Solution structure and dynamics of the A-T tract DNA decamer duplex d(GGTAATTACC)$_2$: implications for recognition by minor groove binding drugs

Clare E. BOSTOCK-SMITH*, Charles A. LAUGHTON† and Mark S. SEARLE*[1]

*Department of Chemistry, University of Nottingham, University Park, Nottingham NG7 2RD, U.K., and †Cancer Research Laboratories, School of Pharmaceutical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, U.K.

The structure of the DNA decamer duplex d(GGTAATTACC)$_2$ has been determined using NMR distance restraints and molecular dynamics simulations of 500 ps to 1 ns in aqueous solution at 300 K. Using both canonical A and canonical B starting structures [root-mean-square deviation (RMSD) 4.6 Å; 1 Å = 10^{-10} m], with and without experimental restraints, we show that all four simulations converge to a similar envelope of final conformations with B-like helical parameters (pairwise RMSD 1.27–2.03 Å between time-averaged structures). While the two restrained simulations reach a stable trajectory after 300–400 ps, the unrestrained trajectories take longer to equilibrate. We have analysed the dynamic aspects of these structures (sugar pucker, helical twist, roll, propeller twist and groove width) and show that the minor groove width in the AATT core of the duplex fluctuates significantly, sampling both wide and narrow conformations. The structure does not have the highly pre-organized narrow minor groove generally regarded as essential for recognition and binding by small molecules, suggesting that ligand binding carries with it a significant component of 'induced-fit'. Our simulations show that there are significant differences in structure between the TpA step (where p = phosphate) and the ApA and ApT steps, where a large roll into the major groove at the TpA step appears to be an important factor in widening the minor groove at this position.

Key words: A-tract structure; DNA conformation; DNA dynamics; molecular dynamics simulations; NMR spectroscopy.

INTRODUCTION

The recognition of specific sequences of DNA by proteins and small molecules is mediated by a combination of factors. These include patterns of hydrogen bond donors and acceptors in the major and minor grooves, and also sequence specific DNA conformation [1–6]. The relative importance of these different factors, with regard to small molecule DNA recognition, is still a matter of some debate. It is clear that sequence strongly influences structure as is evident from the effects of phasing of A–T rich tracts of DNA on DNA bending [7,8]. More recently, X-ray structures of protein–DNA complexes have identified large protein-induced bends that appear to arise from sequence specific kinks (of up to 90°) that occur at 5′-pyrimidine–purine steps [9–11], suggesting that intrinsic dynamic flexibility and ligand-induced deformability are sequence specific phenomena that play an important role in protein–DNA recognition.

Many small molecules that bind in the minor groove of DNA do so by recognizing A–T rich sequences. In high resolution X-ray crystal structures, the high degree of A–T propeller twist in runs of greater than four A–T base pairs results in a particularly deep, narrow minor groove [12,13] that is suited to the binding of planar aromatic ring systems by edge-on insertion into the groove. For example, the polyamide antibiotics netropsin and distamycin, and the *bis*-benzimidazole Hoechst family of compounds, form van der Waals contacts with the walls of the groove [3,4] and bury a large proportion of their hydrophobic surface area [14], factors which appear to account for the major fraction of the ligand binding energy. A recent detailed calorimetric study of the binding of Hoechst 33258 to the dodecamer duplex d(CGCAAATTTGCG)$_2$ has identified a large change in heat capacity (ΔC_p°) on binding that is consistent with the hydrophobic burial model, and agrees well with predictions of ΔC_p° based on calculations of changes in solvent accessible surface areas using empirical relationships derived from hydrocarbon solvent transfer models and protein folding studies [14]. These calculations rely on the availability of high resolution structures of both the drug–DNA complex and drug-free DNA structure to compute accurate changes in buried surface areas when ligands bind to DNA.

DNA footprinting studies have shown that Hoechst 33258 binds strongly to A–T rich sequences, but will not bind across a 5′-TpA step (where p = phosphate) [15], Thus, although 5′-ApT steps behave like runs of poly(dA), anomalous structural features appear to be associated with 5′-TpA steps that are detrimental to drug binding. In several recent low resolution NMR structures we have examined the binding of Hoechst 33258 and a number of analogues to the decamer duplex d(GGTAATTACC)$_2$ [16,17]. The data show that the ligands bind across the central AATT sequence avoiding the TpA steps, consistent with the footprinting data. To rationalize this observation we have set about determining a high resolution solution structure of the free DNA and the drug–DNA complexes from NMR data to examine in detail the DNA conformation at the 5′-TpA steps, and to determine to what extent the conformation of the AATT tract is pre-organized for drug binding in the minor groove. No X-ray data are available on a TAATTA tract for direct comparison; however, a number of structures have been solved that contain the AATT sequence which possess the narrow minor groove synonymous with drug binding [13,18,19].

A number of recent papers from the groups of Darden, Pederson and Kollman have shown that molecular dynamics

Abbeviations used: NOE, nuclear Overhauser effect; R, restrained; RMSD, root-mean-square deviation; U, unrestrained.

[1] To whom correspondence should be addressed (e-mail mark.searle@nottingham.ac.uk).

© 1999 Biochemical Society

calculations, including the use of distance restraints from NMR data, can lead to structures in good agreement with X-ray data [20–22]. These studies have demonstrated that a proper treatment of long range electrostatic interactions using an explicit solvation model with the particle-mesh Ewald method is amongst the most efficient and accurate approaches available, resulting in nanosecond trajectories with a high level of structural stability. Such consistency between X-ray data and dynamics simulations suggests that the latter can now be used reliably to predict DNA sequence-dependent structural features. Cheatham and Kollman have recently demonstrated that in the simulation of DNA, convergence to a common 'B-like' structure can be achieved from different starting structures via complex conformational transitions over a time-scale of approx. 1 ns [20].

In this study we have carried out four molecular dynamics simulations on the decamer duplex $d(GGTAATTACC)_2$ in aqueous solution using two quite different DNA starting geometries, namely the canonical A-DNA and B-DNA forms [root-mean-square deviation (RMSD) 4.6 Å; 1 Å $= 10^{-10}$ m]. Each model was used in two molecular dynamics simulations, one unrestrained and one restrained by experimental nuclear Overhauser effect (NOE) distance data. We show that over a time course of 500 ps of dynamics at 300 K we see convergence of all four simulations to a similar final envelope of conformations with pairwise RMSD values between time-averaged structures of 1.27–2.03 Å; this picture is maintained when the restrained B-form simulation is extended to 1 ns. Auto-correlation plots [23] suggest that while the two restrained simulations relax within 300–400 ps, the unrestrained trajectories take longer to equilibrate. We have analysed dynamic aspects of these structures by determining fluctuations in key conformational features of these duplexes (sugar pucker, helical twist, roll, propeller twist and groove width) for comparison with X-ray structures of related sequences. The groove width in the AATT core of the duplex is shown to fluctuate significantly, sampling both a wide groove and a narrow minor groove compatible with that found in many related X-ray structures. Such dynamic fluctuations suggest that the structure is not highly pre-organized with the narrow minor groove necessary for recognition and binding by small molecules, but that ligand binding carries with it a significant component of 'induced fit'.

MATERIALS AND METHODS

DNA synthesis

The DNA decamer d(GGTAATTACC) was synthesized using standard solid-phase phosphoramidite chemistry, and purified trityl-on by reverse-phase HPLC using triethylammonium acetate buffer (pH 7.0) and an acetonitrile gradient, as described previously [17]. The decamer was shown to be $> 95\%$ pure by ^{1}H NMR spectroscopy. NMR studies were carried out on the decamer duplex in 100 mM NaCl, 10 mM $Na^2H_2PO_4$ at a concentration of 2 mM. The sodium salt of trimethylsilylpropionate was added as an internal reference compound, 0.1 % sodium azide as an anti-bacterial agent and 0.1 % EDTA to complex any heavy metal ions.

NMR analysis

NMR data were collected at 500 MHz on a Bruker DRX500 spectrometer and processed on an R4600PC Silicon Graphics Indy work station using XWINNMR software. Standard phase-sensitive 2D NMR pulse sequences were used throughout, including NOESY, double-quantum-filtered (DQF)-COSY, TOCSY, and jump-and-return NOESY for solvent suppression in 90 % H_2O solutions. NOESY spectra were acquired at mixing times between 50 and 300 ms and 1024 complex data points were collected for each of 512 t_1 increments with 64 transients for each. Spectra were zero filled to 2 k × 1 k prior to Fourier transformation. A total of 634 NOE restraints (317 per strand) were determined from NOESY data collected at 298 K at mixing times of 60, 120 and 200 ms. Interproton distances were determined by integration of NOE data over the full range of mixing times and calibrated to a number of fixed reference distances: deoxyribose H2′–H2″, thymine Me–H6 and cytosine H5–H6. Measured distances were extrapolated back to zero mixing time using linear regression according to the method of Baleja et al. [24]. An error bound of 20 % was added to distances below 3.5 Å and 25 % was added to longer distances. Restraints from a single 100 ms NOESY spectrum in H_2O, which are intrinsically less accurate, were used with error bounds of 35 %.

Molecular dynamics simulations

Structure calculations were carried out on an R10000SC Silicon Graphics work station using AMBER 4.1 [25]. The A-DNA and B-DNA starting structures were generated using the NUCGEN module. Explicit net-neutralizing sodium counterions were placed close to the phosphate groups. The system was solvated to a minimum distance of 5 Å around the solute giving periodic box sizes of approximately 60 Å × 40 Å × 40 Å and 50 Å × 40 Å × 40 Å for the A-DNA and B-DNA starting structures respectively, with 2158 and 2157 water molecules. All simulations were run using the SANDER module, with all bond distances frozen using SHAKE [26] (tolerance 0.0005 Å), a 1 fs timestep, a temperature of 300 K with Berendsen temperature coupling [27] and a time constant of 0.2 ps. A 9 Å cut-off was applied to the Lennard-Jones interactions, and the list of non-bonded atoms was updated every 25 steps. Constant pressure was maintained through isotropic position scaling with a time constant of 0.2 ps. Equilibration was performed by first holding the positions of the DNA and counterions fixed and running 5000 steps of minimization on the solvent, followed by a further 5000 steps in which the whole system was allowed to relax. Then, with the coordinates of the DNA and counterions frozen, the water was subjected to 10 ps molecular dynamics at 100 K. After this initial equilibriation, all subsequent simulations were run using the particle-mesh Ewald method [20–22]. The position restraints on the counterions were removed and the system was subjected to 10 ps molecular dynamics, then heated to 300 K over 5 ps and held there for a further 5 ps. The restraining force constant on the DNA was gradually reduced in successive 10 ps runs to 50, 25, 10, 5, and finally 2 kcal·mol^{-1}·Å^{-2} (8.4 kJ·mol^{-1}·Å^{-2}) before running 100 ps molecular dynamics with no constraints on the DNA, giving a total of 180 ps equilibriation. The system was finally energy minimized ready for introduction of distance restraints. Distance restraints with a force constant of 30 kcal·mol^{-1}·Å^{-2} (126 kJ·mol^{-1}·Å^{-2}) were introduced over a period of 20 ps, with a simultaneous increase in temperature from 1 K to 300 K. Unrestrained runs were subjected to the same temperature rise and following this, all simulations were subjected to 480 ps molecular dynamics with the structure being written every 5 ps. For the restrained B-form structure (B_R) the dynamics simulation was extended to 1 ns.

Average structures were calculated using the CARNAL module of AMBER 4.1 over the final 200 ps of the simulations (an average of 40 structures), or 700 ps in the case of B_R (350 structures). The average structures were energy minimized and RMSDs from the canonical starting structures for each run were calculated using MOLMOL [28]. RMSDs over the trajectories

© 1999 Biochemical Society

were calculated in CARNAL. Structural parameters over the course of the trajectories were analysed in CURVES [29]. Final average NOE restraint violations were as follows: unrestrained A-form structure (A_U) 0.111 Å, B_U 0.118 Å, A_R 0.033 Å, B_R 0.032 Å (after 500 ps), B_R 0.031 Å (after 1 ns). A more detailed analysis of time-averaged restraint violations reveals both A_R and B_R to produce a much better fit to the experimental data. For B_R we observe 11 restraint violations > 0.2 Å; nine of these lie in the range 0.2–0.3 Å and two in the range 0.3–0.5 Å. A_R shows very similar deviations; seven violations lie in the range 0.2–0.3 Å and three in the range 0.3–0.5 Å. In contrast, the unrestrained structures show significant deviations; A_U has 42 time-averaged restraint violations in the range 0.2–1.0 Å, while B_U has 38 in the same range. The coordinates of all calculated structures are available from the authors on request.

RESULTS AND DISCUSSION

Convergence from canonical A-DNA and B-DNA starting structures

Four separate molecular dynamics simulations were run over 500 ps on the fully solvated decamer duplex d(GGTAATTACC)$_2$. Two unrestrained structures (A_U and B_U) were computed from canonical A and canonical B starting structures. Similarly, two structures (A_R and B_R) were calculated from the same canonical A and B starting structures but with a full set of NOE restraints activated. The simulation of B_R was extended to 1 ns. Average structures from each run were computed from the last 200 ps of the 500 ps trajectories and the final 700 ps of the 1 ns trajectory (Figure 1).

RMSD values computed between all possible pairs of averaged structures and canonical A-DNA and B-DNA starting structures are shown in Table 1. The striking observation is that all four structures show a high degree of convergence as measured by pairwise RMSD values (RMSD 1.6 ± 0.3 Å; range 1.27–2.03 Å) irrespective of whether NOE restraints were applied during the calculations. RMSD values deviate to a similar extent from both canonical A-DNA and B-DNA, suggesting structures midway between these canonical forms. However, different convergence criteria, such as autocorrelation plots, demonstrate that while B_R reaches equilibration by 300 ps, and A_R by around 400 ps, A_U and B_U have not fully equilibrated during 500 ps of dynamics, reflecting the role of NOE restraints in guiding the sequence to its final structure. The quality of fit to the restraints also revealed differences between restrained and unrestrained trajectories, the average deviations being 0.03 Å for both restrained simulations, but 0.111 Å and 0.118 Å for A_U and B_U respectively. A more detailed analysis of restraint deviations is presented in methods section.

As observed in previous simulations of DNA over long dynamics trajectories, the structures show some degree of bending towards the major groove which appears to arise from a small *x*-displacement at each step coupled with roll that is on average positive (Table 2). Cheatham and Kollman [20] have suggested that roll probably accounts, in large part, for the calculated RMSD values with respect to the canonical starting structures. However, our own modelling studies suggest that reduced helical twist at each step may be more important. We have examined this more closely by constructing an $A_{10}.T_{10}$ B-form duplex using the nucleic acid building program SCHNArP [30]. All helical parameters except for twist and roll were kept constant, while these latter parameters were varied over the range seen in dynamics simulations. Using a twist of 36° as the reference state, reducing the twist at each base pair step to 34°, 32° and 30° resulted in RMSDs over all base atoms of 0.35, 0.69 and 1.04 Å respectively. By comparison, variations in roll produced smaller RMSDs. Using a reference value of 2.1°, increasing the roll at each base step to 2.6°, 3.6° and 4.6° produced corresponding RMSDs of 0.08, 0.23 and 0.38 Å respectively, showing RMSD to be less sensitive to roll than to twist.

These model studies show that RMSD is more sensitive to small changes in some helicoidal parameters than others questioning whether RMSD is a reliable indicator of convergence or of equilibration. The latter seems to be better assessed by autocorrelation analysis of the dynamics trajectory. In the case of the simulations described here, much lower pairwise RMSD values are evident between the four averaged final structures (1.27–2.03 Å) than between the two different starting structures (4.6 Å). The sequence-dependent variation in the key structural parameters, propeller twist, roll, minor groove width and helical twist averaged over the final 700 ps of the 1 ns B_R simulation, are presented in Figure 2, while sequence averages for these and for additional helical parameters for all four structures are presented in Table 2.

Sugar conformation and repuckering

Deoxyribose conformation provides a clear discrimination between the A-DNA and B-DNA models. The sugar puckers in the A-DNA and B-DNA starting structures are initially quite different but convergence in this parameter is seen after approximately 250 ps of dynamics (Figure 3a). The time scale on which the sugars repucker from the *N* conformer in the A-DNA starting structures to the *S* conformer in the converged structures is very different when NOE restraints are applied, occurring rapidly during the 20 ps period over which the restraints are introduced.

The fluctuations in average sugar pucker during the course of each trajectory provide a dynamic picture of the motions in the double helix that are averaged out in the X-ray 'snapshots' of DNA. The average sugar pucker in all structures is approximately 125° (C1′-exo; see Table 2) but repuckering occurs widely during the dynamics trajectories. For the unrestrained trajectories this repuckering includes a limited population of the classical *N* conformation ($P = 30°$), consistent with dynamic models proposed on the basis of NMR coupling constant data [31], ensemble methods of NOE distance refinement [32] and Raman spectroscopy [33] that support an equilibrium between *N* and *S* conformers. NOE restraints, however, strongly favour the *S* conformer and result in less dynamic fluctuation in the sugar pucker such that the *N* conformer is never sampled. This is evident for both A_R and B_R. As expected the fluctuations from the mean are of larger amplitude for the terminal residues (S.D. $A_R \pm 32°$, $B_R \pm 33°$), consistent with greater flexibility, but even residues at the centre of the sequence show some fluctuation about the mean (S.D. $A_R \pm 15°$, $B_R \pm 15°$).

Helicoidal parameters

Many other helicoidal parameters differ widely between A and B forms of DNA, providing indications of convergence as well as showing sequence-dependent variation. Minor groove width as measured by phosphate–phosphate distances is much narrower in canonical B form DNA than in A form (11.7 Å and 18.4 Å respectively). Convergence is seen within around 50 ps for restrained structures and 250 ps for unrestrained structures (Figure 3b), although some divergence is evident later in the trajectories, indicating that not all motions are sampled in full over such a short time scale. The trajectory for the 1 ns simulation B_R (Figure 3c) shows the dynamic motion of the groove width, where extremes of wide and narrow (from 9 Å to 17 Å) are sampled over the course of the trajectory. Mean values over the

© 1999 Biochemical Society

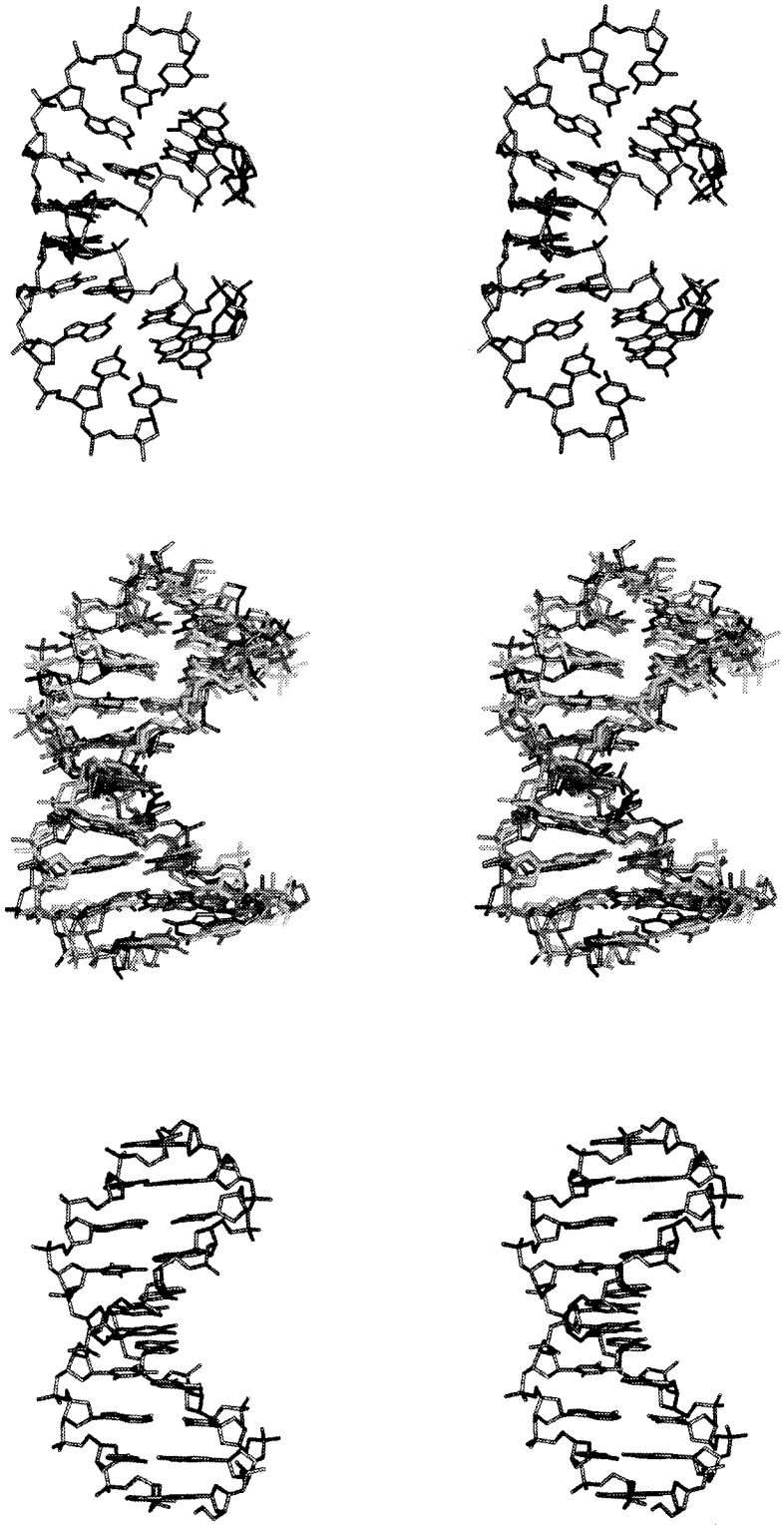

Figure 1 Canonical A, canonical B and calculated structures of d(GGTAATTACC)$_2$

Stereo views of canonical A starting structure (top), RMS best fit of energy minimized time-averaged structures from A_U, A_R, B_U and B_R (as described in the text; middle), and canonical B starting structure (bottom). All atoms except hydrogen atoms are displayed.

© 1999 Biochemical Society

Table 1 RMSD values between calculated and canonical structures

RMSD values calculated for the heavy atoms of the central eight base pairs, for structures averaged over the final 200 ps (A_U, A_R, B_U) or 700 ps (B_R) of the simulations and for canonical A and canonical B forms of DNA.

	A_U	A_R	B_U	B_R	A	B
A_U		1.38	2.03	1.27	2.61	3.37
A_R	1.38		1.67	1.43	3.17	3.10
B_U	2.03	1.67		1.86	3.55	2.40
B_R	1.27	1.43	1.86		2.72	3.33
A	2.61	3.17	3.55	2.72		4.56
B	3.37	3.10	2.40	3.33	4.56	

trajectories of all four simulations range from 10.3 to 15.3 Å across the sequence, in broad agreement with the range 11.7–13.9 Å seen by Cheatham and Kollman [20], and considerably lower than the 18.4 Å distance of canonical A-DNA. Groove width is often narrowed considerably in A–T tracts and indeed a slight narrowing is seen in the centre of the sequence in B_R (Figure 2) and A_R, but while in A_R distances are no less than the 11.7 Å distance of canonical B-DNA, in B_R they are on average slightly greater, reflecting the greater variation seen over the course of the longer dynamics simulation. A_U shows distances of a similar pattern and magnitude to B_R, while B_U is similar in trend but not as symmetrical.

Narrow minor groove width is often seen to be the result of a large negative propeller twist [6,12]. While this is certainly evident in all four structures the only one that shows a pattern of a more negative propeller twist in the centre of the sequence, to correlate with the narrower groove, is B_R (Figure 2). But even here the difference is slight and the most negative values are not confined to the central AATT tract but are also evident at the TpA steps, suggesting that propeller twist may be just one influence on the narrowing of the AATT minor groove [6]. There is no significant change in propeller twist values over the course of the simulation, which implies establishment of this characteristic during the equilibration period. However, in a symmetrical sequence, a symmetrical pattern would be expected. By this measure only B_R would appear to be equilibrated, showing a more even and symmetrical pattern than A_R, A_U and B_U. The asymmetry seen for A_R is likely to be due to low frequency motions that have not been fully sampled over the 500 ps time period, as was observed with the minor groove width. Analysis of B_R at the 500 ps stage also reveals some degree of asymmetry of the structure which is less pronounced in the 1 ns simulation (Figure 3c).

Roll and helical twist show clear sequence-dependent variation. For A_R and B_R, roll is at a maximum of around 20° at the TpA steps (Figure 2) and shows minima of 0° to −10° at purine-pyrimidine steps. Roll at the ApA steps is intermediate between these values. A_U and B_U show a similar but not as pronounced or symmetrical a pattern. A statistical analysis of eight NMR-derived structures [34] shows the same trend in roll values: TpA > ApA > ApT, with a general trend of YR > RR/YY > RY. In a similar study on 38 crystal structures [35] the same observations were made for TpA, ApA and ApT steps. Our values are, however, more greatly polarized than those seen in either of these studies.

Helical twist also shows a sequence variation for A_R and B_R, with maxima at purine–pyrimidine steps and minima at pyrimidine–purine steps. The GpT maximum corresponds to a B-DNA-like value of around 34°, the TpA minimum to an A-DNA-like value of 30° and lower, and the ApT maximum to a

Table 2 Helicoidal parameters for A_U, B_U, A_R and B_R

Mean values, standard deviations and ranges for sugar pucker pseudorotation angle, selected helicoidal parameters and average minor groove distances. Values are taken from the final 200 ps of the 500 ps trajectories and the final 700 ps of the 1ns trajectory. Minimum and maximum values are not the largest and smallest values observed, but represent the residue, base-pair or base pair step for which the range is greatest. These are indicated in parentheses. *, represents the complementary base on the opposite strand; P*x*, phosphorus on nucleotide *x*.

	A_U					B_U					A_R					B_R				
	Mean	S.D.	Min	Max		Mean	S.D.	Min	Max		Mean	S.D.	Min	Max		Mean	S.D.	Min	Max	
Pucker (°)	116.4	29.0	5.5	154	(A8)	125.3	29.8	41.8	185	(T7*)	128.3	21.7	15.4	164	(C10)	128.5	23.1	19.9	170	(C10*)
Twist (°)	30.0	4.22	17.2	40.1	(T7A8)	31.5	5.31	11.9	37.4	(T7A8)	31.3	4.32	24.1	41.5	(C9C10)	31.0	4.20	18.3	47.1	(G1G2)
Roll (°)	3.79	9.02	−21.5	19.1	(G1G2)	2.17	11.1	−7.35	34.2	(T3A4)	3.85	12.4	7.63	47.0	(T3A4)	4.43	10.7	−27.3	25.0	(A8C9)
Tilt (°)	−0.16	6.06	−13.9	14.6	(G1G2)	0.56	4.68	−12.1	11.5	(C9C10)	0.06	6.05	−10.6	9.16	(T3A4)	−0.13	4.88	−10.0	20.1	(G1G2)
Propeller (°)	−7.75	11.8	−25.0	25.7	(G1C10*)	−8.66	15.2	−5.39	51.1	(G1C10*)	−10.1	11.0	−46.1	11.6	(A8T3*)	−9.09	11.0	−49.9	33.5	(G2C9*)
X-displacement (Å)	−2.24	0.67	−3.69	−0.54	(T6A5*)	−1.17	0.49	−2.49	0.19	(A8T3*)	−1.88	0.54	−3.56	−0.31	(T7A4*)	−2.27	0.68	−4.29	−0.12	(G1C10*)
Inclination (°)	−5.79	5.66	−21.8	6.14	(C9G2*)	−6.83	6.39	−19.3	11.76	(G1C10*)	−5.16	4.54	−20.6	5.69	(T6A5)	−2.46	5.26	−22.1	15.0	(T3A8*)
P10*–P5	13.3	0.93	11.2	15.0		12.3	2.1	9.36	16.8		13.0	1.05	11.5	15.1		13.0	1.44	9.69	17.0	
P9*–P6	13.7	1.20	10.2	15.7		11.6	1.58	8.9	14.3		13.5	0.97	11.5	15.9		13.4	1.47	9.44	16.4	
P8*–P7	12.9	1.44	10.4	15.8		10.3	0.91	8.6	12.7		11.6	1.03	9.50	13.6		12.3	1.33	9.75	16.5	
P7*–P8	13.2	1.52	11.0	16.7		11.5	1.05	9.25	14.1		11.1	13.8	8.93	13.6		12.1	1.14	9.38	14.8	
P6*–P9	14.5	1.11	12.2	16.7		13.6	1.33	10.2	16.3		12.8	1.21	10.5	14.9		13.2	1.07	9.74	15.4	
P5*–P10	13.6	1.56	10.0	17.1		15.0	0.96	13.1	16.8		12.5	1.08	10.5	14.8		12.6	1.14	9.59	15.2	

© 1999 Biochemical Society

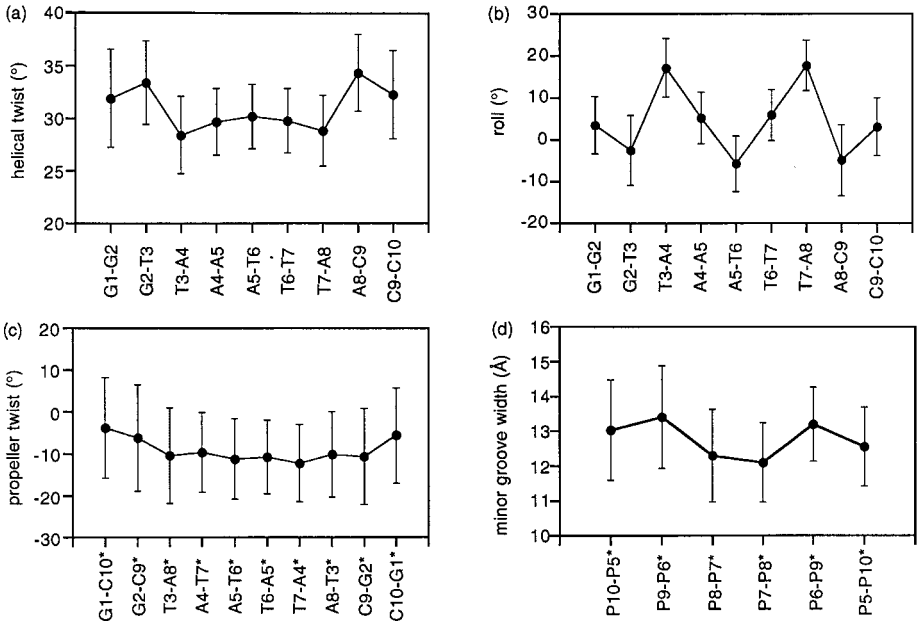

Figure 2 Analysis of helical parameters for d(GGTAATTACC)$_2$

Analysis of mean values for (**a**) helical twist, (**b**) base pair roll, (**c**) propeller twist and (**d**) minor groove width for each base pair/base step of B_R calculated using CURVES [29] over the final 700 ps of the 1 ns molecular dynamics simulations. Error bars reflect fluctuations from the mean value by one standard deviation.

value in between these. Overall the helices are somewhat under-wound compared to B-DNA (mean twists around 31°), an effect clearly visible to the naked eye. A_U and B_U, while showing similar mean values, do not exhibit the symmetrical pattern of A_R and B_R. This again fits with incomplete convergence of the unrestrained trajectories seen in the correlation plots, as with groove width, propeller twist and roll. The agreement between these trends in helical twist and those observed by Ulyanov and James [34] and by Gorin et al. [35] in their statistical analyses is not as good as for roll. Both studies observe the opposite trend in helical twist with TpA > ApT. However, as pointed out by Yanagi et al. [36], base step geometry cannot solely be reliant on those two base pairs, but is influenced by the two flanking steps and perhaps by even longer-range interactions. It is also important to note that there appears to be no simple relationship between helical twist and NMR parameters as the effect of twist on interproton distances can be compensated by other factors such as shear, shift and slide [37]. The introduction of a single NOE restraint can have a significant effect on twist [34]. It is unlikely that helical twist, or any of the other sequence-dependent conformational features, are purely force field-driven as the restrained structures show a clearer sequence-dependent variation than the unrestrained structures. Moreover, the number of distance restraint violations that are observed is significantly greater in the case of the unrestrained structures (see Materials and methods section), indicating that the restraints are instrumental in guiding the structure to its final conformation.

As was noted above, there is a sequence-dependent narrowing of the groove width that does not appear to be solely dependent on propeller twist. It is conceivable that instead it is a result of the patterns of twist and roll. Roll into the minor groove at the central ApT step could account for the narrowing of the groove. In contrast, roll into the major groove at the TpA steps could serve to widen the groove at the ends of the A–T tract. This begs the question, would a CpA, or GpA step exhibit such a large positive roll, and if not, then perhaps the TpA step is the cause of the minor groove of the relatively short A–T tract being wider than expected. Ulyanov and James [34] conclude that roll at CpA steps is greater than at TpA, which in turn is greater than at GpA. Gorin et al. [35] observe that roll is on average slightly less at CpA steps (1.1°) than at TpA steps (2.6°), though still positive, but the step can be divided into two very different conformers, one with a much more positive roll (5.1°), and the other with a large negative roll (−6.3°). It would seem then that if groove width is affected by roll, a GpA rather than a TpA step at the end of an A–T tract could serve to retain the narrowness of the minor groove, but a CpA step could be either more or less detrimental to narrow groove width. It must be pointed out that the number of individual base pair steps looked at is larger in the study of Gorin et al. (averaging 20) than in the Ulyanov and James study (averaging approx. 4).

In general helical twist is seen to be lower in solution structures than in crystal structures [34], in agreement with our results. It is thought that over twisting is a result of crystal packing forces, and an under-wound helix may be a more realistic model of short duplexes in solution. This postulation is backed up by the observation of low twist in a trigonal crystal, the type least likely to be affected by crystal packing [34]. The effects of the various energetic contributions (electrostatic interactions and torsional energies) of the molecular dynamics force field have also been examined for their impact on helical twist; however, no simple correlation emerges, indicating a complex synergy between the various contributions. Small counterions and low ionic strength, corresponding to the simulation conditions used here, are also known to result in helix unwinding. The low counterion concentration used in the simulation could also have a bearing on groove width, allowing increased phosphate–phosphate repulsion which forces the walls of the minor groove apart. Such a correlation between large twist and narrow groove width is apparent for A_R and B_R.

© 1999 Biochemical Society

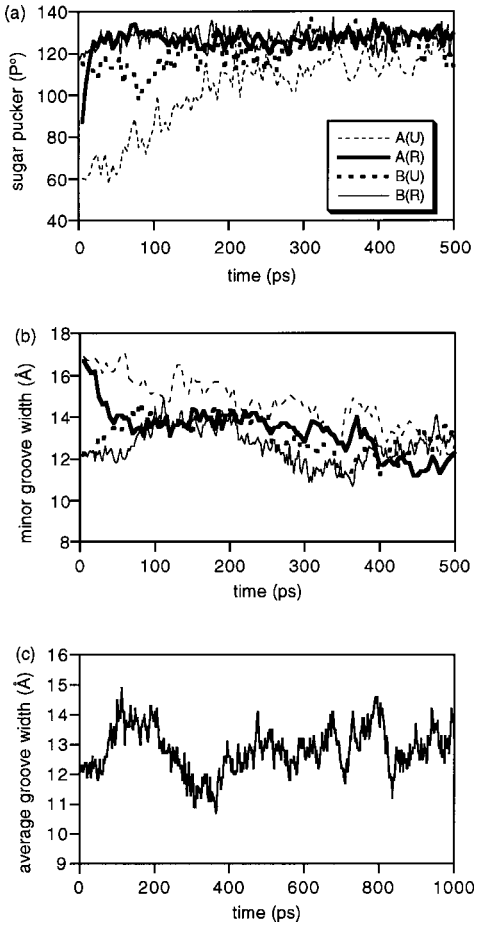

Figure 3 Analysis of sugar conformation and minor groove width for d(GGTAATTACC)$_2$

(**a**) Change in average pseudo-rotation phase angle (sugar pucker) and (**b**) average minor groove width, over the first 500 ps of dynamics of A_U, A_R, B_U and B_R. (**c**) Fluctuation in average minor groove width for B_R over 1 ns of dynamics simulations. The legend shown in (**a**) is also appropriate to (**b**).

Tilt has also been described as a 'fundamental' parameter determining long-range structure in nucleic acids, but to a lesser extent than twist and roll due to its influence on the relatively rigid backbone [34]. In agreement with Cheatham and Kollman's observations rolling is preferred over tilting [20]. Tilt shows no sequence-dependence or difference in value or range between restrained and unrestrained structures, and is close to zero, the value for both canonical A-DNA and B-DNA. *X*-displacement exhibits values between those of canonical A and canonical B forms and is also found to be a significant influence on RMSD values between calculated structures and canonical B-DNA. We have again examined the impact of this parameter on RMSD values using the $A_{10}.T_{10}$ model with a reference state for *x*-displacement of 0.56 Å. Increasing *x*-displacement to 1.56 and 2.56 Å significantly increases the RMSD to 0.83 Å and 1.67 Å respectively. *X*-displacement was seen by Cheatham and Kollman to be the slowest converging parameter during their simulations [20]; however, we do not see any significant change over time. While average values for A_U, A_R and B_R are all similar (-1.88 to -2.27 Å), B_U has an average of -1.17 Å, still close to canonical B form. The magnitude of fluctuations is similar for restrained and unrestrained trajectories, suggesting that this parameter is not greatly influenced by NOE restraints.

Implications of DNA structure and dynamics for drug binding

In several earlier studies we have used the decamer duplex d(GGTAATTACC)$_2$ in drug–DNA recognition studies, and have shown that *bis*-benzimidazole A–T-recognizing minor groove binding agents such as Hoechst 33258 and Hoechst 43254 bind at the centre of this sequence spanning the AATT tract [16,17]. There are two key structural features of this DNA sequence that appear to be important in defining the drug binding site. These are: (i) the nature of the A-tract in terms of its narrowness and potential van der Waals complementarity with minor groove-bound ligands, and (ii) the influence of the TpA steps on the width of the minor groove of the A-tract.

Our molecular dynamics simulations of the structure of d(GGTAATTACC)$_2$ in the absence of any bound ligand give us some insight into the extent to which this sequence is 'predisposed' to bind a ligand in the A–T minor groove. By analysing the molecular dynamics trajectory we obtain a more dynamic picture of the fluctuations in conformation that are taking place. We draw the following conclusions. The minor groove, although on average not especially narrow in the central A–T tract, displays dynamic fluctuations that sample conformations consistent with the narrow grooves found in X-ray structures of A-tract sequences and drug–DNA complexes. The two snapshots taken from the B_R trajectory during the final 700 ps of simulation illustrate this variation (Figure 4). In the top panel, the two structures are viewed into the minor groove, for which average groove widths of 11.2 Å and 13.8 Å are calculated. In the bottom panel, differences in groove depth and curvature are evident for the same two structures when viewed down the canyon of the central AATT sequence. Such a picture of variation in groove topology has also been drawn from molecular dynamics simulations by other groups [38,39]. The intrinsic dynamic nature of the structure suggests that 'induced fit' is a significant factor in drug minor groove recognition in order to maximize intermolecular van der Waals and hydrophobic interactions. On the other hand, a wider A-tract minor groove has been shown to be necessary in accommodating a side-by-side distamycin dimer [5,40], with a difference in groove width of 3.5–4.0 Å between 1:1 and 2:1 binding modes. More generally, conformational heterogeneity is now widely recognized in the structure of nucleic acids from both X-ray and NMR analysis. In particular, RNA shows distinct structural diversity leading to adaptive, induced-fit recognition with small ligands and proteins [41–43]. Protein–DNA interactions provide particularly good examples of induced fit and DNA bending [9–10]; the minor groove binding HMG DNA-binding proteins such as SRY and Lef-1 facilitate this further by intercalating a residue side chain between adjacent bases, causing a large bend and helix unwinding [44].

Previously, footprinting data for Hoechst 33258 have shown that the drug binds preferentially to tracts of more than four A–T base pairs but will not bind across a TpA step [15], suggesting some structural anomaly that may be related to the effects of the TpA step on the intrinsic width and deformability of the minor groove. Our simulations show that there are significant time-averaged differences in structure between the TpA step and the ApA and ApT steps, where a large roll into the major groove at the TpA step appears to be an important factor in accounting for a widening of the minor groove at this position. The TpA effect on intrinsic groove width and ease of deformability may make a significant contribution to its avoidance by *bis*-benzimidazole-based ligands. Previous NMR studies of TpA steps have provided

© 1999 Biochemical Society

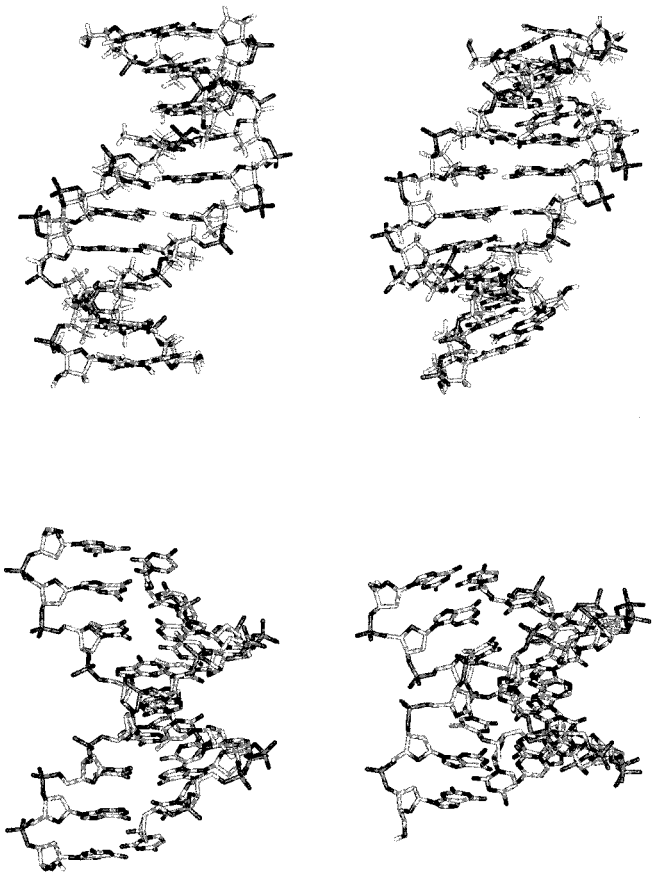

Figure 4 Fluctuations in groove width during molecular dynamics simulations

Snapshots from the 1 ns dynamics simulation of B_R illustrating the variation in minor groove width during the course of the trajectory. Top panel: view into the minor groove; left, average groove width 11.2 Å; right, average groove width 13.8 Å. Bottom panel: view along the canyon of the A-tract for the same two structures.

evidence of fluctuations between two or more conformational states [45,46], although the time scale of these dynamic processes, leading to signal line broadening, are far too long to be accessible through the simulations reported here. The results described have provided some insight into sequence-dependent DNA structure and dynamics which we are currently extending to studies of drug–DNA complexation.

We thank the Engineering and Physical Sciences Research Council for a research studentship for C.E.B.-S., and John Keyte for oligonucleotide synthesis. M.S.S. is grateful to the Department of Chemistry and the Royal Society for financial support. M.S.S. and C.A.L. gratefully acknowledge the University of Nottingham for New Lecturers Awards.

REFERENCES

1 White, S., Szewczyk, J. W., Turner, D. M., David, E. E. and Dervan, P. B. (1998) Nature (London) **391**, 468–471
2 Krugh, T. R. (1994) Curr. Opin. Struct. Biol. **4**, 351–364
3 Neidle, S. (1997) Biopolymers **44**, 105–121
4 Geierstanger, B. H. and Wemmer, D. E. (1995) Annu. Rev. Biophys. Biomol. Struct. **24**, 463–493
5 Wemmer, D. E. and Dervan, P. B. (1997) Curr. Opin. Struct. Biol. **7**, 355–361
6 Shatzky-Schwartz, M., Arbuckle, N. A., Eisenstein, M., Rabinovich, D., Bareket-Samish, A., Haran, T. E., Luisi, B. and Shakked, Z. (1997) J. Mol. Biol. **267**, 595–623
7 Wu, H.-M. and Crothers, D. M. (1986) Nature (London) **308**, 509–513
8 Hagerman, P. J. (1994) Biochim. Biophys. Acta **1131**, 125–132
9 Rice, P. A. (1997) Curr. Opin. Struct. Biol. **7**, 86–93
10 Kim, J. L. and Burley, S. K. (1994) Nat. Struct. Biol. **1**, 638–653
11 Dickerson, R. E. (1998) Nucleic Acids Res. **26**, 1906–1926
12 Nelson, H. M., Finch, J. T., Luisi, B. F. and Klug, A. (1987) Nature (London) **330**, 221–226
13 Coll, M., Frederick, C. A., Wang, A. H.-J. and Rich, A. (1987) Proc. Natl. Acad. Sci. U.S.A. **84**, 8385–8389
14 Haq, T., Ladbury, J. E., Chowdry, B. Z., Jenkins, T. C. and Chaires, J. B. (1997) J. Mol. Biol. **271**, 244–257
15 Portugal, J. and Waring, M. J. (1988) Biochim. Biophys. Acta **949**, 158–168
16 Embrey, K. J., Searle, M. S. and Craik, D. J. (1993) Eur. J. Biochem. **211**, 437–447
17 Bostock-Smith, C. E., Laughton, C. A. and Searle, M. S. (1998) Nucleic Acids Res. **26**, 1660–1667
18 Dickerson, R. E. and Drew, H. R. (1981) J. Mol. Biol. **149**, 761–786
19 Westhof, E. (1987) J. Biomol. Struct. Dyn. **5**, 581–600
20 Cheatham, III, T. E. and Kollman, P. A. (1996) J. Mol. Biol. **259**, 434–444
21 York, D. M., Yang, W., Lee, H., Darden, T. A. and Pederson, L. G. (1995) J. Am. Chem. Soc. **117**, 5001–5002
22 Lee, H., Darden, T. A. and Pederson, L. G. (1995) J. Chem. Phys. **243**, 229–235
23 Swaminathan, S., Ichiye, T., van Gunsteren, W. and Karplus, M. (1982) Biochemistry **21**, 5230–5241
24 Baleja, J. D., Moult, J. and Sykes, B. D. (1990) J. Magn. Reson. **87**, 375–384
25 Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, W. S., Cheatham, III, T. E., Ferguson, D. M., Seibel, G. L., Singh, U. C., Weiner, P. K. and Kollman, P. A. (1995) AMBER 4.1, University of California, San Francisco
26 Rychaert, J. P., Ciccotti, G. and Berendsen, H. J. C. (1977) J. Comput. Phys. **23**, 327–341
27 Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. and Haak, J. R. (1984) J. Comput. Phys. **81**, 3684–3690
28 Koradi, R., Billeter, M. and Wüthrich, K. (1996) J. Mol. Graphics **14**, 51–60
29 Lavery, R. and Sklenar, H. (1998) J. Biomol. Struct. Dynam. **6**, 63–91
30 Lu, X.-J., El Hassan, M. A. and Hunter, C. A. (1997) J. Mol. Biol. **273**, 681–691
31 Rinkel, L. J. and Altona, C. (1987) J. Biomol. Struct. Dyn. **4**, 621–629
32 Ulyanov, N. B., Schnitz, U., Kumar, A. and James, T. L. (1995) Biophys. J. **68**, 13–24
33 Brahms, S., Fritsch, V., Brahms, J. G. and Westhof, E. (1992) J. Mol. Biol. **223**, 455–476
34 Ulyanov, N. B. and James, T. L. (1995) Methods Enzymol. **261**, 90–120
35 Gorin, A. A., Zurkin, V. B. and Olsen, W. K. (1995) J. Mol. Biol. **247**, 34–48
36 Yanagi, K., Prive, G. G. and Dickerson, R. E. (1991) J. Mol. Biol. **217**, 201–214
37 Ulyanov, N. B., Gorin, A. A., Zhurkin, V. B., Chen, B.-C., Sarma, M. H. and Sarma, R. H. (1992) Biochemistry **31**, 3918–3930
38 Duan, Y., Wilkosz, P., Crowley, M. and Rosenberg, J. M. (1997) J. Mol. Biol. **272**, 553–572
39 Young, M., Ravishanker, G. and Beveridge, D. L. (1997) Biophys. J. **73**, 2313–2336
40 Pelton, J. G. and Wemmer, D. E. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 5723–5727
41 Auffinger, P. and Westhof, E. (1998) Curr. Opin. Struct. Biol. **8**, 227–236
42 Patel, D. J., Suri, A. K., Jiang, F., Jiang, L., Fan, P., Kumar, R. A. and Nonin, S. (1997) J. Mol. Biol. **272**, 645–664
43 Westhof, E. and Patel, D. J. (1997) Curr. Opin. Struct. Biol. **7**, 305–309
44 Werner, M. H., Ruth, J. R., Gronenborn, A. M. and Clore, G. M. (1995) Cell **81**, 705–714
45 Kennedy, M. A., Nuutero, S. T., Davis, J. T., Drobny, G. P. and Reid, B. R. (1993) Biochemistry **32**, 8022–8035
46 McAteer, K., Ellis, P. D. and Kennedy, M. A. (1995) Nucleic Acids Res. **23**, 3962–3966

Received 25 February 1999/15 April 1999; accepted 7 June 1999

© 1999 Biochemical Society