10.1098/rsta.2004.1386

THE ROYAL SOCIETY

Deformations of promoter DNA bound to carcinogens help interpret effects on TATA-element structure and activity

By Qing Zhang[1], Suse Broyde[2] and Tamar Schlick[1]

[1]*Department of Chemistry and Courant Institute of Mathematical Sciences, New York University and the Howard Hughes Medical Institute, 251 Mercer Street, New York, NY 10012, USA* (schlick@nyu.edu)

[2]*Department of Biology, New York University, 100 Washington Square East, New York, NY 10003, USA*

Published online 13 May 2004

The TATA-box binding protein (TBP) is required by eukaryotic RNA polymerases for correct transcription initiation. TBP binds to the minor groove of an 8 base pair (bp) DNA-promoter element known as the TATA box and severely bends the TATA box. The promoter-DNA substrate can be damaged by components present in the cell or the environment to produce covalent carcinogen–DNA adducts. These may lead to transcription blockage or unfaithful transcription. Benzo[a]pyrene (BP) is a widespread environmental chemical carcinogen which can be metabolically converted to DNA-reactive enantiomeric (+) and (−)-*anti*-benzo[a]pyrene diol epoxides (BPDEs). Recent experimental studies of a pair of stereoisomeric adenine adducts, derived from (+) and (−)-*anti*-BPDEs, have revealed how these lesions influence the complexation of TBP with the TATA box. Depending on the adduct's location in the TATA box and its stereochemistry, the stability of monomeric TATA–TBP complexes was found to increase or decrease relative to the unmodified DNA. We report here analyses of molecular-dynamics simulations to interpret these findings. Structural analyses of 12 DNA–protein systems representing different combinations of adduct stereoisomer type and placement within the promoter reveal that the location of the adduct within the TATA octamer determines whether the stability of TATA–TBP complexes is increased or decreased. The effect on binding stability can be interpreted in terms of conformational freedom and major-groove space available to BP due to the hydrogen bonds and inserted phenylalanines of the TATA–TBP complex; that is, depending on the position of the adenine to which BP is covalently bound, BP can be accommodated in an intercalated or major-groove orientation with ease or with difficulty (due to interference with TATA–TBP interactions). The unravelled structures and interactions thus reveal the effect of different adduct locations on TATA–TBP complex formation and suggest how transcription initiation may be affected by the presence of a bulky BP.

Keywords: molecular-dynamics simulation; TBP; benzo[a]pyrene diol epoxide; TATA box; stereochemistry; transcription initiation

One contribution of 16 to a Theme 'The mechanics of DNA'.

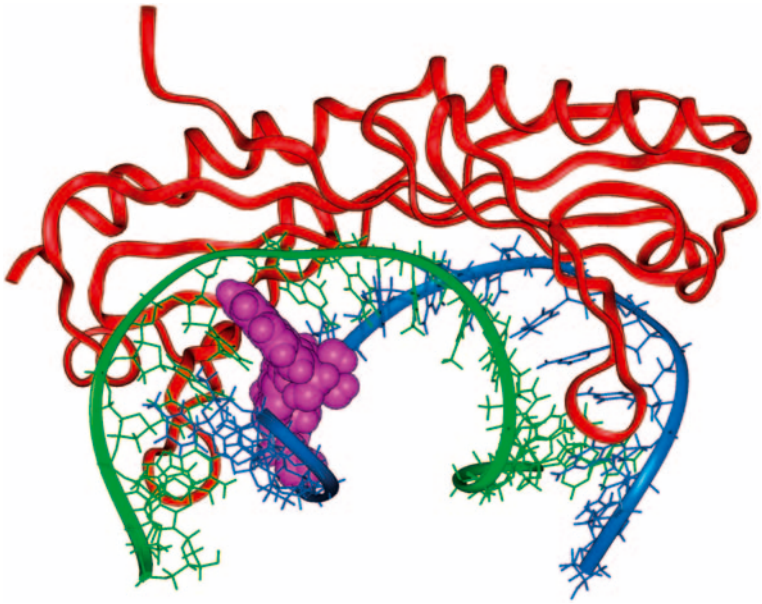

Figure 1. Complex of TBP (red) bound to a BP-modified TATA DNA (blue and green) after a 2.3 ns molecular-dynamics simulation. The BP–adenine adduct is pink (space-filling view) and is positioned at site A_1 (A6 in figure 5).

1. Introduction

Fundamental biological processes involve the interaction of nucleic acids and proteins, the primary molecules of life that make up cell structures and perform essential activities crucial to an organism's life. With the newly found roles for RNAs in regulating gene expression (Ban *et al.* 2000; Doudna & Cech 2002; Hannon 2002; Harms *et al.* 2001; Yusupov *et al.* 2001), nucleic acids emerge as even more powerful regulators of basic biological functions. The proteins encoded by the DNA are generally synthesized by a two-step process: transcription of DNAs to RNAs (Cramer *et al.* 2000, 2001; Gnatt *et al.* 2001; Sentenac 1985) and translation of RNAs to proteins. Transcription in eukaryotes is largely regulated by a complex network of proteins, including RNA polymerases and transcription factors (Kornberg 2001; Nudler 1999; Pugh 2000) as well as recently discovered RNAs (Mattick 2001).

The TATA-box binding protein, or TBP, a component of transcription factor D, is required by all eukaryotic RNA polymerases for correct initiation of transcription of ribosomal, messenger, small nuclear and transfer RNAs (Burley & Roeder 1996). TBP binds to the minor groove of an 8 base pair (bp) DNA-promoter element known as the TATA box (Patikoglou *et al.* 1999). X-ray crystal structures have shown that the TATA box is severely distorted by TBP: the DNA is locally compressed, unwound and severely bent, resulting in a widened minor groove and a compressed major groove (Nikolov *et al.* 1996) (see figure 1). More specifically, two phenylalanines of TBP insert (intercalate) into the first and last base steps of the TATA box, kinking the DNA locally by 52° and 39°, respectively. Another pair of phenylalanines buttress these penetrating phenylalanines and stabilize the kinks through extensive van der Waals contacts with the deoxyribose groups of adenines and thymines. Five hydrogen

bonds are formed between TBP and the central two base pairs of the TATA box. Around 14 hydrogen bonds are formed between TBP and the backbone of the TATA box. These hydrogen bonds and extensive hydrophobic interactions anchor the TATA box strongly to TBP with Watson–Crick base pairing maintained throughout the TATA box.

Far from acting as a 'silent partner' in these regulatory DNA–protein complexes, the DNA's intrinsic deformability plays a crucial role in the selection of DNA TATA elements for transcription association by TBP. From recent dynamics simulations in our group, unravelled sequence-dependent structural, energetic and flexibility properties of DNA important for TBP interactions include overall flexibility, minor-groove widening (with roll, rise and shift increases) at the TATA ends, untwisting within the TATA element (with large rolling at the ends), and relatively low maximal water densities around the DNA (Qian *et al.* 2001). These factors work with the severe deformation induced by the minor-groove binding protein, which kinks the TATA element at the ends and displaces local waters to form stabilizing hydrophobic contacts. Though the preferred bending direction itself is not a significant predictor of activity disposition, certain DNA variants (such as wild-type TATAAAAG and inactive TAAAAAAG) exhibit large preferred bends in directions consistent with their activity or inactivity (major- and minor-groove bends, respectively). These sequence–activity correlations play a role in association–dissociation interactions between TATA–TBP and other biomolecules within the eukaryotic transcription assembly (Strahs *et al.* 2003).

The promoter-DNA substrate can be damaged by components present in the cell or the environment. When such damage occurs to TATA DNA sequences, transcription can be affected. For example, biochemically activated forms of benzo[a]pyrene (BP) can covalently bind to DNA. BP is a representative of a well-known class of chemical carcinogens (polycyclic aromatic hydrocarbons) that are environmental pollutants, present in automobile exhaust, tobacco smoke and as a food contaminant (Geacintov *et al.* 1997; Grimmer 1993; Harvey 1991; Perrin *et al.* 1993; Phillips 1999). When biochemically activated to highly reactive BP diol epoxide (BPDE) molecules (Conney 1982), they can bind chemically to cellular DNA. These lesions have harmful biological effects. They may block transcription (Choi *et al.* 1994; Perlow *et al.* 2002) or replication (Brown & Romano 1991; Hruszkewycz *et al.* 1992; Moore & Strauss 1979), or the BP-damaged base may miscode and produce mutations during replication if not repaired (Geacintov *et al.* 1997; Hanrahan *et al.* 1997; Jelinsky *et al.* 1995; Lenne-Samuel *et al.* 2000; Perlow & Broyde 2001, 2002, 2003; Rechkoblit *et al.* 2002; Wei *et al.* 1993). The resultant faulty proteins, if involved in cell-cycle control, can be responsible for cancer initiation.

Two enantiomeric (i.e. mirror image) diol epoxides are the (+)-(7*R*, 8*S*, 9*S*, 10*R*)-7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene and the (−)-(7*S*, 8*R*, 9*R*, 10*S*) enantiomer (BPDE, metabolically derived from BP). They can react with adenines in DNA to form stereoisomeric 10*S*(+) and 10*R*(−)-*trans*-*anti*-[BP]–N^6 dA covalent DNA adducts, respectively (Cheng *et al.* 1989; Geacintov *et al.* 1997; Meehan & Straub 1979; Szeliga & Dipple 1998; Tan *et al.* 2000; Yan *et al.* 2001) (see figure 2). The conformations of the BP–adenine adducts are governed by three torsion angles: χ (O4′–C1′–N9–C4), α' (N1–C6–N6–C10(BP)), and β' (C6–N6–C10(BP)–C9(BP)) (figure 2). In Tan *et al.* (2000) the sterically allowed conformations are classified into four low-energy domains, as shown in table 1. High-resolution NMR solution studies

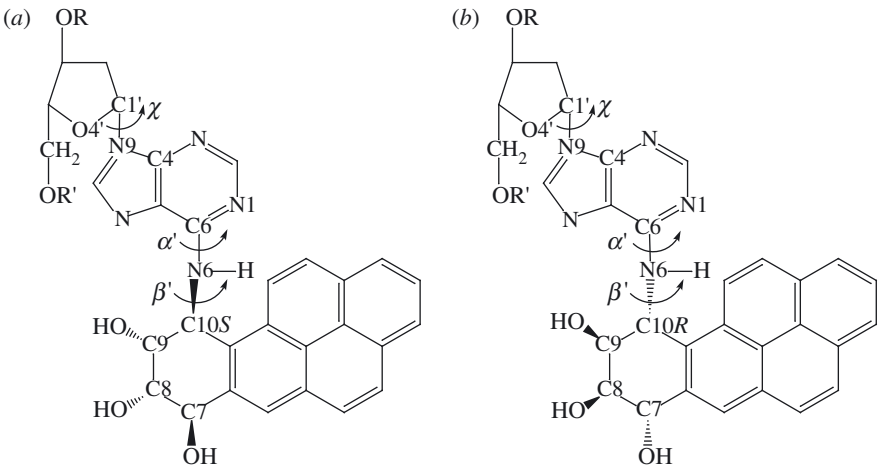

Figure 2. Stereoisomeric BP-modified adenines (*a*) 10*S*(+) and (*b*) 10*R*(−)-*trans-anti*-[BP]–N^6 dA.

Table 1. *Four low-energy conformational domains of* 10*S*(+) *and* 10*R*(−)-*trans-anti*-[BP]–N^6 dA *adducts (Tan et al. 2000)*

(The three torsion angles are χ (O4′–C1′–N9–C4), α' (N1–C6–N6–C10(BP)), and β' (C6–N6–C10(BP)–C9(BP)) (see figure 2).)

			β' (deg)	
domain	χ (deg)	α' (deg)	(+) adduct	(−) adduct
I	0–60	0 ± 35	-90 ± 40	90 ± 40
II	180–360	0 ± 35	-90 ± 40	90 ± 40
III	180–360	180 ± 35	-90 ± 40	90 ± 40
IV	0–60	180 ± 35	-90 ± 40	90 ± 40

Table 2. *The equilibrium constants* K° *(derived from Rechkoblit et al. (2001)) of TBP bound to different BP-modified and unmodified DNA duplexes at the temperature* $T = 277.15$ K

DNA type	K°
unmodified	1.18×10^8
10*S*(+)-*trans*-[A_1]	1.33×10^8
10*R*(−)-*trans*-[A_1]	1.82×10^8
10*S*(+)-*trans*-[A_2]	5.71×10^7
10*R*(−)-*trans*-[A_2]	8.00×10^7

(Schurter *et al.* 1995; Schwartz *et al.* 1997; Yeh *et al.* 1995; Zegar *et al.* 1996) show that BP is classically intercalated between DNA base pairs instead of being exposed in the major or minor groove of the DNA duplex; BP resides on the 3′-side of the modified adenine in the 10*S*(+) adduct and on the 5′-side in the 10*R*(−) adduct; the advantage of BP intercalation is that the hydrophobic pyrenyl moiety of BP can be buried within the helix rather than being exposed to solvent at the helix exterior.

To understand the effects of these two adducts on the binding of TBP to TATA DNAs (see figure 1), we undertook the following simulation study motivated by our group's prior work on DNA on the all-atom (Qian *et al.* 2001; Schlick *et al.* 2000; Strahs & Schlick 2000; Strahs *et al.* 2003) and supercoiled level (Beard & Schlick 2001; Huang & Schlick 2002; Huang *et al.* 2003; Zhang *et al.* 2003) and the recent study by Rechkoblit *et al.* (2001). The recent experimental work examined the influence of the two stereoisomeric BP–adenine adducts in a TATA-promoter sequence on complex formation between human TBP and the adduct-containing DNA. The experimental BP-modified DNA has the 25 bp duplex sequence (underlined is the TATA box):

$$\left|\begin{array}{l} 5'\text{-GAATTCCG } \underline{\text{T}\mathbf{A}_1\text{T AAATA}} \text{ CGTGTCGTG-}3' \\ 3'\text{-CTTAAGGC } \underline{\text{AT }\mathbf{A}_2\text{TTTAT}} \text{ GCACAGCAC-}5' \end{array}\right|.$$

In the experiments (Rechkoblit *et al.* 2001), a BP–adenine adduct—10S(+) or a 10R(−)-*trans*-*anti*-[BP]–N^6 dA—replaced the unmodified adenine at either position A_1 or A_2, to produce four BP-modified DNA duplexes: 10S(+)-*trans*-[A_1], 10R(−)-*trans*-[A_1], 10S(+)-*trans*-[A_2] and 10R(−)-*trans*-[A_2]. The binding of TBP to these four BP-modified, plus one unmodified DNA, was studied by electrophoretic mobility shift assays (EMSA) (Rechkoblit *et al.* 2001). *Intriguingly, the stabilities of the biologically significant monomeric TATA–TBP complexes were found to increase or decrease relative to the unmodified DNA, depending on the adduct's location* (A_1 *or* A_2) *and stereochemistry* (10S(+) *or* 10R(−)) (*table 2*). To interpret the experimental data, we have performed 12 dynamics simulations of varied systems, as described below, to delineate structure/binding/function relationships.

The simulations shed light on the complex systems by providing animated views of biological phenomena to link the experimental static images. Of course, we emphasize the inherent limitations (Schlick 1999, 2003) of state-of-the-art biomolecular modelling and simulations. These include approximate force fields, simplified electrostatics, short time-scales, approximate modelling of the complex physical systems (e.g. finite number of water molecules), limited sampling, and neglect of quantum effects (e.g. bond breaking, or proton-transfer events). Further developments (Schlick 2002, 2003) of force fields, methodologies and computational power are required to simulate longer and more accurate trajectories.

2. Methods

(*a*) *Starting structures*

We first used computer-graphics techniques to construct candidate initial structures for each combination of the adduct's location (A_1 or A_2) and stereochemistry (10S(+) or 10R(−)). We employ the PDB (protein database) (Berman *et al.* 2000) 1 CDW human TATA–TBP complex structure (Nikolov *et al.* 1996) and remodel its DNA to match the experimental sequence (but keep the length of the DNA duplex the same as that in the PDB structure). The remodelled complex has 211 residues in total, with residue numbers 1–32 assigned to the DNA and 33–211 to TBP. Our computer work using the program InsightII (Accelrys) creates starting structures for molecular dynamics based on the sterically allowed domains shown in table 1 and the need to avoid steric clashes. Specifically, we rotate the key torsion angles (χ, α', β') that govern the adduct's conformation within these allowed domains to achieve

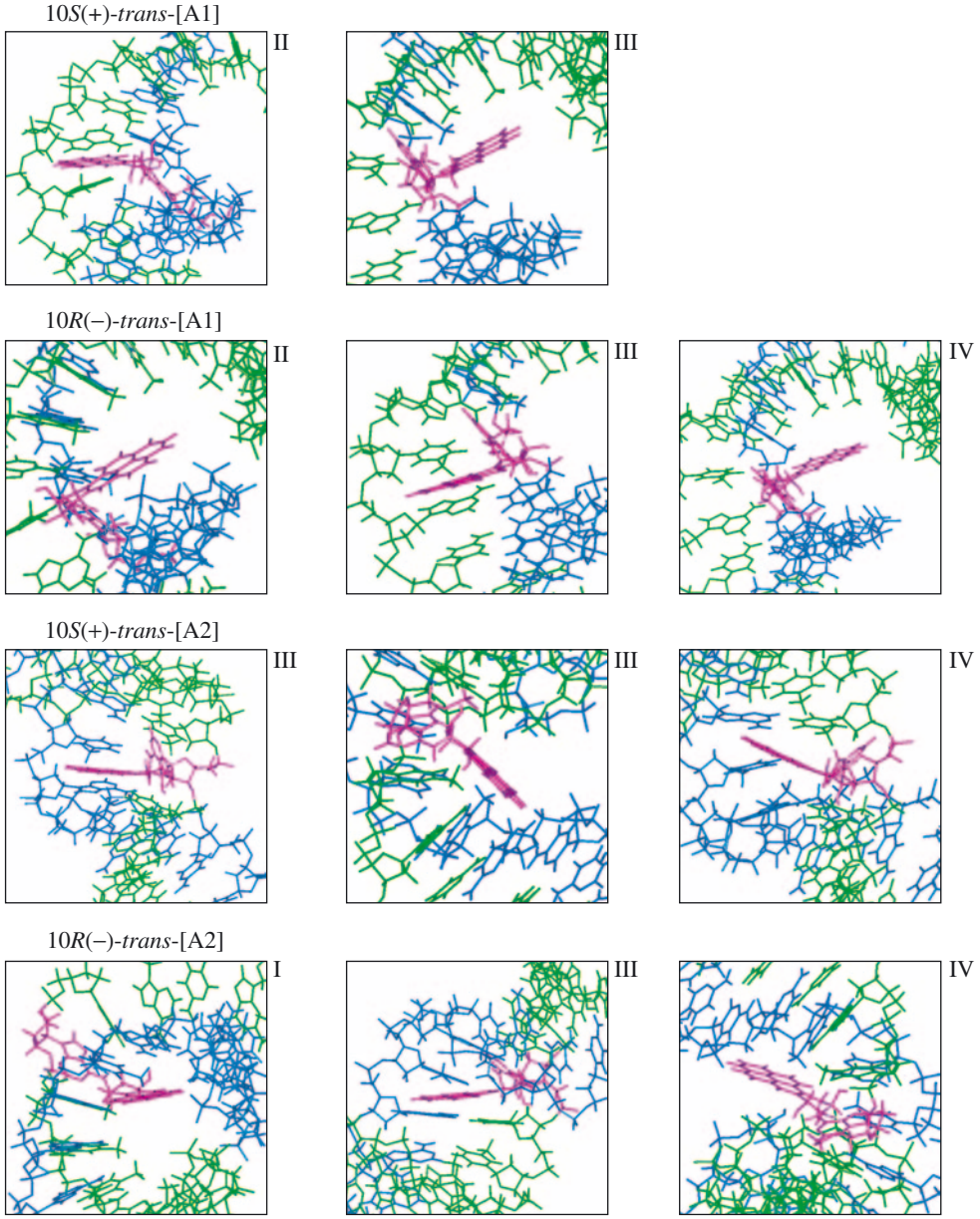

Figure 3. The optimal initial conformations of the four BP–adenine adducts of the DNA duplexes 10*S*(+)-*trans*-[A_1], 10*R*(−)-*trans*-[A_1], 10*S*(+)-*trans*-[A_2] and 10*R*(−)-*trans*-[A_2] in TATA–TBP complexes. Symbols I, II, III, and IV represent the four conformational domains (table 1) of the BP–adenine adducts. The DNA strands are blue and green; the BP–adenine adducts are pink; TBP is not shown.

structures with minimal collisions. These starting structures are given in figure 3 and table 3. In total, we simulate 12 TATA–TBP systems for 2.3 ns each, including the 11 BP-modified and one unmodified TATA–TBP complexes.

Table 3. *Structural changes during molecular dynamics*

before MD					MD average (last 600 ps)				
domain[a]	χ (deg)	α' (deg)	β' (deg)	form[b]	domain	χ (deg)	α' (deg)	β' (deg)	form
10*S*(+)-*trans*-[A_1]									
II	215.8	−32.2	−120.5	int	II	183.8 ± 9.7[c]	−23.6 ± 8.2	−96.6 ± 9.9	int
III	227.7	176.6	−129.1	MG	III	192.9 ± 8.2	168.5 ± 8.8	−104.0 ± 8.8	MG
10*R*(−)-*trans*-[A_1]									
II	222.4	−16.2	90.1	MG	II	210.5 ± 6.7	−20.8 ± 8.6	75.7 ± 6.9	MG
III	222.0	168.8	77.6	int	III	203.9 ± 12.5	170.0 ± 8.8	90.2 ± 8.7	int
IV	47.7	157.1	119.3	MG	IV	23.6 ± 10.8	159.5 ± 8.7	96.0 ± 9.4	MG
10*S*(+)-*trans*-[A_2]									
III	276.9	−159.5	−103.6	int	III	279.9 ± 8.3	202.0 ± 8.2	−81.3 ± 7.1	int
III[d]	190.4	146.5	−97.7	MG	III	232.1 ± 12.3	157.8 ± 8.9	−85.9 ± 10.7	MG
IV	8.0	146.5	−97.7	int	IV	59.1 ± 9.8	157.9 ± 10.5	−95.8 ± 7.6	int
10*R*(−)-*trans*-[A_2]									
I	6.4	−15.3	60.5	MG	I	39.6 ± 10.2	1.3 ± 8.6	51.9 ± 12.0	MG
III	196.9	148.4	83.5	int	III	237.9 ± 13.6	176.4 ± 12.0	69.2 ± 9.4	MG
IV	13.4	148.5	87.3	MG	IV	45.7 ± 9.6	160.9 ± 10.0	85.7 ± 9.9	MG

[a]The domain (I to IV) of BP–adenine adducts corresponds to table 1.
[b]The orientation of BP is classified as intercalation (int) or major groove (MG).
[c]Standard deviations for dihedral angles over the last 600 ps of the production-dynamics stage follow the angle values.
[d]This is a second candidate structure in domain III for the (+) stereoisomer at A_2 according to the sterically allowed domains in the TATA–TBP complex. This system is named as 10*S*(+)-*trans*-[A_2]:III-2.

(*b*) *Force field*

The Cornell force field (Cornell *et al.* 1995) with parm98 parameter set (Cheatham *et al.* 1999) is used to assign parameters for energy minimization and molecular dynamics. The missing partial charges of the BP–adenine adducts are computed with a method (available at http://monod.biomath.nyu.edu/~qzhang/BP-adenine/protocol.txt) similar to that used in the development of assisted model building with energy refinement (AMBER) (Cieplak *et al.* 1995), using AMBER's restrained electrostatic potential (RESP) approach (Bayly *et al.* 1993; Cornell *et al.* 1993) based on the electrostatic potential (ESP) calculated from Gaussian 98 (Frisch *et al.* 1998) using Hartree–Fock calculations with the 6–31G* basis set. A partial charge set is computed for each starting structure employed. Besides the partial charges, other missing force-field parameters (including bond length, bond angle and dihedral-angle parameters) are assigned according to chemically similar atom types already available in the parm98 parameter set (Cheatham *et al.* 1999). These added parameters are available at http://monod.biomath.nyu.edu/~qzhang/BP-adenine/parameters.pdf.

(*c*) *Energy minimization and molecular-dynamics protocol*

We use AMBER 6 (Case *et al.* 1999) to perform energy minimization in preparation for the molecular-dynamics simulations. Energy minimization also eliminates initial unfavourable van der Waals contacts. Specifically, the systems are minimized with 400 steps of steepest descent (SD) followed by 400 steps of conjugate gradient (CG). At this stage, since explicit solvent has not yet been included, we set the dielectric constant to be $\varepsilon = 4R$ to mimic the solvent, where R is the distance between two atoms. The 1-4 electrostatic interactions are scaled by a factor of 1/1.2 for this energy term according to the AMBER protocol. We use 12 Å cutoffs for the non-bonded Lennard-Jones interactions. The solute is restrained using a harmonic force with a weight of 1.0 kcal mol^{-1}.

The system is then solvated with a rectangular box of TIP3 waters (Jorgensen *et al.* 1983) that extends 10.0 Å from the solute. (The water box encloses the solute, and water molecules overlapping the solute are removed. The criterion for removal is as follows: if the distance between any water atom to the closest solute atom is less than the sum of the atoms' van der Waals distances, that molecule is removed. We find no water is added to the TATA–TBP interface, which is consistent with the crystal structure (Nikolov *et al.* 1996).) Counterions are added to neutralize the system and to simulate the experimental ionic strength of 130 mM in the reaction buffer (Rechkoblit *et al.* 2001). To further eliminate van der Waals collisions following the addition of water molecules and ions, the entire system is minimized again using 50 steps of SD followed by 5000 steps of CG. The solute is restrained using a harmonic force with a weight of 50.0 kcal mol^{-1}.

The time-step for the dynamics simulation is 2 fs with the leap-frog Verlet algorithm (Hockney 1970). SHAKE (Ryckaert *et al.* 1977) is used to constrain lengths of bonds involving hydrogens with a tolerance of 10^{-6} Å. The non-bonded list is updated whenever any atom has moved more than 0.5 Å since the last update. The 1-4 electrostatic interactions are scaled by a factor of 1/1.2 following the AMBER protocol. The particle-mesh Ewald (PME) method (Darden *et al.* 1993; Essmann *et*

al. 1995) is used to treat long-range electrostatics interactions. The length of the molecular-dynamics trajectories is 2.3 ns.

For each TATA–TBP system, we follow these five molecular-dynamics stages using AMBER 6 (Case *et al.* 1999) (the first four are equilibration stages).

(i) Heating. Each system is heated from 1.00 to 277.15 K (the EMSA experiment was performed at 4.00 °C (Rechkoblit *et al.* 2001)) with 60 000 steps at constant volume, with an atom-based temperature scaling using the Berendsen coupling algorithm (Berendsen *et al.* 1984) every 0.2 ps. A 10.0 kcal mol^{-1} restraint was placed on the solute.

(ii) Constant pressure simulation I. Each system is equilibrated for 100 000 steps at 277.15 K and constant atmospheric pressure using the Berendsen coupling algorithm (Berendsen *et al.* 1984) with atom-based scaling. The temperature was scaled every 1.0 ps, and a 5.0 kcal mol^{-1} restraint was placed on the solute.

(iii) Constant pressure simulation II. Each system is equilibrated for another 100 000 steps at 277.15 K and constant atmospheric pressure. The temperature was scaled every 2.0 ps, and a 2.0 kcal mol^{-1} restraint was placed on the solute.

(iv) Final equilibration. We employ 200 000 steps at 277.15 K and constant atmospheric pressure. The temperature was scaled every 4.0 ps, and no restraint was placed on the solute.

(v) Production dynamics. We follow 700 000 steps with the same conditions as final equilibration (except that production dynamics uses a molecule-based scaling method instead of atom-based scaling for the pressure scaling of the atomic coordinates).

3. Results

(*a*) *Structural change during molecular dynamics*

In figure 4, we illustrate the results of the 2.3 ns trajectory for the system 10*S*(+)-*trans*-[A_1]:II (figure 3). The figure shows the time dependence of the three torsion angles (χ, α', β') and the structure of the DNA residues near BP at the trajectory's end. The torsion angles have become stable during the equilibration stages (left of red vertical line). Compared with its initial structure in figure 3 (top left), we see that BP of the (+) stereoisomer at A_1 with an intercalated orientation has equilibrated into a stable conformation.

These torsion angles in the remaining 11 systems also become stable within 2.3 ns. Their initial and final values are shown in table 3. We see that no system changes its *domain* (i.e. I, II, III, IV) of the BP–adenine adduct during the dynamics simulation (consistent with these preferred regions deduced previously (Tan *et al.* 2000)), although there are variations as large as 51.1°. However, the BP *orientation* in the system 10*R*(−)-*trans*-[A_2]:III changes from an intercalated to major-groove form. Note that only one BP of the (−) stereoisomers at A_2 was initially intercalated, and that all three systems (initial domains I, III, IV) are major-groove conformers at the trajectory's end. Changes in χ, α', β' within the preferred regions mediate this rearrangement. In fact, only one of the 12 systems (the (−) stereoisomer at A_2 and

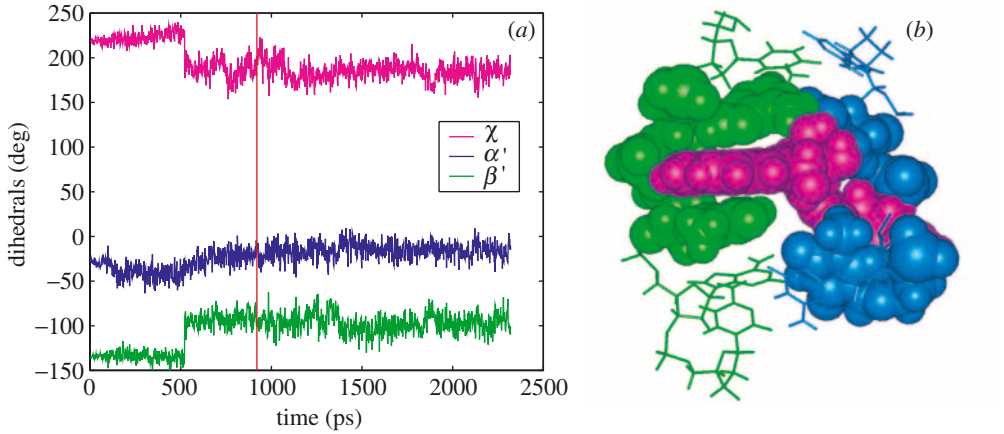

Figure 4. (*a*) the torsion angles (χ, α', β') as a function of dynamics time; the red vertical line separates the equilibration stages and the production-dynamics stage (see § 2). (*b*) The DNA residues near BP in the system 10*S*(+)-*trans*-[A$_1$]:II (after 2.3 ns of molecular dynamics). The two DNA strands are blue and green; the BP–adenine adduct is pink.

Table 4. *RMSD of TATA DNA backbone and TBP backbone of modified systems from unmodified ones*

(The RMSD values are based on the average structures of the last 600 ps of the production-dynamics stage. Underlined values show the lowest RMSD in each column.)

	intercalation family RMSD (Å)		major-groove family RMSD (Å)	
system	TATA	TBP	TATA	TBP
10*S*(+)-*trans*-[A$_1$]: II	0.82	1.22		
III			0.48	1.15
10*R*(−)-*trans*-[A$_1$]: II			<u>0.37</u>	<u>0.96</u>
III	<u>0.70</u>	<u>0.87</u>		
IV			0.45	0.98
10*S*(+)-*trans*-[A$_2$]: III	1.03	1.35		
III-2			1.20	1.40
IV	0.97	1.06		
10*R*(−)-*trans*-[A$_2$]: I			1.29	1.26
III			0.72	1.10
IV			0.88	1.05

domain III) changes the mode of BP conformation; all others started as intercalated remain so, and those which started as major-groove conformers remain there. We interpret this change from intercalated to major-groove orientation below.

Although the BP–adenine adducts have become stable within 2.3 ns, the complexes still exhibit slight increases in root mean square deviation (RMSD) (compared with

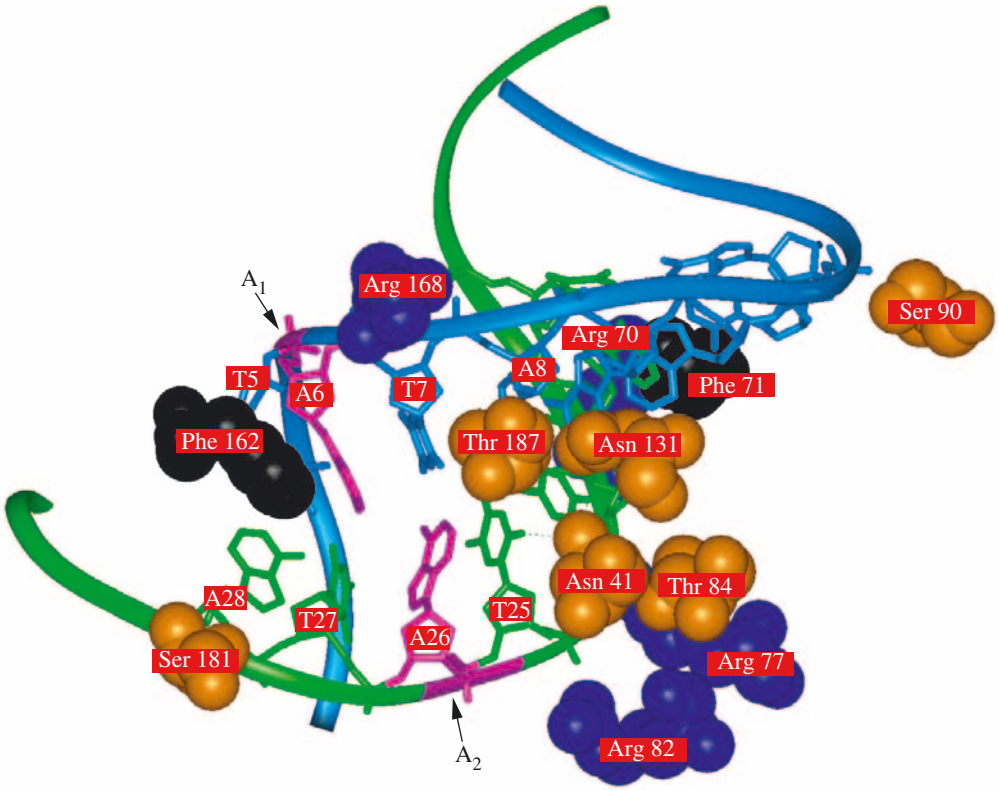

Figure 5. TATA–TBP interface with the 8 bp TATA box and the amino acids contributing to hydrogen bonds and insertion of phenylalanines. Amino acids are shown in space-filling rendering, DNA residues are in stick and hydrogens are not shown. Phenylalanines (Phe) are black, asparagines (Asn), threonines (Thr) and serines (Ser) are orange, and arginines (Arg) are dark blue. The DNA strand with A_1 (A6) is blue; the DNA strand with A_2 (A26) is green; both A_1 and A_2 within them are highlighted in pink (arrows).

initial state) during the last 600 ps of production dynamics. For example, the unmodified complex rises 0.07 Å in RMSD within the final 600 ps. This is likely due to the limited time-scale of the simulations.

(*b*) *Evaluation of structure conservation*

To evaluate which BP-modified systems conserve the original unmodified structure most closely, we compute the RMSD of the TATA DNA backbones and the TBP backbones of the modified systems from the unmodified ones in table 4. The systems are grouped into two families according to their BP's orientations: intercalation and major groove (intercalation of BP lengthens the DNA backbone). In the intercalation family, 10*R*(−)-*trans*-[A_1]:III has the lowest RMSD from the TATA DNA backbone and the TBP backbone. In the major-groove family, 10*R*(−)-*trans*-[A_1]:II has the lowest RMSD. Our structural analysis on 10*R*(−)-*trans*-[A_1]:III shows that this system conserves the original structure best: the two phenylalanines that kink the TATA DNA and the two phenylalanines that constrain the DNA bending have moved little

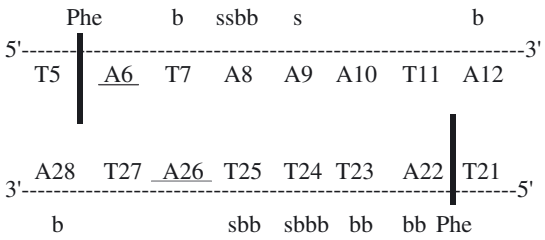

Figure 6. Diagram of major TATA–TBP interactions (hydrogen bonds and insertion of phenylalanines): 'b' represents a hydrogen bond on the DNA backbone; 's' represents a hydrogen bond on a DNA base; and two phenylalanines (Phe) insert the DNA at A6 and A22.

during the dynamics. The results clearly indicate that the (−) stereoisomer at A_1 distorts the structure least, followed by the (+) stereoisomer at A_1.

The A_2 systems 10*S*(+)-*trans*-[A_2] and 10*R*(−)-*trans*-[A_2] distort the structure most, regardless of which stereoisomer of the BP–adenine adduct and whether its BP is intercalated or exposed to the major groove.

(i) In the 10*S*(+)-*trans*-[A_2]:III complex, one phenylalanine that inserts at A_1 is pushed back towards the TBP by 1.3 Å, the entire DNA backbone is distorted, and the BP aromatic ring is distorted.

(ii) In the 10*S*(+)-*trans*-[A_2]:III-2 complex (domain III), all four phenylalanines are pushed back toward the TBP and the entire DNA backbone is distorted.

(iii) In the 10*S*(+)-*trans*-[A_2]:IV complex, the DNA backbone around A_2 is distorted, affecting the other TATA DNA bases upstream of A_2, and the BP aromatic ring is markedly distorted.

(iv) In the 10*R*(−)-*trans*-[A_2]:I complex, the DNA backbone is distorted with adenine A_2 largely stretched to allow the bulky BP to fit into the major groove.

(v) In the 10*R*(−)-*trans*-[A_2]:III complex, the DNA backbone with A_1 is drawn away from TBP, with the two phenylalanines near A_1 pushed back toward the TBP; the thymine downstream of A_1 loses its Watson–Crick hydrogen bonds. (10*R*(−)-*trans*-[A_2]:III is the only system where the BP's orientation is changed during dynamics: from intercalation to major groove.)

(vi) In the 10*R*(−)-*trans*-[A_2]:IV complex, the DNA backbone near A_2 is largely distorted, the thymine at the downstream of A_1 loses its Watson–Crick hydrogen bonds, and the base of the thymine at the upstream of A_2 is rotated by 23.7° (χ (O4′–C1′–N1–C2) changes from −94.6° to −118.3°).

(*c*) *Structural explanations*

In figure 5 we examine the TATA–TBP interface with the 8 bp TATA box and the amino acids contributing to hydrogen bonds, and insertion of phenylalanines in the average structure over the last 600 ps production simulation corresponding to the unmodified TATA–TBP system. A diagram of the major TATA–TBP interactions (hydrogen bonds and insertion of phenylalanines) is also provided in figure 6, where we denote hydrogen bonds with 'b' (on DNA backbone) and 's' (on

Table 5. *Hydrogen bonds between the TATA box and TBP*

(We compute for each hydrogen bond an occupancy (percent of time the hydrogen bond is intact) within the last 600 ps of the production-dynamics stage of the unmodified TATA–TBP system. The criterion of a hydrogen bond is that the donor–acceptor distance is not greater than 3.35 Å and the donor–hydrogen-acceptor bond angle is greater than or equal to 135°. Hydrogen bonds with occupancy less than 25% are not reported. Each acceptor and donor is listed with its residue type, residue number, and atom name; atom names are in AMBER nomenclature (Cornell *et al.* 1995). All acceptors are found to be DNA adenine or thymine atoms.)

hydrogen bonds on DNA backbone			hydrogen bonds on DNA bases		
acceptor	donor	occupancy (%)	acceptor	donor	occupancy (%)
T7: O3′	Arg 168: NH2	47	A8: N3	Thr 187: OG1	97
A8: O2P	Arg 168: NH1	100	A8: N3	Asn 131: ND2	30
A8: O2P	Arg 168: NH2	72	A9: N3	Asn 131: ND2	52
A12: O2P	Ser 90: OG	100	T24: O2	Asn 41: ND2	100
A22: O2P	Arg 70: NH1	36	T25: O2	Asn 41: ND2	94
A22: O3′	Arg 70: NH1	51			
T23: O2P	Arg 70: NH1	68			
T23: O2P	Arg 70: NH2	99			
T24: O2P	Arg 77: NH2	100			
T24: O2P	Arg 77: NE	58			
T24: O2P	Thr 84: OG1	98			
T25: O2P	Arg 82: NH1	86			
T25: O2P	Arg 82: NH2	52			
A28: O2P	Ser 181: OG	100			

DNA base). These bonds anchor the DNA tightly to TBP, and the two phenylalanines (Phe) insert and kink the DNA at A6 (A_1) and A22; the two phenylalanines (not shown) reside over T11 and A27 and are stabilized by van der Waals interactions.

The hydrogen bonds listed in table 5 correspond to the same structure shown in figure 5. All the hydrogen-bond acceptors are in the TATA box (note that the TATA box adenines and thymines cannot serve as donors since they have no hydrogen on oxygens (O2 of thymines) nor nitrogens (N3 of adenines) in the minor groove). The hydrogen bonds on the DNA bases concentrate on the central two base pairs (A8–T25, A9–T24) of the 8 bp TATA box. The hydrogen bonds on the DNA backbone are mostly located on the central TATA residues with two strong hydrogen bonds at the two ends (A12 and A28). Hydrophobic interactions (not shown) are mostly from valines (Val) and leucines (Leu) and on the adenines (A). The hydrophilic part (three hydroxyls) of BP is always near or in the DNA major groove and has no chance to interact with the minor-groove-bound TBP, regardless of whether BP is intercalated or exposed to the major groove. Weak water-mediated hydrogen bonds are not shown either as they are not sensitive to the addition of BP to the TATA box due to the mobility of water molecules.

First, from these interactions we can interpret why it is difficult for a BP to adopt an intercalated orientation at the A_2 binding position. When a BP binds to A6 (A_1)

it could potentially intercalate either between A28 and T27 or between T27 and A26 without difficulty because the bases of A28 and T27 are flexible and can move to the $3'$ end. Now, when a BP binds instead to A26 (A_2) it can hypothetically intercalate either between A6 and T7 or between T7 and A8. However, since A6 (A_1) is tightly constrained by a phenylalanine, and A8 is strongly constrained by two hydrogen bonds on its base, a BP at A26 (A_2) cannot easily adopt an intercalated orientation without distorting the DNA backbone and hence the TBP backbone. Thus, it must resort to exposing its hydrophobic aromatic ring to the solvent (present in the binding experiments and simulations), and this is not energetically favourable. The more facile intercalation at A_1 also increases the stacking between BP and DNA bases and thus stabilizes the system (Yan *et al.* 2001).

Second, we can interpret why it is also difficult for a BP to adopt a major-groove orientation at the A_2 binding position. This emerges from the limited space in the major groove for BP bonded at A_2. The N6 groups (N6, H61, H62) of the adenines in the TATA box point to the major groove. In the BP-modified systems, H62 is replaced by a carbon C10 of the bulky BP (figure 2). Thus, the major-groove orientation adopted by four of the six BPs at A_2 (see tables 3 and 4) also comes at a structural cost (see below).

To further interpret the more facile BP fit at A_1 compared to A_2, we compute the exposed surface area of the N6 groups of A_1 and A_2 in the unmodified system; the exposed surface area can indicate which N6 group extends further into the major groove to possibly accommodate the bulky BP. Our calculation shows the N6 group of A_1 has an exposed surface area of 16.3 $Å^2$ but that of A_2 has only 6.6 $Å^2$.

Figure 7 shows the DNA residues near A_1 (A6) and A_2 (A26) in the average structure of the unmodified system from the last 600 ps of the production-dynamics stage. The two C7 groups (C7, H71, H72, H73; C7, or C5M, bonds to C5 of thymine) from both T5 and T25 limit the available space of the major groove for the bulky BP bonded at both A_1 and A_2. However, the angle (C7–H62–C7) at A_1 is 115.2°, while at A_2 it is only 93.0°, as shown in figure 7. The smaller angle at A_2 narrows the major-groove space for the bulky BP bonded at A_2. Thus, BP at A_2 must distort the DNA backbone to fit in the major groove (RMSD of the DNA backbone is therefore larger). The basic reason for the smaller exposed surface area and the smaller angle at A_2 is the phenylalanine insertion at A_1, since it pushes the N6 group of A_1 towards the major groove and also kinks the DNA to bend its upstream bases towards the major groove.

Thus, our structural analyses revealing the greater distortion at A_2 compared with A_1 are consistent with the experimental result: BP at A_1 slightly increases the binding affinity between TBP and the TATA DNA, while BP at A_2 decreases the binding affinity. This is true regardless of adduct stereochemistry. Our results also indicate that any structural effect of stereochemistry is more modest than the positional effect, also in accord with the experiments.

4. Summary

Within the well-recognized caveats of biomolecular-dynamics simulations, our results offer atomic-level structural explanations for the puzzling difference in binding affinities observed experimentally for the bulky BP bound to two different adenines of the TATA-promoter octamer (table 2) (Rechkoblit *et al.* 2001). The disparate behaviour

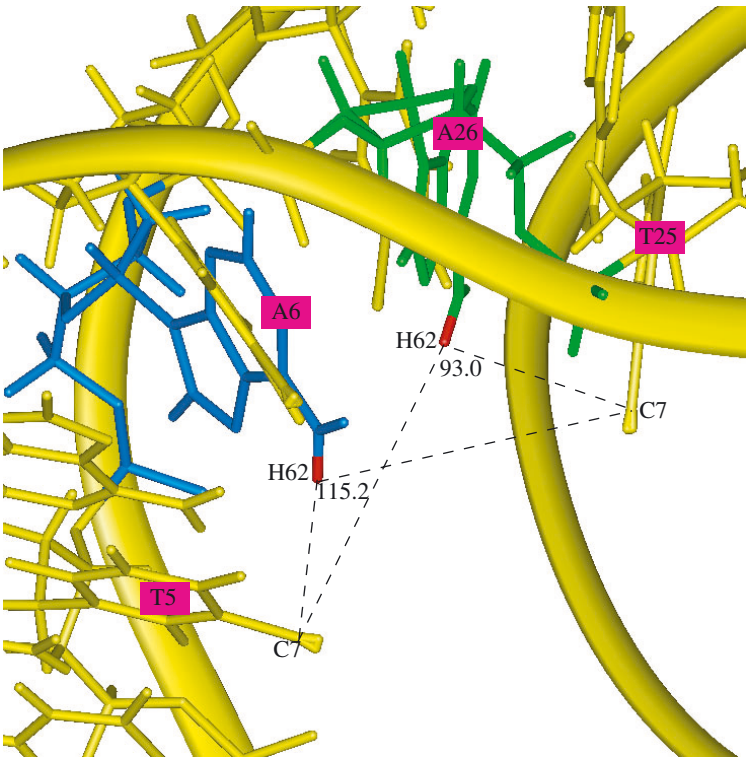

Figure 7. DNA major groove from the average structure of the unmodified system. A_1 (A6) is blue, A_2 (A26) is green, the rest of the residues are yellow, hydrogens H62 of A_1 and A_2 are red, carbons C7 (C5M; bonded to C5 of thymine) of T5 and T25 are labelled, the angle (C7–H62–C7) at A_1 is 115.2°, the angle (C7–H62–C7) at A_2 is 93.0°.

at A_1 and A_2 is caused by differences in conformational freedom available to BP when bound to A_1 or A_2; differences result from the TATA DNA's deformation when complexed to TBP and hence the available major-groove space. The two related factors lead to the following outcome. A BP bound at A_1 can easily adopt both the intercalated and major-groove orientations, while a BP bound at A_2 can only adopt intercalated or major-groove forms under penalty, due to distortion to the DNA backbone as well as the TBP backbone. Thus, a bulky BP at A_2 decreases complex stability. The hydrogen bonds and inserted phenylalanines of the TATA–TBP complex (see figures 5, 6) explain these effects by structural considerations.

Although the experimental equilibrium constants have only a two-fold difference between A_1- and A_2-modified systems, our molecular-dynamics study has given a consistent explanation. Further work, including energetic analysis, is required to determine preferred BP orientations based on both adduct stereochemistry and position. Our studies might ultimately help to explain how carcinogenic substances can interfere with transcription initiation, a problem of fundamental importance to biology and medicine.

The authors thank N. E. Geacintov for providing additional information regarding the experiments, P. Cieplak for helping with the computation of AMBER partial charges, M. Wang and S. Ding for helping with Gaussian, R. A. Perlow for discussing AMBER's energy minimization,

X. Huang for discussing electrophoresis, S. Yan and M. Wu for discussing building starting structures and analysing trajectories, and L. Yang, D. Strahs, L. Zhang and K. Arora for discussing various aspects of molecular-dynamics simulations. We also appreciate helpful responses on the AMBER mail reflector (especially from D. Case). Support from the National Institutes of Health (R01 GM55164 to T.S., CA28038 and CA75449 to S.B.) is gratefully acknowledged. Acknowledgment is also made to the donors of the American Chemical Society Petroleum Research Fund (39115-AC4 to T.S.) for partial support of this research. Computations were supported by National Computational Science Alliance under MCA99S021N and used on the NCSA SGI Origin2000.

References

Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. 2000 *Science* **289**, 905–920.

Bayly, C. I., Cieplak, P., Cornell, W. D. & Kollman, P. A. 1993 *J. Phys. Chem.* **97**, 10 269–10 280.

Beard, D. A. & Schlick, T. 2001 *Struct. Folding Design* **9**, 105–114.

Berendsen, H. J. C., Postma. J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. 1984 *J. Chem. Phys.* **81**, 3684–3690.

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. 2000 *Nucl. Acids Res.* **28**, 235–242.

Brown, W. C. & Romano, L. J. 1991 *Biochemistry* **30**, 1342–1350.

Burley, S. K. & Roeder, R. G. 1996 *A. Rev. Biochem.* **65**, 769–799.

Case, D. A. (and 20 others) 1999 *AMBER 6*. San Francisco, CA: University of California.

Cheatham, T. E., Cieplak, P. & Kollman, P. A. 1999 *J. Biomol. Struct. Dynam.* **16**, 845–862.

Cheng, S. C., Hilton, B. D., Roman, J. M. & Dipple, A. 1989 *Chem. Res. Toxicol.* **2**, 334–340.

Choi, D. J., Marino-Alessandri, D. J., Geacintov, N. E. & Scicchitano, D. A. 1994 *Biochemistry* **33**, 780–787.

Cieplak, P., Cornell, W. D., Bayly, C. & Kollman, P. A. 1995 *J. Comput. Chem.* **16**, 1357–1377.

Conney, A. H. 1982 *Cancer Res.* **42**, 4875–4917.

Cornell, W. D., Cieplak, P., Bayly, C. I. & Kollman, P. A. 1993 *J. Am. Chem. Soc.* **115**, 9620–9631.

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. & Kollman, P. A. 1995 *J. Am. Chem. Soc.* **117**, 5179–5197.

Cramer, P., Bushnell, D. A., Fu, J., Gnatt, A. L., Maier-Davis, B., Thompson, N. E., Burgess, R. R., Edwards, A. M., David, P. R. & Kornberg, R. D. 2000 *Science* **288**, 640–649.

Cramer, P., Bushnell, D. A. & Kornberg, R. D. 2001 *Science* **292**, 1863–1876.

Darden, T., York, D. & Pedersen, L. 1993 *J. Chem. Phys.* **98**, 10 089–10 092.

Doudna, J. A. & Cech, T. R. 2002 *Nature* **418**, 222–228.

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. 1995 *J. Chem. Phys.* **103**, 8577–8593.

Frisch, M. J. (and 56 others) 1998 *Gaussian 98*, revision A.7. Pittsburgh, PA: Gaussian Inc.

Geacintov, N. E., Cosman, M., Hingerty, B. E., Amin, S., Broyde, S. & Patel, D. J. 1997 *Chem. Res. Toxicol.* **10**, 111–146.

Gnatt, A. L., Cramer, P., Fu, J., Bushnell, D. A. & Kornberg, R. D. 2001 *Science* **92**, 1876–1882.

Grimmer, G. 1993 In *Proc. 13th Int. Symp. Polynuclear Aromatic Hydrocarbons*. London: Gordon and Breach.

Hannon, G. J. 2002 *Nature* **418**, 244–251.

Hanrahan, C. J., Bacolod, M. D., Vyas, R. R., Liu, T., Geacintov, N. E., Loechler, E. L. & Basu, A. K. 1997 *Chem. Res. Toxicol.* **10**, 369–377.

Harms, J., Schluenzen, F., Zarivach, R., Bashan, A., Gat, S., Agmon, I., Bartels, H., Franceschi, F. & Yonath, A. 2001 *Cell* **107**, 292–302.

Harvey, R. G. 1991 *Polycyclic aromatic hydrocarbons*. Cambridge University Press.

Hockney, R. W. 1970 *Meth. Computat. Phys.* **9**, 136–211.

Hruszkewycz, A. M., Canella, K. A., Peltonen, K., Kotrappa, L. & Dipple, A. 1992 *Carcinogenesis* **13**, 2347–2352.

Huang, J. & Schlick, T. 2002 *J. Chem. Phys.* **117**, 8573–8586.

Huang, J., Zhang, Q. & Schlick, T. 2003 *Biophys. J.* **85**, 804–817.

Jelinsky, S. A., Liu, T., Geacintov, N. E. & Loechler, E. L. 1995 *Biochemistry* **34**, 13 545–13 553.

Jorgensen, W. L., Chandreskhar, J., Madura, J. D., Imprey, R. W. & Klein, M. L. 1983 *J. Chem. Phys.* **79**, 926–935.

Kornberg, R. D. 2001 *Biol. Chem.* **382**, 1103–1107.

Lenne-Samuel, N., Janel-Bintz, R., Kolbanovskiy, A., Geacintov, N. & Fuchs, R. 2000 *Mol. Microbiol.* **38**, 299–307.

Mattick, J. S. 2001 *EMBO Rep.* **2**, 986–991.

Meehan, T. & Straub, K. 1979 *Nature* **277**, 410–412.

Moore, P. & Strauss, B. S. 1979 *Nature* **278**, 664–666.

Nikolov, D. B., Chen, H., Halay, E. D., Hoffmann, A., Roeder, R. G. & Burley, S. K. 1996 *Proc. Natl Acad. Sci. USA* **93**, 4862–4867.

Nudler, E. 1999 *J. Mol. Biol.* **288**, 1–12.

Patikoglou, G. A., Kim, J. L., Sun, L., Yang, S., Kodadek, T. & Burley, S. K. 1999 *Genes Dev.* **13**, 3217–3230.

Perlow, R. A. & Broyde, S. 2001 *J. Mol. Biol.* **309**, 519–536.

Perlow, R. A. & Broyde, S. 2002 *J. Mol. Biol.* **322**, 291–309.

Perlow, R. A. & Broyde, S. 2003 *J. Mol. Biol.* **327**, 797–818.

Perlow, R. A., Kolbanovskii, A., Hingerty, B. E., Geacintov, N. E., Broyde, S. & Scicchitano, D. A. 2002 *J. Mol. Biol.* **321**, 29–47.

Perrin, J. L., Poirot, N., Liska, P., Hanras, C., Theinpont, A. & Felix, G. 1993 In *Proc. 13th Int. Symp. Polynuclear Aromatic Hydrocarbons*. London: Gordon and Breach.

Phillips, D. H. 1999 *Mutat. Res.* **443**, 139–147.

Pugh, B. F. 2000 *Gene* **255**, 1–14.

Qian, X., Strahs, D. & Schlick, T. 2001 *J. Mol. Biol.* **308**, 681–703.

Rechkoblit, O., Krzeminksy, J., Amin, S., Jernström, B., Louneva, N. & Geacintov, N. E. 2001 *Biochemistry* **40**, 5622–5632.

Rechkoblit, O., Zhang, Y., Guo, D., Wang, Z., Amin, S., Krzeminsky, J., Louneva, N. & Geacintov, N. E. 2002 *J. Biol. Chem.* **277**, 30 488–30 494.

Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. 1977 *J. Comput. Phys.* **23**, 327–341.

Schlick, T. 1999 *J. Comput. Phys.* **151**, 1–8.

Schlick, T. 2002 *Molecular modeling: an interdisciplinary guide.* Springer.

Schlick, T. 2003 *Biophys. J.* **85**, 1–4.

Schlick, T., Beard, D. A., Huang, J., Strahs, D. & Qian, X. 2000 *IEEE Comput. Sci. Engng* **2**, 38–51.

Schurter, E. J., Yeh, H. J., Sayer, J. M., Lakshman, M. K., Yagi, H., Jerina, D. M. & Gorenstein, D. G. 1995 *Biochemistry* **34**, 1364–1375.

Schwartz, J. L., Rice, J. S., Luxon, B. A., Sayer, J. M., Xie, G., Yeh, H. J., Liu, X., Jerina, D. M. & Gorenstein, D. G. 1997 *Biochemistry* **36**, 11 069–11 076.

Sentenac, A. 1985 *CRC Crit. Rev. Biochem.* **18**, 31–90.

Strahs, D. & Schlick, T. 2000 *J. Mol. Biol.* **301**, 643–663.

Strahs, D., Barash, D., Qian, X. & Schlick, T. 2003 *Biopolymers* **69**, 216–243.

Szeliga, J. & Dipple, A. 1998 *Chem. Res. Toxicol.* **11**, 1–11.

Tan, J., Geacintov, N. E. & Broyde, S. 2000 *J. Am. Chem. Soc.* **122**, 3021–3032.

Wei, S. J., Chang, R. L., Bhachech, N., Cui, X. X., Merkler, K. A., Wong, C. Q., Hennig, E., Yagi, H., Jerina, D. M. & Conney, A. H. 1993 *Cancer Res.* **53**, 3294–3301.

Yan, S., Shapiro, R., Geacintov, N. E. & Broyde, S. 2001 *J. Am. Chem. Soc.* **123**, 7054–7066.

Yeh, H. J., Sayer, J. M., Liu, X., Altieri, A. S., Byrd, R. A., Lakshman, M. K., Yagi, H., Schurter, E. J., Gorenstein, D. G. & Jerina, D. M. 1995 *Biochemistry* **34**, 13 570–13 581.

Yusupov, M. M., Yusupova, G. Z., Baucom, A., Lieberman, K., Earnest, T. N., Cate, J. H. & Noller, N. F. 2001 *Science* **292**, 883–896.

Zegar, I. S., Kim, S. J., Johansen, T. N., Horton, P. J., Harris, C. M., Harris, T. M. & Stone, M. P. 1996 *Biochemistry* **35**, 6212–6224.

Zhang, Q., Beard, D. A. & Schlick, T. 2003 *J. Comput. Chem.* **24**, 2063–2074.