,G.P. and Reid,B.R. (1993) *Biochemistry*, **32**, 8022–8035.

3. Spiellmann,H.P. (1998) *Biochemistry*, **37**, 16863–16876.
4. Leporc,S., Mauffret,O., El Antri,S., Convert,O., Lescot,E., Tevanian,G. and Fermandjian,S. (1998) *J. Biomol. Struct. Dyn.*, **16**, 639–649.
5. Schmitz,U., Sethson,I., Egan,W.M. and James,T.L. (1992) *J. Mol. Biol.*, **227**, 510–531.
6. Lane,A.N. and Lefevre J.F. (1994) In James,T.L. (ed.), *Methods in Enzymology*. Academic Press, San Diego, Vol. 239, pp. 596–619.
7. Lingbeck,J., Kubinec,M.G., Miller,J., Reid,B.R., Drobny,G.P. and Kennedy,M.A. (1996) *Biochemistry*, **35**, 719–734.
8. Jacobson,A., Leupin,W., Liepinsh,E. and Otting,G. (1996) *Nucleic Acids Res.*, **24**, 2911–2918.
9. McAteer,K., Ellis,P.D. and Kennedy,M.A. (1995) *Nucleic Acids Res.*, **23**, 3962–3966.
10. Lane,A.N., Jenkins,T.C. and Frenkiel,T.A. (1997) *Biochim. Biophys. Acta*, **1350**, 205–220.
11. Leonard,G.A., Booth,E.D., Hunter,W.N. and Brown,T. (1992) *Nucleic Acids Res.*, **20**, 4753–4759.
12. Bertrand,H., Ha-Duong,T., Fermandjian,S. and Hartmann,B. (1998) *Nucleic Acids Res.*, **26**, 1261–1267.
13. Goodsell,D.S., Kaczor-Grzeskowiak,M. and Dickerson,R.E. (1994) *J. Mol. Biol.*, **239**, 79–96.
14. El Hassan,M.A. and Calladine,C.R. (1998) *J. Mol. Biol.*, **282**, 331–343.
15. Suzuki,M. and Yagi,N. (1995) *Nucleic Acids Res.*, **23**, 2083–2091.
16. Dickerson,R.E. (1998) *Nucleic Acids Res.*, **26**, 1906–1926.
17. Chaoui,M., Derreumaux,S., Mauffret,O. and Fermandjian,S. (1999) *Eur. J. Biochem.*, **259**, 877–886.
18. Dornberger,U., Flemming,J. and Fritzsche,H. (1998) *J. Mol. Biol.*, **284**, 1453–1463.
19. Leijon,M., Zdunek,J., Fritzsche,H., Sklenar,H. and Graäslund,A. (1995) *Eur. J. Biochem.*, **234**, 832–842.
20. Tonelli,M., Ragg,E., Bianucci,A.M., Lesiak,K. and James,T.L. (1998) *Biochemistry*, **37**, 11745–11761.
21. Lefebvre,A., Mauffret,O., El Antri,S., Monnot,M., Lescot,E. and Fermandjian,S. (1995) *Eur. J. Biochem.*, **229**, 445–454.
22. Brünger,A.T., Clore,G.M., Gronenborn,A.M., Saffrich,R. and Nilges,M. (1993) *Science*, **261**, 328–331.
23. Schmitz,U. and James,T.L. (1995) In James,T.L. (ed.), *Methods in Enzymolology*. Academic Press, San Diego, Vol. 261, pp. 3–44.
24. Marion,D. and Wüthrich,K. (1983) *Biochem. Biophys. Res. Commun.*, **113**, 967–974.
25. Griesinger,C., Otting,G., Wüthrich,K. and Ernst,R.R. (1988) *J. Am. Chem. Soc.*, **110**, 7870.
26. Kellogg,G.W. and Schweitzer,B.I. (1993) *J. Biomol. NMR*, **3**, 577–595.
27. Piotto,M., Saudek,V. and Sklenar,V. (1992) *J. Biomol. NMR*, **2**, 661–666.
28. Lavery,R. (1988) In Olson,W.K., Sarma,M.H. and Sundaralingam,M. (eds), *Structure and Expression.* Adenine Press, New York, Vol. III, pp. 191–211.
29. Lavery,R., Zakrzewska,K. and Pullman,A. (1984) *J. Comp. Chem.*, **5**, 363–373.
30. Lavery,R., Sklenar,H., Zakrzewska,K. and Pullman,B. (1986) *J. Biomol. Struct. Dyn.*, **3**, 989–1014.
31. Cornell,W.D., Cieplak,P., Bayly,C.I., Gould,I.R., Merz,K.M., Ferguson,D.M., Spellmeyer,D.C., Fox,T., Caldwell,J.W. and Kollman,P.A. (1995) *J. Am. Chem. Soc.*, **117**, 5179–5197.
32. Lefebvre,A., Mauffret,O., Lescot,E., Hartmann,B. and Fermandjian,S. (1996) *Biochemistry*, **35**, 12560–12569.
33. Lefebvre,A., Mauffret,O., Hartmann,B., Lescot,E. and Fermandjian,S. (1995) *Biochemistry*, **34**, 12019–12028.
34. Lavery,R. and Sklenar,H. (1988) *J. Biomol. Struct. Dyn.*, **6**, 63–91.
35. Poncin,M., Hartmann,B. and Lavery,R. (1992) *J. Mol. Biol.*, **226**, 775–794.
36. Kim,S.-G. and Reid,B.R. (1992) *Biochemistry*, **31**, 12103–12116.
37. Boelens,R., Köning,T.M.G., Van der Marel,G.A., Van Boom,J.H. and Kaptein,R. (1989) *J. Magn. Reson.*, **82**, 290–308.
38. Banks,K.M., Hare,D.R. and Reid,B.R. (1989) *Biochemistry*, **28**, 6996–7010.
39. Huang,P. and Eisenberg,M. (1992) *Biochemistry*, **31**, 6518–6532.
40. Baleja,J.D., Germann,M.W., van de Sande,J.-H. and Sykes,B.D. (1990) *J. Mol. Biol.*, **215**, 411–428.
41. Schweitzer,B.I., Mikita,T., Kellogg,G.W., Gardner,K.H. and Beardsley,G.P. (1994) *Biochemistry*, **33**, 11460–11475.
42. Saenger,W. (1984) *Principles of Nucleic Acid Structure*. Springer Verlag, New York.
43. Ryckaert,J.-P., Ciccotti,G. and Berendsen,H.J.C. (1977) *J. Comp. Phys.*, **23**, 327–341.
44. Kumar,R.A., Ikemoto,N. and Patel,D.J. (1997) *J. Mol. Biol.*, **265**, 173–186.
45. Kumar,R.A., Ikemoto,N. and Patel,D.J. (1997) *J. Mol. Biol.*, **265**, 187–201.
46. Brünger,A.T. (1992) *X-PLOR, version 3.1. A System for X-Ray Crystallography and NMR.* Yale University Press, New Haven, CT.
47. Hosur,R.V., Ravikumar,M., Chary,K.V.R., Sheth,A., Govil,G., Zu-Kun,T. and Miles,H.T. (1986) *FEBS Lett.*, **205**, 71–76.
48. Wüthrich,K. (1986) *NMR of Proteins and Nucleic Acids.* John Wiley & Sons, New York.
49. Gorenstein,D.G. (1994) *Chem. Rev.*, **94**, 1315–1338.
50. Allain,F.H.-T. and Varani,G. (1997) *J. Mol. Biol.*, **267**, 338–351.
51. Weisz,K., Shafer,R.H., Egan,W. and James,T.L. (1994) *Biochemistry*, **33**, 354–366.
52. Mujeeb,A., Kerwin,S.M., Kenyon,G.L. and James,T.L. (1993) *Biochemistry*, **32**, 13419–13431.
53. Ulyanov,N.B. and James,T.L. (1995) In James,T.L. (ed.), *Methods in Enzymology.* Academic Press, San Diego, Vol. 261, pp. 90–120.
54. Shui,X., McFail-Isom,L., Hu,G.G. and Williams,L.D. (1998) *Biochemistry*, **37**, 8341–8355.
55. Peck,L.J. and Wang,J.C. (1981) *Nature*, **292**, 375–378.
56. Rhodes,D. and Klug,A. (1980) *Nature*, **286**, 573–578.
57. Lam,S.L. and Au-Yeung,S.C. (1997) *J. Mol. Biol.*, **266**, 745–760.
58. Gray,B.N., Owen,E.A. and Keniry,M.A. (1994) *Eur. J. Biochem.*, **226**, 115–124.
59. McAteer,K., Jing,Y., Kao,J., Taylor,J.S. and Kennedy,M.A. (1998) *J. Mol. Biol.*, **282**, 1013–1032.
60. Brahms,S., Fritsch,V., Brahms,J.G. and Westhof,E. (1992) *J. Mol. Biol.*, **223**, 455–476.
61. Dickerson,R.E. and Chiu,T.K. (1997) *Biopolymer*, **44**, 361–403.
62. Kim,Y., Geiger,J.H., Hahn,S. and Sigler,P.B. (1993) *Nature*, **365**, 512–527.
63. Schultz,S.C., Shields,G.C. and Steitz,T.A. (1991) *Science*, **253**, 1001–1007.
64. Timsit,Y., Vilbois,E. and Moras,D. (1991) *Nature*, **354**, 167–170.
65. Nelson,H.C.M., Finch,J.T., Luisi,B.F. and Klug,A. (1987) *Nature*, **330**, 221–226.
66. Shatzky-Schwartz,M., Arbuckle,N.D., Eisenstein,M., Rabinovich,D., Bareket-Samish,A., Haran,T.E., Luisi,B.F. and Shakked,Z. (1997) *J. Mol. Biol.*, **267**, 595–623.
67. Sponer,J., Florian,J., Hobza,P. and Leszczynski,J. (1996) *J. Biomol. Struct. Dyn.*, **13**, 827–833.
68. Sponer,J., Leszczynski,J. and Hobza,P. (1996) *J. Biomol. Struct. Dyn.*, **14**, 117–135.
69. Luisi,B., Orozco,M., Sponer,J., Luque,F.J. and Shakked,Z. (1998) *J. Mol. Biol.*, **279**, 1123–1136.
70. Travers,A.A. (1985) In Lilley,D.M.J. (ed.), *DNA–Protein Structural Interaction.* IRL Press, Oxford University Press, Oxford, New York, Tokyo, pp. 49–68.
71. Leroy,J.L., Chatretier,E., Kochoyan,M. and Guéron,M. (1988) *Biochemistry*, **27**, 8894–8898.
72. Sprous,D., Young,M.A. and Beveridge,D.L. (1995) *J. Mol. Biol.*, **285**, 1623–1632.