

Stereochemistry and Position-Dependent Effects of Carcinogens on TATA/TBP Binding

Qing Zhang and Tamar Schlick

Department of Chemistry and Courant Institute of Mathematical Sciences, New York University, New York, New York

ABSTRACT The TATA-box binding protein (TBP) is required by eukaryotic RNA polymerases to bind to the TATA box, an eight-basepair DNA promoter element, to initiate transcription. Carcinogen adducts that bind to the TATA box can hamper this important process. Benzo[a]pyrene (BP) is a representative chemical carcinogen that can be metabolically converted to highly reactive benzo[a]pyrene diol epoxides (BPDE), which in turn can form chemically stereoisomeric BP-DNA adducts. Depending on the TATA-bound adduct's location and stereochemistry, TATA/TBP binding can be decreased or increased. Our previous study interpreted the location-dependent effect in terms of conformational freedom and major-groove space available to BP. Here we further explore specific structural changes of the TATA/TBP complex to help interpret the stereochemical effect in terms of the flexibility of the TATA bases that frame the intercalated adduct. Thermodynamic analyses using molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) yield large standard deviations, which make the computed binding free energies the same within the error bars and point to current limitations of free energy calculations of large and highly charged systems like DNA/protein complexes.

INTRODUCTION

Transcription in eukaryotes involves a large range of RNA polymerases, transcription factors, and other agents (1–3). The TATA-box binding protein (TBP), a component of transcription factor II D, is required by all eukaryotic RNA polymerases for correct initiation of transcription of ribosomal, messenger, small nuclear, and transfer RNAs (4). TBP has a phylogenically conserved 180-residue core segment, which contains two imperfect direct repeats and supports the protein's biological activities. Upon binding to the TATA box—an A/T-rich 8-basepair (bp) DNA in the DNA promoter region (5)—TBP causes local compaction of the TATA box, unwinding, and severe bending, which in turn results in a widened minor groove and a compressed major groove (6) (see Fig. 1). Specifically, TBP's two phenylalanines insert into the first and last base steps of the TATA box, kinking the DNA locally by 52° and 39°, respectively (6). The kinks are accompanied by another pair of phenylalanines, which form extensive van der Waals contacts with the second and seventh basepairs of the TATA box.

Our previous molecular dynamics (MD) study (7) showed that 12 hydrogen bonds form between TBP and the central two basepairs of the TATA box and another seven between TBP and the backbone of the rest of the TATA box residues (including two strong interactions at the two ends). Ingeniously, these hydrogen bonds and extensive hydrophobic interactions stabilize the complex's kinks as well as the TATA/TBP binding (see hydrogen bond contribution in (8)),

therefore allowing the normal Watson-Crick basepairings to be maintained, despite the severely bent TATA DNA.

The TATA/TBP binding and transcription initiation can be affected if the TATA box is damaged by chemical carcinogens. For example, DNA is often affected by benzo[a]pyrene (BP), a chemical carcinogen in the class of polycyclic aromatic hydrocarbons that is a ubiquitous component of environmental pollutants, like automobile exhaust, tobacco smoke, and food contaminants (9–11). BP can be metabolically activated to highly reactive benzo[a]pyrene diol epoxide (BPDE) molecules, which in turn can form BP-DNA adducts (12). If unrepaired by cellular DNA repairing machinery, these adducts can lead to transcription blocking (13,14) or to carcinogenic mutations if bypassed (9,15–19).

BPDE has two enantiomers (Fig. 1): (+)-(7*R*,8*S*,9*S*,10*R*)-7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene and (–)-(7*S*,8*R*,9*R*,10*S*)-7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene. 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 (9,20–22). High-resolution NMR solution studies (23–26) show that BP has conformational preferences in these two adducts: it is classically intercalated between DNA basepairs, instead of being exposed in the solvent-accessible major or minor groove of the DNA duplex (thus, the hydrophobic pyrenyl moiety of BP can be buried within the 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.

An experimental study by Rechkoblit and co-workers (27) has shown that the TATA/TBP binding affinity can be decreased or slightly increased (Table 1) by the BP-adenine adducts in the TATA box, depending on the adduct location (A₁ or A₂) and stereochemistry (10*S*(+) or 10*R*(–)), as

Submitted September 13, 2005, and accepted for publication November 21, 2005.

Address reprint requests to T. Schlick, E-mail: schlick@nyu.edu.

Qing Zhang's present address is Dept. of Molecular Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd., Mail Drop MB-5, La Jolla, CA 92037.

© 2006 by the Biophysical Society

0006-3495/06/03/1865/13 \$2.00

doi: 10.1529/biophysj.105.074344

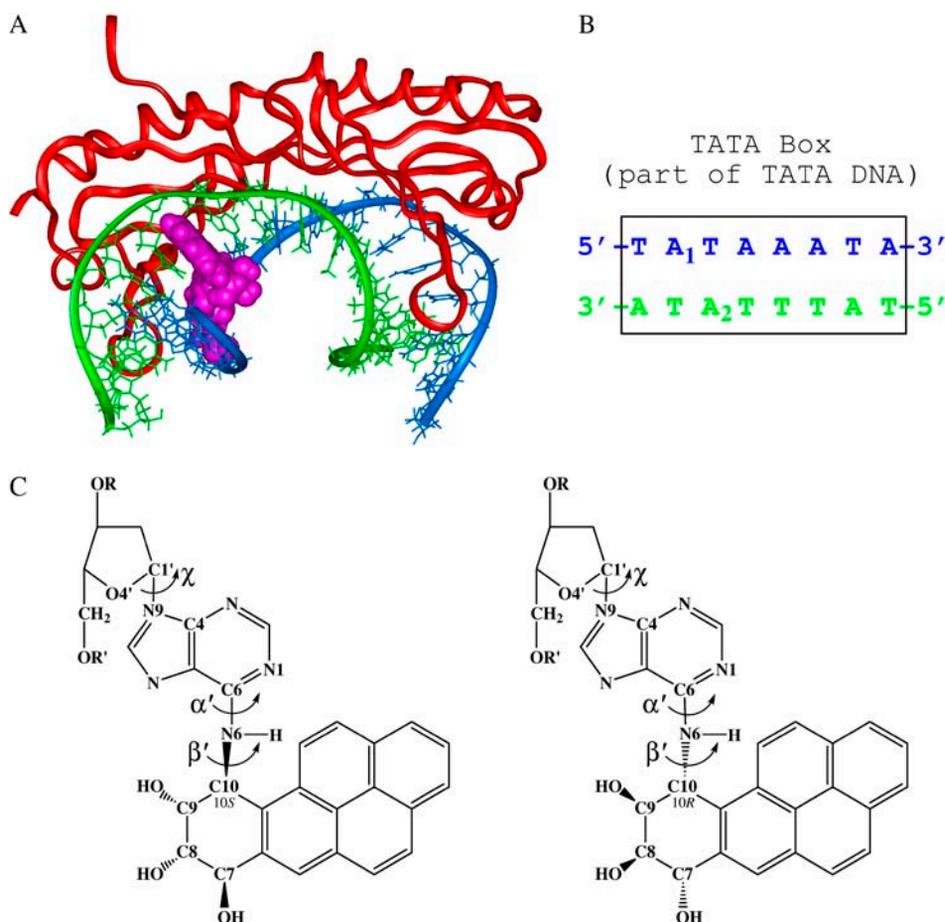


FIGURE 1 TATA/TBP/BP complex. (A) Complex of TBP (red) bound to a BP-modified 16-bp TATA DNA (blue and green) after a 2.32-ns molecular dynamics simulation. The BP-adenine adduct is pink (spacefilling view) and is positioned at site A_1 . (B) TATA box, in the center of the TATA DNA, is shown with the A_1 and A_2 BP-adenine adduct locations. (C) Stereoisomeric BP-adenine adducts: 10S(+)-*trans-anti*-(BP)- N^6 -dA (left) and 10R(-)-*trans-anti*-(BP)- N^6 -dA (right).

shown in Fig. 1. The combination of the adduct location and stereochemistry leads to four systems studied by the experimental work: 10S(+)-*trans*-(A_1), 10R(-)-*trans*-(A_1), 10S(+)-*trans*-(A_2), and 10R(-)-*trans*-(A_2). The first two systems (BP at A_1) were found to increase slightly the TATA/TBP binding affinity, while the other two (BP at A_2) decrease the binding affinity (Table 1). The adduct stereochemistry plays a minor role for changing the binding affinity; the 10R(-) stereoisomer at both locations shows stronger TATA/TBP binding affinity than the 10S(+) stereoisomer.

Our previous MD simulation study (7) interpreted the effect of the adduct location on TATA/TBP binding in terms

TABLE 1 Experimental equilibrium constants K° and binding free energies ΔG° of TBP bound to BP-modified and unmodified TATA DNA duplexes

TATA DNA	K°	ΔG° (kcal/mol)	$\Delta\Delta G^\circ$ (kcal/mol)
Unmodified	1.18×10^8	-10.23	0.00
10S(+)- <i>trans</i> -(A_1)	1.33×10^8	-10.30	-0.07
10R(-)- <i>trans</i> -(A_1)	1.82×10^8	-10.47	-0.24
10S(+)- <i>trans</i> -(A_2)	5.71×10^7	-9.84	0.39
10R(-)- <i>trans</i> -(A_2)	8.00×10^7	-10.02	0.21

K° and ΔG° are derived from Rechkoblit et al. (27). $\Delta\Delta G^\circ$ is the relative binding free energy of a BP-modified system compared to the unmodified one. Experimental temperature is 277.15 K.

of the conformational freedom and major-groove space available to BP due to the hydrogen bonds and phenylalanine insertions at the TATA/TBP interface. That is, we found that BP at A_1 can easily adopt an intercalated orientation between DNA bases or a major-groove orientation in the DNA major groove without much distortion of the TATA or TBP backbone; on the other hand, the A_2 location does not allow BP to easily adopt either orientation without distorting the backbones.

Here we further study the effect of the adduct location on TATA/TBP binding by examining the adduct-dependent structural changes of the TATA/TBP complex. We also interpret the effect of the adduct stereochemistry on TATA/TBP binding. To derive the thermodynamic effects of BP modifications and to correlate the TATA/TBP complex's structural changes with energetics, we perform free energy analyses on all systems.

METHODS

Initial structures

With InsightII (Accelrys, San Diego, CA), we construct the initial TATA/TBP/BP structure by combining the crystal structure of TATA/TBP complex (Protein DataBank ID: 1CDW) (6) and the NMR structure of a BP-modified DNA structure (PDB ID: 1AGU) (26). The TATA DNA has the same length

(16 bp) as that in the 1CDW PDB structure, but the initial structure is remodeled to match the experimental sequence (see below; the TATA box is underlined). Sequence remodeling is conducted on the DNA bases only (not on sugars or backbone) and causes only one contact change between the DNA and TBP, as shown in Appendix S1 in the Supplementary Material. The remodeled structure contains 32 DNA residues (#1–32) and 179 TBP residues (#33–211) with the BP covalently binding to the DNA.

Experimental sequence: 5'-GAATTCGG TATAAATA CGTGTCGTG-3'
 PDB sequence: 5'-CTGC TATAAAG GCTG-3'
 Model sequence: 5'-TCCG TATAAATA CGTG-3'

For each combination of the adduct location (A_1 or A_2) and stereochemistry (10S(+) or 10R(-)), we determine the best orientation of the adduct in its TATA/TBP/BP complex structure based on its four low-energy domains (I, II, III, and IV; see Table 2) (28). These domains, defined by the three torsion angles in Fig. 1, account for the conformational preferences of the two stereoisomeric adducts: BP is classically intercalated and residing on the 3'-side of the modified adenine in the 10S(+) adduct and on the 5'-side in the 10R(-) stereoisomer. The best orientation we find produces the minimal steric collision between the adduct and the TATA/TBP complex.

A total of 11 initial structures of BP-modified TATA/TBP complexes are constructed; an unmodified complex is used as reference. In total, the 12 starting structures are listed in Table 3 (see also Fig. 3 and Table 5 of Zhang et al. (7)) along with their BP-adenine adduct's domains (I, II, III, or IV) and BP's orientations (intercalated between DNA bases or exposed to DNA major groove). The initial structure of free TBP is then taken from the PDB structure of the TBP-DNA complex, by removing the DNA, and then removing the crystal water molecules that are more than 10 Å away from TBP. The remaining water molecules are retained as they may be important for TBP. The initial structure of each free TATA DNA is taken from its corresponding TATA/TBP complex structure by removing TBP, counterions, and water molecules after the complex has been minimized with a water box and counterions (see Energy Minimization and MD Simulation, below). Therefore, totally 25 systems are simulated, including 12 TATA/TBP complexes, 12 free TATA DNA, and one free TBP (Table 3), as required for the thermodynamic analyses.

Force field

The force-field parameters of the TATA DNA and the TBP protein are assigned using the Cornell et al. (29) force field with *parm98* parameter set (30). The missing partial charges and three bond angles of the BP-adenine adducts are computed using our partial charge protocol (see Appendix), based on the RESP approach (31–33) in AMBER (34), and according to chemically similar bond angles in the *parm98*; these computed force-field parameters are listed in Appendix S2 in the Supplementary Material.

Energy minimization and MD simulation

Energy minimization and MD simulations were performed in AMBER 6 (34). Energy minimization eliminates unfavorable van der Waals contacts in the initial structures. Each system is then solvated with a rectangular box of TIP3 water molecules (35) and counterions (neutralizing the system and

TABLE 3 Simulated systems and simulation lengths

12 TATA/TBP complexes (2.32 ns each)	12 free TATA DNA (3.34 ns each)	1 free TBP (2.32 ns)
Unmodified	Unmodified	
10S(+)- <i>trans</i> -(A ₁): II (int*)	10S(+)- <i>trans</i> -(A ₁): II	
10S(+)- <i>trans</i> -(A ₁): III (MG [†])	10S(+)- <i>trans</i> -(A ₁): III	
10R(-)- <i>trans</i> -(A ₁): II (MG [†])	10R(-)- <i>trans</i> -(A ₁): II	
10R(-)- <i>trans</i> -(A ₁): III (int*)	10R(-)- <i>trans</i> -(A ₁): III	
10R(-)- <i>trans</i> -(A ₁): IV (MG [†])	10R(-)- <i>trans</i> -(A ₁): IV	Free TBP
10S(+)- <i>trans</i> -(A ₂): III (int*)	10S(+)- <i>trans</i> -(A ₂): III	
10S(+)- <i>trans</i> -(A ₂): III-2 [‡] (MG [†])	10S(+)- <i>trans</i> -(A ₂): III-2	
10S(+)- <i>trans</i> -(A ₂): IV (int*)	10S(+)- <i>trans</i> -(A ₂): IV	
10R(-)- <i>trans</i> -(A ₂): I (MG [†])	10R(-)- <i>trans</i> -(A ₂): I	
10R(-)- <i>trans</i> -(A ₂): III (int*)	10R(-)- <i>trans</i> -(A ₂): III	
10R(-)- <i>trans</i> -(A ₂): IV (MG [†])	10R(-)- <i>trans</i> -(A ₂): IV	

*BP is initially intercalated between DNA bases.

[†]BP is initially exposed to DNA major groove.

[‡]The second possible conformation of this adduct in domain III.

reproducing the experimental ionic strength). Energy minimization is performed again to eliminate van der Waals collisions due to the addition of water molecules and counterions. MD simulations then follow with four equilibration stages (heating, constant pressure I, constant pressure II, final equilibration) and one production dynamics stage (7). Equilibration focuses on the solvent while production dynamics collects structural changes of the solute. The simulations employ the leap-frog Verlet algorithm (36) with a time step of 2 fs. Electrostatic interactions are computed using the particle-mesh Ewald method (37,38).

For each of the 12 TATA/TBP systems, energy minimization on the initial structure is performed with 400 steps of steepest descent (SD) followed by 400 steps of conjugate gradient (CG). The system is then solvated with a rectangular box of TIP3 water molecules that extends 10.0 Å from the solute and contains 37 Na⁺ and 23 Cl⁻ for neutralizing the system and mimicking the experimental ionic strength of 130 mM. To relax the added solvent, the entire system is minimized again using 50 steps of SD followed by 5000 steps of CG. The four equilibration stages take 460,000 steps (920 ps) and the production dynamics stage takes 700,000 steps (1400 ps); the total length of the trajectory is then 2.32 ns. More details of the energy minimization and MD simulation of the 12 TATA/TBP systems are described in Zhang et al. (7).

For the free TBP, a water box is added with a buffer size of (12.0, 10.0, 10.0) Å. The buffer size in the X direction is larger than those in the Y and Z directions because TBP may expand in this direction without bound DNA. Then 14 Na⁺ and 30 Cl⁻ are added to neutralize the system and satisfy the experimental ionic strength (130 mM). The free TBP system undergoes the same procedure of minimizations and dynamics as those TATA/TBP systems (7) and has a trajectory length of 2.32 ns too.

For the 12 corresponding free TATA DNA systems, each DNA is taken from its corresponding TATA/TBP complex after the complex is minimized with a water box and counterions. The TBP, counterions, and all water molecules are then removed. Each DNA is minimized using 400 steps of SD followed by 400 steps of CG with a 1.0 kcal/mol restraint on the DNA's non-end atoms. A water box is added with a buffer size of (10.0, 15.0, and 10.0) Å.

TABLE 2 Low-energy conformational domains of 10S(+) and 10R(-)-*trans-anti*-(BP)-N⁶-dA adducts

Domain	χ (°) (O4'-C1'-N9-C4)	α' (°) (N1-C6-N6-C10(BP))	β' (°) (C6-N6-C10(BP)-C9(BP))	
			(+) 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

The larger buffer in the Y direction allows for the possibility that the bent DNA may become straight during dynamics. Thirty Na^+ are then added to neutralize the system. Minimization follows using 50 steps of SD and 5000 steps of CG with a 50.0 kcal/mol restraint only on DNA. The 12 DNA systems undergo fewer steps of equilibration stages because these systems have less solvent than those of the TATA/TBP complex systems. The numbers of equilibration steps are: 10,000 (heating), 20,000 (constant pressure simulation I), 20,000 (constant pressure simulation II), and 20,000 (final equilibration). More steps of production dynamics (1,600,000) were used to allow the bent DNA to become stable. The length of each TATA DNA trajectory is 3.34 ns.

Molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) free energy analyses

Molecular mechanics Poisson-Boltzmann surface area (MM-PBSA) (39–42) is a free energy evaluation methodology that has been employed in a variety of applications (see Discussion). The MM-PBSA free energy is estimated from molecular mechanical energy E_{MM} , solvation free energy G_{SOL} , and vibrational, rotational, and translational entropies S ,

$$G = E_{\text{MM}} + G_{\text{SOL}} - TS, \quad (1)$$

$$E_{\text{MM}} = E_{\text{int}} + E_{\text{vdw}} + E_{\text{ele}}, \quad (2)$$

$$G_{\text{SOL}} = G_{\text{pb}} + G_{\text{np}}, \quad (3)$$

where T is the temperature; E_{int} is internal energy, $E_{\text{int}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}}$, the sum of bond, angle, and dihedral energies; E_{vdw} is van der Waals energy; E_{ele} is electrostatic energy; G_{pb} is the electrostatic solvation free energy, computed from the finite difference Poisson-Boltzmann method in the DelPhi program (43,44); G_{np} is nonpolar solvation free energy, $G_{\text{np}} = \gamma\text{SASA} + b$, $\gamma = 0.00542 \text{ kcal}/\text{\AA}^2$, $b = 0.92 \text{ kcal/mol}$, and SASA is the solvent-accessible surface area, which is estimated using Sanner's algorithm implemented in the MSMS software (45); and the solute entropy S is approximated with normal mode analysis (46) using the *Nmode* module in AMBER 7 (47). The MM-PBSA free energy computations are conducted in AMBER 7 (47).

Binding free energies can be computed from triplet-trajectory analysis or single-trajectory analysis. Triplet-trajectory analysis involves computing free energies for the complex (TATA/TBP) as well as for the free ligand (TATA DNA) and the free receptor (TBP) after they are simulated, respectively (three trajectories); the respective binding free energy is the free energy of the complex minus the free energies of the free ligand and the free receptor. This method is theoretically accurate but computationally expensive. Single- rather than triplet-trajectory analysis only uses the trajectory of the complex; the binding free energy is thus expressed as the free energy of the complex minus the free energies of the bound ligand and the bound receptor in the complex. This method is faster but does not consider possible ligand and receptor conformational changes going from their free states to the bound states. We employ both methods for proper assessment. In addition, we employ a computationally inexpensive analysis method by calculating only the local molecular mechanics interaction energies between the TATA DNA and TBP residues that are within 5 \AA from the carcinogen.

Triplet-trajectory analysis

For each TATA/TBP complex and for the free TBP, the last 600-ps trajectory of the production dynamics stage is used for free energy analyses: 60 snapshots at a 10-ps interval for computation of enthalpy ($H = E_{\text{MM}} + G_{\text{SOL}}$) and 10 snapshots at 60-ps interval for computation of entropy. For each free TATA DNA, the last 1600-ps trajectory of the production dynamics stage is used for free energy analyses: 80 snapshots at a 20-ps interval for computation of enthalpy and 10 snapshots at a 160-ps interval for computation of entropy. The normal mode entropy analyses are estimates

and computationally expensive, so only 10 snapshots are selected for the analyses. (With 50 snapshots for the unmodified DNA system, we obtained the nearly identical entropy as in the 10-snapshot analysis; therefore we use the more computationally feasible 10-snapshot entropy analysis for all systems.)

Single-trajectory analysis

For each TATA/TBP complex and its bound TATA DNA and bound TBP, the last 600-ps trajectory of the production dynamics stage of the complex is used for free energy analyses: 60 snapshots at a 10-ps interval for computation of enthalpy and 10 snapshots at 60-ps interval for computation of entropy.

Temperature T is set to 277.15 K to mimic the experimental temperature (27). For calculating E_{MM} , the dielectric constant is set to 1.0. For calculating G_{SOL} , the salt concentration is set to 0.13 M to mimic the experimental ionic strength (27); the internal dielectric constant is set to 1.0 and external to 80.0. For calculating entropy, each structure is minimized using the conjugate gradient method with a distance-dependent dielectric constant ($\epsilon = 4R$, where R is the interatomic distance) until the root-mean-square of the elements in the gradient vector was $<10^{-4} \text{ kcal/mol}/\text{\AA}$. Its entropy is then computed and averaged on all the 10 minimized structures.

RESULTS

Structural changes of TATA/TBP complex due to BP modifications

In our previous study (7), we ranked the structural distortions of the TATA DNA backbones and the TBP backbones of the BP-modified systems compared to the unmodified one as follows: minimal distortions for the $10R(-)$ stereoisomer at A_1 , moderate distortion for the $10S(+)$ stereoisomer at A_1 , and greatest distortions for the two stereoisomers at A_2 . We also interpreted this adduct-location-dependent effect in terms of conformational freedom and major-groove space available to BP due to the hydrogen bonds and phenylalanine insertions at the TATA/TBP interface: at A_2 , BP cannot easily intercalate between the DNA bases and has a smaller space in the DNA major groove (7).

To examine the specific BP-induced structural changes of the TATA/TBP complex, we visualize these changes by superimposing each BP-modified complex on the unmodified one in Fig. 2. The structural changes are summarized in Table 4. These structural changes are consistent with our RMSD analyses (7): BP at A_2 distorts the TATA/TBP complex, while BP at A_1 nearly conserves the complex, and agree with the experimental finding (27) that BP at A_2 decreases the TATA/TBP binding affinity while BP at A_1 slightly increases the binding affinity. The slight increase of binding affinity by BP at A_1 suggests that the ease of intercalation by BP at A_1 increases the stacking energy in the severely bent TATA DNA and stabilizes the overall binding between TATA and TBP. (Although we remodeled the DNA sequence and our systems had limited simulation time, BP has been stabilized early in simulations in all BP-modified systems; in Appendix S3 in the Supplementary Material, BP's three torsion angles (χ , α' , β') as a function of

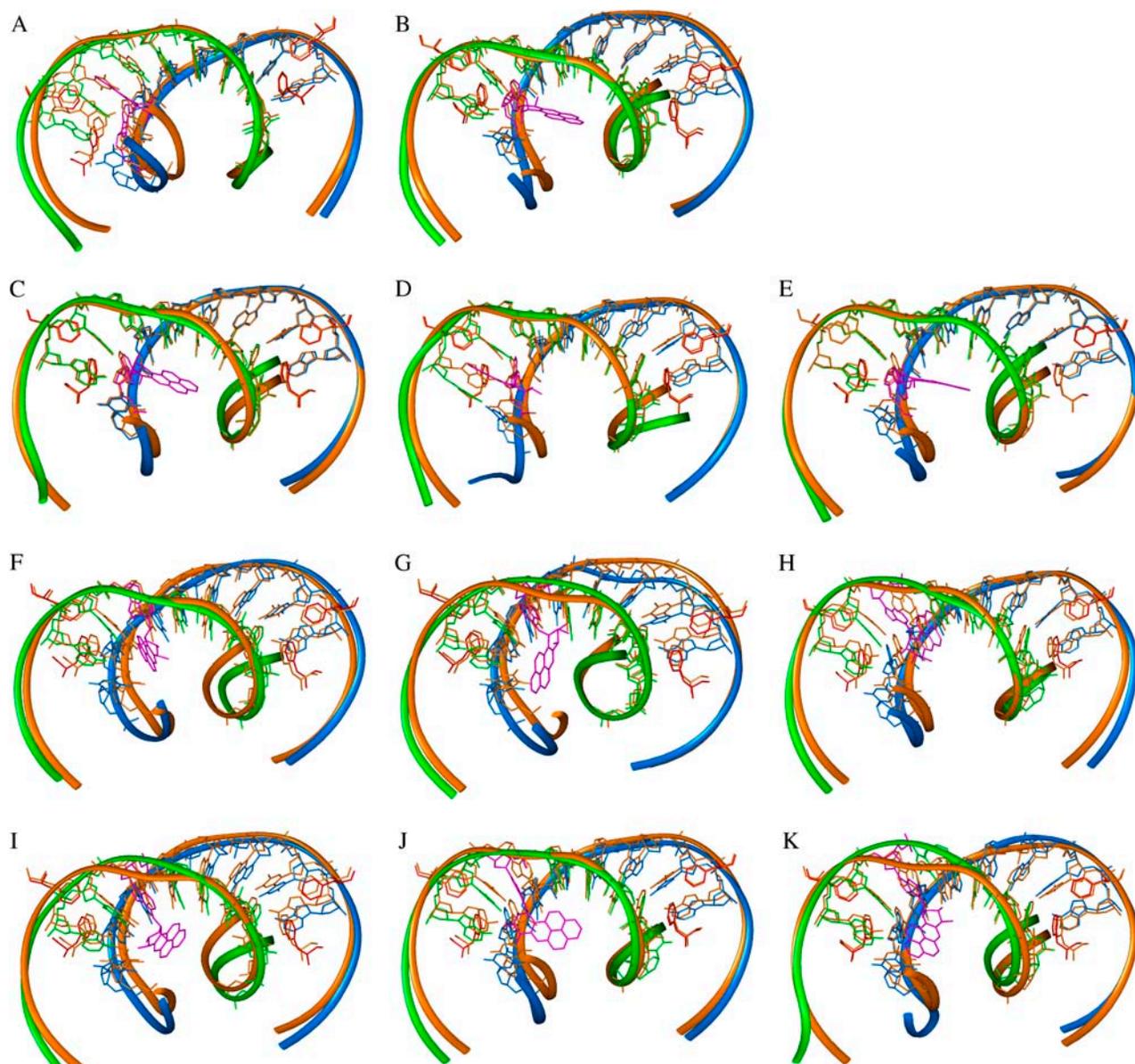


FIGURE 2 Superimposition of BP-modified and unmodified TATA/TBP complexes. Both structures are averaged over the last 600 ps of production dynamics. The DNA strands are blue and green in the BP-modified complex and orange in the unmodified complex; the BP-adenine adduct is pink; TBP residues are not shown except four phenylalanines; the four phenylalanines are red in the BP-modified complex and orange in the unmodified complex; hydrogens are not shown. The two phenylalanines (Phe-162 at the *second left* and Phe-71 at the *second right*) insert and kink the DNA at A6 (A_1) and A22 (A_2); the other two phenylalanines reside over T11 and A27 and are stabilized by van der Waals interactions. (A) $10S(+)-trans-(A_1)$: II versus unmodified. (B) $10S(+)-trans-(A_1)$: III versus unmodified. (C) $10R(-)-trans-(A_1)$: II versus unmodified. (D) $10R(-)-trans-(A_1)$: III versus unmodified. (E) $10R(-)-trans-(A_1)$: IV versus unmodified. (F) $10S(+)-trans-(A_2)$: III versus unmodified. (G) $10S(+)-trans-(A_2)$: III-2 versus unmodified. (H) $10S(+)-trans-(A_2)$: IV versus unmodified. (I) $10R(-)-trans-(A_2)$: I versus unmodified. (J) $10R(-)-trans-(A_2)$: III versus unmodified. (K) $10R(-)-trans-(A_2)$: IV versus unmodified.

simulation time are plotted for the first BP-modified system $10S(+)-trans-(A_1)$: II.)

Stereochemical effect of BP at A_1 on TATA/TBP binding

The experimental work (27) showed that the TATA/TBP binding affinity changes are also dependent on the stereochemistry of the BP-adenine adducts (Table 1). Of the two

adducts, the $10R(-)$ stereoisomer gives more favorable binding affinity changes than the $10S(+)$ stereoisomer at both A_1 and A_2 . However, as shown here, it is difficult for BP at A_2 to adopt either the intercalated orientation or the major-groove orientation. Thus, we interpret here the stereochemistry-dependent effect of BP at A_1 on TATA/TBP binding. BP at A_1 can adopt both the intercalated orientation and major-groove orientation. In the former arrangement, the hydrophobic pyrenyl moiety of BP is buried between DNA

TABLE 4 Structural changes of TATA/TBP complex due to BP modifications

System	Fig.	Phenylalanines* of TBP	TATA DNA	BP
10S(+)- <i>trans</i> -(A ₁): II	2 A	Phe-162 [†] is pushed back to TBP by 1.65 Å; the rest are stable.	Backbone in the TATA region is pushed toward TBP; the remaining part remains stable.	Hydroxyls are in the solvent-accessible DNA major groove; hydrophobic aromatic rings are intercalated between DNA bases.
	III 2 B	All stable.	Backbone in the TATA region is pushed slightly toward TBP; the remaining part remains stable.	Exposed to the solvent-accessible DNA major groove.
10R(-)- <i>trans</i> -(A ₁): II	2 C	All stable.	Backbone is reserved; all TATA residues (except A ₁) are positionally identical with those from the unmodified system.	Exposed to the solvent-accessible DNA major groove.
	III 2 D	All stable.	Backbone in the TATA region is pushed slightly toward TBP; the remaining part remains stable.	Hydroxyls are in the solvent-accessible DNA major groove; hydrophobic aromatic rings are intercalated between DNA bases.
	IV 2 E	All stable.	Backbone is perfectly conserved; all TATA residues (except A ₁) are positionally similar to those from the unmodified system.	Exposed to the solvent-accessible DNA major groove.
10S(+)- <i>trans</i> -(A ₂): III	2 F	All pushed back to TBP.	Backbone is distorted.	Hydroxyls are in the solvent-accessible DNA major groove; hydrophobic aromatic rings are partly intercalated between DNA bases.
	III-2 2 G	All moved slightly.	Backbone is severely distorted.	Hydroxyls are in contact with the central TATA box; hydrophobic aromatic rings are in the DNA major groove, but are close to the beginning of the TATA box.
	IV 2 H	All pushed back to TBP.	Backbone around A ₂ is distorted; backbone in the TATA region is pushed slightly toward TBP.	Hydroxyls are in the solvent-accessible DNA major groove; hydrophobic aromatic rings are distorted.
10R(-)- <i>trans</i> -(A ₂): I	2 I	Phe-162 is pushed back to TBP by 1.88 Å; the rest are moved slightly.	Backbone is distorted; A ₂ is largely stretched to allow bulky BP to stay in DNA major groove.	Exposed to the solvent-accessible DNA major groove.
	III 2 J	Phe-162 is pushed back to TBP by 1.32 Å; the rest are slightly pushed back toward TBP.	Backbone is slightly distorted; thymine (downstream of A ₁) loses its Watson-Crick hydrogen bonds.	Exposed to the solvent-accessible DNA major groove; the only system in which BP's orientation changes during MD: from intercalated to major-groove form.
	IV 2 K	All moved slightly.	Backbone is distorted, especially at the A ₂ residue; thymine (downstream of A ₁) loses its Watson-Crick hydrogen bonds; glycosidic torsion χ of the thymine (upstream of A ₂) is rotated by 23.7°.	Exposed to the solvent-accessible DNA major groove.

*Referring to the four phenylalanines that interact with the TATA DNA and are shown in red in Fig. 2.

[†]Referring to the phenylalanine that inserts the first basepair step of the TATA box.

bases and forms favorable stacking interactions between the aromatic moiety and DNA bases; in the latter orientation, the hydrophobic moiety will be exposed to the hydrophilic solvent-accessible major groove. Thus the intercalated orientation is favored and represents the major population in the BP-modified TATA/TBP complex solution. This is also consistent with the experimental result that BP at A₁ slightly increases TATA/TBP binding (due to the favorable stacking interaction in the intercalated orientation).

High-resolution NMR solution studies (23,24,26) show that intercalated BP resides on the 3'-side of the modified adenine in the 10S(+) adduct and on the 5'-side in the

10R(-) adduct. This agrees with our simulations of the two systems with BP at A₁ adopting the intercalated orientation: 10S(+)-*trans*-(A₁): II (Fig. 2 A) and 10R(-)-*trans*-(A₁): III (Fig. 2 D).

We diagram the positions of the two stereoisomeric adducts in Fig. 3; the 10S(+) BP intercalates between DNA residues T27 and A26 (A₂), and the 10R(-) BP between A28 and T27. We also diagram the major TATA/TBP interactions in the figure for interpreting the stereochemistry-dependent effect. The major interactions between TBP and the TATA box include hydrogen bonds and phenylalanine insertions computed from the last 600-ps production dynamics simulation

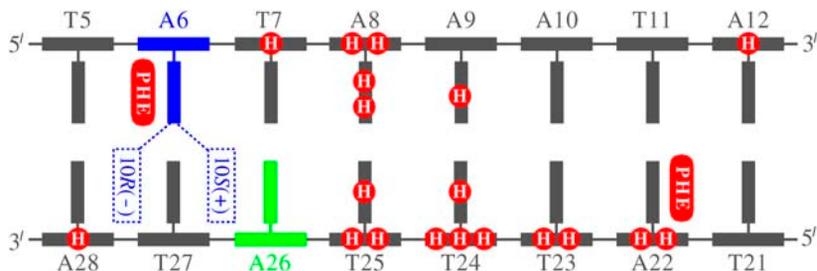


FIGURE 3 Diagram of major TATA/TBP interactions and possible intercalation orientations of BP at A_1 (A6). The TATA box residues are shown in gray, with A6 (A_1) and A26 (A_2) highlighted in blue and green, respectively. Each DNA residue is represented by one horizontal bar as backbone and one vertical bar as base. The major TATA/TBP interactions include hydrogen bonds (white H letters on red background) and phenylalanine insertions (white PHE letters on red background). Possible intercalation orientations of BP at A_1 (A6) are shown with dashed boxes, and the stereochemistry of the BP-adenine adducts (10S(+), 10R(-)) are labeled in the dashed boxes. The relatively weak interactions (hydrophobic interactions and water-mediated hydrogen bonds) are not diagrammed.

of the unmodified TATA/TBP complex (7). The two phenylalanines kink the DNA at A6 (A_1) and A22.

For the 10S(+) BP, the bulky aromatic rings intercalating between the bases of T27 and A26 (A_2) require more space between these two bases. Thus the bases of T27 and A26 are pushed to their 3' side and 5' side, respectively. T27's base lacks hydrogen bonds with TBP or with its 3'-side neighbor A28's base; thus T27's base has the flexibility required to be pushed to its 3' side. A26's base has no hydrogen bonds with TBP either but its 5'-side neighbor T25's base has one, making the pushing of A26's base to its 5' side relatively difficult. For the 10R(-) BP between A28 and T27, the intercalation is relatively easy. A28's base has no hydrogen bonds with TBP and neither do its 3'-side neighbors (outside the 8-bp TATA box; not diagrammed); thus A28's base can be pushed to its 3' side. T27's base also has some flexibility to be pushed to its 5' side because T27's base or its 5'-side neighbor (A26's base) has no hydrogen bonds with TBP.

Thus, we have interpreted the stereochemistry-dependent of BP at A_1 in terms of the flexibility of the TATA bases that frame the intercalated adduct due to the hydrogen bonds between the TATA box and TBP. It is relatively more easy for the 10R(-) BP to adopt the intercalated orientation than the 10S(+) BP, as found experimentally (Table 1). Note that the above analysis does not consider the Watson-Crick hydrogen bonds, as each basepair in the TATA box maintains two. Although the binding affinity difference due to the stereochemistry is very small (-10.30 vs. -10.47 kcal/mol; see the binding free energies in Table 1), our simulations together with structural analyses have helped interpret the small stereochemistry-dependent effect of BP modification at A_1 on TATA/TBP binding.

Binding free energy analyses

The experimental binding free energies are derived from the experimental equilibrium constants (27) and listed in Table 1. The derivation procedure is described in Appendix S4 in the Supplementary Material. We compute the theoretical binding affinities from our MD trajectories mainly using the

following two methods: triplet-trajectory analysis and single-trajectory analysis. The other method (local molecular mechanics interaction analysis) yields large variations in the interaction energies (Appendix S5 in the Supplementary Material) and is thus unable to interpret the global binding affinity differences shown in Table 1.

Triplet-trajectory analysis

Analyses based on separate trajectories are theoretically more accurate for deriving binding free energies, since the conformational changes of the TATA/DNA and TBP upon binding are considered. Although the free TBP and DNA are initially taken from the complexes (see Methods), their free energies have stabilized during the simulation time in which the free energies are computed, as shown in Appendix S6 in the Supplementary Material. Table 5 shows the free energies $G(\text{TATA/TBP})$ of TATA/TBP complexes, free energies

TABLE 5 Binding free energies from triplet-trajectory analysis

System	$G(\text{TATA/TBP})$	$G(\text{TATA})$	ΔG	$\Delta\Delta G$
Unmodified	-15176 ± 44	-7008 ± 23	-34 ± 58	0
10S(+)-trans-(A_1): II	-15086 ± 41	-6942 ± 20	-10 ± 55	+24
III	-15108 ± 32	-6952 ± 21	-22 ± 49	+12
10R(-)-trans-(A_1): II	-15087 ± 40	-6942 ± 20	-11 ± 54	+23
III	-15130 ± 40	-6957 ± 19	-39 ± 54	-5
IV	-15069 ± 41	-6929 ± 23	-6 ± 56	+28
10S(+)-trans-(A_2): III	-15143 ± 34	-6971 ± 20	-38 ± 50	-4
III-2	-15123 ± 38	-6964 ± 22	-25 ± 53	+9
IV	-15103 ± 45	-6979 ± 21	10 ± 58	+44
10R(-)-trans-(A_2): I	-15101 ± 40	-6930 ± 23	-37 ± 55	-3
III	-15131 ± 44	-6953 ± 22	-44 ± 58	-10
IV	-15110 ± 42	-6956 ± 19	-20 ± 55	14

$G(\text{TATA/TBP})$ is the free energy of a TATA/TBP complex, $G(\text{TATA})$ is the free energy of corresponding free TATA DNA, ΔG is the binding free energy between TBP and TATA DNA, and $\Delta\Delta G$ is the relative binding free energy of a BP-modified TATA/TBP complex compared to the unmodified one. The free energy of free TBP $G(\text{TBP})$ is -8134 ± 30 kcal/mol. Underlined are the lowest energy in each of the four combinations of carcinogen location (A_1 or A_2) and stereochemistry (10S(+) or 10R(-)). All energies have the unit of kcal/mol and are shown in both average and standard deviation (except $\Delta\Delta G$).

$G(\text{TATA})$ of corresponding free TATA DNA, free energy $G(\text{TBP})$ of free TBP, binding free energies ΔG between TATA DNA and TBP, and relative binding free energies $\Delta\Delta G$ of the BP-modified TATA/TBP complexes compared to the unmodified complex. We see that the ΔG values are dominated by $G(\text{TATA/TBP})$ and have relative binding free energies $\Delta\Delta G$ ranging from -10 to 44 kcal/mol. The $\Delta\Delta G$ values are within the standard deviations of ΔG (49 – 58 kcal/mol), which means all systems have nearly the same binding free energy. This agrees with the experimental binding free energies that have less than 1 kcal/mol differences (Table 1). However, the computed standard deviations are also too large to distinguish the binding free energies from one another.

Single-trajectory analysis

Table 6 provides the results from the single-trajectory analysis. The molecular mechanics energy changes ΔE_{MM} are dominated by the electrostatic energy changes ΔE_{ele} , and the solvation free energy changes ΔG_{SOL} are dominated by the electrostatic solvation free energy changes ΔG_{pb} . The binding free energies ΔG are not dominated by any one term, and their standard deviations (12 – 18 kcal/mol) are much smaller than those from the triplet-trajectory analysis (49 – 58 kcal/mol; Table 5). Most $\Delta\Delta G$ values are within the standard deviations of ΔG . If we consider only the systems that have the lowest relative binding free energy (underlined in the last column of Table 6) in each of the four combinations of adduct location (A_1 or A_2) and stereochemistry ($10S(+)$ or $10R(-)$), their relative binding free energies are much closer (7 , 7 , 5 , and 10 kcal/mol) and well within their standard deviations. This small spread in magnitude is consistent

with the experimental binding free energy differences (<1 kcal/mol; Table 1).

Therefore, both the triplet-trajectory analysis and the single-trajectory analysis reproduce the experimental relative binding free energies within the computed standard deviations. However, these standard deviations are large especially in the (theoretically more accurate) triplet-trajectory analysis. We analyze this problem in details in Discussion.

DISCUSSION

Geacintov and co-workers (27) have shown that TATA/TBP binding affinity is altered with a bulky BP present in the TATA box and that the binding affinity decreases or slightly increases depending on the location and stereochemistry of the BP-adenine adduct. Although the changes of binding affinity are <1.0 kcal/mol, our MD studies help interpret such small changes in terms of conformational freedom and major-groove space available to BP due to the hydrogen bonds and phenylalanine insertions at the TATA/TBP interface. That is, our structural and flexibility analyses revealed that BP at A_2 has difficulty to adopt either intercalated or major-groove orientation without distorting the TATA and TBP backbones. For BP at A_1 , both orientations are easily adopted. Furthermore, the $10R(-)$ stereoisomer at A_1 can more easily adopt the intercalated orientation due to the larger flexibility of the two DNA bases, between which the BP is intercalated (BP in the $10R(-)$ stereoisomer thus has larger conformational freedom). Thus, MD simulations, together with structural analyses, help relate atomic-level structure/flexibility effects to global binding effects.

MM-PBSA has been applied to studying binding free energies between protein and small ligand (48 – 57), protein

TABLE 6 Binding free energies from single-trajectory analysis

System	ΔE_{ele}	ΔE_{vdw}	ΔE_{MM}	ΔG_{np}	ΔG_{pb}	ΔG_{SOL}	ΔH	$-T\Delta S$	ΔG	$\Delta\Delta G$
Unmodified	-6721 ± 108	-199 ± 6	-6920 ± 107	-11 ± 0.4	6793 ± 98	6782 ± 98	-138 ± 11	67 ± 13	-71 ± 17	<u>0</u>
$10S(+)$ - <i>trans</i> -(A_1): II	-6540 ± 79	-197 ± 6	-6737 ± 80	-11 ± 0.3	6609 ± 83	6598 ± 83	-139 ± 17	75 ± 6	-64 ± 18	<u>7</u>
III	-6472 ± 78	-201 ± 7	-6672 ± 81	-11 ± 0.3	6554 ± 74	6543 ± 74	-129 ± 10	77 ± 14	-52 ± 18	<u>19</u>
$10R(-)$ - <i>trans</i> -(A_1): II	-6579 ± 62	-199 ± 5	-6777 ± 62	-11 ± 0.3	6656 ± 56	6645 ± 56	-132 ± 9	78 ± 9	-54 ± 13	<u>17</u>
III	-6508 ± 79	-205 ± 6	-6713 ± 82	-11 ± 0.3	6591 ± 76	6580 ± 76	-133 ± 8	69 ± 10	-64 ± 18	<u>7</u>
IV	-6599 ± 74	-198 ± 7	-6797 ± 74	-11 ± 0.3	6676 ± 68	6665 ± 68	-132 ± 9	70 ± 8	-62 ± 12	<u>9</u>
$10S(+)$ - <i>trans</i> -(A_2): III	-6566 ± 128	-198 ± 7	-6764 ± 131	-11 ± 0.3	6642 ± 119	6631 ± 119	-133 ± 14	70 ± 6	-63 ± 15	<u>8</u>
III-2	-6471 ± 152	-202 ± 7	-6673 ± 154	-11 ± 0.4	6551 ± 138	6540 ± 138	-133 ± 17	67 ± 5	-66 ± 18	<u>5</u>
IV	-6589 ± 76	-201 ± 7	-6790 ± 78	-11 ± 0.4	6670 ± 71	6659 ± 71	-131 ± 10	76 ± 10	-55 ± 14	<u>16</u>
$10R(-)$ - <i>trans</i> -(A_2): I	-6534 ± 90	-197 ± 5	-6731 ± 90	-11 ± 0.2	6607 ± 83	6596 ± 83	-135 ± 10	74 ± 7	-61 ± 12	<u>10</u>
III	-6656 ± 62	-194 ± 7	-6850 ± 61	-11 ± 0.4	6733 ± 57	6712 ± 57	-138 ± 8	83 ± 11	-55 ± 14	<u>16</u>
IV	-6588 ± 75	-209 ± 6	-6797 ± 77	-12 ± 0.4	6679 ± 70	6667 ± 70	-130 ± 10	75 ± 15	-55 ± 18	<u>16</u>

ΔE_{ele} , ΔE_{vdw} , and ΔE_{int} (zeros; not shown) are electrostatic energy, van der Waals energy, and internal energy changes. ΔE_{MM} is the molecular mechanics energy change and $\Delta E_{\text{MM}} = \Delta E_{\text{ele}} + \Delta E_{\text{vdw}} + \Delta E_{\text{int}}$. G_{np} and G_{pb} are nonpolar and electrostatic solvation free energy changes. ΔG_{SOL} is solvation free energy change and $\Delta G_{\text{SOL}} = \Delta G_{\text{np}} + \Delta G_{\text{pb}}$. ΔH is enthalpy change and $\Delta H = \Delta E_{\text{MM}} + \Delta G_{\text{SOL}}$. T is the absolute temperature and ΔS is the entropy change. ΔG is the binding free energy and $\Delta G = \Delta E_{\text{MM}} + \Delta G_{\text{SOL}} - T\Delta S$. $\Delta\Delta G$ is the relative binding free energy of a BP-modified system compared to the unmodified one. Underlined are the lowest energy in each of the four combinations of carcinogen location (A_1 or A_2) and stereochemistry ($10S(+)$ or $10R(-)$). All energies have the unit of kcal/mol and are shown in both average and standard deviation (except $\Delta\Delta G$). Although $\Delta H = \Delta E_{\text{MM}} + \Delta G_{\text{SOL}}$, the large standard deviations of ΔE_{MM} and ΔG_{SOL} do not lead to large standard deviations of ΔH , which means fluctuations of these two terms are mostly canceled out.

and protein (58–62), protein and peptide (63,64), small molecule and small molecule (65,66), RNA and peptide (67), RNA and small ligand (68), DNA and small ligand (69), and protein and RNA (70). To the best of our knowledge, our case is the first MM-PBSA binding free energy study between protein and DNA. Though our free energy calculations reproduce the experimental relative binding free energies within the computed standard deviations, these standard deviations are large, especially in the theoretically more accurate triplet-trajectory analysis, which considers the conformational changes of the ligand (TATA DNA) and the receptor (TBP) upon binding. Wang and co-workers underscore several sources of error in MM-PBSA: the force field, MD sampling, solvation free energy estimate of PBSA, and the entropy estimated by normal mode analyses (54). They also suggest that the MM-PBSA approach may not work as well in calculating the absolute binding free energy for charged ligands as for neutral systems (54). In the following, we focus on discussing the two error sources—MD sampling and electrostatic solvation free energy.

Although experimental results are condition-dependent, they are averaged in time and space for many molecular systems. In contrast, a MD trajectory only mimics behavior of one molecule within a limited time span in the large thermally accessible conformation space. This causes large standard deviations in the free energy evaluation. Fig. 4 shows relative TATA/TBP electrostatic interaction energies of four BP-modified systems compared to the unmodified one. The wide

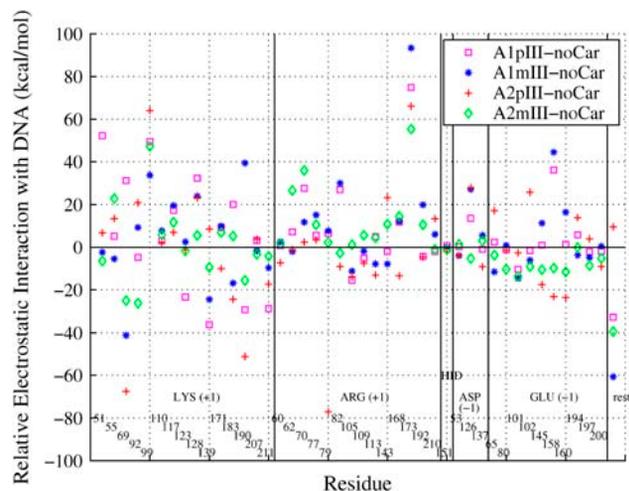


FIGURE 4 Relative electrostatic interaction energies of TBP residues with TATA DNA. Each spot indicates the relative TBP residue/DNA electrostatic interaction energy of a carcinogen-modified system compared to the unmodified one (noCar). The four carcinogen-modified systems are 10S(+)-*trans*-(A₁): III (A1pIII); 10R(-)-*trans*-(A₁): III (A1mIII); 10S(+)-*trans*-(A₂): III (A2pIII); and 10R(-)-*trans*-(A₂): III (A2mIII). The energies are averaged over the last 600 ps of production dynamics. The TBP residues are listed in six groups separated by vertical lines: positively charged (+1 *e*) lysine (LYS); positively charged (+1 *e*) arginine (ARG); neutral histidine (HID); negatively charged (-1 *e*) aspartic acid (ASP); negatively charged (-1 *e*) glutamic acid (GLU); and the rest are residues. All the groups, except the last one, are also provided with residue numbers.

variations of interactions between the TBP residues and the TATA/DNA indicate that there are large random movements by the flexible TBP residues. Large fluctuations in binding free energies are also reported in a recent protein-ligand association free-energy study by Swanson and co-workers (53). These fluctuations explain why freezing the atoms outside binding regions yields good binding free energies for large systems (54,56). Without freezing, two trajectories for the same molecule may produce two quite different binding free energies. For example, one trajectory of our 10S(+)-*trans*-(A₂): III complex produces a -63 kcal/mol binding free energy and another gives -75 kcal/mol, although they are within the standard deviations of 15 kcal/mol. (Please note that freezing may not be appropriate if the frozen parts should undergo conformational changes upon binding.) Increasing MD sampling to overcome large energy fluctuations remains difficult due to the high cost of MD simulations and free-energy analyses of large systems.

The electrostatic solvation free energy G_{pb} (main part of solvation free energy) obtained by solving the Poisson-Boltzmann equation (PBE) is another large error source in MM-PBSA. One reason is that the implicit treatment of water and counterions in the PBE is still a step away from describing the atomic-level explicit solvent environment. Wagoner and Baker find that polar solvation forces generated by PBE systematically overestimate the magnitude of the explicit solvent forces (71) (their work also shows that SASA-based nonpolar forces have no significant correlation with nonpolar explicit solvent forces). Another reason stems from uncertainty regarding dielectric constants. The solute dielectric constant ϵ_{in} , which we set to its vacuum value of 1.0, can vary over a wide range (1.0–20.0) according to different studies (discussed in (72)). However, the variation in ϵ_{in} has a dramatic effect on the electrostatic solvation free energies G_{pb} . For example, G_{pb} of our 10R(-)-*trans*-(A₁): III complex is -8518 kcal/mol if $\epsilon_{in} = 1.0$, and -2019 kcal/mol if $\epsilon_{in} = 4.0$. In a comprehensive review (73), Schutz and Warshel conclude that the dielectric constant is not a universal constant but a parameter that depends on the model used. The more explicit the model, the smaller the dielectric constant that is needed. The dielectric constant can be as small as 1.0 if all contributions are treated explicitly and accurately, including protein permanent dipole, protein reorganization energy, and water penetration, etc. We also note that the solvent dielectric constant ϵ_{out} , which we set to 80, is dependent on temperature (74).

In sum, developing a more accurate and faster free energy calculation method still is a distance away. Our TATA/TBP systems are both large (211 residues) and highly charged (-30 *e* on the DNA), which makes our systems difficult for binding free-energy evaluation. The large standard deviations may also be partly caused by the limited trajectories and timescales in our simulations (one 2.32-ns trajectory for each of the 12 TATA/TBP complexes and the free TBP; one 3.34-ns trajectory for each of the 12 free TATA DNA) due to the large set and sizes of the systems and limited computing

resources available. Still, despite these limitations, both the triplet-trajectory analysis and the single-trajectory analysis produce the relative binding free energies (Table 5 and Table 6) consistent with the experimental values (Table 1) within the computed standard deviations. Thus, structurally distinct molecules may have similar binding free energies. The single-trajectory analysis (Table 6) yields much smaller standard deviations than the triplet-trajectory analysis (Table 5) because fewer random movements (Fig. 4) are introduced in the former, which actually only calculates the interactions at the TATA/TBP interfaces. This is similar to the case in which freezing the atoms outside the binding regions can give better binding free energies for large systems (54,56). Therefore, besides improvements in the MM-PBSA methodology, future binding free energy calculations of DNA/protein complexes require longer simulation time, multiple trajectories for each system, and restraints on noninterface residues (if they do not undergo conformational changes upon binding). Employing multiple trajectories for each system can also help to identify the timescale limitation on a system due to the deficiencies in force fields and simulation protocols (75,76).

CONCLUSION

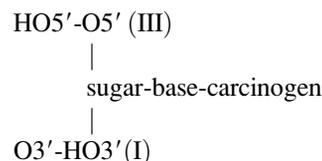
This study reveals the specific position and stereochemistry-dependent structural changes of the TATA/TBP complex due to BP binding to the TATA box. Together with our former analyses (7), we can interpret the location-dependent effect and stereochemistry-dependent effect (for BP at A₁) of BP modifications on TATA/TBP binding in terms of the conformational freedom and major-groove space available to BP due to the hydrogen bonds and phenylalanine insertions at the TATA/TBP interface. Thermodynamic analyses using MM-PBSA produce large standard deviations, which make the computed binding free energies the same within the error bars. Although it agrees with the experimental binding free energy differences (<1 kcal/mol), it suggests the need for MM-PBSA methodology improvements and better sampling for evaluating binding free energies of large and highly charged systems like DNA/protein complexes. More generally, the unraveled structures and interactions in these studies can help investigate how transcription initiation may be affected by a bulky BP chemically binding to the TATA box.

APPENDIX: PROCEDURE OF COMPUTING PARTIAL CHARGES OF CARCINOGEN-MODIFIED B-DEOXYNUCLEOTIDES

Step 1: Build two molecules (InsightII)

To make our partial charges compatible with AMBER's force field (29), we calculate partial charges of the following two molecules, similar to the way AMBER (34) does. Finally we will combine the partial charges of these two molecules in Step 5. We use InsightII (Accelrys) to build these two molecules.

Nucleoside with carcinogen



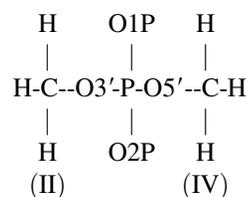
Cap 5' and 3'-end (no phosphate group).

Set dihedral angles:

$$\begin{aligned} \beta(\text{HO5}'\text{-O5}'\text{-C5}'\text{-C4}') &= 180.0^\circ \\ \varepsilon(\text{C4}'\text{-C3}'\text{-O3}'\text{-HO3}') &= 180.0^\circ \\ \gamma(\text{O5}'\text{-C5}'\text{-C4}'\text{-C3}') &= 58.5(\text{DA}, \text{DA5}, \text{DA3}) \\ &= 58.9(\text{DC}, \text{DC5}, \text{DC3}) \\ &= 58.5(\text{DG}, \text{DG5}, \text{DG3}) \\ &= 58.4(\text{DT}, \text{DT5}, \text{DT3}) \end{aligned}$$

No bumps.

Dimethylphosphate (DMP)



Set dihedral angles:

$$\begin{aligned} \text{C-O3}'\text{-P-O5}' &= -73.1^\circ \\ \text{O3}'\text{-P-O5}'\text{-C} &= -73.1^\circ \end{aligned}$$

(Cieplak et al. (33) used positive dihedral angles. But in regular A- and B-DNA, these dihedral angles are negative, i.e., (g⁻, g⁻) conformation.)

No bumps.

Step 2: Optimize geometry on each molecule (Gaussian)

The geometry optimization of both molecules will be performed in next step as the electrostatic potentials (ESP) are calculated simultaneously in Gaussian (77). The time of Gaussian calculation with optimization option will be quite long (~1 week) on the nucleoside with carcinogen. (The optimized structures can be used for building topology files for AMBER, if one needs to do molecular dynamics.) If you want to calculate the partial charges on a specific structural conformation, you may skip this step.

Step 3: Calculate electrostatic potential (ESP) on each molecule (Gaussian)

How to build Gaussian input files can be found in the Gaussian Manual (77). Following are a few parameters which have to be set in Gaussian input files:

Route section (HF/6-31G*, ESP output, optimization; remove "opt" if you want to calculate the partial charges on a specific structural conformation): # HF/6-31G* pop = (minimal,mk) opt maxdisk = 134217728 iop(6/33 = 2).

Charge = 0 (nucleoside with carcinogen, if carcinogen has zero charge), or = -1 (dimethylphosphate).

Spin multiplicity = $N_{\text{up}} + 1$ (N_{up} is the number of unpaired electrons; ground state).

Step 4: Check interactions in optimized structures (InsightII)

Gaussian does not know where bonds, angles, and dihedral angles are if we do not tell in the input files. So the optimized structures may have unreasonable interactions. We use InsightII to check bumps in the optimized structures. If all the interactions are reasonable, we can go to next step. Otherwise, we have to restart from Step 1 (or the current optimized structures) by manually adjusting the structures to avoid further unreasonable interactions.

Step 5: Calculate partial charges of both molecules together (RESP)

Now we need to combine ESP of these two molecules to calculate the partial charges of the carcinogen modified nucleotide within two stages of AMBER's RESP. See AMBER Manual (34) for how to build the input files for RESP. The partial charges of the phosphate group (P, O1P, and O2P) will be taken from the dimethylphosphate (DMP).

First stage

Charge constraints:

$I + II = 0$ (DX5, $X = A, C, G,$ or T). Partial charge of O3' will be taken from DMP), or $III + IV = 0$ (DX3, Partial charge of O5' will be taken from DMP), or $I + II = III + IV = 0$ (DX, Partial charges of O3' and O5' will be taken from DMP).

O1P, O2P in phosphate group are equivalenced.

H in NH_2 groups are equivalenced.

Hyperbolic restraints with a force constant $a = 0.0005$ are applied to all heavy atoms, and hydrogens are not restrained.

Second stage

CH_2 and CH_3 in nucleoside are refit; other atoms' charges are kept.

Hyperbolic restraints with a force constant $a = 0.001$ are applied to only the carbons of CH_2 and CH_3 in nucleoside, and hydrogens are not restrained.

H in CH_2 or CH_3 in nucleoside are equivalenced.

Notes

The final total charge of the carcinogen modified nucleotide will be exactly -1.0000 for DX types ($X = A, C, G,$ or T), if the carcinogen has zero charge. This procedure can be easily applied to A-deoxynucleotides and A-ribonucleotides with Cieplak et al. (33). PERL programs and a sample are available at <http://monod.biomath.nyu.edu/~qzhang/Research.htm>.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at <http://www.biophysj.org>.

We thank Nicholas E. Geacintov for providing additional information of the experimental work, Piotr Cieplak for discussing computation of AMBER partial charges, Shixiang Yan for discussing calculation of binding free energies, and Linjing Yang for discussing the simulation protocol.

Support from the National Institutes of Health (No. R01 GM55164) is gratefully acknowledged. Acknowledgment is also made to the donors of the American Chemical Society Petroleum Research Fund (No. 39115-AC4) for partial support of this research. Computations were supported by National Computational Science Alliance (NCSA) under grant No. MCA99S021N and utilized on the NCSA SGI Origin2000.

REFERENCES

- Kornberg, R. D. 2001. The eukaryotic gene transcription machinery. *Biol. Chem.* 382:1103–1107.
- Nudler, E. 1999. Transcription elongation: structural basis and mechanisms. *J. Mol. Biol.* 288:1–12.
- Pugh, B. F. 2000. Control of gene expression through regulation of the TATA-binding protein. *Gene.* 255:1–14.
- Burley, S. K., and R. G. Roeder. 1996. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.* 65: 769–799.
- Patikoglou, G. A., J. L. Kim, L. Sun, S. H. Yang, T. Kodadek, and S. K. Burley. 1999. TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* 13:3217–3230.
- Nikolov, D. B., H. Chen, E. D. Halay, A. Hoffman, R. G. Roeder, and S. K. Burley. 1996. Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. Natl. Acad. Sci. USA.* 93:4862–4867.
- Zhang, Q., S. Broyde, and T. Schlick. 2004. Deformations of promoter DNA bound to carcinogens help interpret effects on TATA-element structure and activity. *Philos. Transact. A Math. Phys. Eng. Sci.* 362: 1479–1496.
- Pardo, L., M. Campillo, D. Bosch, N. Pastor, and H. Weinstein. 2000. Binding mechanisms of TATA box-binding proteins: DNA kinking is stabilized by specific hydrogen bonds. *Biophys. J.* 78:1988–1996.
- Geacintov, N. E., M. Cosman, B. E. Hingerty, S. Amin, S. Broyde, and D. J. Patel. 1997. NMR solution structures of stereoisomeric covalent polycyclic aromatic carcinogen-DNA adduct: principles, patterns, and diversity. *Chem. Res. Toxicol.* 10:111–146.
- Grimmer, G., G. Dettbarn, and J. Jacob. 1993. Biomonitoring of polycyclic aromatic hydrocarbons in highly exposed coke plant workers by measurement of urinary phenanthrene and pyrene metabolites (phenols and dihydrodiols). *Int. Arch. Occup. Environ. Health.* 65:189–199.
- Phillips, D. H. 1999. Polycyclic aromatic hydrocarbons in the diet. *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* 443:139–147.
- Conney, A. H. 1982. Induction of microsomal-enzymes by foreign chemicals and carcinogenesis by polycyclic aromatic-hydrocarbons—G.H.A. Clowes Memorial Lecture. *Cancer Res.* 42:4875–4917.
- Choi, D. J., D. J. Marinoaleessandri, N. E. Geacintov, and D. A. Scicchitano. 1994. Site-specific benzo[a]pyrene diol epoxide-DNA adducts inhibit transcription elongation by bacteriophage T7-RNA polymerase. *Biochemistry.* 33:780–787.
- Perlow, R. A., A. Kolbanovskii, B. E. Hingerty, N. E. Geacintov, S. Broyde, and D. A. Scicchitano. 2002. DNA adducts from a tumorigenic metabolite of benzo[a]pyrene block human RNA polymerase II elongation in a sequence- and stereochemistry-dependent manner. *J. Mol. Biol.* 321:29–47.
- Lenne-Samuel, N., R. Janel-Bintz, A. Kolbanovskiy, N. E. Geacintov, and R. P. P. Fuchs. 2000. The processing of a benzo(a)pyrene adduct into a frameshift or a base substitution mutation requires a different set of genes in *Escherichia coli*. *Mol. Microbiol.* 38:299–307.
- Perlow, R. A., and S. Broyde. 2003. Extending the understanding of mutagenicity: structural insights into primer-extension past a benzo [a]pyrene diol epoxide-DNA adduct. *J. Mol. Biol.* 327:797–818.
- Perlow, R. A., and S. Broyde. 2002. Toward understanding the mutagenicity of an environmental carcinogen: structural insights into nucleotide incorporation preferences. *J. Mol. Biol.* 322:291–309.

18. Perlow, R. A., and S. Broyde. 2001. Evading the proofreading machinery of a replicative DNA polymerase: induction of a mutation by an environmental carcinogen. *J. Mol. Biol.* 309:519–536.
19. Wei, S. J., R. L. Chang, N. Bhachech, X. X. Cui, K. A. Merkler, C. Q. Wong, E. Hennig, H. Yagi, D. M. Jerina, and A. H. Conney. 1993. Dose-dependent differences in the profile of mutations induced by (+)-7R,8S-dihydroxy-9S,10R-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene in the coding region of the hypoxanthine (guanine) phosphoribosyltransferase gene in Chinese hamster V-79 cells. *Cancer Res.* 53:3294–3301.
20. Cheng, S. C., B. D. Hilton, J. M. Roman, and A. Dipple. 1989. DNA adducts from carcinogenic and noncarcinogenic enantiomers of benzo[a]pyrene dihydrodiol epoxide. *Chem. Res. Toxicol.* 2:334–340.
21. Meehan, T., and K. Straub. 1979. Double-stranded DNA stereoselectively binds benzo(a)pyrene diol epoxides. *Nature.* 277:410–412.
22. Szeliga, J., and A. Dipple. 1998. DNA adduct formation by polycyclic aromatic hydrocarbon dihydrodiol epoxides. *Chem. Res. Toxicol.* 11:1–11.
23. Schurter, E. J., H. J. C. Yeh, J. M. Sayer, M. K. Lakshman, H. Yagi, D. M. Jerina, and D. G. Gorenstein. 1995. NMR solution structure of a nonanucleotide duplex with a dG mismatch opposite a 10R adduct derived from *trans* addition of a deoxyadenosine *n*-6-amino group to ((-)-(7S,8R,9R,10S)-7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene. *Biochemistry.* 34:1364–1375.
24. Schwartz, J. L., J. S. Rice, B. A. Luxon, J. M. Sayer, G. Xie, H. J. C. Yeh, X. Liu, D. M. Jerina, and D. G. Gorenstein. 1997. Solution structure of the minor conformer of a DNA duplex containing a dG mismatch opposite a benzo[a]pyrene diol epoxide/dA adduct: glycosidic rotation from *syn* to *anti* at the modified deoxyadenosine. *Biochemistry.* 36:11069–11076.
25. Yeh, H. J. C., J. M. Sayer, X. H. Liu, A. S. Altieri, R. A. Byrd, M. K. Lakshman, H. Yagi, E. J. Schurter, D. G. Gorenstein, and D. M. Jerina. 1995. NMR solution structure of a nonanucleotide duplex with a dG mismatch opposite a 10S adduct derived from *trans* addition of a deoxyadenosine *n*-6-amino group to (+)-(7R,8S,9S,10R)-7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene—an unusual *Syn* glycosidic torsion angle at the modified dA. *Biochemistry.* 34:13570–13581.
26. Zegar, I. S., S. J. Kim, T. N. Johansen, P. J. Horton, C. M. Harris, T. M. Harris, and M. P. Stone. 1996. Adduction of the human *n*-Ras codon 61 sequence with (-)-(7S,8R,9R,10S)-7,8-dihydroxy-9,10-epoxy-7,8,9,10-tetrahydrobenzo[a]pyrene: structural refinement of the intercalated *SRSR*(61,2) (-)-(7S,8R,9S,10R)-N6-(10-(7,8,9,10-tetrahydrobenzo[a]pyrenyl))-2'-deoxyadenosyl adduct from 1H NMR. *Biochemistry.* 35:6212–6224.
27. Rechkoblit, O., J. Krzeminsky, S. Amin, B. Jernstrom, N. Louneva, and N. E. Geacintov. 2001. Influence of bulky polynuclear carcinogen lesions in a TATA promoter sequence on TATA binding protein—DNA complex formation. *Biochemistry.* 40:5622–5632.
28. Tan, J., N. E. Geacintov, and S. Broyde. 2000. Principles governing conformations in stereoisomeric adducts of bay region benzo[a]pyrene diol epoxides to adenine in DNA: steric and hydrophobic effects are dominant. *J. Am. Chem. Soc.* 122:3021–3032.
29. Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force-field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.
30. Cheatham, T. E., P. Cieplak, and P. A. Kollman. 1999. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dynam.* 16:845–862.
31. Bayly, C. I., P. Cieplak, W. D. Cornell, and P. A. Kollman. 1993. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges—the RESP model. *J. Phys. Chem.* 97:10269–10280.
32. Cornell, W. D., P. Cieplak, C. I. Bayly, and P. A. Kollman. 1993. Application of RESP charges to calculate conformational energies, hydrogen-bond energies, and free energies of solvation. *J. Am. Chem. Soc.* 115:9620–9631.
33. Cieplak, P., W. D. Cornell, C. Bayly, and P. A. Kollman. 1995. Application of the multimolecule and multiconformational RESP methodology to biopolymers—charge derivation for DNA, RNA, and proteins. *J. Comput. Chem.* 16:1357–1377.
34. Case, D. A., D. A. Pearlman, J. W. Caldwell, T. E. Cheatham, W. S. Ross, C. Simmerling, T. Darden, K. M. Merz, R. V. Stanton, A. Cheng, J. J. Vincent, M. Crowley, et al. 1999. AMBER 6. University of California, San Francisco, CA.
35. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.
36. Hockney, R. W. 1970. The potential calculation and some applications. *Methods Comput. Phys.* 9:136–211.
37. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald—an *N*-log(*N*) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
38. Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.
39. Cheatham, T. E., J. Srinivasan, D. A. Case, and P. A. Kollman. 1998. Molecular dynamics and continuum solvent studies of the stability of polyG-polyC and polyA-polyT DNA duplexes in solution. *J. Biomol. Struct. Dyn.* 16:265–280.
40. Kollman, P. A., I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham III. 2000. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33:889–897.
41. Jayaram, B., D. Sprous, M. A. Young, and D. L. Beveridge. 1998. Free energy analysis of the conformational preferences of A and B forms of DNA in solution. *J. Am. Chem. Soc.* 120:10629–10633.
42. Srinivasan, J., T. E. Cheatham, P. Cieplak, P. A. Kollman, and D. A. Case. 1998. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate—DNA helices. *J. Am. Chem. Soc.* 120:9401–9409.
43. Honig, B., and A. Nicholls. 1995. Classical electrostatics in biology and chemistry. *Science.* 268:1144–1149.
44. Nicholls, A., and B. Honig. 1991. A rapid finite-difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comput. Chem.* 12:435–445.
45. Sanner, M. F., A. J. Olson, and J. C. Spehner. 1996. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers.* 38:305–320.
46. Srinivasan, J., J. Miller, P. A. Kollman, and D. A. Case. 1998. Continuum solvent studies of the stability of RNA hairpin loops and helices. *J. Biomol. Struct. Dyn.* 16:671–682.
47. Case, D. A., D. A. Pearlman, J. W. Caldwell, T. E. Cheatham, J. Wang, W. S. Ross, C. Simmerling, T. Darden, K. M. Merz, R. V. Stanton, A. Cheng, J. J. Vincent, et al. 2002. AMBER 7. University of California, San Francisco, CA.
48. Laitinen, T., J. A. Kankare, and M. Perakyla. 2004. Free energy simulations and MM-PBSA analyses on the affinity and specificity of steroid binding to antiestradiol antibody. *Proteins.* 55:34–43.
49. von Langen, J., K. H. Fritzemeier, S. Diekmann, and A. Hillisch. 2005. Molecular basis of the interaction specificity between the human glucocorticoid receptor and its endogenous steroid ligand cortisol. *Chem. Biol. Chem.* 6:1110–1118.
50. Hou, T., L. Zhu, L. Chen, and X. Xu. 2003. Mapping the binding site of a large set of quinazoline type EGF-R inhibitors using molecular field analyses and molecular docking studies. *J. Chem. Inf. Comput. Sci.* 43:273–287.
51. Diaz, N., D. Suarez, K. M. Merz, Jr., and T. L. Sordo. 2005. Molecular dynamics simulations of the TEM-1 β -lactamase complexed with cephalothin. *J. Med. Chem.* 48:780–791.
52. Kuhn, B., and P. A. Kollman. 2000. Binding of a diverse set of ligands to avidin and streptavidin: an accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem.* 43:3786–3791.
53. Swanson, J. M., R. H. Henchman, and J. A. McCammon. 2004. Revisiting free energy calculations: a theoretical connection to MM/

- PBSA and direct calculation of the association free energy. *Biophys. J.* 86:67–74.
54. Wang, J., P. Morin, W. Wang, and P. A. Kollman. 2001. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J. Am. Chem. Soc.* 123:5221–5230.
55. Perakyla, M., and N. Nordman. 2001. Energetic analysis of binding of progesterone and 5 β -androstane-3,17-dione to anti-progesterone antibody DB3 using molecular dynamics and free energy calculations. *Protein Eng.* 14:753–758.
56. Huo, S., J. Wang, P. Cieplak, P. A. Kollman, and I. D. Kuntz. 2002. Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: insight into structure-based ligand design. *J. Med. Chem.* 45:1412–1419.
57. Nordman, N., J. Valjakka, and M. Perakyla. 2003. Analysis of the binding energies of testosterone, 5 α -dihydrotestosterone, androstenedione and dehydroepiandrosterone sulfate with an antitestosterone antibody. *Proteins.* 50:135–143.
58. Wu, Y., Z. Cao, H. Yi, D. Jiang, X. Mao, H. Liu, and W. Li. 2004. Simulation of the interaction between ScyTx and small conductance calcium-activated potassium channel by docking and MM-PBSA. *Biophys. J.* 87:105–112.
59. Luo, C., L. Xu, S. Zheng, X. Luo, J. Shen, H. Jiang, X. Liu, and M. Zhou. 2005. Computational analysis of molecular basis of 1:1 interactions of NRG-1 β wild-type and variants with ErbB3 and ErbB4. *Proteins.* 59:742–756.
60. Wang, W., and P. A. Kollman. 2000. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *J. Mol. Biol.* 303:567–582.
61. Huo, S., I. Massova, and P. A. Kollman. 2002. Computational alanine scanning of the 1:1 human growth hormone-receptor complex. *J. Comput. Chem.* 23:15–27.
62. Adekoya, O. A., N. P. Willassen, and I. Sylte. 2005. The protein-protein interactions between SMPI and thermolysin studied by molecular dynamics and MM/PBSA calculations. *J. Biomol. Struct. Dyn.* 22:521–531.
63. Suenaga, A., M. Hatakeyama, M. Ichikawa, X. Yu, N. Futatsugi, T. Narumi, K. Fukui, T. Terada, M. Taiji, M. Shirouzu, S. Yokoyama, and A. Konagaya. 2003. Molecular dynamics, free energy, and SPR analyses of the interactions between the SH2 domain of Grb2 and ErbB phosphotyrosyl peptides. *Biochemistry.* 42:5195–5200.
64. Wang, W., W. A. Lim, A. Jakalian, J. Wang, J. Wang, R. Luo, C. I. Bayly, and P. A. Kollman. 2001. An analysis of the interactions between the Sem-5 SH3 domain and its ligands using molecular dynamics, free energy calculations, and sequence analysis. *J. Am. Chem. Soc.* 123:3986–3994.
65. Choi, Y., and S. Jung. 2004. Molecular dynamics (MD) simulations for the prediction of chiral discrimination of *n*-acetylphenylalanine enantiomers by cyclomaltoheptaose (β -cyclodextrin, β -CD) based on the MM-PBSA (molecular mechanics-Poisson-Boltzmann surface area) approach. *Carbohydr. Res.* 339:1961–1966.
66. Bonnet, P., C. Jaime, and L. Morin-Allory. 2002. Structure and thermodynamics of α -, β -, and γ -cyclodextrin dimers. Molecular dynamics studies of the solvent effect and free binding energies. *J. Org. Chem.* 67:8602–8609.
67. Reyes, C. M., R. Nifosi, A. D. Frankel, and P. A. Kollman. 2001. Molecular dynamics and binding specificity analysis of the bovine immunodeficiency virus BIV Tat-TAR complex. *Biophys. J.* 80:2833–2842.
68. Gouda, H., I. D. Kuntz, D. A. Case, and P. A. Kollman. 2003. Free energy calculations for theophylline binding to an RNA aptamer: comparison of MM-PBSA and thermodynamic integration methods. *Biopolymers.* 68:16–34.
69. Spackova, N., T. E. Cheatham III, F. Ryjacek, F. Lankas, L. Van Meervelt, P. Hobza, and J. Sponer. 2003. Molecular dynamics simulations and thermodynamics analysis of DNA-drug complexes. Minor groove binding between 4',6-diamidino-2-phenylindole and DNA duplexes in solution. *J. Am. Chem. Soc.* 125:1759–1769.
70. Reyes, C. M., and P. A. Kollman. 2000. Structure and thermodynamics of RNA-protein binding: using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. *J. Mol. Biol.* 297:1145–1158.
71. Wagoner, J., and N. A. Baker. 2004. Solvation forces on biomolecular structures: a comparison of explicit solvent and Poisson-Boltzmann models. *J. Comput. Chem.* 25:1623–1629.
72. Fogolari, F., A. Brigo, and H. Molinari. 2003. Protocol for MM/PBSA molecular dynamics simulations of proteins. *Biophys. J.* 85:159–166.
73. Schutz, C. N., and A. Warshel. 2001. What axe the dielectric “constants” of proteins and how to validate electrostatic models? *Proteins Struct. Funct. Genet.* 44:400–417.
74. Hasted, J. B. 1973. *Aqueous Dielectrics*. Chapman and Hall, London.
75. Auffinger, P., and E. Westhof. 1998. Simulations of the molecular dynamics of nucleic acids. *Curr. Opin. Struct. Biol.* 8:227–236.
76. Cheatham III, T. E., and M. A. Young. 2000. Molecular dynamics simulation of nucleic acids: successes, limitations, and promise. *Biopolymers.* 56:232–256.
77. Frisch, M. J., G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, et al. 1998. Gaussian 98. Gaussian, Inc., Pittsburgh, PA.