

INTERACTION OF TATA-BOX BINDING PROTEIN (TBP) WITH DNA: A MOLECULAR DYNAMICS SIMULATION STUDY

SUSIL J. SILVA¹ and ALEXANDER D. MACKERELL JR.

¹ Department of Chemistry, University of Sri Jayewardenepura, Nugegoda

² Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland, Baltimore, MD 21201, USA

(Received: 18 February 2002 ; accepted: 30 January 2003)

Abstract: The TATA binding protein (TBP) is a DNA-binding protein required by all three eukaryotic nuclear RNA polymerases for transcription initiation. Crystal structures of several TBP-DNA complexes showed that TBP binds to the minor groove of the TATA box, unwinds the TATA box, and bends the DNA towards the major groove. It has previously been shown both *in vitro* and *in vivo* that some A/T point mutations in the TATA core sequence have strong effects on transcription activity and that the induced degree of bending of TATA box mutants correlates with the stability of TBP-DNA complexes. In order to obtain structural, binding, energetics and dynamics changes between wild type (active) and mutated (inactive) TATA-box sequences, molecular dynamic simulations on three TBP-DNA complexes with the TATA sequences TATAAAAG, TAAAAAAG and TATGAAAG have been performed. The small average root mean square deviation (RMSD) between simulated structures and X-ray crystal structure of TBP-DNA complex indicated that the simulated TBP-DNA structures are very similar to the X-ray crystallographic structures. The extent of DNA bending in three simulated complexes is found to be very similar. The most favourable interaction energy between DNA and TBP was found in the wild type complex. The present results suggest that the specificity of the TBP for the TATA box is due to the sum of collective interactions rather than specific interactions between nucleotides and amino acids.

Keywords: molecular dynamics, TATA-box, TATA-box bindingprotein, transcription

INTRODUCTION

The TATA box binding protein (TBP) is a DNA-binding protein required by all three eukaryotic RNA polymerases¹ (Pol I, II and II). The role of TBP in transcription has been studied extensively in the context of the RNA pol II, where TBP has been shown to recognize and bind a conserved promoter element known as TATA box, consensus sequence TATAAAA.² TBP binds several variants of TATA box sequences.^{3,4} The affinity of the TBP-TATA box interaction has been proposed to contribute to promoter strength both *in vivo*⁴ and *in vitro*.⁵ Wobbe *et al.*⁴ analyzed the DNA sequence requirements for TATA element function by assaying the transcriptional activities of 25 promoters, including those representing each of the 18 single-point mutants of the consensus sequence. Their results showed that only two mutant elements, TAAAAAA and TATGAAA, had undetectable activity and that the others had activities ranging from 2 to 75% of the activity of the consensus sequence.

* Corresponding author

The structures of several TBP-DNA complexes have been determined by X-ray crystallography.⁶⁻¹⁰ All these studies revealed that TBP binds to the minor groove of the TATA box, unwinds the TATA box and bends the DNA towards the major groove. Starr *et al.*¹¹ analyzed the extent, location and direction of the DNA bend induced by the TBP upon complexation using gel electrophoresis. They observed that the induced degree of bending of TATA box mutants correlates with the stability of TBP-DNA complexes. Among the various mutant sequences, the sequence TAAAAAA showed the shortest lifespan and smallest relative transcription activity.

In a recent study, Patikoglou *et al.*¹⁰ determined the X-ray crystal structure of wild type TBP recognizing ten naturally occurring TATA elements including the consensus sequence. They concluded that TBP binding induced the same deformation in the core promoter independent of the sequence and the transcription activity of the TATA elements. Their attempt to crystallize TBP-DNA complexes containing TAAAAAA and TATGAAA was unsuccessful. Contrary to these results, in another recent study by Parkhurst *et al.*¹² observed variable DNA bends in solution in the TBP-bound adenovirus major late promoter (AdMLP) and eight naturally occurring variant TATA sequences. They observed a strong correlation between bend angles and relative basal transcription activity.

Several molecular dynamics simulation studies have been performed on the TATA box containing DNA sequences by various other groups. Flatters *et al.*¹³ suggested that the TATA box region adopts an A-like configuration and undergoes bending related to that seen in the TBP-DNA complex. Using a TATA box and three mutant DNA sequences, de Souza *et al.*¹⁴ demonstrated that DNA sequence dependent curvature and flexibility play a significant role in TBP recognition, binding and transcription activity. A significant change in time-averaged structure and in dynamic behaviour was observed in the MD simulation study on a DNA oligomer containing the sequence TAAAAAA as compared to TATAAAA using the same MD protocol.¹⁵ Molecular dynamics simulation of seven DNA dodecamers containing various TBP binding sequences led Pastor¹⁶ *et al.* to suggest that alternating pyrimidine-purine (YR) sequences are preferred for TBP binding because YR steps achieve the average high rise required for the kinks at steps 1 and 7.¹⁶ Hinge bending and twisting was observed in the simulation performed for the TBP-DNA complex by Miaskiewicz *et al.*¹⁷

The mode of interaction between TBP and its binding site is also interesting from a purely structural perspective. It is obvious that more detailed information of structural and dynamical aspects of interaction between TBP and DNA is necessary for understanding the molecular basis of TBP activities in TATA containing promoters. The computational methods, in particular molecular dynamics (MD) simulations can be employed to uncover an atomic detailed picture of TBP-DNA interactions in solution.

In the present study some of the questions concerning TBP-DNA interaction using MD simulations are discussed.

COMPUTATIONAL METHODS

Initial coordinates and preparation of systems: The three DNA sequences and the amino acid residues in the DNA binding domain (DBD) of TBP along with their numbering schemes used in this study are shown in Figure 1. The nomenclature used is based on Starr *et al.*¹¹ The initial coordinates of WT DNA and DNA-protein complex were obtained from the X-ray crystal structure of the TBP-DNA complex (NDB code PDT009). The atomic coordinates of WT, 29A and 28G were built with B-form using the SYBYL programme. The starting geometries of 29A and 28G TBP-DNA complexes were generated by replacing the T6-A9 and A7-T8 base pairs in WT TBP-DNA with A6-T9 and G7-C8 respectively. The 5' phosphate groups at the end of the strands were removed and replaced with hydroxyl moieties. The resulting structures were overlaid with a pre-equilibrated solvent box containing the CHARMM TIP3P water,¹⁸ sodium and chlorine ions. All the solvent atoms within 1.6 Å of any DNA non-hydrogen atom and 2.8 Å from any protein non-hydrogen atom were deleted. In addition, solvent atoms within 5.0 Å from the minor groove non-hydrogen atoms in the TATA box region were also deleted. For each system the size of the solvent box is determined according to the maximum and minimum coordinates of the solute non-hydrogen atoms in each direction. The edges of the solvent box were extended a minimum of 8-9 Å along each direction. The ionic strength of 0.78M was maintained via addition of NaCl beyond the solute.

MD Simulations: MD simulations were performed using the CHARMM programme^{19,20} in the NPT ensemble at 300 K. The new CHARMM27 all-atom force field²¹ was used in all simulations. All calculations were performed using SHAKE²² to constrain covalent bonds involving hydrogen, allowing for an integration time step of 0.002 ps in the integration of the equation of motion. The image atoms were generated using the CRYSTAL²³ in CHARMM. The electrostatic interactions were treated via the particle mesh Ewald (PME) method.²⁴ The PME charge grid spacing was approximately 1.0 Å and the calculations were performed using real space cutoff 10 Å. The value of the screening parameter (κ) used in all PME calculations was 0.34. The Lenard-Jones interactions are evaluated with 10 Å atom-based cutoff. All calculations were carried out applying periodic boundary conditions. The resulting systems were energy minimized using 100 steepest decent (SD) steps keeping all solute non-hydrogen atoms fixed. The energy minimized systems were then subjected to 50 ps isothermal, constant volume MD simulation with all solute atoms harmonically constrained with mass weighted force constant of 5 kcal/mol/ Å to equilibrate the solvent around solute. This system was energy minimized using 200 ABNR(Adopted Basis Newton Raphson) steps and the resulting coordinate set was used to initiate the production trajectory. Simulations were done for 2.0ns saving trajectory for every 1.0 ps. For all simulations, the first 500 ps portion of the trajectory

was considered as the equilibration, with analysis performed on the remainder of the trajectory.

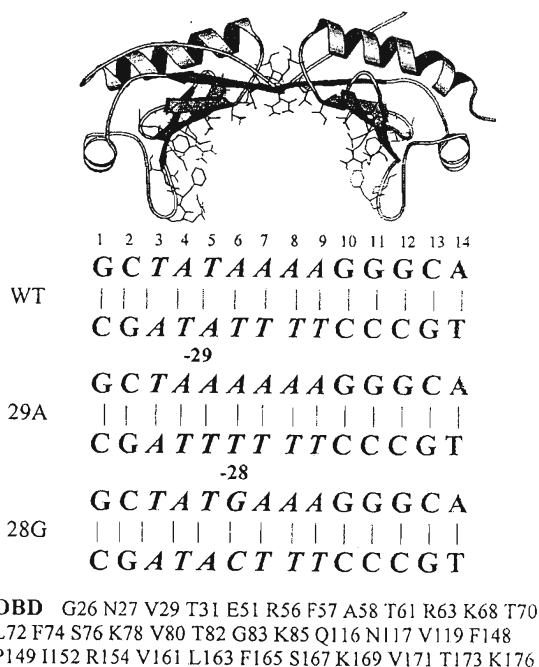


Figure 1: The structure of the TBP with DBD residues in stick figures (top), three DNA sequences (middle) used in the molecular dynamics simulations. The TATA box sequences are in italics and two mutations are labelled according to Starr *et al.* The DBD residues with single letter codes and residue numbers are also included (bottom).

Analysis of trajectories: The root mean square deviation (RMSD) is calculated using the CORREL command in CHARMM for entire trajectory in each simulation. The helicoidal and phosphate backbone structural parameters were calculated using the FREEHELIX programme.²⁵ The first 500ps of each trajectory is considered as the equilibration time and the analysis was carried out on the last 1500 ps of simulations. The averages and standard errors were calculated by dividing the data into five blocks of 300 ps each and carrying out statistical analysis on the block average values.

RESULTS

RMSD Analysis

DNA Simulations: Analysis of the trajectories for the base pair hydrogen bond distances revealed that in all cases the Watson-Crick duplex structure of DNA was

maintained throughout the simulations. The root mean square deviation (RMSD) for each simulation was calculated with respect to the A and B form of DNA and X-ray crystallographic structure of DNA in the TBP-DNA complex. Figure 2 shows the variation of RMSD as a function of time for three DNA simulations. The lack of long-term drift to higher values after 500ps in RMSD suggested that the simulations are equilibrated. The calculated average RMSD and fluctuation, with respect to the canonical B-DNA conformation over 2000ps interval, is 3.4 ± 1.7 , 2.8 ± 0.6 and 2.5 ± 0.3 Å for WT, 29A and 28G respectively. With respect to the canonical A-DNA it is 4.6 ± 0.9 , 5.0 ± 0.5 and 5.5 ± 0.4 Å for WT, 29A and 28G respectively. In order to test the influence of the starting conformation on the dynamic behaviour of 14-mer, another simulation was carried out for WT DNA under the same conditions as described in the methods section but using the initial geometry of DNA from the X-ray crystal structure of the TBP-DNA complex. The key feature of this WT trajectory is the departure of DNA from the distorted form to the B form within 500-700ps. Apart from the local fluctuations around 1100-1300ps in WT and 400-600ps in 29A, Figure 2 shows that DNA remained in the B-conformation throughout the trajectory. In order to investigate the structural changes in the TATA box region during the course of simulation, the RMSD of this region was calculated with respect to canonical A and B-DNA and experimental DNA structures. The calculated RMSD values for WT, 29A and 28G are 2.5 ± 1.1 , 2.0 ± 0.3 and 1.8 ± 0.3 Å with respect to B-DNA and 3.0 ± 0.6 , 3.3 ± 0.3 and 3.6 ± 0.3 Å with respect to A-DNA and 4.1 ± 1.0 , 4.7 ± 0.3 and 5.0 ± 0.3 Å with respect to the experimental DNA structure. These results confirmed that TATA box region of all three 14-mers also remained in the B-DNA conformation. Our prediction that the average conformation adopted by 14-mers and also the TATA box region is closer to the canonical B-DNA is contradictory to the previous molecular dynamics results. Flatters and co-workers¹³ showed that one TATA box sequence, that of the *S. cerevisiae* TBP-DNA complex,⁷ inherently adopts an A-like conformation in the TBP binding region. Four MD simulations (WT and three mutants 27T, 30T and 29A) starting with the canonical B form were performed on a dodecamer containing TATA box sequence by de Souza *et al.*¹⁴ Their RMSD results showed that WT and 27T adopt an A form whereas 30T and 29A adopt a B form.

TBP-DNA Simulations: Including the WT and two single point mutants of the TATA consensus sequence, 2ns length MD simulations were performed on the TBP-DNA complex. Two mutants were selected based on the result that they show no transcriptional activity both in *vitro* and in *vivo*.⁴ All three simulations were initiated from the X-ray crystal structure of the TBP-DNA complex.⁶ In order to investigate how the DNA and protein dynamics change throughout the simulations, the RMSD values with respect to their X-ray crystal structures and canonical A and B forms were calculated. As shown in Figure 2, the RMSD values predict that the DNA structure stays closer to the experimental structure. The average RMSD values with respect to the experimental structure for the WT, 29A and 28G trajectories were 1.6, 1.7 and 1.9 Å respectively. The RMSD values of the TATA box segment with respect to the experimental structure were also calculated. The values obtained

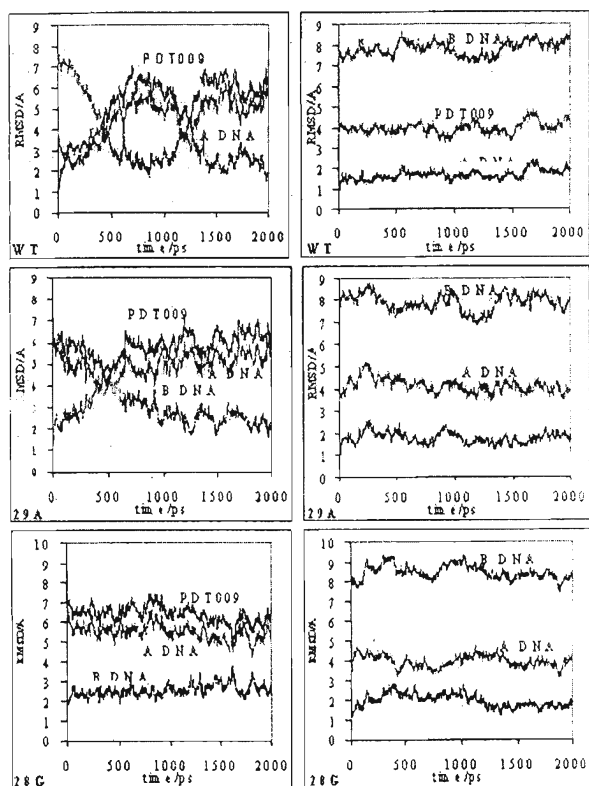


Figure 2: The variation of RMSD as a function of time for DNA alone (left) and TBP-DNA (right) simulations.

for WT, 29A and 28G are 1.2, 1.2 and 1.2 Å respectively. These results indicated that the average simulated structures of DNA in TBP-DNA complexes in the condensed phase are very close to the X-ray structure of DNA in the TBP-DNA complex. The average RMSD values of the TBP backbone atoms with respect to the experimental structure for WT, 29A and 28G are 1.57, 1.49 and 1.54 Å respectively, and for side-chain atoms are 2.07, 2.18 and 2.06 Å respectively. All these results suggested that a globally stable TBP-DNA interaction exists throughout the trajectory for all three complexes. In order to understand the dynamical behaviour of the amino acids in the DNA binding domain (DBD), the RMSD of each amino acid in the DBD was calculated with respect to the crystal structure. The results shown in Figure 3 revealed that amino acids in the DBD behave almost similarly in all three complexes despite the fact of single point mutations in the TATA box segments of 29A and 28G. The RMSD values of backbone and side-chain confirmed that structural interactions between DBD of TBP and DNA observed in the crystal structure are well retained during the simulations. In the previous MD simulation on TBP and TBP-DNA complex, Miaskiewicz *et al.*¹⁷ observed hinge-bending motions in TBP resulting in a closure of the DNA binding surface. They also observed the loss of saddle shape of

TBP as seen in the crystal structure of TBP and TBP-DNA complexes during the simulation. The simulation results on TBP and TBP-DNA complexes in the present study did not show such a hinge-bending motion. The visualization of snapshots at various time points confirmed that the saddle shape of TBP is well maintained throughout each simulation. The analysis of RMSD values of each amino acid in TBP with respect to the crystal structure (data not shown) further confirmed the present observations.

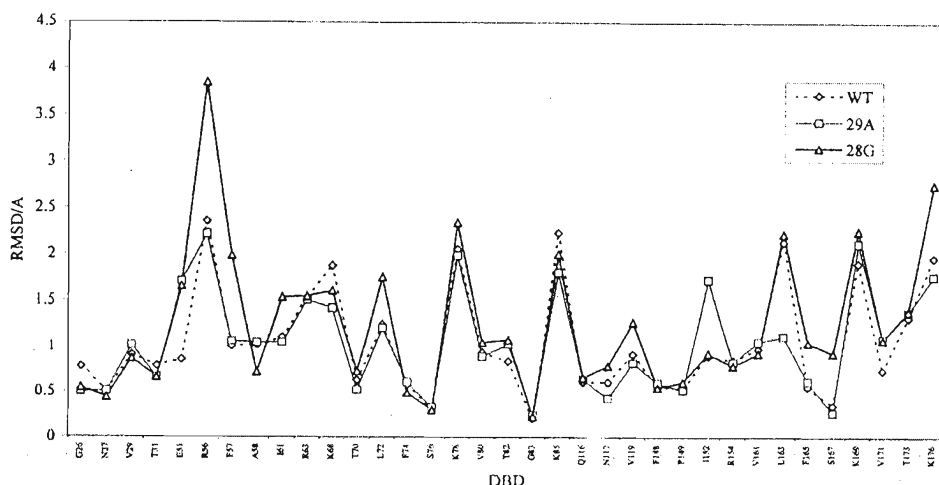


Figure 3: The RMSD of each amino acid in the DBD with respect to the X-ray crystal structure of TBP in the TBP-DNA complex.

Helicoidal Parameters

In order to understand how the DNA conformation varies during the simulation, helicoidal parameters were calculated using the FREEHELIX²⁵ programme. The analysis on the 500-2000ps region of each trajectory was performed. The results presented here include local parameters, which describe the relative position of successive base pair steps. The average base step and base pair parameters of DNA alone and TBP-DNA simulations are shown in Figures 4 and 5 respectively. The examination of base pair roll values showed the sequence dependant effects. The large positive roll values were observed at T-A and T-G base steps in WT and 28G 14-mers and at T-A base step in 29A 14-mer. The small roll values can be seen for A-T and A-A steps. These results are consistent with the values observed in crystal²⁶ and solution NMR structures and also in MD simulated structures¹⁴ of DNA oligomers. As seen in Figure 5, large inclination values are observed for the base pairs in the TATA box sequence compared to the flanking regions in all three 14-mers. The X-ray crystal structures of TBP-DNA complexes showed an average inclination value about 50° for the TATA box sequence.²⁷ Guzikevich-Guerstein *et al.*²⁷ suggested that this large unusual inclination allows the unwinding of DNA

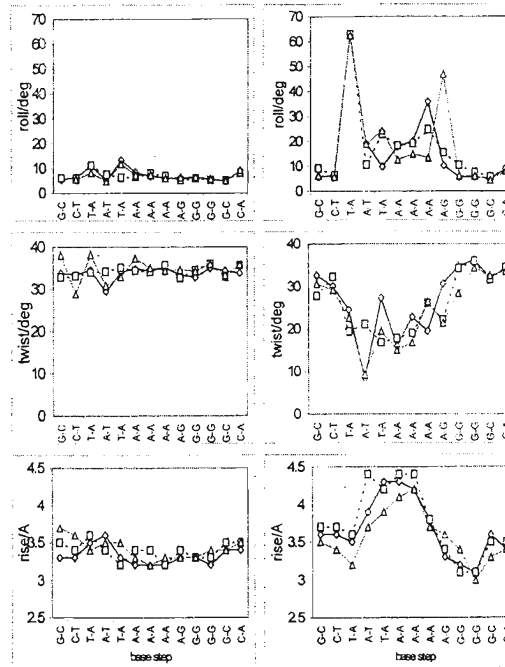


Figure 4: The average base step parameters of DNA alone (left) and TBP-DNA (right) simulations. \diamond - WT, \square - 29A and \triangle - 28G

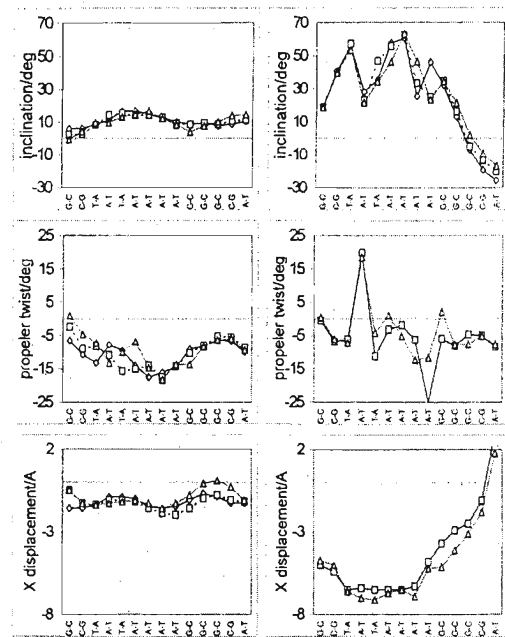


Figure 5: The base pair parameters of DNA alone (left) and TBP-DNA (right) simulations. \diamond - WT, \square - 29A and \triangle - 28G

bases in the TATA box. The average helical twist value lies between A and B forms for all three 14-mers. However, rise values reflect the characteristic B-DNA structure. The variation of the average propeller twist value for 14 base pairs is shown in Figure 5. The large negative propeller twist values in the TATA box sequence are a feature characteristic to A-tracts. Highly twisted base pairs were observed in the X-ray crystal structures of short DNA oligomers containing A-tracts.²⁸

X-ray crystallographic studies of aTBP⁶⁻⁹ with DNA have revealed major DNA deformations such as unwinding and bending upon complex formation. The structure of the eight base pair TATA box differs significantly from both A and B form DNA. Guzikovich-Guerstein *et al.*²⁷ suggested a novel conformation, named TA-DNA, for the TATA box. The flanking DNA on either side of the TATA box adopts the canonical B-form. Here a detailed analysis of DNA helicoidal parameters observed throughout the 2ns trajectory of WT, 29A and 28G complexes is presented. The calculated results are compared with the corresponding values found in the aTBP-DNA and TFIIB-aTBP-DNA²⁹ crystal structure. The selected average base step parameters calculated using the FREEHELIX programme is shown in Figure 5. The kinks at the T-A (-31):A-T(-30) and A-T(-25):G-C(-24) steps in the aTBP-DNA crystal structure due to the partial intercalation of phenylalanine rings are reflected clearly in large positive roll angles. Similar but somewhat reduced roll angles were also seen in the crystal structure of the TFIIB-aTBP-DNA. Figure 5 shows a large positive roll angle at the T-A(-31):A-T(-30) step in all three simulated complexes suggesting the existence of the first kink as observed in the crystal structures. However, the position of the second kink in the simulated structures depends on the sequence. It is moved one base step up in the 3' to 5' direction in WT and retained in the same position in 28G simulation as in the crystal structures. There is no significant kink observed in 29A simulation at A-T(-25):G-C(-24) step. As seen in Figure 5, the roll of the base steps in the center of the TATA box also contributes to the bending. The small roll angles in the two flanking regions correspond to the typical B-DNA conformation. In contrast to roll, tilt is small and plays no significant role in DNA bending. The helical twist values for the TATA box sequence are lower than the 36 expected for undeformed B-DNA. The simulated and two crystal structures exhibit similar helical twist patterns in 14-mer. In general, the average helicoidal parameters of the three simulated complexes agree quite well with the corresponding values obtained from the crystal structures of aTBP-DNA and TFIIB-aTBP-DNA.

The backbone torsion angles for the TATA box region of the simulated structures agree well with the crystal structures as shown in Table 1. It is very clear that torsion angles among the simulated structures are almost indistinguishable regardless of the single point mutations. The torsion angles of the TATA box region are close to the values typical of A-DNA, whereas those of the flanking regions are close to the B-DNA. The average pseudo-rotation angle for the TATA box region varies between C_{3'}-endo to O_{4'}-exo and the flanking regions show C2'-endo sugar puckers corresponding to B-DNA.

Table 1: Average backbone torsion angles in degrees (averaged over strand 1 and 2)

	α	β	γ	δ	ϵ	ξ	χ
DNA							
WT	299.6±0.3	170.7±0.7	51.6±0.5	129.5±1.1	191.0±0.7	252.4±1.5	254.5±1.7
29A	299.0±0.4	171.6±0.7	52.5±0.8	127.2±2.2	190.0±0.7	257.0±1.8	246.4±1.9
28G	298.3±0.8	170.5±0.5	53.0±0.5	126.2±1.8	191.0±0.6	256.5±1.7	246.2±1.5
A-DNA	313.7	275.3	207.9	45.5	82.6	177.5	205.8
B-DNA	313.2	214.0	36.4	156.4	155.0	264.9	262.1
TBP-DNA							
WT	292.5±0.7	173.6±0.5	51.7±0.3	98.0±0.5	199.7±0.4	275.4±1.1	240.8±0.6
29A	292.6±0.9	170.6±0.8	53.2±0.7	92.2±1.1	204.2±0.8	277.8±1.1	236.1±1.1
28G	290.0±0.9	172.0±0.9	54.0±0.8	97.0±2.0	201.5±0.8	273.1±1.3	240.7±0.8
pdt009	293.1	165.8	51.2	110.3	209.3	257.6	243.9

Table 2: Minor groove width in Å

Base pair	DNA alone			TBP-DNA			
	WT	29A	28G	WT	29A	28G	PDT009
A4-T11	8.1±0.1	8.7±0.1	9.5±0.2	11.0±0.2	11.2±0.3	10.3±0.4	9.7
T5-A10	7.8±0.1	8.4±0.1	8.7±0.2	12.0±0.1	12.2±0.1	11.9±0.1	11.2
A6-T9	8.9±0.3	7.8±0.3	8.7±0.3	12.0±0.1	12.4±0.1	11.9±0.1	12.3
A7-T8	8.1±0.1	7.7±0.3	8.7±0.3	12.1±0.3	12.4±0.1	13.1±0.1	12.9
A8-T7	7.6±0.3	7.4±0.2	7.9±0.3	11.5±0.2	12.0±0.2	12.4±0.1	12.3
A9-T6	7.0±0.2	7.3±0.1	7.5±0.2	12.0±0.1	12.2±0.1	12.1±0.1	12.0
G10-C5	6.8±0.2	7.4±0.3	7.6±0.2	11.2±0.1	11.5±0.2	11.6±0.1	11.4
G11-C4	7.6±0.1	8.2±0.4	8.3±0.2	10.0±0.1	10.5±0.2	11.4±0.1	9.7
G12-C3	8.0±0.1	8.1±0.3	8.4±0.2	7.3±0.1	8.2±0.2	9.8±0.1	8.7
C13-G2	7.8±0.2	7.7±0.2	8.1±0.3	6.7±0.1	7.5±0.2	8.7±0.1	8.6
A14-T1	7.8±0.2	7.6±0.2	7.7±0.3	7.5±0.2	7.7±0.1	8.7±0.2	9.0
A-DNA	11.2						
B-DNA	5.9						

Minor Groove Width

It has been shown by X-ray crystallographic results that when TBP binds to DNA the minor groove width of DNA is increased. We have calculated the minor groove width of DNA in free and complex simulations by measuring the smallest P-P distance across the strands according to the procedure described by Hassan *et al.*³⁰ The results are tabulated in Table 2. The minor groove width of WT TBP-DNA crystal structure and canonical A and B-forms of DNA were also measured using the same procedure and listed in Table 2. The results showed that the minor groove width of free DNA varies between canonical A and B-DNA values. The A₆-tract in 29A DNA showed a clear decrease in minor groove width in the 5'-3' direction. Similar results were obtained by Lavery *et al.* in the MD simulation of the same 14-mer.¹⁵

DNA bending

The degree of TBP induced bending of the DNA duplex has shown to be a significant aspect of its transcription activity. We have calculated the degree of DNA bending as the angle between second and thirteenth base pair normal vectors using the FREEHELIX programme. The variation of bend angle as a function of time for three TBP-DNA simulations is shown in Figure 6. Figure 6 shows that the DNA duplex in three complexes remains bent throughout the simulation. The average bend angles of DNA in WT, 29A and 28G DNA simulation are 12.6 ± 0.3 , 16.0 ± 0.3 and 11.3 ± 0.2 and for the TBP-DNA simulations are 76.8 ± 0.2 , 75.0 ± 0.2 and 67.7 ± 0.3 respectively. The same bend angle in X-ray crystal structure of WT TBP-DNA complex⁶, measured using the FREEHELIX program is 74.

Interaction Energy

In order to understand how TBP interacts with DNA, we have calculated the interaction energy between the TBP and DNA. The results shown in Table 3 revealed that the interaction between DBD of TBP and TATA box region of DNA contributed 87.3% in WT, 87.6% in 29A and 97.3% in 28G to the total interaction energy which is calculated as the sum of electrostatic and van der Waals components of the energy. The remaining percentage of the interaction energy accounts for the rest. On average, the total interaction energy between DBD and TATA box comprises 91% electrostatic energy and 9% van der Waals energy. These results suggested that the long-range electrostatic interactions play an important role in stabilizing the TBP-DNA complex. As observed in X-ray crystallography, the total van der Waals interaction energy between TBP and DNA is primarily due to the interaction between DBD and TATA box region. The calculated percentage van der Waals interaction energy between DBD and TATA region in WT, 29A and 28G are 82, 83 and 81% respectively. The calculated total interaction energy between DBD and TATA box region of 29A and 28G are 167 and 2 kcal/mol larger than WT. The total interaction energy between

TBP and DNA revealed that the TBP-DNA interaction in WT is favoured over 29A and 28G by 195 and 140 kcal/mol respectively. This extra stabilization gained by the WT over the 29A and 28G contributes favourably to the overall free energy of formation of the complex.

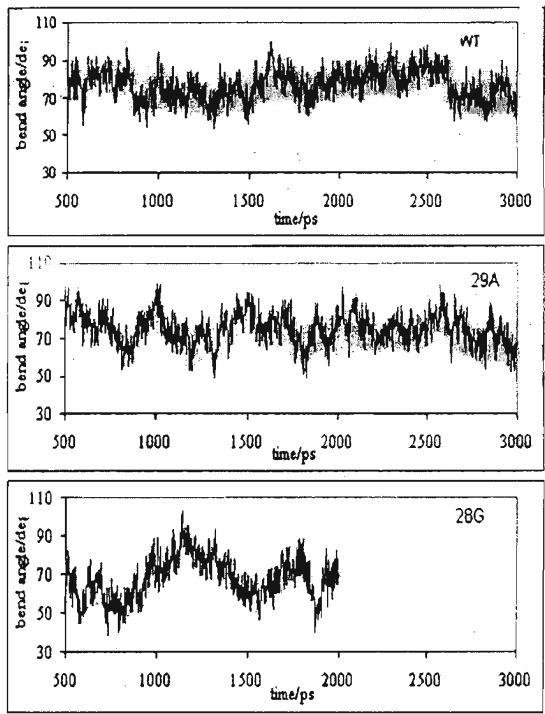


Figure 6: The variation of bend angle as a function of time for TBP-DNA simulations

Table 3: Interaction energy (kcal mol⁻¹) between the DNA and the TBP. The standard error is given as a subscript.

	ΔE^a				ΔE^b			
	Total	TATA-DBD ELEC	VDW		Total	TBP-DNA ELEC	VDW	
WT	-1244 ₂₄	-1144 ₂₄	-100.3 _{0.7}	124	-1338 ₁₈	-1221 ₁₉	-117.1 _{1.1}	195
29A	-1120 ₄₁	-1018 ₄₀	-102.3 _{0.7}		-1143 ₃₆	-1026 ₃₅	-117.4 _{0.9}	
28G	-1199 ₂₈	-1088 ₂₇	-110.5 _{0.8}	45	-1198 ₂₅	-1070 ₂₅	-127.5 _{0.8}	14

^{a,b} total interaction energy difference between mutant and wild type complexes DBD DNA binding domain

Table 4: Interaction energy (kcal mol⁻¹) between the side chain and backbone of the TBP and the TATA box region. Standard error is given as a subscript.

	WT			28G			29A		
	total	e lec	vdw	total	elec	vdw	Total	elec	vdw
Base-bkb ^a	-80.2 _{9.1}	-75.7 _{9.2}	-4.6 _{0.2}	-6.4 _{15.7}	-1.7 _{15.7}	-4.7 _{0.1}	-54.3 _{8.4}	-48.7 _{8.3}	-5.6 _{0.2}
Base-sdc ^a	-64.0 _{5.0}	-27.7 _{4.5}	-36.2 _{0.4}	-33.9 _{7.8}	2.5 _{7.6}	-36.4 _{0.3}	-94.8 _{4.2}	-50.2 _{4.6}	-44.6 _{0.5}
Total	-144 ₇	-103 ₈	-40.80 _{0.5}	-40.2 ₁₈	0.9 _{18.4}	-41.1 _{0.3}	-149 ₁₀	-98.8 ₁₀	-50.2 _{0.5}
Phos-bkb ^a	-9.4 _{11.9}	-5.3 _{11.9}	-4.1 _{0.1}	19.7 _{22.3}	23.6 _{22.7}	-3.9 _{0.3}	-4.2 _{24.0}	0.8 ₂₄	-4.9 _{0.1}
Phos-sdc ^a	-1313 ₂₄	-1297 ₂₄	-16.4 _{0.3}	-1262 ₁₆	-1246 ₁₆	-15.8 _{0.6}	-1264 ₂₉	-1247 ₂₈	-17.2 _{0.5}
Total	-1323 ₂₀	-1302 ₂₁	-20.5 _{0.4}	-1242 ₂₂	-1222 ₂₂	-19.7 _{0.7}	-1268 ₂₄	-1246 ₂₄	-22.1 ₉
Suga-bkb ^a	-12.8 _{2.7}	0.1 _{2.7}	-12.8 _{0.2}	-21.7 _{8.5}	-9.2 _{8.5}	-12.4 _{0.2}	33.2 _{7.5}	46.1 _{7.5}	-12.9 _{0.3}
Suga-sdc ^a	180 ₄	219 ₅	-39.0 _{0.4}	184 ₅	224 ₅	-40.4 _{0.3}	181 ₉	220 ₁₀	-39.1 _{0.5}
Total	167 ₄	219 ₄	-51.8 _{0.4}	162 ₇	215 ₈	-52.8 _{0.2}	214 ₉	266 ₉	-52.0 _{0.7}

^a bkb-backbone, sdc-side chain

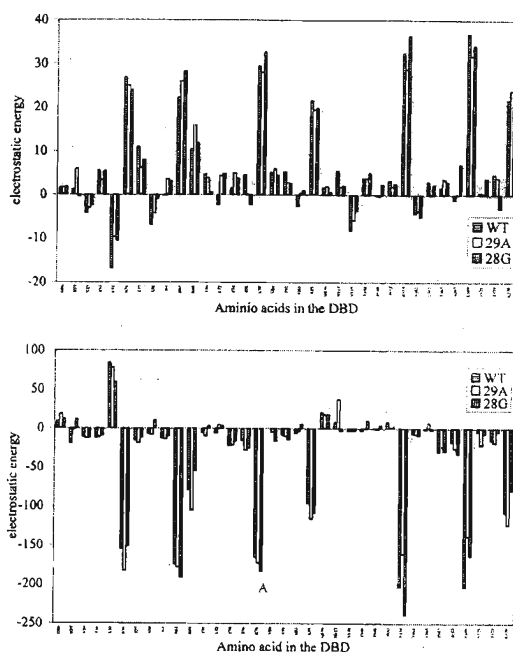


Figure 7: The electrostatic interaction energy (in kcal/mol) between amino acids in the DBD and A) phosphate atoms and B) sugar atoms in the TATA box

To get further insight into the DBD and TATA box interaction, the interaction energy was calculated by dividing the DBD into side-chain and backbone and TATA box into minor groove, major groove, phosphate and sugar regions. The results given in Table 4 showed that the electrostatic interaction between phosphate atoms in the TATA box region and side-chain of amino acids in the DBD dominates the total interaction energy. The sugar atoms in the TATA box showed repulsive interaction with the side-chain amino acids. The results also showed that minor and major groove, phosphate and sugar atoms interact favourably with the side-chain amino acid atoms via van der Waals interactions.

Further analysis of interaction between DNA and TBP was carried out by calculating the interaction energy between the TATA box region and each amino acid in the DBD. Figure 7 shows the electrostatic interaction energy between phosphate and sugar atoms of the TATA box and each amino acid in the DBD. Only nine out of 35 amino acids in the DBD of TBP interact significantly with the phosphate atoms. The repulsive interaction between Glu51 and phosphate is due to the two negatively charged groups. Arg56, arg63, lys78, lys85, arg154, lys169 and lys176 showed attractive interactions with phosphate atoms and are due to the positively charged N atoms in arginine and lysine residues in DBD and negatively charged O atoms in phosphate. The second major contribution to the electrostatic

interaction energy comes from sugar and amino acid residues in DBD. The interaction between backbone sugar atoms and the amino acid atoms is repulsive in all three complexes. It is interesting to note out that the same nine residues also interact with the sugar atoms with opposite signs. It is important to note here that any mutation carried out in the TATA box would not change the electrostatic interactions discussed above. As shown in Figure 7, almost similar electrostatic interaction energy patterns were observed for all three systems investigated in this study.

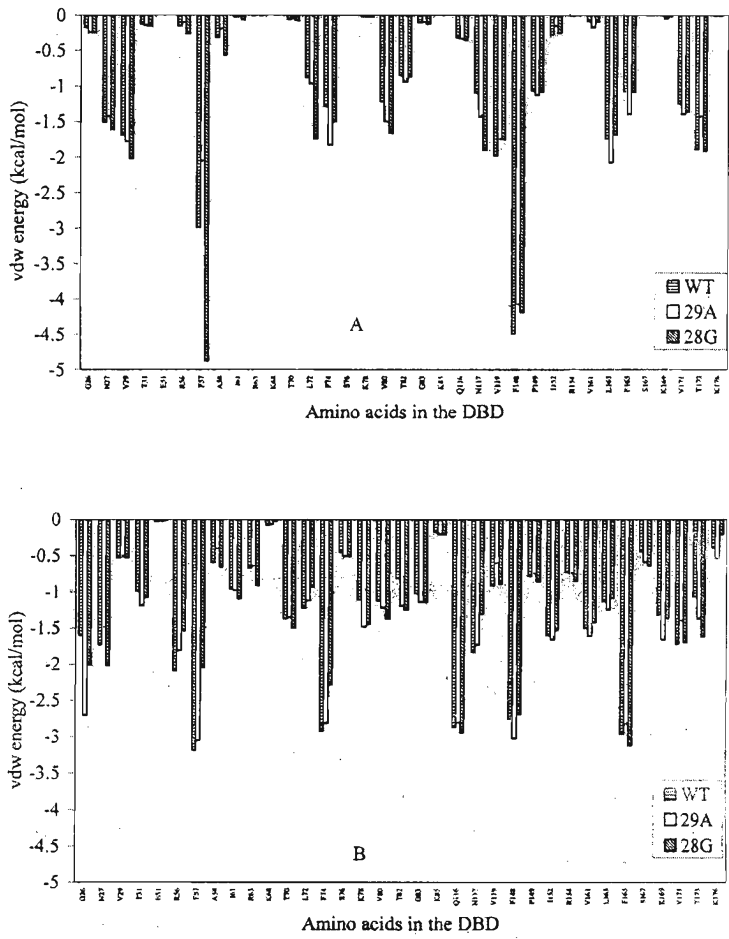


Figure 8: Van der Waals interaction energy between amino acids in the DBD and A) minor groove atoms and B) sugar atoms in the TATA box.

As described above, amino acid residues in the DBD made van der Waals contact with all four regions of TATA box. Figure 8 shows the distribution of two major contributions, interaction between minor groove and sugar atoms with each amino acid in the DBD. However, as mentioned above this energy is only a small part of the total interaction energy. The comparison of van der Waals interaction

energies of WT, 29A and 28G is shown in Figure 9. The comparison of van der Waals interaction energy between amino acid residues and base atoms in minor and major groove is particularly important because this result can be used to investigate the sequence specific interactions. The results presented in Figure 9 revealed that the amino acid residues in the DBD of TBP interact similarly with three TATA box sequences except for Phe57. These results, therefore, suggest that van der Waals interaction between DBD and TATA box is not sequence specific.

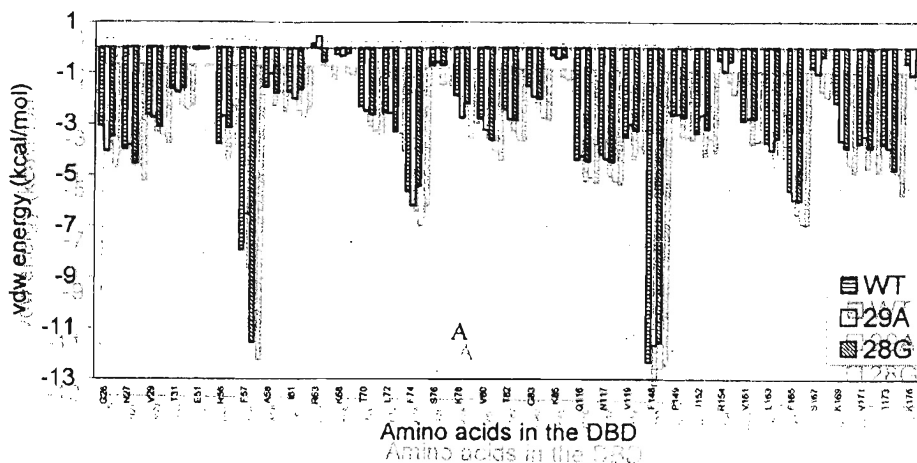


Figure 9: Van der Waals interaction energy between amino acids in the DBD and TATA box region.

Hydration Properties

Water plays a key role in DNA-protein interaction. It has been shown that the interface between TBP and DNA is very hydrophobic.^{6,9} In order to understand the activity of water hydrating the DNA in the MD simulation of naked DNA and TBP-DNA complex, the diffusion constants, hydration numbers and rotational correlation times were calculated. First, the radial distribution function of water oxygen and the solute site were obtained to determine the position of hydration shells around a given solute site. The results showed only one hydration shell located at 3.5 Å around the minor groove and major groove of TATA region of both complexed and uncomplexed DNA. One sharp peak at 3.3 Å and one broad peak at 6.5 Å were found between water oxygen and phosphate and sugar atoms. Hydration properties were calculated for the TATA box region of both DNA and TBP-DNA. The hydration numbers are presented in Table 5. Comparison of the results shown in Table 5 clearly show that the interface formed between the minor groove of TATA binding region and DBD is almost dehydrated. The average hydration number of water around backbone sugar is 1.3 in the TATA box region of DNA and this number is reduced to about 0.4 in TBP-DNA suggesting that dehydration also occurred around sugar

atoms during the complex formation. A small decrease in hydration number around major groove and phosphate backbone atoms was observed.

It is expected that the formation of the TBP-DNA complex would decrease the mobility of water around DNA. The diffusion constants calculated for the water molecules in the first hydration shell are given in Table 6. The results indicated that the mobility of water is decreased when TBP is bound to the DNA.

Table 5: Hydration number of water around TATA box region of DNA

Site	DNA			DNA-TBP		
	WT	29A	28G	WT	29A	28G
Minor	1.27±0.01	1.25±0.01	1.32±0.01	0.30±0.01	0.26±0.02	0.28±0.02
Major	1.19±0.01	1.25±0.01	1.23±0.01	1.10±0.02	1.22±0.02	1.11±0.01
Phosphate	2.20±0.01	2.20±0.01	2.20±0.01	2.03±0.01	2.08±0.00	2.04±0.01
Sugar	1.27±0.01	1.26±0.01	1.30±0.01	0.35±0.02	0.36±0.02	0.39±0.02

Table 6: Diffusion constants of water around aTATA box region of DNA

Site	DNA			DNA-TBP		
	WT	29A	28G	WT	29A	28G
Minor	0.20±0.03	0.28±0.04	0.31±0.03	0.07±0.02	0.07±0.03	0.11±0.06
Major	0.31±0.03	0.41±0.10	0.34±0.03	0.22±0.03	0.18±0.03	0.15±0.02
Phosphate	0.39±0.01	0.39±0.02	0.37±0.02	0.29±0.02	0.30±0.01	0.29±0.02
Sugar	0.22±0.04	0.31±0.04	0.28±0.03	0.16±0.06	0.25±0.04	0.23±0.06

The rotational correlation time of water within the first hydration shell of the minor and major grooves, phosphate and sugar was calculated and the results are given in Table 7. In general, all three rotational correlation times in the TBP-DNA complexes were increased as compared to the free DNA. The large increased in the rotational correlation time in the first hydration shell of minor groove and sugar indicated the involvement of water-mediated interactions between DNA and TBP.

CONCLUSIONS

- Experimental studies by Starr and co-workers¹¹ found that the magnitude of DNA bending is proportional to both the stability of the DNA-TBP complex and the level of transcriptional activity. The calculated average bend angle of WT DNA-TBP complex in the present study lies between the values obtained from the X-ray crystal structure and gel electrophoresis. Comparison of RMS differences and bend angles from the MD simulations for the three DNA-TBP complexes show the WT and two mutant DNA structures to be similar, although a small decrease in bending is calculated. The calculated bend angle for the 29A complex is larger than the experimental value from gel electrophoresis. This result and the measured lifespan of the 29A complex being 1 minute indicates that the 3ns simulation time is not enough to observe complete relaxation of the mutant complexes.

Table 7: Rotational correlation times (ps) of water around TATA box region

	DNA			TBP-DNA		
	wag	twist	rock	wag	twist	rock
WT						
Minor	3.30±0.03	6.93±1.16	2.91±0.27	6.84±0.77	16.96±2.25	6.50±0.75
Major	2.65±0.11	3.99±0.23	2.12±0.09	5.39±0.87	7.20±0.74	4.17±0.51
Phosphate	2.17±0.05	3.15±0.12	1.72±0.05	3.26±0.19	4.82±0.15	2.59±0.0.12
Sugar	3.36±0.41	6.95±1.20	2.93±0.38	6.41±1.19	14.21±2.28	6.13±1.27
29A						
Minor	3.16±0.16	5.73±0.32	2.71±0.13	15.28±4.21	24.00±3.40	12.39±3.05
Major	2.53±0.08	3.83±0.11	2.04±0.08	3.66±0.21	5.72±0.25	3.10±0.19
Phosphate	2.15±0.03	3.08±0.05	1.67±0.02	2.92±0.07	4.39±0.08	2.34±0.05
Sugar	3.12±0.16	5.52±0.26	2.63±0.11	5.33±0.65	9.84±0.62	4.61±0.46
28G						
Minor	2.97±0.19	5.58±0.46	2.60±0.17	7.73±0.37	17.09±1.96	7.18±0.41
Major	2.33±0.08	3.37±0.17	1.88±0.08	4.33±0.18	7.00±0.21	3.44±0.08
Phosphate	2.14±0.03	3.17±0.05	1.72±0.03	3.39±0.09	5.00±0.10	2.72±0.05
Sugar	2.96±0.18	5.50±0.43	2.59±0.16	5.81±0.38	12.40±0.83	5.36±0.36

- 2 Analysis of the DNA-TBP interaction energies shows them to be most favourable for the WT as compared to both mutants. The less favourable interaction energies of the mutants are suggested to lead to their decreased binding and transcriptional activity. It is also viewed as being consistent with the decreased bending observed experimentally.
- 3 Component analysis of the DNA-TBP interaction energies indicates that the contributions to the less favourable energetics in the mutants are delocalized over both the DNA and the protein. This result suggests that the specificity of the TBP for the TATA box is due to the sum of collective interactions rather than specific interactions involving individual nucleotides and/or protein amino acid residues.
- 4 Patikoglou and co-workers¹⁰ have predicted that the introduction of a G-C base pair at the fourth position of the TATA box would interfere with the contact between the Val119 and the minor groove surface due to the exocyclic amino group of the guanine located in the minor groove. However, we did not observe unfavourable contacts in the 28G simulation suggesting that the flexibility of the TBP allows for similar protein-DNA contacts as in the WT complex.

Reference

- 1 Sentenac A. (1985). Eukaryotic RNA polymerases *CRC. Critical Reviews of Biochemistry*, **18**: 31.
- 2 Roeder R. (1991). The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly, *Trends Biochem Sci.*, **16** : 402.
- 3 Hahn S., Buratowski S., Sharp P. A. & Guarente L. (1989). Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and non-consensus DNA sequences. *Proceedings to the National Academy of Science, USA*, **86**: 5718.
- 4 Wobbe C. R. & Struhl K. (1990). Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription *in vitro* *Mol. Cell. Bio.*, **10**: 3859.
- 5 Colgan J. A. M. & J. L. (1992). TFIID can be rate limiting *in vivo* for TATA containing but not TATA lacking RNA polymerase II promoters. *Gens Dev.* **6**: 304.
- 6 Kim J. L., Nikolov D. Hassan, M. A. E. & Calladine C. R. (1998). Two distinct modes of protein-induced bending in DNA. *Journal of Molecular Biology* **282**,

- 331B. Burley, S. K. (1993). Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature* **365**: p. 520.
- 7 Kim Y., Geiger J. H., Hahn S. & Sigler P. B. (1993). Crystal structure of a yeast TBP/TATA-box complex. *Nature* **365**: p. 512.
- 8 Juo Z. S., Chiu T. K., Leiberman P. M., Baikalov I., Berk A. J. & Dickerson R. E. (1996). How proteins recognize the TATA box. *Journal of Molecular Biology* **261**: p. 239.
- 9 Nikolov D. B., Chen H., Halay E., Hoffmann, A., Roeder R. & Burley S. (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. *Proceedings of the National Academy of Science USA* **93**: p. 4862.
- 10 Patikoglou G. A., Kim J. L., Sun L., Yang S., Kodadek T. & Burley S. K. (1999). TATA element recognition by the TATA box-binding protein has been conserved throughout evolution. *Genes Dev.* **13**: p. 3217.
- 11 Starr D. B., Hoopes B. C. & Hawley D. K. (1995). DNA bending is an important component of site specific recognition by the TATA binding protein. *Journal of Molecular Biology* **250**: p. 434.
- 12 Parkhurst K. M., Parkhurst L. J., Wu J., Brenowitz M. & Powell R. M., (2001). DNA bends in TATA-binding protein-TATA complexes in solution are DNA sequence dependent. *J. Biol. Chem.*, **276**: p. 14614.
- 13 Flatters D., Young M., Beveridge D. L. & Lavery R. (1997). Conformational properties of the TATA-box binding sequence of DNA. *Journal of Biomolecular Structure Dynamics* **14**: p. 757.
- 14 de Souza N. & Ornstein R. L. (1998). Inherent DNA curvature and flexibility correlate with TATA box functionality. *Biopolymers*, **46**: 403.
- 15 Flatters D. & Lavery R. (1998). Sequence dependent dynamics of TATA-box binding sites. *Biophysical Journal* **75**: p. 372.
- 16 Pastor N., Pardo L. & Weinstein H. (1997). Does TATA matter? A structural exploration of the selectivity determinants in its complexes with TATA box-binding protein. *Biophys. J.* **73**: p. 640.
- 17 Miaskiewicz K. & Ornstein R. (1996). DNA binding by TATA-box binding protein (TBP): A molecular dynamics computational study. *J. Biomol. Struct. Dynam.* **13**: p. 593.

- 18 Jorgensen W. L., Chandrasekhar J., Madura J. D., Impey R. W. & Klein M. L. (1983). *Journal of Chemical Physics*, **79**: p. 926.
- 19 Brooks B. R., Brucoleri R. E., Olafson B. D., States D. J., Swaminathan S. & Karplus M. (1983). CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *Journal of Computational Chemistry* **4**: 187.
- 20 MacKerell A. D., Jr. Brooks B., Brooks C. L., III, Nilsson L., Roux B., Won Y. & Karplus M. (1998). *CHARMM: The Energy Function and Its Parameterization with an Overview of the Program*; Schleyer, P. v. R., Ed.; Wiley: New York.
- 21 Foloppe N. & MacKerell Jr, A. D. (2000). All atom empirical force field for nucleic acids 1) Parameter optimization based on small molecule and condensed phase macromolecular target data, *J. Comput. Chem.*, **21**: 86.
- 22 Ryckaert J. P., Ciccotti G. & Berendsen H. J. C. (1977). *J. Comput. Phys.* **23**: 327.
- 23 Field M. J. & Karplus M. (1992). *CRYSTAL: Program for Crystal Calculations in CHARMM*; Harvard University: Cambridge, MA
- 24 Darden T. A., York D. & Pedersen L. G. (1993). Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems, *J. Chem. Phys.* **98**: 10089.
- 25 Dickerson R. E. (1998). DNA bending: the prevalence of kinetics and the virtues of normality *Nucleic Acids Research* **26**: 1906.
- 26 Goodsell D. S., Grzeskowiak M. K. & Dickerson R. E. (1994). The crystal structure of C-C-A-T-T-A-A-T-G-G. Implications for bending of B-DNA at TA steps, *J. Mol. Biol.*, **239**: 79.
- 27 Guzikevich-Guerstein & Shakked Z. (1996). A novel form of the DNA double helix imposed on the TATA-box by the TATA-binding protein, *Nature Structural Biology* **3**: 32.
- 28 Crothers D. M. (1998). DNA curvature and deformation in protein-DNA complexes: A step in the right direction. *Proceedings of National Academy of Science USA*, **95**: 15163.
- 29 Nikolov D. B., Chen H., Halay E. D., Usheva A. A., Hisatake K., Lee D. K., Roeder R. G. & Burley S. K. (1995). Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature*, **377**: 119.

- 30 Hassan M. A. E. & Calladine C. R. (1998). Two distinct modes of protein-induced bending in DNA. *Journal of Molecular Biology*, **282**: 331.