Crystal structure of a human TATA box-binding protein/TATA element complex

Dimitar B. Nikolov*, Hua Chen*†, Elaine D. Halay*†, Alexander Hoffmann‡§, Robert G. Roeder‡, and Stephen K. Burley*†¶

*Laboratories of Molecular Biophysics, ‡Laboratory of Biochemistry and Molecular Biology, and †Howard Hughes Medical Institute, The Rockefeller University, 1230 York Avenue, New York, NY 10021

ABSTRACT   The TATA box-binding protein (TBP) is required by all three eukaryotic RNA polymerases for correct initiation of transcription of ribosomal, messenger, small nuclear, and transfer RNAs. The cocrystal structure of the C-terminal/core region of human TBP complexed with the TATA element of the adenovirus major late promoter has been determined at 1.9 Å resolution. Structural and functional analyses of the protein–DNA complex are presented, with a detailed comparison to our 1.9-Å-resolution structure of *Arabidopsis thaliana* TBP2 bound to the same TATA box.

Eukaryotes have three distinct RNA polymerases (forms I, II, and III) that catalyze transcription of nuclear genes (1). Despite their structural complexity, these multisubunit enzymes require sets of auxiliary proteins known as general transcription initiation factors to initiate transcription from corresponding class I, II, and III nuclear gene promoters (2–5). TATA box-binding protein (TBP), first identified as a component of the class II initiation factor TFIID, participates in transcription by all three nuclear RNA polymerases (reviewed in ref. 6). TBP has a phylogenetically conserved, 180-residue C-terminal portion, containing two direct repeats flanking a highly basic segment known as the basic repeat. The C-terminal/core region of TBP binds to the TATA consensus sequence (TATAa/tAa/t) with high affinity and slow off rate, recognizing minor groove determinants and inducing a dramatic DNA deformation. The N-terminal portion of TBP varies in length, shows little or no conservation among different organisms, and is largely unnecessary for transcription in certain yeast strains.

The role of TBP in transcription initiation and its regulation are best understood for genes transcribed by RNA polymerase II (reviewed in refs. 3 and 4). In this setting, TBP is tightly associated with other polypeptides known as TBP-associated factors (reviewed in ref. 6). This multiprotein complex (TFIID) is a general initiation factor (7) that binds to the TATA element, coordinating accretion of class II initiation factors (TFIIA, -B, -D, -E, -F, -G, -H, and -I) and RNA polymerase II (pol II) into a functional preinitiation complex (PIC) (reviewed in refs. 3 and 8). Although incapable of mimicking TFIID *in vivo* (at least in higher eukaryotes; reviewed in ref. 6), recombinant TBP alone is competent for PIC assembly and basal transcription in the presence of the other general class II factors (9). TBP engages in physical and functional interactions with the general initiation factors TFIIA and TFIIB, the C terminus of the large subunit of pol II, some negative cofactors (NC1, NC2, DR1) that inhibit PIC formation, some transcriptional activators, and an initiator-binding factor (TFII-I) that may be important for transcription initiation from TATA-less promoters.

Recently, we reported the x-ray crystal structures of TBP isoform 2 (TBP2) from *Arabidopsis thaliana* (10, 11), a complex of TBP2 with the TATA element of the adenovirus major late promoter (AdMLP) (12, 13), and a ternary complex of TBP2 with human TFIIB and the AdMLP TATA box (14). TBP2 displays a novel, two-domain DNA-binding fold, and the protein's approximate twofold intramolecular symmetry generates a saddle-shaped structure dominated by a curved antiparallel β-sheet, forming a concave DNA binding surface. TATA element recognition proceeds via an induced-fit mechanism, involving subtle conformational changes at the level of the protein and an unprecedented DNA distortion. The seat of the saddle presents a large convex surface to which other proteins bind during transcription initiation. TFIIB recognizes the preformed TBP–DNA complex via protein–protein and protein–DNA interactions. The N-terminal portion of TFIIB forms the downstream surface of the ternary complex, where it could fix the transcription start site. The remaining surfaces of TBP and the TFIIB can interact with TBP-associated factors, other class II initiation factors, and transcriptional activators and coactivators.

We now report the crystal structure of human C-terminal/core TBP (hTBPc) recognizing the AdMLP TATA element. The 1.9-Å-resolution structure permits detailed structural and functional analyses of the protein–DNA interactions and provides further insights into how TBPs from different organisms recognize and direct transcription initiation from TATA-containing promoters.

MATERIALS AND METHODS

Crystallization. hTBPc residues 155–335 (15) fused with an additional 20 N-terminal amino acids (MGSSHHHHHHSS-GLVPRGSH) was overexpressed in *Escherichia coli* [BL21-(DE3)pLysS] using the T7 RNA polymerase system (16). Cells grown at 30°C to an absorbance of 0.7 at 595 nm and induced with 0.5 mM isopropyl β-D-thiogalactopyranoside for 3 h were harvested by low-speed centrifugation. Lysis was performed by three cycles of freeze–thaw and DNA digestion was performed by DNase I. The soluble fraction was applied directly to a Ni^{2+} ion affinity column and washed with a buffer containing increasing amounts of imidazole. hTBPc with an estimated homogeneity of 95% was then eluted from the resin with a buffer containing 100 mM EDTA. After overnight dialysis, removal of the histidine-containing N-terminal sequence was
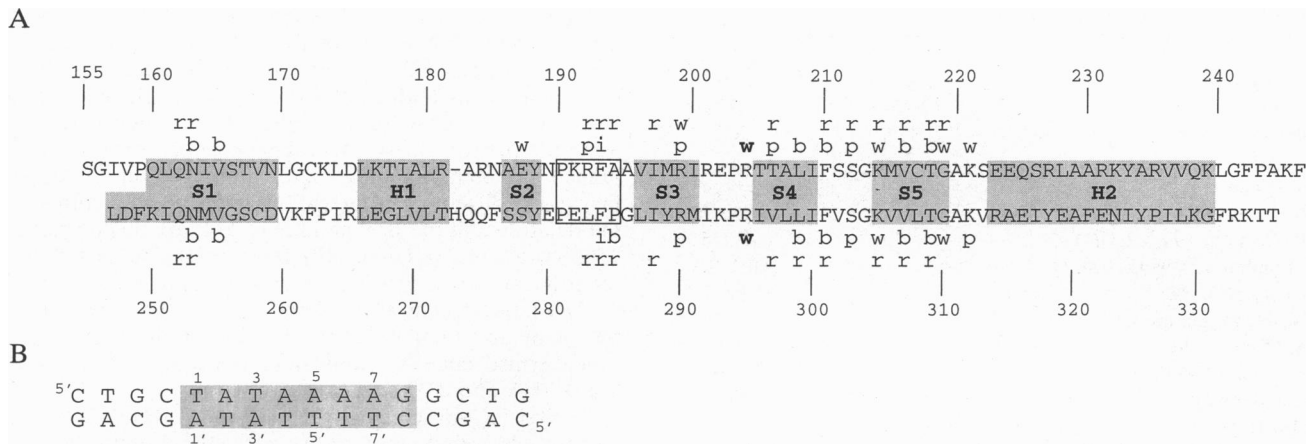
A

B

FIG. 1.   (*A*) Sequence of the 180-amino acid conserved C-terminal region of human TBP with the two direct repeats aligned one above the other. α-Helix and β-strand regions are shaded and labeled (' denotes second direct repeat), and the two "stirrups" are boxed. Residues interacting with DNA are labeled according to interaction type: b, base; p, phosphate; r, ribose; i, intercalating phenylalanine; w, water mediated. Protein numbering is that of full-length human TBP (15) and corresponds to the *A. thaliana* numbering plus 136 and *Saccharomyces cerevisiae* numbering plus 94. A full conversion scheme can be found in ref. 10. (*B*) Sequence of 16-base-pair oligonucleotide used for cocrystallization, with the 8-base-pair TATA box of the AdMLP shaded. DNA numbering scheme is identical to that used in ref. 13. Position 1 of the TATA box corresponds to base pair −31 in the AdMLP, and ' denotes bottom strand.

performed by trypsin cleavage of the peptide bond between R and G in the above sequence to yield hTBPc-(155–335) plus 4 N-terminal amino acids—namely, GSHM. A final cation-exchange chromatography step to remove the protease and any remaining contaminants yielded material of homogeneity estimated to be 98% or greater. Mass spectrometry documented that the TBP used for crystallization was neither modified nor further proteolyzed during expression and purification. The measured molecular mass was 20,829 ± 5, compared with 20,825 calculated from the predicted amino acid sequence. Further control experiments were performed to ensure that the protein used for the crystallographic study was biochemically active and indistinguishable from recombinant yeast TBP. Gel-retardation analyses, performed as described (11), confirmed that the purified protein had full TBP DNA-binding activity (data not shown). Photon correlation spectroscopy (PCS) documented that TBP is a dimer in solution in the absence of DNA and undergoes a dimer-to-monomer transition upon binding to oligonucleotides containing the TATA-consensus sequence (data not shown). The PCS experiments were performed with a dp801 molecular size detector (Molecular Solutions, Charlottesville, VA) under a wide range of experimental conditions. hTBPc was monodisperse and dimeric in the absence of DNA. Addition of a duplex oligonucleotide containing the TATA-consensus sequence yielded a hTBPc–DNA complex consisting of 1 molecule of TBP and 1 molecule of duplex oligonucleotide.

The crystallization DNA (Fig. 1) was synthesized and purified as described (12), and the hTBPc–DNA complex was prepared to a final concentration of 0.5 mM in 200 mM KCl/250 mM ammonium acetate/5 mM MgCl₂/5 mM CaCl₂/10 mM dithiothreitol/10% (vol/vol) glycerol/2% (vol/vol) ethylene glycol/40 mM Tris·HCl, pH 8.5. The complex is relatively insoluble under lower salt conditions, enabling crystals to be grown at 4°C by vapor diffusion against a reservoir containing 8–12% (vol/vol) glycerol, 20 mM dithiothreitol, 40 mM Tris·HCl, pH 8.5. After macroseeding, crystals with typical dimensions of 0.6 × 0.4 × 0.05 mm³ grew in space group P2₁2₁2 (*a* = 45.8 Å, *b* = 78.0 Å, *c* = 97.4 Å), containing one protein–DNA complex per asymmetric unit.

Data Collection and Structure Determination. Diffraction measurements were carried out at −150°C using a refrigerated nitrogen gas delivery system. The data were collected using CuKα radiation from a rotating anode x-ray source equipped with a Rigaku RAXIS-IIc imaging plate area detector. Crystals

were transiently "cryoprotected" in a 10% glycerol/25% ethylene glycol/1 mM MgCl₂/5 mM Tris·HCl, pH 8.5, mixture before freezing on a platform created by a loop of ophthalmological suture material (10–0 Ethilon, Ethicon). Oscillation photographs were integrated, scaled and merged using DENZO and SCALEPACK. Molecular replacement (MR) was carried out with X-PLOR (17) using the structure of the TBP2–DNA complex (12, 13). Patterson correlation refinement of the rotation function search maxima yielded a solution with correlation coefficient 0.216. A translation search in X–Y–Z, gave a final MR solution 14.2σ above the mean peak level. Rigid body refinement of the search model, followed by positional and simulated annealing refinement with X-PLOR, gave R_{factor} = 0.26 at 3.0 Å resolution. At this point, $(2|F_{observed}| - |F_{calculated}|)$ and $(|F_{observed}| - |F_{calculated}|)$ difference maps were calculated and the correct amino acids for the human TBP and the full DNA model were built using O (18). Subsequently, the resolution limit was increased in 0.2-Å steps with positional refinement and, beyond 2.6 Å, tightly restrained individual B-factors were refined for each nonhydrogen atom. After simulated annealing at 2.2 Å resolution, missing side chains and tightly bound water molecules were built into appropriate electron density features, and the resolution limit was extended to 1.9 Å. No restraints were placed on sugar pucker, DNA backbone torsion angles, or Watson–Crick hydrogen bonding. In addition, charges on all atoms were set to zero in order to minimize bias from the Coulomb potential. The final refinement model has a crystallographic R_{factor} of 18.9% (Table 1) and consists of amino acids 152–334, the 16-base-pair oligonucleotide, and 285 water molecules. Average temperature factors are 31 Å² for the protein, 35 Å² for the DNA, and 44 Å² for the water molecules. The electron density for the polypeptide backbone is everywhere continuous at 1.2σ in a $(2|F_{observed}| - |F_{calculated}|)$ difference Fourier synthesis. PROCHECK (19) revealed no disallowed backbone torsion angle combinations and main-chain and side-chain structural parameters consistent with or better than those expected at 1.9 Å resolution (overall *G* factor = 0.2). The DNA electron density is well defined throughout.

RESULTS

Fig. 2 illustrates the structure of the hTBPc–TATA element complex, which is very similar to the *A. thaliana* (12, 13) and yeast (21) cocrystal structures. The rms deviation (rmsd)

Table 1. Crystallographic summary

Resolution, Å	18.0–1.90	1.97–1.90
Observations	214,715	7,700
Reflections	28,256	2,758
Completeness, %	98.6	90.3
R_{sym}, %	5.4	22.7
Refinement		
Resolution, Å		6.0–1.90
Reflections ($|F| > 2\sigma\,|F|$)		26,476
Completeness work/test, %		88.4/8.7
R_{factor}/R_{free}, %		18.9/25.8
Nonhydrogen atoms		2,375
H_2O molecules		285
rms bond length, Å		0.010
rms bond angle, °		1.68
rms B-factor bonded atoms		
main chain/side chains, Å^2		2.06/2.74

Analysis of the 1.9-Å-resolution refinement of the TBP–TATA box complex. Summary of data collection and refinement statistics. $R_{sym} = \Sigma|I - \langle I\rangle|/\Sigma I$, where I = observed intensity and $\langle I\rangle$ = average intensity from multiple observations of symmetry-related reflections. $R_{factor} = \Sigma||F_{observed}| - |F_{calculated}||/\Sigma|F_{observed}|$, where $|F_{observed}|$ = observed structure factor amplitude and $|F_{calculated}|$ = calculated structure factor amplitude from the refined model. R_{free} is calculated in the same way using 9% of the observed structure factor amplitudes selected at random. rms bond lengths and angles are the respective rmsd values from ideal values, and rms B-factor bonded atoms are the rmsd between B values of covalently bonded atomic pairs.

between the plant and human complexes for equivalent α-carbon and TATA box C1' atomic positions is 0.59 Å. The hTBPc–TATA box interface encompasses 8 base pairs of DNA (Fig. 1*B*) and the protein's concave surface (Figs. 1*A* and 2). Complex formation involves severe distortion of the TATA box (reviewed in ref. 13). The DNA is simultaneously unwound, doubly kinked, and smoothly bent toward the major groove, creating a greatly exposed minor groove surface that is complementary to TBP's concave DNA-binding surface. Despite approximate twofold symmetry of both TBP and many TATA elements, TBP and TFIID bind TATA boxes in a directional manner during transcription initiation. As in the plant and yeast TBP–DNA complex structures (12–14, 21), hTBPc orients itself on the DNA with its first and second direct repeats interacting with the 3' and 5' halves of the 8-base-pair TATA box, respectively.

Protein Structure. hTBPc shares 82% sequence identity with the C-terminal/core region of *A. thaliana* TBP2 and their structures are nearly identical (rmsd for equivalent α-carbon atomic positions = 0.60 Å). The two imperfect direct repeats of hTBPc adopt quasi-identical folds (Fig. 2, rmsd for equivalent α-carbon atomic positions = 0.92 Å). There are only two significant differences between the two halves of hTBPc, which

occur at the end of helix H1 (where there is a single amino acid deletion in the first repeat relative to the second) and at amino acid 326 in helix H2' (where a proline residue introduces a bend in the helix that is not present in H2). The single amino acid deletion appears to destabilize the turn between H1 and S2 in the first repeat, as this region is poorly ordered in both human and plant complexes and, depending on crystal packing, adopts different conformations in the two structures. Both the deletion and the proline in helix H2' are phylogenetically conserved (reviewed in ref. 10). It is remarkable that, although the sequence identity is higher for the first direct repeat (87% vs. 77%), the structural similarity between the core regions of the plant and human TBP is higher for the second direct repeat (rmsd values for equivalent α-carbon atomic positions are 0.70 Å and 0.27 Å, respectively).

Analysis of the solvent structure in the hTBPc–DNA complex reveals buried waters at two of the three sites found in the high-resolution structures of uncomplexed (10) and TATA-bound TBP2 (13). As predicted (13), the third water-filled cavity is not present in human TBP because it is obliterated by a substitution (Thr-317 → Leu) in the hydrophobic core of the second direct repeat.

DNA Structure. Figs. 2 and 3 illustrate the dramatically distorted structure of the oligonucleotide in the ternary complex. The structure of the 8-base-pair TATA box (TATA-AAAG), although somewhat A-like, differs significantly from both A- and B-form DNA. Watson–Crick base pairing is maintained throughout, despite being unwound by 117° over the 7-base steps of the recognition site. Unwinding is compensated by induction of ≈1/3rd of a turn of a right-handed superhelix, leaving the linking number unchanged and directing the DNA through a complicated path resulting in a bend of 80° (Fig. 3). The difference in overall DNA trajectory in the plant and human TBP cocrystal structures is due to slight differences in the kink at the 3' end of the TATA element (discussed below) and to DNA stacking within the hTBPc–DNA crystals. As in the plant and yeast TBP–DNA complex structures (12–14, 21), the flanking DNA on either side of the TATA box adopts canonical B-form.

A detailed analysis of the DNA conformation is presented in Table 2 with the corresponding values from our 1.9-Å-resolution refinement of the TBP2–DNA cocrystal structure (13). The rmsd between the 8-base-pair TATA elements in the high-resolution human and plant complex structures is 0.55 Å for all nonhydrogen atoms and 0.35 Å for C1' atoms. Severe bending of the TATA box and widening of its minor groove is accomplished by a combination of large, positive base-pair roll coupled with decreased twist (Table 2), resulting in a smooth bend of the helix toward the major groove. Within the central 6 base pairs of the TATA box, the DNA undergoes an ≈90° bend and 90° twist; 52° and 39° kinks at the first and last base steps of the TATA box direct the upstream and downstream

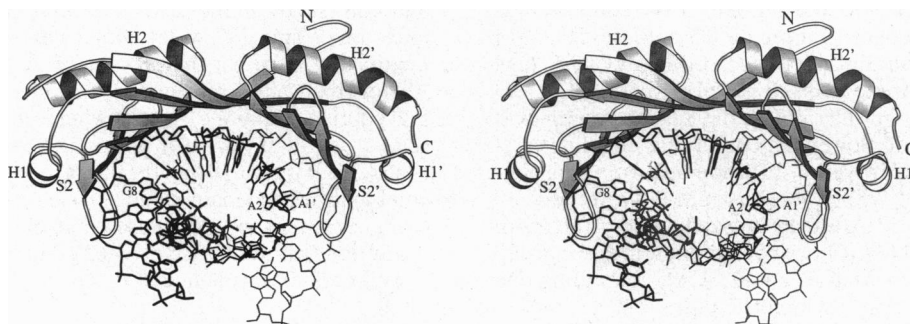FIG. 2. MOLSCRIPT (20) stereodrawing of the hTBPc–TATA box complex viewed perpendicular to the approximate twofold symmetry axis. hTBPc is displayed as a ribbon and DNA is shown as a stick. The coding strand of the DNA is drawn in black and the noncoding strand is gray. The 5' end of the TATA box enters from the right and interacts with the second direct repeat of TBP, whose secondary structural elements are denoted with '.
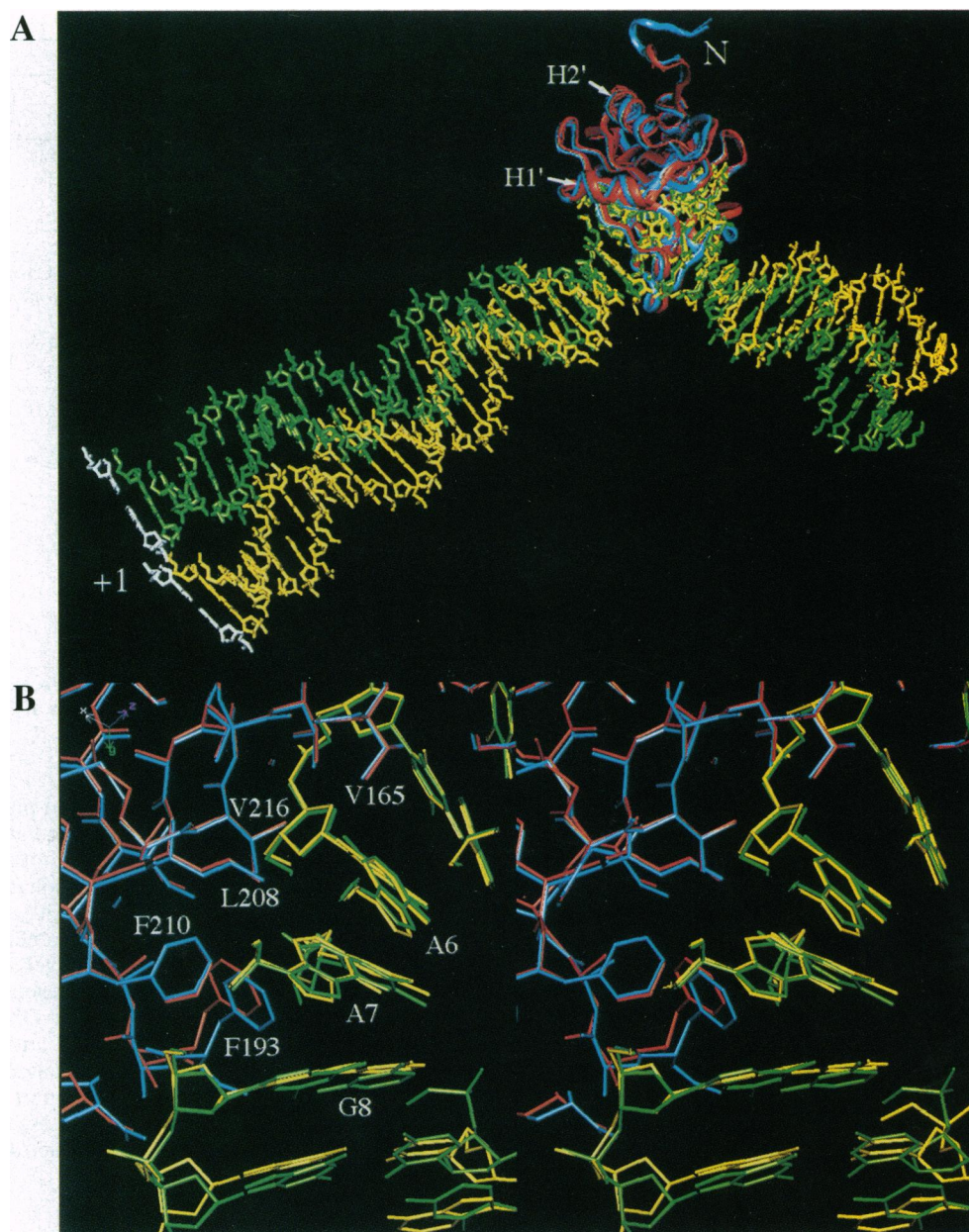
Cell Biology: Nikolov *et al.*

Proc. Natl. Acad. Sci. USA 93 (1996)   4865

FIG. 3. Superimposed hTBPc (red) and *A. thaliana* TBP2 (blue) interacting with the AdMLP (yellow and green, respectively). The C1′ atomic positions of the 6 central base pairs of the TATA box were used for alignment of the complexes. (*A*) View of the complex at 90° with respect to the axis of approximate intramolecular twofold symmetry from the side of TBP's second stirrup. DNA is depicted as a stick figure, with hypothetical, linear B-form DNA extensions modeled at both the 5′ and 3′ ends of the x-ray structure oligonucleotide. Transcription start site is identified as the white base pair labeled +1. Angle between incoming and outgoing DNA helices is ≈110°, when viewed in projection. (*B*) Stereoview of protein–DNA interface showing the most significant difference in TATA element binding—namely, Phe-193 inserting deeper in the ApG base step in the plant complex than in the human complex. Key protein residues and DNA bases are labeled.

DNA in opposite directions, imposing an overall bend of 80° and a lateral displacement of 18 Å in the trajectory of the DNA axis. Bending of the TATA box toward the major groove approximates adenine N^6 and thymine O^4 atoms from neighboring base pairs, forming a series of bifurcated hydrogen bonds in the 4-base-pair A-tract. Compression of the major groove is also stabilized by an extensive hydrogen-bonded network of water molecules, as seen with TBP2 (12, 13).

TBP–DNA Interactions and TATA Element Recognition. Fig. 1 identifies amino acids in hTBPc that interact with DNA. With the exception of Arg-204 and Arg-295 (which make water-mediated DNA backbone contacts, denoted with **w** in Fig. 1*A*), these residues are conserved among yeast, *A. thaliana*, and human (reviewed in ref. 10). Consequently, the hTBPc–DNA contacts are essentially identical to those seen in

the plant and yeast cocrystal structures (12, 13, 21). Once again, the TBP–TATA box minor groove interface is dominated by van der Waals interactions between nonpolar atoms and between nonpolar and polar atoms. Of the 3100 Å^2 of surface area buried at the interface, 74% is hydrophobic (protein, 79%; DNA, 70%), compared to 40–50% for most protein–DNA complexes (24). There is no crystallographic evidence of water molecules in the hTBPc–TATA box interface. Moreover, there are no obvious cavities within the interface. Seven of the 8 base pairs of the AdMLP TATA element protein participate in one or more minor groove base edge interactions with side chains projecting from the protein's antiparallel β-sheet (Fig. 2).

The most impressive side-chain–base interactions within the minor groove are provided by phenylalanine residues, Phe-284

Table 2. DNA parameters

	C G	T A	G C	C G	T A	A T	T A	A T	A T	A T	A T	G C	G C	G C	C G	T A	G G	B	A
Helical twist, °	34.3	30.7	33.7	35.7	**16.9**	**19.8**	**24.4**	**11.6**	**18.7**	**23.3**	**20.5**	**35.3**	35.7	33.4	45.0			36.0	30.7
TBP2–TATA			32.7	32.5	**18.3**	**19.9**	**23.7**	**14.0**	**19.2**	**22.8**	**29.0**	**31.2**	37.0	23.8	32.0				
Roll, °	7.0	11.0	1.1	2.0	**52.5**	**20.1**	**7.5**	**20.1**	**22.6**	**16.4**	**39.1**	**4.7**	−0.2	4.6	5.6			0.9	11.4
TBP2–TATA			−5.4	3.1	**40.2**	**21.4**	**11.4**	**25.2**	**22.0**	**16.0**	**44.8**	**4.4**	3.0	−1.9	2.8				
Rise, Å	3.06	3.31	3.44	3.38	**5.90**	**3.76**	**3.05**	**3.43**	**3.99**	**3.56**	**5.26**	**3.44**	3.61	3.27	3.29			3.38	3.44
TBP2–TATA			3.44	3.15	**5.36**	**3.93**	**3.06**	**3.33**	**3.82**	**3.52**	**5.74**	**3.28**	3.48	3.48	3.11				
Tilt, °	3.43	−3.86	−2.42	6.57	**−3.51**	**−0.61**	**0.47**	**−1.54**	**2.69**	**4.56**	**1.14**	**−1.75**	−0.88	0.94	−6.48			0.0	0.0
TBP2–TATA			0.68	4.57	**0.19**	**2.26**	**4.20**	**−1.98**	**−0.97**	**2.01**	**−3.65**	**−1.85**	−0.79	1.95	4.17				
Slide, Å	−0.69	−0.35	−1.41	−1.18	**−0.87**	**−0.72**	**1.72**	**1.16**	**1.05**	**−0.51**	**0.29**	**−0.57**	−1.36	−0.74	−0.68			0.1	−1.92
TBP2–TATA			−1.59	−0.54	**−0.83**	**−0.80**	**1.66**	**1.16**	**1.04**	**0.59**	**0.18**	**−0.61**	−0.90	−1.13	0.13				
P-P distance, Å		6.47	6.72	6.68	**5.39**	**6.32**	**5.51**	**6.05**	**5.75**	**6.20**	**5.72**	**6.86**	6.87	6.35	7.14			7.0	5.9
	6.91	6.31	6.92	6.63	**5.85**	**6.13**	**6.04**	**5.92**	**5.92**	**6.37**	**6.28**	**6.80**	6.98	6.97					
TBP2–TATA					**5.54**	**5.46**	**6.20**	**5.60**	**6.03**	**5.83**	**6.11**	**5.99**	6.87	6.83	6.65	6.64			
				6.65	**6.66**	**5.74**	**6.15**	**6.14**	**5.94**	**6.03**	**6.42**	**5.19**	6.38	6.40	5.78				
Minor groove width, Å			6.15	6.25	**5.38**	**7.19**	**7.84**	**9.07**	**8.90**	**9.09**	**8.24**	**8.00**	7.19	5.60	4.55	4.90		3.97	6.18
TBP2–TATA				7.34	**6.47**	**7.35**	**7.90**	**9.10**	**8.83**	**9.07**	**8.44**	**7.64**	7.28	5.90	6.17				
Propeller twist, °	−18.6	−10.3	−14.3	−18.8	**−13.1**	**11.5**	**−13.9**	**−3.9**	**−8.2**	**−16.6**	**−27.0**	**2.1**	−3.5	−15.8	−7.1	−18.4		−13.3	−7.5
TBP2–TATA			−26.7	−25.7	**−14.6**	**−14.2**	**−28.8**	**−16.8**	**−27.2**	**−27.5**	**−39.6**	**−10.9**	−25.0	−21.5	13.1				
Buckle, °	−13.9	−2.1	8.9	2.7	**−6.1**	**−38.2**	**−29.2**	**−2.1**	**18.5**	**24.9**	**32.9**	**9.8**	10.4	−5.3	−2.1	3.2		0.0	0.0
TBP2–TATA			−11.8	16.9	**−12.6**	**−27.6**	**−24.1**	**−3.6**	**14.6**	**22.3**	**30.4**	**6.2**	7.8	−0.9	2.8	7.7			

Properties of the bound DNA. DNA structural parameters calculated with the program CURVES (22) using a local helix axis. DNA parameters for the 8-base-pair TATA box are in boldface. Corresponding values are given for the *A. thaliana* TBP2–DNA complex at 1.9 Å resolution (13) and for A- and B-form DNA. Minor groove widths were calculated as defined (23).

and Phe-193, that insert into the first and last base steps of the TATA box, respectively, twice kinking the DNA (Figs. 2 and 3). In addition, Phe-210 and Phe-301 buttress the penetrating phenylalanines and stabilize the kinks through extensive van der Waals contacts with the ribose groups of A(7) and T(2'), respectively. The most interesting difference in DNA binding between hTBPc and plant TBP involves Phe-193, which is inserted 1.0 Å further into the base step in the plant complex when compared with the human complex (Fig. 3). This subtle difference leads to a slightly smaller kink in the 3' end of the TATA box (39° vs. 44°) and a similar but different overall trajectory of the DNA generated by hTBPc (Fig. 3). The rmsd for the 6 central TATA element base pairs (i.e., ATAAAA between the phenylalanine insertions) of our high-resolution plant and human cocrystal structures is 0.21 Å for C1' atoms (vs. 0.35 Å if all 8 base pairs of the TATA box are aligned). For reference, the estimated precision for the atomic positions in these two structures is ≈0.2 Å.

Only five protein–base hydrogen bonds occur in the hTBPc–DNA complex, leaving 13/17 possible hydrogen bond acceptors on the minor groove edges of the bases unsatisfied. All five hydrogen bonds are donated to the two quasi-equivalent pairs of N^3 and O^2 acceptors of the central two A·T base pairs of the AdMLP TATA box. Asn-163 donates a pair of hydrogen bonds to the O^2 atoms of T(4') and T(5'), twofold-related Asn-253 donates a pair of hydrogen bonds to the N^3 atoms of A(4) and A(5), and Thr-309 from strand S5' also donates a hydrogen bond to the N^3 of A(4). The remaining TBP–base contacts include van der Waals interactions between nonpolar groups and between polar and nonpolar groups (see ref. 13 for a complete description of base contacts). The first 7 base pairs of the AdMLP TATA element make minor groove base edge interactions with hTBPc side chains projecting from strands S1, S1', S4, S4', S5, and S5' and the C-terminal stirrup (Fig. 1A). The 8th base pair G·C(8) does not make any base–side-chain contacts.

Three arginine side chains make direct contacts to the phosphate backbone: Arg-192 with the phosphate oxygens of T(6') and T(7'), Arg-199 forms a salt bridge with the phosphate oxygen of T(5'), and twofold-related Arg-290 interacts with the phosphate groups of A(4) and T(3). Two lysine side chains make direct contacts to the DNA backbone, including Lys-221 to O3' of T(4') and Lys-312 to a phosphate oxygen of A(6). Lys-221 also makes a water-mediated contact. Four other lysine residues participate in a total of seven water-mediated contacts to six different phosphate oxygens, including Lys-204, Lys-214, Lys-295, and Lys-305. These salt bridges are probably required for partial charge neutralization within the TATA box, permitting the relatively short intrastrand phosphate–phosphate distances that result from bending of the DNA. Additional direct side chain–DNA backbone contacts are provided by Thr-206, Ser-212, and Ser-303. Lys-204, Gly-219, and Gly-310 make water-mediated contacts via carbonyl oxygen atoms. With the exception of the 5' phosphates on each strand of the TATA box and those of A(2) and T(2'), all TATA box phosphate groups interact directly or through ordered water molecules with hTBPc.

DISCUSSION

Although the *A. thaliana*, yeast core, and human core TBP–TATA element cocrystal structures differ slightly, they demonstrate a common induced-fit mechanism of protein–DNA recognition. Our current model involves both the TATA box and to a lesser extent TBP undergoing conformational changes, which result in approximation of two expansive hydrophobic surfaces that are complementary in shape, charge, and polarity. TBP–DNA association kinetics have been studied by various techniques (25–27), yielding results consistent with simultaneous binding and bending with a single second-order rate constant of $\approx 10^5$ $\text{M}^{-1}\cdot\text{s}^{-1}$. In addition, a novel chemical probe has demonstrated that promoter distortion transiently extends beyond the confines of the TATA box during TBP binding (28).

Many different TATA elements, even those containing G·C or C·G base pairs, can be recognized by TBP and direct pol II-mediated transcription initiation (29). High-resolution, x-ray crystallographic studies of *A. thaliana* TBP2 bound to more than a dozen variants of the AdMLP TATA box demonstrated that all these functional TATA elements undergo the same dramatic conformational change on molecular recognition (G. Patikoglou, J. L. Kim, and S.K.B., unpublished data). These

data and the similarities of the *A. thaliana*, yeast core, and human core TBP–TATA element cocrystal structures suggest that most, if not all, functional TATA boxes are capable of undergoing the same structural deformation and making a tight, long-lived, transcriptionally active TBP–DNA complex. Conversely, a subset of AdMLP TATA element variants cannot function in transcription (29). At least one of these sequences (TGTAAAA) can be deformed by a mutant form of TBP (30). However, the resulting minor groove surface is probably not complementary to the DNA-binding surface of wild-type TBP, effectively precluding stable complex formation and transcription from this TATA element variant. Sequence-dependent flexibility of the DNA must also play a role in TBP binding. The kinked TpA and ApG base steps in the recognition site represent two of the least enthalpically stable base steps (31), minimizing the penalty due to loss of base stacking at these sites.

During transcription initiation, TBP recognizes the TATA box in a directional manner, although both TBP's DNA-binding surface and the minor groove face of the AdMLP TATA box are nearly twofold symmetric. It was suggested that binding polarity could be due, at least in part, to asymmetry in the deformability of the two domains of TBP and the two halves of the recognition site (13). Our structure provides additional support for this assertion. The second direct repeats of human and plant TBP bind to the 5' end of the AdMLP TATA element in identical fashion, whereas the first direct repeats show some significant structural differences (Fig. 3). Asymmetry of the phylogenetically conserved charge distribution between the two halves of TBP may also help define its orientation on the TATA box as proposed (21). The structure of the TFIIB–TBP–DNA ternary complex (14) revealed that TFIIB recognizes both protein and DNA features of the preformed TBP–DNA complex, preserving the TBP-induced DNA deformation and stabilizing the TBP–TATA box complex. In this orientation, TFIIB recognizes the stirrup of the second direct repeat of TBP, positioning the N-terminal portion of TFIIB near the transcription start site where it can act as a precise spacer or bridge between pol II and TBP. Thus, TFIIB may stabilize the correct orientation of TBP binding to TATA elements in class II nuclear gene promoters, further ensuring the directionality of transcription.

1. Sentenac, A. (1985) *CRC Crit. Rev. Biochem.* **18**, 31–90.
2. Reeder, R. (1992) in *Transcription Regulation*, eds. McKnight, S. & Yamamoto, K. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 315–348.
3. Roeder, R. G. (1991) *Trends Biochem. Sci.* **16**, 402–408.
4. Moldonado, E. & Reinberg, D. (1995) *Curr. Opin. Cell. Biol.* **7**, 352–361.
5. Gabrielson, O. & Sentenac, A. (1991) *Trends Biochem. Sci.* **16**, 412–416.
6. Burley, S. K. & Roeder, R. G. (1996) *Annu. Rev. Biochem.* **65**, 769–799.
7. Matsui, T., Segall, J., Weil, P. & Roeder, R. (1980) *J. Biol. Chem.* **255**, 11992–11996.
8. Zawel, L. & Reinberg, D. (1993) *Prog. Nucleic Acids Res. Mol. Biol.* **44**, 67–108.
9. Buratowski, S., Hahn, S., Guarente, L. & Sharp, P. A. (1989) *Cell* **56**, 549–561.
10. Nikolov, D. B. & Burley, S. K. (1994) *Nature Struct. Biol.* **1**, 621–637.
11. Nikolov, D. B., Hu, S.-H., Lin, J., Gasch, A., Hoffmann, A., Horikoshi, M., Chua, N.-H., Roeder, R. G. & Burley, S. K. (1992) *Nature (London)* **360**, 40–46.
12. Kim, J. L., Nikolov, D. B. & Burley, S. K. (1993) *Nature (London)* **365**, 520–527.
13. Kim, J. L. & Burley, S. K. (1994) *Nature Struct. Biol.* **1**, 638–653.
14. Nikolov, D. B., Chen, H., Halay, E., Usheva, A., Hisatake, K., Lee, D., Roeder, R. G. & Burley, S. K. (1995) *Nature (London)* **377**, 119–128.
15. Hoffmann, A., Sinn, E., Yamamoto, T., Wang, J., Roy, A., Horikoshi, M. & Roeder, R. G. (1990) *Nature (London)* **346**, 387–390.
16. Studier, F. W., Rosenberg, A. H., Dunn, J. J. & Dubendorff, J. W. (1990) *Methods Enzymol.* **185**, 60–89.
17. Brünger, A. T. (1992) x-plor *Manual* (Yale Univ., New Haven, CT), Version 3.1.
18. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991) *Acta Crystallogr. A* **47**, 110–119.
19. Laskowski, R. J., Macarthur, M. W., Moss, D. S. & Thornton, J. M. (1993) *J. Appl. Crystallogr.* **26**, 283–290.
20. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24**, 946–950.
21. Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993) *Nature (London)* **365**, 512–520.
22. Lavery, R. & Sklenar, H. (1989) *J. Biomol. Struct. Dyn.* **6**, 655–667.
23. Stofer, E. & Lavery, R. (1993) *Biopolymers* **34**, 337–346.
24. Winkler, F. K., Banner, D. W., Oefner, C., Tsernoglou, D., Brown, R. S., Heathman, S. P., Bryan, R. K., Martin, P. D., Petratos, K. & Wilson, K. S. (1993) *EMBO J.* **12**, 1781–1795.
25. Hoopes, B., LeBlanc, J. & Hawley, D. (1992) *J. Biol. Chem.* **267**, 11539–11546.
26. Perez-Howard, G., Weil, P. & Beechem, J. (1995) *Biochemistry* **34**, 8005–8017.
27. Parkhurst, K., Brenowitz, M. & Parkhurst, L. (1996) *Biochemistry*, in press.
28. Sun, D. & Hurley, L. (1995) *Chem. Biol.* **2**, 457–469.
29. Wobbe, C. & Struhl, K. (1990) *Mol. Cell. Biol.* **10**, 3859–3867.
30. Strubin, M. & Struhl, K. (1992) *Cell* **68**, 721–730.
31. Ornstein, R. L., Rein, R., Breen, D. L. & MacElroy, R. D. (1978) *Biopolymers* **17**, 2341–2361.