Journal of Computer-Aided Molecular Design, 8 (1994) 307–322
ESCOM

J-CAMD 244

Molecular dynamics simulations of oligonucleotides in solution: Visualisation of intrinsic curvature

Osmar Norberto de Souza and Julia M. Goodfellow*

Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, U.K.

Received 17 June 1993
Accepted 27 November 1993

Key words: Molecular dynamics; Oligonucleotide conformation; Intrinsic curvature

SUMMARY

We have undertaken molecular dynamics simulations on the d(CGCAAAAAAGCG)•d(CGCTTTTTTGCG) dodecamer in solution. In this study, we focus on aspects of conformation and dynamics, including the possibility of cross-strand hydrogen bonds. We compare our results with those from crystallography as well as infrared, Raman and NMR spectroscopy and cyclization kinetics. Our method of analysis allows us to visualise the curvature of the helix as a function of time during the simulation. We find that the major distortions of the helix axis path occur at the junctions between the (essentially straight) A-tract and the CG- and GC-tracts, although at one junction this is due to hyperflexibility (i.e., regions of high flexibility with no preferred direction of curvature), while at the other junction a static curvature is found (i.e., a preferred, sustained direction of curvature).

INTRODUCTION

There is increasing evidence that oligonucleotide sequences may adopt curved conformations, either because of sequence-directed intrinsic bending or on interaction with proteins [1,2]. Such helix curvature affects recognition properties by changing structural features, e.g. groove width. Bending models can be divided into two main types – the wedge [3–5] and junction [6,7] models. The former involves a series of small independent roll or tilt components between adjacent base pairs, whereas the latter involves specific junctions at which bending occurs with effectively 'straight' regions in between.

The analysis of oligonucleotide crystal structures has not provided an unambiguous model. These data show that A-tracts themselves are likely to be fairly rigid, with good base stacking and bifurcated hydrogen bonds [8–10], although the role of these bifurcated hydrogen bonds has been questioned [11–13]. Moreover, one structure has provided evidence that bending of oligo-

*To whom correspondence should be addressed.

nucleotides in crystals is due to packing interactions in the lattice and it is not some intrinsic sequence-dependent effect [10].

An alternative way to approach the study of DNA curvature is through the use of computer simulation techniques. One such technique is molecular dynamics (MD) [14–16], which has the capability of modelling the structure and dynamics of DNA at the atomic level. Such methods are now used routinely to study DNA conformation and dynamics and applications include the study of drug–DNA interactions [17–19], DNA hydration [20–23], mispairs and modified bases within duplex DNA [24,25].

In this study, we have used MD simulations to investigate aspects of conformation and dynamics of the d(CGCAAAAAAGCG)•d(CGCTTTTTTGCG) dodecamer in solution, a sequence whose structure has been solved by X-ray crystallography [9]. In this paper, we focus on visualisation and analysis of intrinsic curvature, as well as a description of the major conformational features.

METHODS

The initial model of the d(CGCAAAAAAGCG)•d(CGCTTTTTTGCG) double helix in the canonical B-form (B82) was model built, using the coordinates from the Arnott fibre diffraction model [26] and contained 538 atoms, including polar hydrogens atoms. A model of the same sequence but in the A-form (A82) was also built, using the same procedure for comparison. The aqueous environment of DNA was modelled by adding 22 hexa-hydrated sodium counterions to the structure using the standard AMBER procedure. We used the united-atom force field parameters [27] with a distance-dependent dielectric function. The use of a distance-dependent dielectric function and hydrated counterions has previously been found to reproduce experimental data and maintain helix stability [23]. Energy minimisation, MD and analyses of hydrogen bonds and sugar pucker were performed using AMBER 3.0 Rev. A [28], implemented on our in-house CONVEX C220 and on the Convex at ULCC. The MD trajectories were visualised with the programs QUANTA on a Silicon Graphics IRIS workstation and FRODO on a PS330 Evans & Sutherland. Scheme 1 shows the nomenclature used for this dodecamer.

We started by energy-minimising our initial model, using conjugate gradient minimisation until the rms energy gradient was <0.1 kcal mol^{-1} Å^{-1}. The initial velocities were taken from a Maxwell–Boltzmann distribution at 10 K. In the first 3 ps the system was heated from 10 to 298 K by coupling to a stochastic heat bath, with a temperature coupling parameter of 0.25 ps^{-1}. The system was then allowed to equilibrate for 47 ps and the simulation continued for another 100 ps, during which time averaged properties were analysed. The temperature during the simulation was 298 ± 5 K. The SHAKE algorithm was used with an integration time step of 2 fs and a residue-based nonbonded cutoff of 12 Å. The nonbonded pair list was updated every 12 time steps. A total of 3000 structures were saved for the MD trajectory analyses.

5'	C1	G2	C3	A4	A5	A6	A7	A8	A9	G10	C11	G12
	|	|	|	|	|	|	|	|	|	|	|	|
3'	G24	C23	G22	T21	T20	T19	T18	T17	T16	C15	G14	C13

Scheme 1. Numbering of the simulated sequence.

For the analyses and display of helicoidal parameters and curvature, we used the CURVES software [29,30] and parameters were displayed using DIALS and WINDOWS [31]. In this representation the backbone torsion angles are displayed using DIALS, following the usual convention with 0° at the top and increasing in the clockwise direction. Standard nomenclature is used [32], with the sequence running from 5' (top of figure) to 3' (bottom of figure). The parameters describing the detailed conformation of the base pairs are displayed along the horizontal axis in the WINDOW fashion, with the time running vertically in each plot, increasing from bottom to top. The CURVES algorithm minimises a function that describes simultaneously the deflection of the helix axis from linearity and the change in the helicoidal parameters at each base step. As a consequence, we do not necessarily obtain a straight helix axis, but rather a distorted or curved one.

We have measured possible curvature in the following ways:

(1) The length of helical molecules may be characterised by their helix end-to-end distance, EED, and the contour or path length, PL [34,35]. The difference, DPATH, between PL and EED has been used as a high-resolution measure of DNA curvature [31]. If the DNA sequence were straight, then the helix axis would also be straight and DPATH would be zero. A positive DPATH means that the helix axis is distorted from linearity; the greater the difference, the greater the distortion or curvature. This parameter can be used to study the curvature of the whole sequence or parts (with a minimum of three base pairs) thereof.

(2) The angle between the first and last helical axis segment (ANG1) [29,30] and the mean angle between successive local axes (ANG2) [36] have also been used to estimate the molecular curvature of DNA fragments. In an ideal, uniformly curved DNA fragment, ANG1 should equal ANG2, multiplied by the number of local axes [36].

(3) The visualisation of curvature is obtained by projecting the helix axis path on the X-Z and Y-Z planes. The dodecamer fragment is divided into three subfragments, CGC (i.e., CG-tract), A_6 (i.e., A-tract) and GCG (i.e., GC-tract). In doing so not only can we see the curvature pattern of the sequence along the dynamics trajectory, but also the relative orientation of these three subfragments at their interfaces or junctions.

(4) The direction of curvature can also be analysed by calculation of the base axis junction parameters such as ROLL (the angle between mean base-pair planes when each base pair rotates about its long axis), TILT (the angle between mean base-pair planes when each base pair rotates about its short axis), ATIP (the orientation of the global axis about the long base-pair directions), and AX (the dislocation of the global axis along the short base-pair direction), at each base-pair step with respect to a straight reference axis.

RESULTS

Equilibration

At the beginning of the simulation the dodecamer is not in an equilibrium configuration, as can be seen from the changes which occur in the individual terms of the potential energy function during the first 50 ps. However, the plot of the total potential energy during the simulation (Fig. 1) shows that the system remains energetically stable after our extended equilibration period of 50 ps.

The overall root-mean-square (rms) deviation also provides information on the internal changes which the system undergoes during the simulation. In order to compare the dynamic

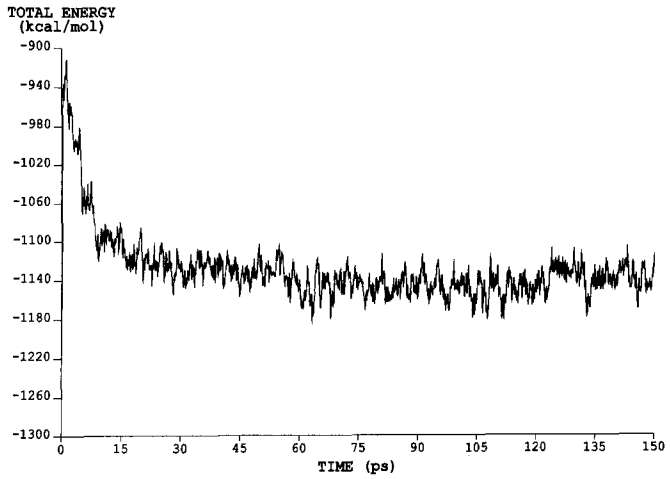

Fig. 1. Total potential energy (kcal mol^{-1}) versus length of simulation (ps).

structure with the initial configuration, B82, and the canonical A-DNA conformation, A82, we superimpose the simulated conformations on these structures, having optimised the overlap by rotating and translating the structures. Polar hydrogen atoms and the counterions were not included in this calculation, because of their inherent large motional amplitudes. The rms deviation is monitored throughout the simulation period (Fig. 2). It can be seen that at the start of the simulation, the initial conformation is very different from the canonical A-form, with an rms of 5 Å. This changes during the course of the simulation, as the deviation of the dynamic structure from the canonical B-form (thicker line in Fig. 2) not only increases but also approaches a state intermediate between the B- and A-forms of the dodecamer; this is confirmed by the second graph (thinner line). The final rms deviations at the end of the simulation run are 3.04 and 3.05 Å for the B- and A-forms, respectively.

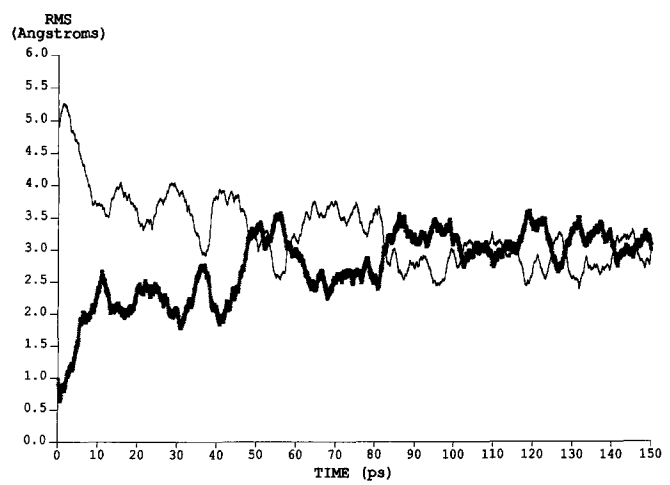

Fig. 2. Root-mean-square deviation (Å) from the canonical B-form (thick line) and the A-form (thin line) as a function of the length of simulation (ps).

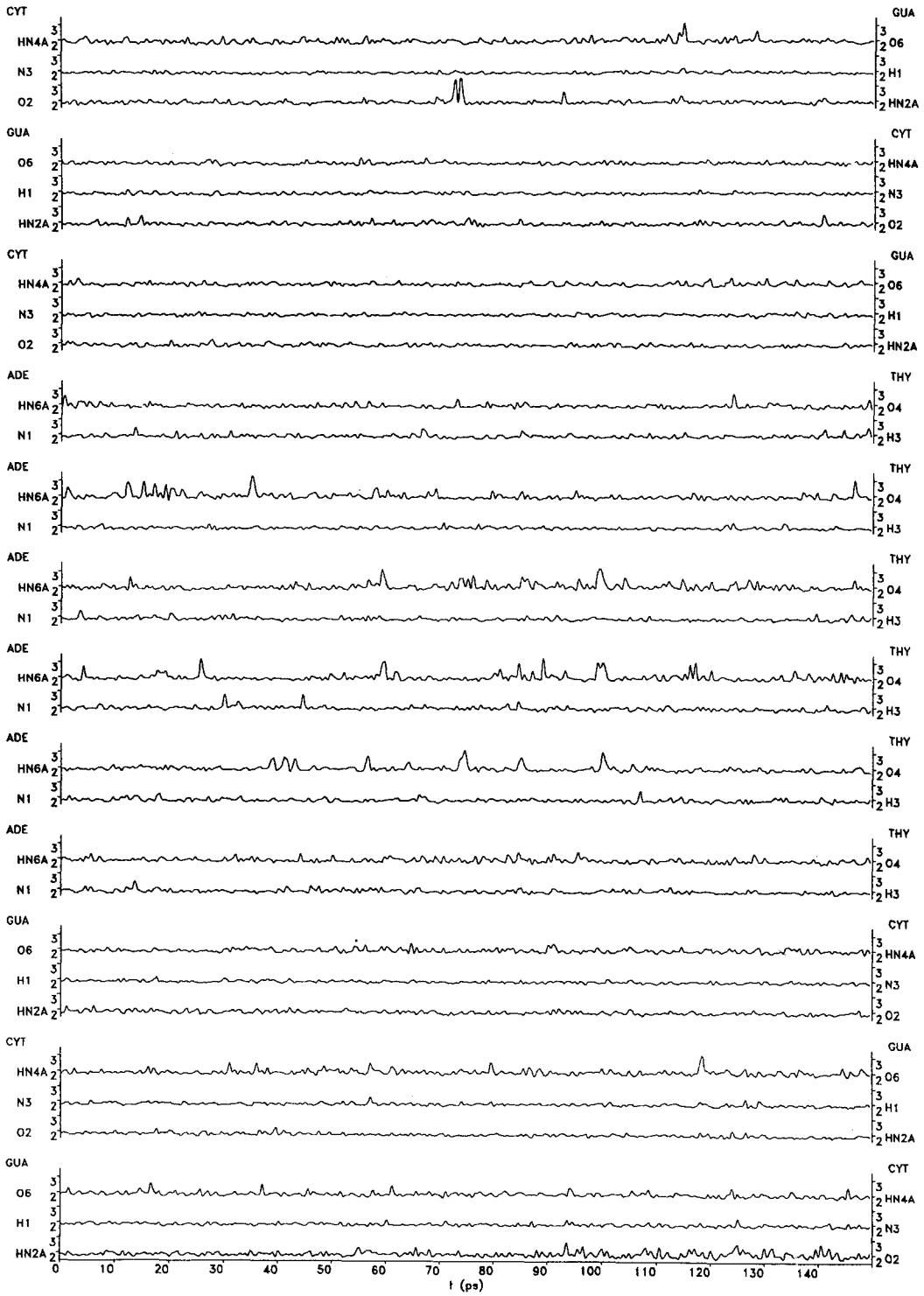

Fig. 3. Time course (ps) of the length of the hydrogen bonds (Å) between base pairs.

Hydrogen bonds

The Watson–Crick hydrogen bonds are fairly stable throughout the time course of the simulation, with only a few occasional and transient disruptions (Fig. 3). The inner hydrogen bonds of the C•G base pairs at either end of the dodecamer are the most stable. It is noticeable, however, that the central four A•T base-pair hydrogen bonds fluctuate in length more than the surrounding base pairs. We also looked at the geometry of the five possible major groove cross-strand hydrogen bonds (CSHB) in the A-tract (i.e., an oligo(dA)•oligo(dT) stretch) of this dodecamer (Fig. 4). They show large fluctuations, even when we use a longer than normal distance cutoff of 3.0 Å, as used by others [33]. Using 2000 snapshots we find that for each of the five hydrogen bonds studied, between 55 and 82% are within this cutoff. The average values of the Watson–Crick N6-(A)/O4-(T) distances are very typical, being between 2.87 and 2.99 Å (and between 1.80 and 2.04 Å for N6-H-(A)/O4-(T)). In contrast, the cross-strand diagonal N6-(A)/O4-(T) distances are much longer, i.e., between 2.99 and 3.34 Å (and 2.66–2.97 Å for N6-H-(A)/O4-(T)).

Sugar puckering and backbone torsion angles

In strand 1, the C4'-C3' exocyclic sugar torsion angle, δ, is relatively stable throughout the simulation period. Its conformational preference is essentially +anticlinal, a form intermediate between t and g^+ of canonical B-DNA and A-DNA, respectively. Strand 2 is not so stable, though for most residues its conformational preference is as described above for strand 1. Exceptions are C15, T16 and T18, which exhibit C4'-C3' torsion angle values in the vicinity of g^+. Towards the end of the simulation, T19 undergoes a transition from g^+ to +anticlinal.

The sugar puckers show the most flexibility of all parameters, although the amplitude is low and stable. The sugar puckers are predominantly C2'-endo in strand 1, with the exception of C1 which prefers the C3'-endo conformation. In strand 2 the sugar puckers give way to other conformations than C2'-endo. C15, T16, T18 and T19 adopt the C3'-endo conformation, although T19 repuckers to C2'-endo at the end of the simulation period. Table 1 shows the percentages of different sugar conformations for each residue during the data collection period. Overall, the

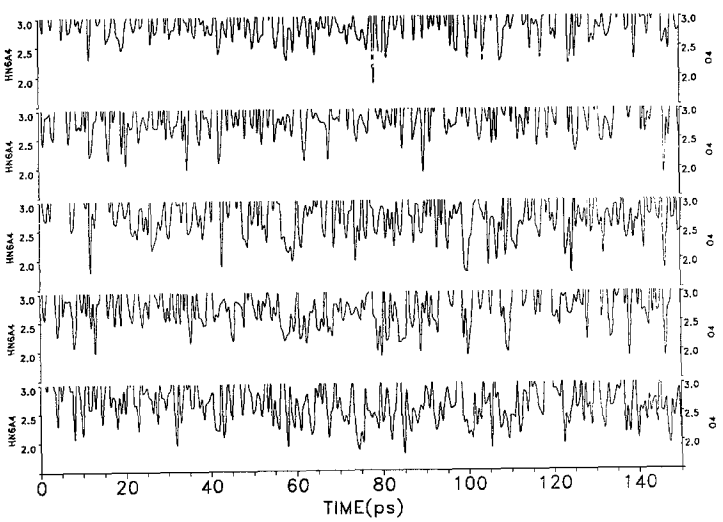

Fig. 4. Time course (ps) of the cross-strand hydrogen bonds (Å).

TABLE 1
PERCENTAGES OF SUGAR PUCKER CONFORMATIONS FOR EACH RESIDUE

Residue	Sugar conformation				
	C3'-endo	O1'-endo	C2'-endo	O1'-exo	C1'-endo
Strand I					
C1	**43.8**	40.4	12.4	0.0	3.4
G2	3.5	4.1	**89.3**	1.5	1.7
C3	0.0	0.3	**99.6**	0.2	0.0
A4	2.9	5.8	**48.3**	27.9	15.3
A5	0.0	8.8	**89.1**	2.1	0.0
A6	0.0	0.7	**99.3**	0.0	0.0
A7	0.4	11.3	**60.4**	27.7	0.3
A8	17.7	34.1	**47.4**	0.8	0.1
A9	0.1	12.8	**86.9**	0.3	0.0
G10	0.0	19.6	**80.4**	0.1	0.0
C11	3.8	3.5	**76.3**	16.5	0.1
G12	8.8	9.7	**78.8**	0.7	2.1
Strand II					
C13	0.6	33.1	**63.4**	3.0	0.0
G14	0.0	33.5	**66.5**	0.0	0.0
C15	**40.4**	6.7	36.7	0.6	15.7
T16	**69.5**	11.4	12.9	0.0	6.2
T17	1.8	9.7	**88.6**	0.0	0.0
T18	**46.5**	30.0	21.8	0.0	1.7
T19	**54.5**	20.0	21.2	0.0	4.4
T20	14.8	16.0	**68.8**	0.3	0.2
T21	0.2	30.3	**69.6**	0.0	0.0
G22	1.9	**50.2**	47.8	0.2	0.0
C23	0.4	14.9	**84.5**	0.3	0.0
G24	3.5	40.4	**55.9**	0.1	0.3

The major conformation is highlighted in bold.

majority of the sugar puckers is in the C2'-endo conformation. The exceptions are the sugars of C15, T16, T18 and T19, which adopt preferably a C3'-endo conformation.

The C3'-O3' torsion ε and the 3'-phosphodiester torsion ζ oscillate about t and g^- (the high end of g^-), respectively. There are, however, a few conformational transitions that deserve comments. After the equilibration period, ε and ζ of C3 behave in a manner consistent with a B_{II} conformation (g^-,t). At about 30 ps, ε of A6 changes abruptly from t → g^- while ζ simultaneously changes from ca. g^- to ca. t. At about 60 ps, ζ of C23 changes from t → g^-. T21 exhibits a g^- → t → g^- transition for ζ, not accompanied by a change in ε, which is very stable. The transition observed for A6 is essentially a permanent B_I → B_{II} transition. Overall, ε and ζ behave in a manner consistent with canonical B-DNA.

The α torsion is g^- in most cases; two deviations are observed. In strand 1, a transitory g^+ → t → g^+ transition is seen for α of A4 at about 35 ps. Torsion angle α of A7 changes from t to g^+. In strand 2, α of C15 and T18 goes from g^- → t and from g^- → +ac, respectively. β and γ are relatively stable. β shows one major deviation, i.e., t → g^+ for C11 at about 30 ps. γ shows a g^+ → t type of deviation for A4, A7 and C11 in strand 1 at the start of the simulation. In strand 2, γ of T18 undergoes a g^+ → t transition at about 70 ps. The expected correlated transition of

α and β $(g^+,g^+) \rightarrow (g^-,t)$ is observed for G2, A9 and G24 at about 25, 50 and 60 ps, respectively. For C11, a $t \rightarrow g^-$ transition in α is accompanied by $t \rightarrow g^+$ in β. The correlated motion between α and γ is observed for T18. It is a crankshaft rotation $(g^-,g^+) \rightarrow (t,t)$ about the O5'-C5' bond.

Flexibility

The individual rms displacement from the averaged atomic position, which provides information on the motional properties of individual atoms or groups of atoms in the system, is shown in Fig. 5. The largest deviations correspond to the atoms in the terminal base pairs on each end of the dodecamer. This may reflect the absence of the crystal environment in our simulation. The atoms of the central four base pairs show smaller rms deviations than the average for the whole sequence. Overall, the rms deviations are largest for the phosphate group and smallest for the base atoms, with those of the furanose ring being intermediate in value.

We have noted that within the A-tract the sugar flexibility decreases from 5' to 3' in strand 1 and is larger than those of the complementary sugars in strand 2. In the latter, the flexibility increases from 5' to 3', although these differences are small. From the rms deviations in Table 2, it can be seen that the phosphate groups at the junctions tend to be more flexible than those of neighbouring base pairs, whereas the atoms of the bases at the junction tend to be less flexible.

Curvature

The parameters describing the helical axis have been obtained from CURVES [29,30], an algorithm that minimises a function describing simultaneously the deflection of the helical axis from linearity and the change in the helicoidal parameters at each base step. This leads to a picture in which the helix axis may be curved. We use the difference (DPATH) between the path length and the end-to-end distance, together with projections of the helix axis path on the X-Z and Y-Z planes, in order to visualise the curvature of the oligonucleotide and to ascertain which regions of the fragment contribute most to the curvature. Analysis of DPATH (Fig. 6) for the whole dodecamer shows that a major change occurs after 40 ps during the equilibration period.

However, the DPATH parameter for the whole dodecamer does not tell us if the fragment is uniformly curved or if there are different contributions to the curvature from different regions of the sequence. This local curvature can be found by calculating the DPATH parameter for

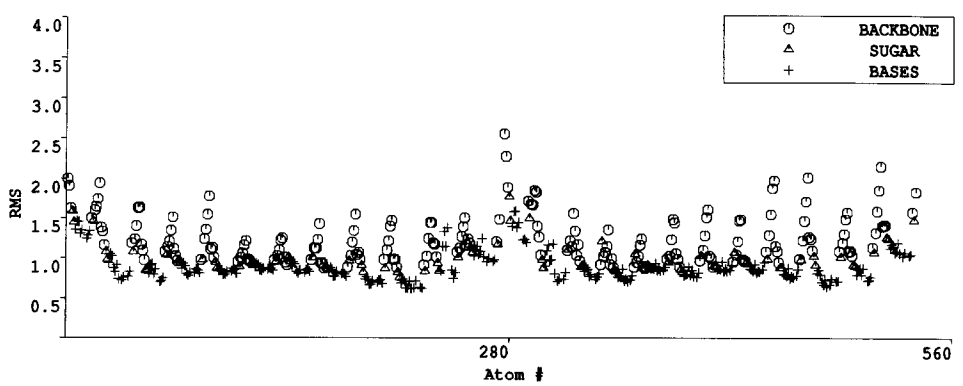

Fig. 5. Root-mean-square deviation (Å) from the average position during the data collection period for the backbone phosphate, sugar and base atoms.

TABLE 2
ROOT-MEAN-SQUARE DEVIATIONS (Å) AROUND AVERAGE POSITIONS FOR THE DODECAMER

Base pair[a]	B1	B2	R1	R2	P1	P2
C1•G24	1.38	1.24	1.60	1.28	–	1.88
G2•C23	0.90	0.94	1.12	0.95	1.78	1.47
C3•G22	0.89	0.82	0.95	1.01	1.56	1.75
A4•T21	0.90	0.79	0.97	0.93	1.36	1.82
A5•T20	0.87	0.85	0.96	0.93	1.56	1.40
A6•T19	0.90	0.81	0.94	0.92	1.15	1.48
A7•T18	0.90	0.78	0.96	0.88	1.19	1.40
A8•T17	0.85	0.81	0.89	0.89	1.27	1.15
A9•T16	0.79	0.71	0.89	0.86	1.37	1.20
G10•C15	0.73	0.89	0.83	0.95	1.37	1.36
C11•G14	1.03	0.95	1.07	1.10	1.39	1.80
G12•C13	1.11	1.43	1.14	1.83	1.40	–
Aver12[b]	0.94	0.92	1.03	1.04	1.40	1.52
Averc4[c]	0.88	0.81	0.94	0.91	1.30	1.36
Averj1[d]	0.90	0.81	0.96	0.97	1.46	1.79
Averj2[e]	0.76	0.80	0.86	0.91	1.37	1.28

[a] B1 and B2 refer to bases in strands 1 and 2; R1 and R2 refer to sugar atoms in strands 1 and 2 and P1 and P2 refer to phosphates in strands 1 and 2.
[b] Aver12 is the average rmsd over 12 bases and sugars and 11 phosphate groups in each strand.
[c] Averc4 is the average rmsd for the central four bases.
[d] Averj1 is the average rmsd for the first junction (base pairs C3•G22 and A4•T21).
[e] Averj2 is the average rmsd for the second junction (base pairs A9•T16 and G10•C15).

different regions, i.e., the initial six-base-pair CGA-tract, the central A-tract, and the final six-base-pair AGC-tract. Then we find that curvature is mainly due to distortions within the final AGC-tract, and that the A-tract is only slightly deflected from linearity.

In order to emphasize the local geometry of these DNA fragments, we again divide the dodecamer into three regions and calculate the local axis of curvature as described above. Little change in the conformation of the central A-tract is seen over 150 ps (Fig. 7). What is clearly visible is the relative inclination and dislocation of the helical axis at the interface of the three

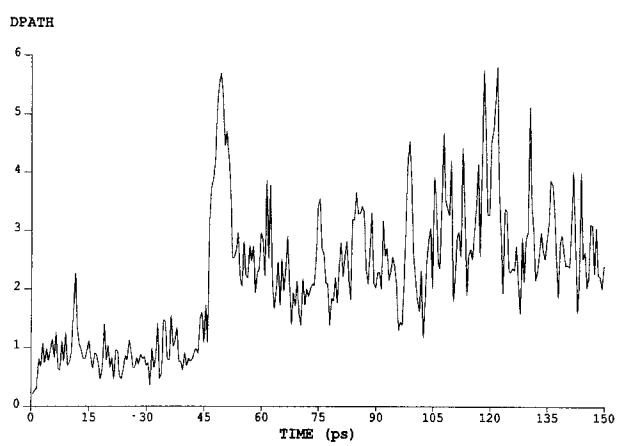

Fig. 6. Time course of the difference between path length and end-to-end distance, DPATH (Å), for the whole sequence as a function of the length of simulation (ps).

A

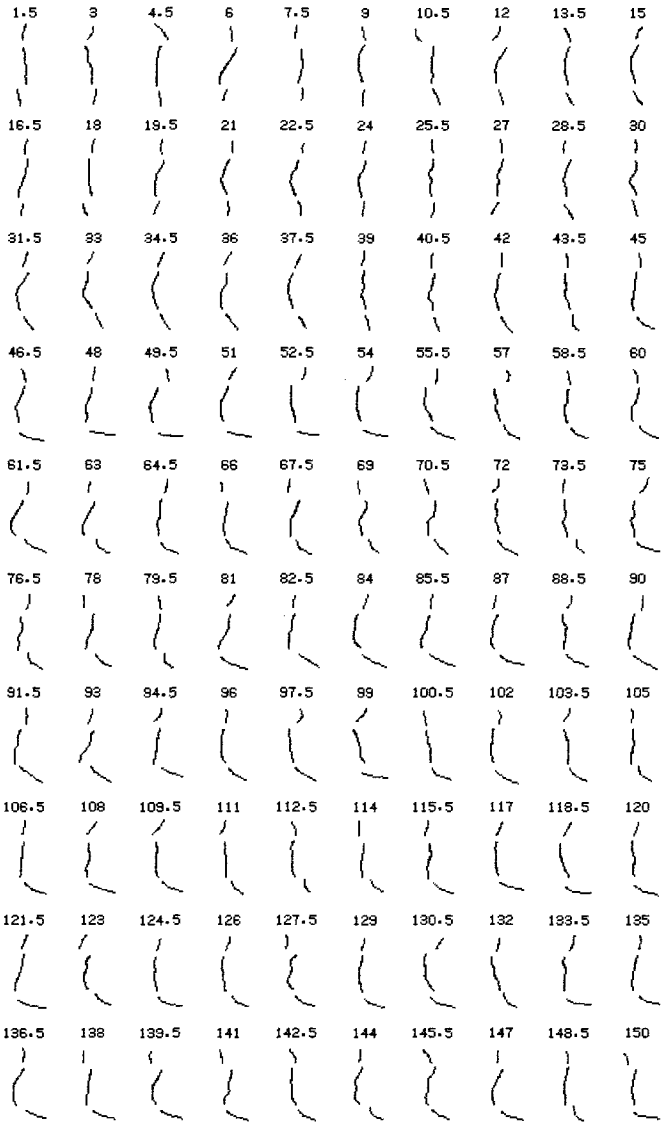

Fig. 7. (A) The X-Z projection and (B) the Y-Z projection of the dodecamer helix, visualising the intrinsic curvature of the initial CG-tract, the central A-tract and the terminal GC-tract.

regions, i.e., at the junctions. The initial fragment can be seen to adopt a range of different inclinations with respect to the central tract, such that on average it might appear straight. In contrast, the final fragment tends to adopt a similar inclination with respect to the central A-tract throughout the data collection period.

Although the major contributions to distortions at the end of the helix come from the bottom junction (junction 2) and the GC-tract itself, we can see some minor degrees of distortion within the A-tract, particularly during the data collection period. The distinct behaviour of the central AAA triplets confers a kinked appearance to the A-tract, which is due to nonuniform changes

B

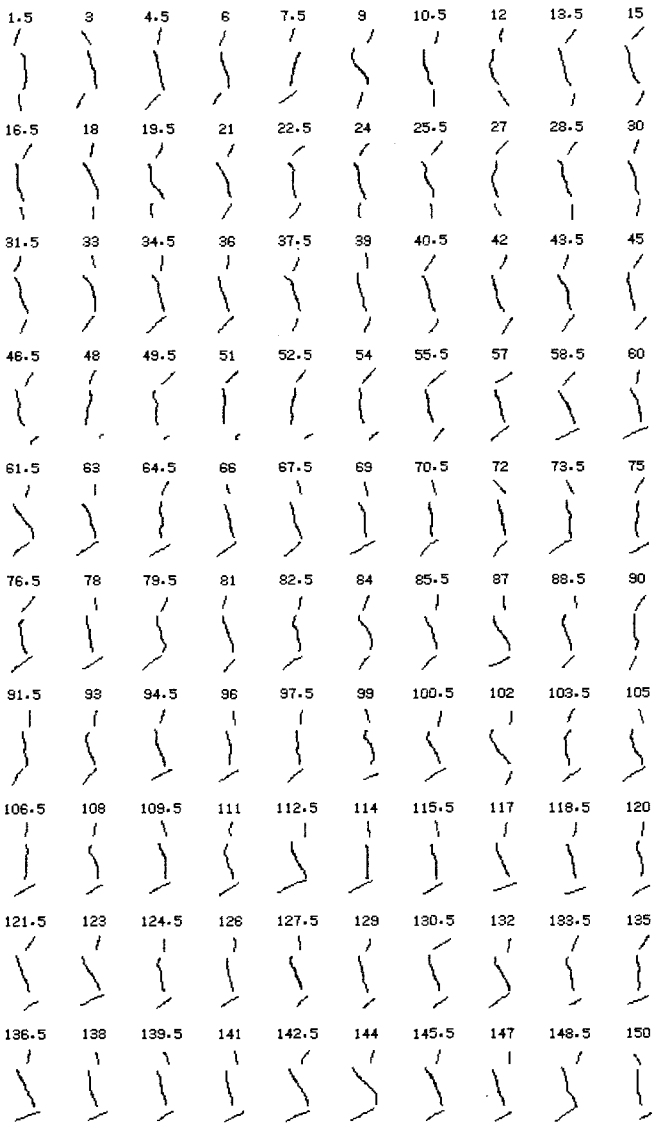

Fig. 7. (continued).

in the helical and axis parameters, as we shall see below. Another interesting feature is the trend of the helix path within the initial CG- and final GC-tracts to have similar inclinations with respect to the central A-tract, leading to an ∫ shaped curve (see Fig. 7B).

We have also looked at parameters at the base-step level (Table 3). The ROLL angle is mainly positive, with the base step narrowing towards the major groove and therefore the dodecamer in these steps curves towards the major groove. The effect of the small negative ROLL value is to attenuate the curvature in the CG-tract. In the final GC-tract, the negative ROLL values curve the double helix towards the minor groove. The large ROLL at step 5 makes the A-tract region look kinked at its centre. In contrast, the TILT angle is small and negative, except at steps 2 and

TABLE 3
AXIS BENDING AND JUNCTION PARAMETERS

Base pair	AX[a] (Å)	AY[b] (Å)	ATIP[c] (deg)	TILT[d] (deg)	ROLL[e] (deg)
C1•G24	0.01	–0.17	3.87	–2.42	7.25
G2•C23	0.29	–0.17	–2.97	2.56	3.92
C3•A22	–0.46	0.01	3.48	–2.92	4.59
A4•A21	0.09	–0.01	4.67	–0.56	2.80
A5•A20	0.32	–0.21	6.89	–0.58	10.66
A6•A19	–0.35	–0.05	1.71	–3.98	–0.12
A7•A18	–0.23	–0.27	7.56	–1.97	5.85
A8•A17	–0.16	0.05	9.82	–3.60	6.21
A9•G16	0.15	0.07	8.62	–5.31	4.76
G10•C15	1.00	0.12	–8.72	2.88	–13.26
C11•G14	–0.91	–0.19	–7.58	–1.42	–4.64

[a] AX: axis dislocation in the x-direction.
[b] AY: axis dislocation in the y-direction.
[c] ATIP: axis inclination about y-axis.
[d] TILT: the angle between mean base-pair planes when each base pair rotates about its short axis.
[e] ROLL: the angle between mean base-pair planes when each base pair rotates about its long axis.

10 where it becomes positive. The initial CG-tract is therefore almost straight, though inclined with respect to a straight global helix axis through the structure.

Inspection of the helix axis bending parameters shows that changes in ATIP and AX occur mainly near or at the junctions. Within the central A-tract, ATIP increases monotonically up to step 6, where it is small compared to the values of the flanking steps. It then continues to increase and becomes negative in the last two steps. AX is very small and positive for the first two steps. It becomes relatively large at step 3, which corresponds to the first junction. Within the A-tract, AX is small and positive, except for the last three steps. At the bottom junction (junction 2), AX is small and negative. AX has high values with opposite signs at steps 10 and 11.

The angle between the local helix axes at the ends of the dodecamer (ANG1) [29,30] has been used also to measure the total molecular curvature of a DNA fragment. In this simulation the average value of this parameter is about 45°. The mean angle (ANG2) between successive local helix axes [36] is also considered a good indicator of the total molecular curvature and has an average value of 8.5°. The ANG2 parameter indicates that the dodecamer is not smoothly curved; the larger values are found in the final five base steps of the helix.

In summary, our analysis of curvature for this sequence shows that the largest deviations of the helix axis from linearity occur near or at the junctions between the A-tract and the CG- and GC-tracts. Secondly, we find that the A-tract is essentially straight. Finally, we observe that there is a slight overall curvature towards the major groove with averages of 8.5° for ANG2 and 45° for ANG1 over the dodecamer.

DISCUSSION AND CONCLUSIONS

It is of interest to compare the results of our MD simulation of the dodecamer d(CGCAAAAAAGCG)•d(CGCTTTTTTGCG) with a variety of available experimental data. We find that the 30 Watson–Crick hydrogen bonds are generally well preserved throughout the simulation period. A few occasional and transient disruptions are seen in the outer hydrogen

bonds within some CG and GC base pairs. We also see slight disruption in the Watson–Crick hydrogen bonds between N6(A) and O4(T), i.e., atoms which also take part in the possible cross-strand hydrogen bonds. We find that the DNA fragment becomes curved with distortions at the junctions between the initial CG-tract and final GC-tract and the central A-tract.

There are now several crystal structures of oligonucleotides which contain an A-tract with three or more adenines [8–10,13,37]. These can be used to assess the importance of the possible cross-strand hydrogen bonds. Some of these data are summarised in Table 4 and we have compared our results with them. Our results, representing averages over 2000 structures saved during 100 ps of the data collection period, agree well with the experimental data if we consider the distance between the cross-strand N6(A) and O4(T) heavy atoms. In a recent paper [37], cross-strand hydrogen bonds were found to be in the range 2.9–3.4 Å. However, the use of more specific geometric criteria, based on the sum of van der Waals radii [13], found only *one weak* cross-strand hydrogen bond. Previous simulations on a poly(dA)•poly(dT) decamer [11] have clearly shown that the lifetimes of cross-strand hydrogen bonds are considerably shorter than those of Watson–Crick hydrogen bonds. Thus, whether a cross-strand hydrogen bond exists depends on the authors definition, but our data are consistent with experimentally determined structures.

We have also shown that the dynamic structure in our simulation resides in an intermediate state between the A- and B-forms of the double helix. The displacement of the base pairs (XDISP) towards the minor groove is another important indicator of the intermediate state adopted by the d(CGCAAAAAAGCG)•d(CGCTTTTTTGCG) dodecamer in this simulation. Such behaviour is expected [19,32,38] taking into account that we, like many others, make use of a dielectric constant to approximate the role of solvent. In this respect, it is interesting that we find the use of a dielectric constant $\varepsilon = r$, together with hydrated counterions, to be sufficient to maintain the stability of the helix without restraining the first or last residues. In previous simulations, we have found this stability maintained over 1 ns simulations.

In our simulation, all the sugar puckers for the adenine residues display a C2'-endo conformation, while three out of six thymine sugars exhibited a C3'-endo pucker. This is in good agreement with theoretical calculations on the lowest energy state of dA_6•dT_6 [39] that found the sugars attached to adenine in a C2'-endo conformation and those attached to thymine in a C3'-endo conformation. Nelson et al. [9] noted some heteronomity and found that the torsion angles in the thymine strand are different and vary more widely than those in the adenine strand. These results are inconsistent with the heteronomous model for poly(dA)•poly(dT) [26], based on model

TABLE 4
A COMPARISON OF THREE-CENTER HYDROGEN BONDS IN VARIOUS OLIGONUCLEOTIDE X-RAY STRUCTURES AND IN COMPUTER SIMULATIONS

	N6(A)–O4(T) distances (Å)		
Reference	Min.	Max.	Avg.
Nelson et al. [9]	3.1	3.4	3.2
Coll et al. [8]	2.7	3.2	3.0
DiGabriele et al. (up) [10]	3.1	3.7	3.4
(down) [10]	2.8	3.3	3.1
MD simulation [11]	3.11	3.23	3.15
This simulation	2.99	3.36	3.19

building fit to fibre diffraction data, which found the sugars of the adenine and thymine strands in the C3'-endo and C2'-endo conformations, respectively. IR data [38] taken of cast films of dA_5dT_5 and poly(dA)•poly(dT) showed that when the relative humidity was decreased, a 'gradual evolution of the geometry' of both polymers into a form intermediate between B- and A-type double helices occurred. The adenine in these DNAs adopts a C3'-endo conformation. Raman [38] studies in solution found 30% of the adenine sugars in the C3'-endo conformation. In summary, the IR and Raman studies mentioned above support the original heteronomous model [26]. However, another recent Raman investigation on the dynamic structure of adenine- and thymine-containing DNA in solution [40] showed that the thymine sugar puckers are displaced from C2'-endo towards O1'-endo, while the adenine sugars predominantly adopt the C2'-endo conformation. Solution NMR [41,42] studies are in complete disagreement with any heteronomous model. One study [41] showed that the adenosines in the oligomer dA_6•dT_6 are in the C2'-endo conformation. In the former study nothing is said about the sugar conformation in the poly(dT) strand. The other study [42] found the sugars in both strands of poly(dA)•poly(dT) in the C2'-endo conformation, characteristic of 'pure' B-DNA. Thus, there are a range of conflicting experimental results, such that one must conclude that sugar pucker is dependent on environment. Our MD results are consistent with theory and recent NMR data.

The flexibility of individual nucleotides is in good qualitative agreement with experimental data [9] and with other molecular dynamics studies [18–25], in that phosphates are more mobile than sugars and sugars are more mobile than bases. However, we also find that the flexibility of the A-tract sugars tends to decrease from 5' to 3' in the adenine strand but increase from 5' to 3' in the complementary thymine strand. This behaviour can be related to hydroxyl radical cutting patterns of kinetoplast DNA [43–45] which are found to be in phase with the adenine tracts but with a rate of cutting which decreases from 5' to 3' and within the poly(dA) strand. This decrease has been attributed to narrowing of the minor groove, which reduces the accessibility of the backbone H4' atom to the OH-radical. However, the reduced flexibility of the sugar moiety within the poly(dA) strand from 5' to 3' may also reduce the accessibility of the H4' atom to the OH-radical. The increased flexibility of the sugar residues in the 5' to 3' direction of the thymine strand would increase the accessibility of the corresponding sugars to the OH-radical, thus promoting an increase in the cutting rate in that direction. Again, this is consistent with the experimental data [43–45].

On average, the CG- and GC-tracts are more flexible than the A-tract, although this may be due at least in part to the location of these regions at the termini. Even given this flexibility, the hydrogen bonds in C•G base pairs are retained throughout the simulation, with the innermost of the three being the most stable. The relatively small flexibility of the A-tract is confirmed by the DPATH parameter, which indicates that only very small changes in the A-tract length occur during the simulation period. The extra rigidity in the A-tract region and the unusual conformational rigidity of poly(dA)•poly(dT) in fibres and in solution [9] may well be due to the formation of cross-strand hydrogen bonds across the major groove of the A-tract, resulting from the high propeller twist of AA/TT steps in A-tracts containing more than two adenines [8–10]. As a result, one of the two Watson–Crick hydrogen bonds between adenine and thymine bases is less stable than the other, i.e., the one in the major groove that is implicated in cross-strand hydrogen bonding.

In order to visualise any possible curvature, we display projections of the helix axis path onto the X-Z and Y-Z planes. Our simulation agrees with the crystallographic study [9] in that the

most abrupt changes of the helix axis direction are found at step 2 (G2•C23/C3•G22) and at step 10 (G10•C15/C11•G14). However, our simulations indicate that the junction close to step 2 is intrinsically flexible but the one close to step 10 shows almost static curvature (over 100 ps). As a consequence, the helix axis path of the *average* structure appears almost straight up to the lower junction, at which it curves smoothly. However, analysis of the junction and bending axis parameters such as ROLL, TILT, ATIP and AX (Table 3) suggests a small net curvature towards the major groove at the first and second junctions. Along the GC-tract the curvature kinks towards the minor groove. The average ROLL directions for purine-pyrimidine and pyrimidine-purine steps found during the simulation are supported by previous energy calculations [1,46].

Our attempts to estimate the net curvature imparted to our dodecamer resulted in the values 45° and 8.5° for ANG1 and ANG2, respectively. The large difference between these two values is a consequence of the nonplanarity of the curvature in this dodecamer. Actually, the helix axis follows a superhelix path in going from the first to the last base pair. The high value of ANG1 is mainly due to the relatively high values of ROLL, ATIP and AX at steps 10 and 11, which produce a kink in the GC-tract (at base pair 11). For crystallographic dodecamers, ANG1 is about 19°, although a recent crystal structure of another dodecamer [48] that crystallises in a different lattice showed an overall bending of 30°. Despite the significant differences between simulation and crystallographically derived values for the curvature, we find that the DNA fragment curves in the direction of the major groove.

Estimations from cyclization kinetics experiments [47] showed that an A_6-tract curves the DNA helix by 17° to 21° and that the curvature is in the direction of the minor groove in the centre of the A-tract [49]. In fact, estimates of curvature produced by phased A-tracts have varied from about 11° to 28° [49]. Such biochemical experiments cannot easily distinguish between wedge, junction or any other models of curvature. We find that there is almost static curvature at the A9•T16/G10•C15 junction in this oligonucleotide. In contrast, the top junction involving C3•G22/A4•T21, which is responsible for the major distortions in the helix axis during the first 45 ps of the equilibration period, subsequently becomes a region with high flexibility but without preferred direction of curvature. Our results show that in this dodecamer curvature occurs essentially at the junctions, but with a small wedge-like component of neighbouring base steps. Further simulations are underway to see if this picture is still true over longer (i.e., nanosecond) time scales.

ACKNOWLEDGEMENTS

O.N.S. thanks the National Research Council of Brazil (CNPq-RHAE) for a Ph.D. scholarship. J.M.G. gratefully acknowledges a Wellcome Research Leave Fellowship and SERC support for both supercomputing resources at ULCC and for in-house computer resources. We would like to thank Dr. L. Cruzeiro-Hansson for all her help during this project.

REFERENCES

1 Travers, A., Annu. Rev. Biochem., 58 (1989) 427.
2 Hagerman, P.J., Annu. Rev. Biochem., 59 (1990) 755.
3 Bolshoy, A., McNamara, P., Harrington, R.E. and Trifonov, E.N., Proc. Natl. Acad. Sci. USA, 88 (1991) 2312.
4 Ulanovsky, L., Bodner, M., Trifonov, E.N. and Choder, M., Proc. Natl. Acad. Sci. USA, 83 (1986) 862.
5 Trifonov, E.N. and Sussman, J.L., Proc. Natl. Acad. Sci. USA, 77 (1980) 3816.

6 Levene, S.D. and Crothers, D.M., J. Biomol. Struct. Dyn., 1 (1983) 429.
7 Wu, H.-M. and Crothers, D.M., Nature, 308 (1984) 509.
8 Coll, M., Frederick, C.A., Wang, A.H.-J. and Rich, A., Proc. Natl. Acad. Sci. USA, 84 (1987) 8385.
9 Nelson, H.C.M., Finch, J.T., Bonaventura, F.L. and Klug, A., Nature, 330 (1987) 221.
10 DiGabriele, A.D., Sanderson, M.R. and Steitz, T.A., Proc. Natl. Acad. Sci. USA, 86 (1989) 1816.
11 Fritsch, V. and Westhof, E., J. Am. Chem. Soc., 113 (1991) 8271.
12 Diekmann, S., Mazzarelli, J.M., McLaughlin, L.W., Von Kitzing, E. and Travers, A.A., J. Mol. Biol., 225 (1992) 729.
13 Edwards, K.J., Brown, D.G., Spink, N., Skelly, J.V. and Neidle, S., J. Mol. Biol., 226 (1992) 1161.
14 Goodfellow, J.M. and Williams, M.A., Curr. Opin. Struct. Biol., 2 (1992) 211.
15 Van Gunsteren, W.F. and Berendsen, H.J.C., Angew. Chem., Int. Ed. Engl., 29 (1990) 992.
16 MacCammon, J.A. and Harvey, S.C., Dynamics of Proteins and Nucleic Acids, Cambridge University Press, Cambridge, 1986.
17 Feuerstein, B.G., Pattabiraman, N. and Marton, L.J., Nucleic Acids Res., 17 (1989) 6883.
18 Herzyk, P., Neidle, S. and Goodfellow, J.M., J. Biomol. Struct. Dyn., 10 (1992) 97.
19 Boehncke, K., Nonella, M., Schulten, K. and Wang, A.H.-J., Biochemistry, 30 (1991) 5465.
20 Van Gunsteren, W.F., Berendsen, H.J.C., Geurtsen, R.G. and Zwinderman, H.R.J., Ann. New York Acad. Sci., 482 (1986) 287.
21 Zielinski, T.J. and Shibata, M., Biopolymers, 29 (1990) 1027.
22 Seibel, G., Singh, U.C. and Kollman, P.A., Proc. Natl. Acad. Sci. USA, 82 (1985) 6537.
23 Singh, U.C., Weiner, S.J. and Kollman, P.A., Proc. Natl. Acad. Sci. USA, 82 (1985) 755.
24 Cruzeiro-Hansson, L., Swan, P.F., Pearl, L. and Goodfellow., J.M., Carcinogenesis, 13 (1992) 2067.
25 Parker, K., Cruzeiro-Hansson, L. and Goodfellow, J.M., J. Chem. Soc. Faraday Trans., 89 (1993) 2637.
26 Arnott, S., Chandrasekaran, R., Hall, I.H. and Puigjaner, L.C., Nucleic Acids Res., 2 (1983) 1493.
27 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta Jr., S. and Weiner, P., J. Am. Chem. Soc., 106 (1984) 765.
28 a. Singh, U.C., Weiner, P.K., Caldwell, J. and Kollman, P.A., AMBER 3.0, University of California at San Francisco, San Francisco, CA, 1986.
b. Seibel, G.L., AMBER 3.0, Rev. A, University of California at San Francisco, San Francisco, CA, 1989.
29 Lavery, R. and Sklenar, H., J. Biomol. Struct. Dyn., 6 (1988) 63.
30 Lavery, R. and Sklenar, H., J. Biomol. Struct. Dyn., 6 (1989) 655.
31 Ravishanker, G., Swaminathan, S., Beveridge, D.L., Lavery, R. and Sklenar, H., J. Biomol. Struct. Dyn., 6 (1989) 669.
32 Saenger, W., Principles of Nucleic Acid Structure, Adenine Press, New York, NY, 1988.
33 Mohan, S. and Yathindra, M., J. Biomol. Struct. Dyn., 9 (1991) 113.
34 Elias, H.-G., Macromolecules, Vol. 1., Plenum Press, New York, NY, 1984, pp. 89–150.
35 Tanford, C., Physical Chemistry of Macromolecules, Wiley, New York, NY, 1961.
36 Bansal, M. and Bhattacharayya, D., In Sarma, R.H. and Sarma, M.H. (Eds.) Structure & Methods, Vol. 3, Adenine Press, New York, NY, 1990, pp. 139–153.
37 Gordon, A.L., Booth, E.D., Hunter, W.N. and Brown, T., Nucleic Acids Res., 20 (1992) 4753.
38 Taillandier, E., Ridoux, J.-P., Liquier, J., Leupin, W., Denny, W.A., Wang, Y., Thomas, G.A. and Peticolas, W.L., Biochemistry, 26 (1987) 3361.
39 Rao, S. and Kollman, P.A., J. Am. Chem. Soc., 107 (1985) 1611.
40 Brahms, S., Fritsch, V., Brahms, J.G. and Westhof, E., J. Mol. Biol., 223 (1992) 455.
41 Roy, S., Borah, B., Zon, G. and Cohen, J.S., Biopolymers, 26 (1987) 525.
42 Behling, R. and Kearns, D.R., Biochemistry, 25 (1986) 3335.
43 Tullius, T.D. and Dombroski, B.A., Science, 230 (1985) 679.
44 Burkhoff, A.M. and Tullius, T.D., Cell, 48 (1987) 935.
45 Price, M.A. and Tullius, T.D., Biochemistry, 32 (1993) 127.
46 Zhurkin, V.B., Gorin, A.A., Charakhchyan, A.A. and Ulyanov, N.B., In Beveridge, D. and Lavery, R. (Eds.) Theoretical Biochemistry and Molecular Biophysics, Vol. 1, Adenine Press, New York, NY, 1990, pp. 411–431.
47 Crothers, D.M., Drak, J., Kahn, J.D. and Levene, S.D., Methods Enzymol., 212 (1992) 3.
48 DiGabriele, A.D. and Steiz, T.A., J. Mol. Biol., 231 (1993) 1024.
49 Crothers, D.M., Haran, T.E. and Nadeau, J.G., J. Biol. Chem., 265 (1990) 7093.