

## Selective Binding of the TATA Box-Binding Protein to the TATA Box-Containing Promoter: Analysis of Structural and Energetic Factors

Leonardo Pardo,<sup>\*,#</sup> Nina Pastor,<sup>\*</sup> and Harel Weinstein<sup>\*</sup>

<sup>\*</sup>Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York, New York 10029 USA and <sup>#</sup>Laboratorio de Medicina Computacional, Unidad de Bioestadística, Facultad de Medicina, Universidad Autónoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

**ABSTRACT** We report the results of an energy-based exploration of the components of selective recognition of the TATA box-binding protein (TBP) to a TATA box sequence that includes 1) the interaction between the hydrophobic Leu, Pro, and Phe residues of TBP with the TA, AT, AA, TT, and CG steps, by *ab initio* quantum mechanical calculations; and 2) the free energy penalty, calculated from molecular dynamics/potential of mean force simulations, for the conformational transition from A-DNA and B-DNA into the TA-DNA form of DNA observed in a complex with TBP. The GTAT, GATT, GAAT, and GTTT tetramers were explored. The results show that 1) the discrimination of TA, AT, AA, TT, or CG steps by TBP cannot rest on their interaction with the inserting Phe side chains; 2) the steric clash between the bulky and hydrophobic Pro and Leu residues and the protruding —NH<sub>2</sub> group of guanine is responsible for the observed selectivity against any Gua-containing basepair; 3) the Pro and Leu residues cannot selectively discriminate among TA, AT, AA, or TT steps; and 4) the calculated energy required to achieve the TA-DNA conformation of DNA that is observed in the complex with TBP appears to be a key determinant for the observed selectivity against the AT, AA, and TT steps. The simulations also indicate that only the TA step can form a very efficient interbase hydrogen bond network in the TA-DNA conformation. Such an energetically stabilizing network is not achievable in the AA and TT steps. While it is viable in the AT step, structural constraints render the hydrogen bonding network energetically ineffective there.

### INTRODUCTION

Transcription by RNA polymerase II is initiated by the binding of the transcription factor (TF) IID to the consensus sequence TATA<sub>t</sub>/aA<sub>t</sub>/aX (TATA box) of the core promoter [see Burley and Roeder (1996); Pugh (1996) for a review]. TFIID is a multiprotein complex consisting of the TATA box-binding protein (TBP) and TBP-associated factors. The TBP subunit of TFIID provides the specificity for the TATA box sequence, usually located 30 basepairs upstream of the transcriptional start site. The three-dimensional structures of TBP complexed with the TATA box (Juo et al., 1996; Kim et al., 1993a, b; Nikolov et al., 1996), and of the trimeric complexes between TBP/TATA box and TFIIA (Geiger et al., 1996; Tan et al., 1996) or TFIIB (Kosa et al., 1997; Nikolov et al., 1995) have been reported. In all of these crystal structures, TBP binds to the minor groove of the TATA element of the core promoter. In contrast to TBP, which undergoes only a very small conformational change upon binding, the DNA is clearly unwound and there is a striking bending of the DNA structure toward the major groove in the TBP/DNA complexes. This new DNA structure has been termed TA-DNA (Guzikevich Guerstein and Shakked, 1996). The protein-DNA interface is primarily formed by van der Waals contacts, hydrogen bonds, and salt

bridges. van der Waals contacts occur between a series of conserved Val, Leu, and Pro residues and the minor groove edges of the bases, and between the side chains of two pairs of conserved Phe residues and the first two (TA) and last two (t/aX) basepairs of the TATA box. Insertion of these Phe residues between two basepairs produces noticeable kinks. There are very few hydrogen-bonding interactions between DNA and TBP in the crystal structures. These involve two Asn and two Thr side chains interacting with the central two basepairs (At/a) of the TATA box. Finally, the negatively charged phosphate moieties of the DNA are contacted, either directly or water-mediated, by Arg, Lys, and Ser residues.

Electrophoretic mobility shift assays (Wong and Bateman, 1994) showed that TBP has a clear preference for AT-rich DNA sequences (nanomolar affinity), and that it can discriminate among A:T and T:A basepairs despite their similar chemical pattern in the minor groove. Moreover, there is a clear specificity for TA, especially at the first step of the TATA box, since TBP also binds the nonconsensus TAAAAA sequence, with only fivefold lower affinity than for TATAAA or TATATA (Bernues et al., 1996). The molecular basis for the above conclusions is still poorly understood (Pastor et al., 1997). The steric contact between the conserved Val, Leu, and Pro residues of TBP with the protruding —NH<sub>2</sub> group of guanine is considered the key determinant for selection against any Gua-containing base step (Juo et al., 1996; Kim et al., 1993a, b; Kosa et al., 1997). In addition, gel electrophoretic methods revealed that DNA bending might be an important component of site-specific recognition by TBP (Starr et al., 1995), indicating

Received for publication 16 March 1998 and in final form 16 June 1998.

Address reprint requests to Dr. Harel Weinstein, Department of Physiology and Biophysics, Mount Sinai School of Medicine, One Gustave L. Levy Place, Box 1218, New York, NY 10029. Tel.: 212-241-7018; Fax: 212-860-3369; E-mail: hweinstein@inka.mssm.edu.

© 1998 by the Biophysical Society

0006-3495/98/11/2411/11 \$2.00

that the selectivity determinants for TBP binding might also rely on the energetics of DNA bending, which contribute significantly to the overall binding affinity of TBP (Pastor et al., 1997).

In a previous study we used molecular dynamics (MD) simulations of seven DNA dodecamers to study the ease of deformation as a dynamic determinant for recognition (Pastor et al., 1997). The results showed that alternating YR sequences achieve the high *rise* required at the first and last step of the TATA box for TBP binding, whereas RY sequences attain the low *twist* and high positive *roll* needed for TBP binding. However, these conformational preferences for TA-DNA-like properties of various DNA sequences are not the sole determinant for the sequence specificity of TBP. Rather, recognition depends on the interplay between the energy required for the conformational change of DNA upon TBP binding, and the appropriate protein-DNA surface complementarity. Here, we have explored elements of the structural and energetic basis of the selectivity of TBP for the first and most conserved TA step of the TATA box, compared with the other and less common AT, TT, AA, and CG steps. We report results from 1) the potential of mean force (PMF) calculations of the DNA tetramers GTAT, GATT, GAAT, and GTTT with explicit water molecules and counterions that address the free energy penalty required for the conformational transition from A- and B-DNA into the TA-DNA form of DNA observed in

a complex with TBP (Kim et al., 1993b); and 2) the ab initio quantum mechanical calculation of the interaction between the hydrophobic Leu, Pro, and Phe residues of TBP with the different TA, AT, AA, TT, and CG steps to explore the discriminant role that these interactions can have in establishing sequence selectivity.

## METHODS

### Atomic coordinates of TBP/DNA complexes

The atomic coordinates and notation of *Saccharomyces cerevisiae* TBP (sce) complexed with a 12 basepair duplex containing the TATA box of the CYC1 promoter (Kim et al., 1993b) are employed throughout the manuscript. The computed TBP/DNA complex was solvated with a spherical cap of TIP3P waters of 31 Å radius centered on the center of mass of the region of the complex encompassed between Gua-1 to Ade-7, Thy-21 to Cyt-29, Ser-61 to Val-71, and Gln-158 to Met-240. For this region of the complex with the spherical cap and of the water molecules, the energy was minimized (5000 steps) with the Sander module of AMBER 4.1 (Pearlman et al., 1995) and the all-atom force field (Cornell et al., 1995). The resulting system is shown in Fig. 1 (used as reference in the text and named TA<sub>ref</sub>). The atomic coordinates of the other modeled systems were obtained from the appropriate replacements: for the TT step (named TT<sub>ref</sub>), the Ade-3:Thy-27 basepair was replaced by Thy:Ade; for the AA step (named AA<sub>ref</sub>), the Thy-2:Ade-28 basepair was replaced by Ade:Thy; for the AT step (named AT<sub>ref</sub>), Thy-2:Ade-28 and Ade-3:Thy-27 basepairs were replaced by Ade:Thy and Thy:Ade, respectively; and for the CG step (named CG<sub>ref</sub>), Thy-2:Ade-28 and Ade-3:Thy-27 basepairs were replaced by Cyt:Gua and Gua:Cyt, respectively. The molecular model systems were constructed by



FIGURE 1 Molecular model of *S. cerevisiae* TBP (sce) complexed with 12 basepair duplex containing the TATA box of the CYC1 promoter (Kim et al., 1993b), solvated with a spherical cap of TIP3P waters of 31 Å radius centered on the center of mass of the atoms formed by Gua-1 to Ade-7, Thy-21 to Cyt-29, Ser-61 to Val-71, and Gln-158 to Met-240. The system shown is obtained after energy minimization of the crystallographic coordinates in the water sphere. It serves to model the macromolecular structure of AT, AA, TT, and CG steps (see Methods). The figure was created using MOLSCRIPT (Kraulis, 1991).

substituting the geometry of the bases in such a manner that the torsion angles of the glycosyl bond ( $\chi$ ) were not altered. Energy minimization of the macromolecular models thus obtained was performed as described above, and referred to in the text as optimized: i.e.,  $TT_{\text{opt}}$ ,  $AA_{\text{opt}}$ ,  $AT_{\text{opt}}$ ,  $CG_{\text{opt}}$ , respectively.

## A model of protein/DNA interaction

The only direct contacts between TBP and the DNA basepairs Thy-2: Ade-28 and Ade-3:Thy-27 of the TATA box occur through the side chains of Phe-190, Pro-191, Leu-205, and Phe-207 (Kim et al., 1993b). These residues will be referred to as the recognition sites for the first step of the TATA box. To characterize these interactions between TBP fragments and DNA bases with *ab initio* quantum mechanical calculations, a model system was constructed consisting of the following molecular fragments: the free bases Ade, Cyt, Gua, or Thy (the sugar moiety was replaced by a hydrogen); and of the side chains of Leu, Phe, or Pro in which the backbone was replaced by a hydrogen (Pro included both the  $C_\alpha$  and NH moieties of the backbone). The *ab initio* quantum mechanical calculations were used to explore the interaction energies in this model of the selective recognition to consider energy terms that are not well represented in molecular mechanics calculations. Thus, aromatic-aromatic interactions and resonant H-bond networks are described better with the quantum mechanical methods. Because these interactions may participate in determining the selectivity of TBP for TA steps, we performed energy calculations at a level of theory (including MP2 treatment of electron correlation effects) that is capable, in principle, of describing these interactions properly.

The structure optimizations of the isolated residues were performed with Hartree-Fock level calculations and the 6-31G\* basis set. The molecular fragments were positioned at the corresponding recognition sites of the reference structures ( $TA_{\text{ref}}$  or  $AT_{\text{ref}}$  or  $AA_{\text{ref}}$  or  $TT_{\text{ref}}$  or  $CG_{\text{ref}}$ ) and of the optimized structures ( $AT_{\text{opt}}$  or  $AA_{\text{opt}}$  or  $TT_{\text{opt}}$  or  $CG_{\text{opt}}$ ) by superimposing the heavy atoms. The energies of interaction,  $E_{\text{int}}$ , between each basepair of DNA and each side chain of TBP, were evaluated with the 6-31G\* basis set at the MP2 level of correlation energy. All the quantum mechanical calculations were performed with the GAUSSIAN94 system of programs (Frisch et al., 1995).

## Determination of the free energy change in the A-DNA to TA-DNA ( $A \rightarrow TA$ ) and B-DNA to TA-DNA ( $B \rightarrow TA$ ) transitions by a potential of mean force (PMF) calculation

The DNA tetramers GTAT, GATT, GAAT, and GTTT were built in the canonical conformations of both A-DNA and B-DNA (Arnott and Hukins, 1972), which are common starting points for simulations with the AMBER package (Cheatham III and Kollman, 1996; Cieplak et al., 1997). To achieve electroneutrality of the systems, counterions (a total of six sodium ions) were included, positioned initially at a distance of 3.5 Å from each P atom. The DNA and the counterions were placed in a rectangular box containing Monte Carlo-equilibrated TIP3P water (1340 and 1210 water molecules for A-DNA and B-DNA, respectively). For energy minimizations, the DNA atoms were kept fixed initially, while the sodium ions and water molecules were energy minimized (500 steps), heated (from 0 to 300 K in 15 ps) and equilibrated (from 15 to 100 ps) at constant pressure with isotropic scaling. The final box sizes after equilibration were  $\sim 37 \times 37 \times 31$  Å for the A-DNA conformation, and  $36 \times 33 \times 32$  Å for B-DNA, resulting in a final density of  $\sim 1.04$  g cm $^{-3}$  in both cases. Subsequently, the entire systems were energy minimized (500 steps), heated (from 0 to 300 K in 15 ps), and equilibrated (from 15 to 250 ps) at constant volume. During these 250 ps of equilibration the torsion angles  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\zeta$ , and  $\chi$ , and the Watson-Crick basepairing, were maintained close to either the initial A-DNA or B-DNA conformations ( $\pm 5^\circ$  for torsional angles and  $\pm 0.1$  Å for hydrogen bond distances) with flat harmonic restraints (64 kcal mol $^{-1}$  rad $^{-1}$  for torsional angles and 64 kcal mol $^{-1}$  Å $^{-1}$  for hydrogen

bond distances). This procedure produces similar equilibrated initial conformations for the four DNA tetramers compared in the calculation, thus allowing for direct comparisons of the free energy differences between them. The PMF calculations for the  $A \rightarrow TA$  and  $B \rightarrow TA$  transitions were carried out by changing in 51 windows the torsional angles  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ ,  $\zeta$ , and  $\chi$ , and the Watson-Crick hydrogen bond distances from the values found in either the A-DNA or B-DNA conformations to those found in the tetrameric reference structures ( $TA_{\text{ref}}$  or  $TT_{\text{ref}}$  or  $AA_{\text{ref}}$  or  $AT_{\text{ref}}$ ). The simulations were conducted at constant volume so the Helmholtz free energies ( $\Delta F$ ) are reported. The convergence of the simulation was checked by systematically increasing the number of steps of data collection in each window. Thus, the simulation run for the  $B \rightarrow TA$  transition for the GTAT tetramer was carried out for 1020 ps: 51 windows  $\times$  (5 ps of equilibration + 15 ps of data collection), for 1530 ps:  $51 \times (5 + 25)$ , and for 2040 ps:  $51 \times (5 + 35)$ . No significant differences were found in the plot of the free energy changes as a function of  $\lambda$  (see Fig. 3 and Results and Discussion) between the runs with simulation times of 1530 ps and 2040 ps. Moreover, the hysteresis of the calculation, at the simulation time of 1530 ps, was not significant (see Fig. 3 and Results and Discussion). Thus, the free energy changes of the  $A \rightarrow TA$  and  $B \rightarrow TA$  transitions, for the GTAT, GATT, and GAAT tetramers, were subsequently obtained with simulations times of 1530 ps:  $51 \times (5 + 25)$ . The  $A \rightarrow TA$  and  $B \rightarrow TA$  transitions for the GTTT tetramer were run in 81 windows, with simulations time of 2025 ps:  $81 \times (3 + 22)$ , given the difficulty of achieving the  $TT_{\text{ref}}$  structure in only 51 windows (see Results and Discussion). It is important to note that MD/PMF simulations have been applied successfully to study the free energy of the stacking-unstacking process of the 16 naturally occurring deoxyribodinucleoside monophosphates (Norberg and Nilsson, 1995). The MD/PMF simulations were run with the Sander and Gibbs modules of AMBER 4.1 (Pearlman et al., 1995), the all-atom force field (Cornell et al., 1995), SHAKE bond constraints, a 2 fs time step, constant temperature of 300 K coupled to a heat bath, and the thermodynamic integration method to evaluate the free energy change. All solute-solute interactions were evaluated, and an 8 Å cutoff was applied to the solute-water and water-water interactions (the default cutoff scheme in AMBER 4.1).

## Evaluation of the relative stability of the TA, TT, AA, and AT steps

The *ab initio* optimized structures of the free bases Ade and Thy, in which the sugar moiety was replaced by a hydrogen (see above), were positioned at Thy-2:Ade-28 and Ade-3:Thy-27 of  $TA_{\text{ref}}$ ; Ade-2:Thy-28 and Thy-3:Ade-27 of  $AT_{\text{opt}}$ ; Ade-2:Thy-28 and Ade-3:Thy-27 of  $AA_{\text{opt}}$ ; and Thy-2:Ade-28 and Thy-3:Ade-27 of  $TT_{\text{opt}}$ , by superimposing the heavy atoms. The energy of the systems and the energies of interaction between bases 2 and 3,  $E_{\text{int}}(2/3)$ , and between bases 27 and 28,  $E_{\text{int}}(27/28)$ , were evaluated with the 6-31G\* basis set at the MP2 level of correlation energy. All the quantum mechanical calculations were performed with the GAUSSIAN94 system of programs (Frisch et al., 1995).

## RESULTS AND DISCUSSION

### The protein-DNA surface complementarity

#### The TA base step

The direct contacts between TBP and the first and most conserved TA step of the TATA box (Kim et al., 1993b) involve the side chains of Phe-190 and Phe-207 interacting with basepairs Thy-2:Ade-28 and Ade-3:Thy-27, Pro-191 interacting with Thy-2:Ade-28, and Leu-205 interacting with Ade-3:Thy-27. Reflecting on the ability of the Phe interactions to discriminate among the basepair steps, and on the possibility of specific discrimination among TA, AT,



AA, or TT steps by the Pro and Leu side chains, Table 1 reports the calculated energies of interaction.

The  $E_{\text{int}}$  values in Table 1 (*top*) represent interaction energies between each side chain of TBP and each TA basepair of DNA in their reference conformation (see  $\text{TA}_{\text{ref}}$  in Table 1 and Methods). Fig. 2 *a* depicts the molecular model of the protein-DNA interaction employed in the calculations with the TA step reported in Table 1 (*top*). Phe-190 is intercalated in the TA step, producing a strong aromatic-aromatic interaction with both Thy-2:Ade-28 ( $-4.4$  kcal/mol) and Ade-3:Thy-27 ( $-7.0$  kcal/mol) basepairs. Phe-207 is only partially inserted, and it stacks against the sugar moiety of Thy-27. The interaction with the sugar moiety is not included, as it is common to all sequences and is therefore not likely to contribute to sequence selectivity. The partial insertion causes a significant interaction with Ade-3:Thy-27 ( $-3.1$  kcal/mol) and a lesser interaction with Thy-2:Ade-28 ( $-0.7$  kcal/mol). Notably, the van der Waals interactions of Pro-191 and Leu-205 with the minor groove edges of the bases are also significant ( $-1.6$  kcal/mol for both residues).

Because all four modeled side chains interact in a similarly favorable manner with the Thy-2:Ade-28 and Ade-3:Thy-27 basepairs of the TATA box, the selectivity determinants of TBP binding might rely both on the Phe side chains engaged in the aromatic-aromatic interaction with the TA step, and on the van der Waals contacts of the molecular surface of the hydrophobic residues with the TA step.

#### The AT, AA, TT, and CG steps

The atomic coordinates of the AT, AA, TT, and CG steps were all constructed in such a manner that the conformation

of the DNA bases (defined by the  $\chi$  torsional angle) was not altered (see reference structures in Methods). Thus, the overall TBP/DNA conformations of the TA, AT, AA, TT, and CG steps were assumed to be equivalent, although some have not been found in complexes with TBP. The detailed elements of the interaction (e.g., aromatic-aromatic, van der Waals) can thus be examined for their relative contribution to selective binding, assuming they would acquire a similar conformation in the complex with TBP.

Table 1 (*top*) shows the calculated values of  $E_{\text{int}}$  between the side chains of TBP and the various basepairs at positions 2:28 and 3:27 (see  $\text{AT}_{\text{ref}}$ ,  $\text{AA}_{\text{ref}}$ ,  $\text{TT}_{\text{ref}}$ , and  $\text{CG}_{\text{ref}}$ ). Substitution of Thy-2:Ade-28 by basepairs that do not bind TBP (Ade-2:Thy-28 or Cyt-2:Gua-28) has only a modest effect in the interaction with Phe-190 and Phe-207 (see Table 1). Similarly, substitution of Ade-3:Thy-27 by Thy-3:Ade-27 [only bound efficiently by hyperthermophiles (Kosa et al., 1997)] or the nonbound Gua-3:Cyt-27 basepairs, does not significantly change the interaction with the aromatic ring of both Phe residues [see Table 1 (*top*)]. Thus, the observed selectivity toward the TA step cannot be explained by these aromatic-aromatic interactions.

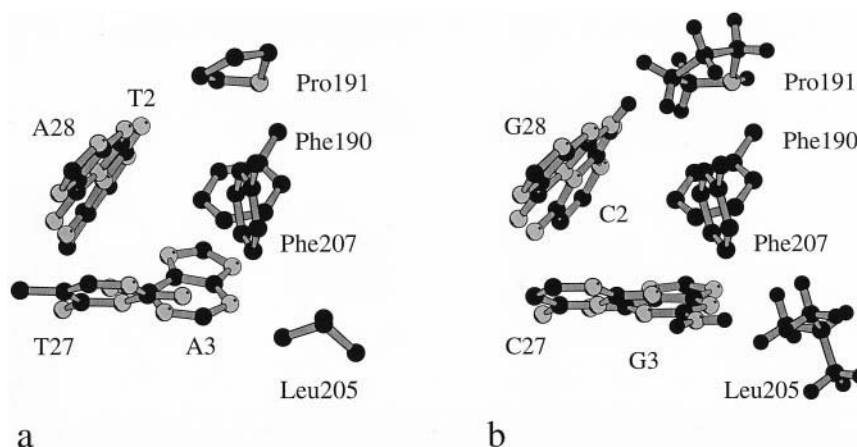
The situation is different in the analysis of the van der Waals contacts between Pro-191 and Leu-205 with the minor groove edges of the bases (see Fig. 2 *b* for the molecular model of the CG step). The calculated energy of interaction between Pro-191 and the nonbound Cyt-2:Gua-28 basepair is repulsive [1.5 kcal/mol, see Table 1 (*top*)] due to the steric contact with the protruding  $-\text{NH}_2$  group of guanine; this effect is strongest in the interaction between Leu-205 and the nonbound Gua-3:Cyt-27 (32.9 kcal/mol). In contrast,  $E_{\text{int}}$  values for the interactions between Pro-191 and the non-TBP-binding Ade-2:Thy-28

**TABLE 1** Energies of interaction between basepairs 2:28 and 3:27 of the TA, AT, AA, TT, and CG steps of DNA and the side chains of Phe-190, Pro-191, Leu-205, and Phe-207 of TBP for the reference and optimized structures, calculated ab initio at the MP2/6-31G\* level of theory

Structure	Basepair	Phe-190 (kcal/mol)	Phe-207 (kcal/mol)	Pro-191 (kcal/mol)	Leu-205 (kcal/mol)
<b>Reference</b>					
$\text{TA}_{\text{ref}}$ , $\text{TT}_{\text{ref}}$	Thy-2:Ade-28	<b>-4.4</b>	<b>-0.7</b>	<b>-1.6</b>	
$\text{AA}_{\text{ref}}$ , $\text{AT}_{\text{ref}}$	Ade-2:Thy-28	-3.4 (1.0)	-0.7 (0.0)	-1.5 (0.1)	
$\text{CG}_{\text{ref}}$	Cyt-2:Gua-28	-3.0 (1.4)	-0.6 (0.1)	1.5 (3.1)	
$\text{TA}_{\text{ref}}$ , $\text{AA}_{\text{ref}}$	Ade-3:Thy-27	<b>-7.0</b>	<b>-3.1</b>		<b>-1.6</b>
$\text{AT}_{\text{ref}}$ , $\text{TT}_{\text{ref}}$	Thy-3:Ade-27	-5.8 (1.2)	-3.5 (-0.4)		-0.8 (0.8)
$\text{CG}_{\text{ref}}$	Gua-3:Cyt-27	-7.1 (-0.1)	-3.3 (-0.2)		32.9 (34.5)
<b>Optimized</b>					
$\text{TA}_{\text{ref}}$	Thy-2:Ade-28	<b>-4.4</b>	<b>-0.7</b>	<b>-1.6</b>	
$\text{AT}_{\text{opt}}$	Ade-2:Thy-28	-3.3 (1.1)	-0.8 (-0.1)	-2.3 (-0.7)	
$\text{AA}_{\text{opt}}$	Ade-2:Thy-28	-3.3 (1.1)	-0.9 (-0.2)	-1.4 (0.2)	
$\text{TT}_{\text{opt}}$	Thy-2:Ade-28	-4.2 (0.2)	-0.7 (0.0)	-1.5 (0.1)	
$\text{CG}_{\text{opt}}$	Cyt-2:Gua-28	-3.8 (0.6)	-0.6 (0.1)	-0.4 (1.2)	
$\text{TA}_{\text{ref}}$	Ade-3:Thy-27	<b>-7.0</b>	<b>-3.1</b>		<b>-1.6</b>
$\text{AT}_{\text{opt}}$	Thy-3:Ade-27	-4.2 (2.8)	-3.0 (0.1)		-1.3 (0.3)
$\text{AA}_{\text{opt}}$	Ade-3:Thy-27	-5.1 (1.9)	-2.1 (1.0)		-1.6 (0.0)
$\text{TT}_{\text{opt}}$	Thy-3:Ade-27	-4.0 (3.0)	-3.0 (0.1)		-1.4 (0.2)
$\text{CG}_{\text{opt}}$	Gua-3:Cyt-27	-4.9 (2.1)	-2.7 (0.4)		1.7 (3.3)

Numbers in parentheses show the difference in energies of interaction relative to the TA step (in bold).

FIGURE 2 Molecular models used in the calculation of relative selectivity for TBP: (a) for the first and most conserved TA step of the TATA box, and (b) for the CG step. The systems are composed of the DNA basepairs Thy-2: Ade-28 and Ade-3:Thy-27 in (a), or Cyt-2: Gua-28 and Gua-3:Cyt-27 in (b), with the sugar moiety replaced by a hydrogen, and models of the TBP side chains Phe-190, Pro-191, Leu-205, and Phe-207, in which the backbone was replaced by a hydrogen (Pro included both the C $\alpha$  and NH moieties of the backbone).



basepair, or between Leu-205 and the hyperthermophile-bound Thy-3:Ade-27, are similar to those for TBP-binding pairs Thy-2:Ade-28 or Ade-3:Thy-27 ( $-1.6$  vs.  $-1.5$  kcal/mol or  $-1.6$  vs.  $-0.8$  kcal/mol, respectively).

The results described above are from calculations with protein/DNA models frozen in the orientation (between the basepairs of DNA and the aromatic or hydrophobic residues of TBP) that correspond to the TA step. When the geometries of the complexes are optimized (see Methods) the values of  $E_{\text{int}}$  obtained for the various TBP/DNA models change little (see  $AT_{\text{opt}}$ ,  $AA_{\text{opt}}$ ,  $TT_{\text{opt}}$ , and  $CG_{\text{opt}}$  in Table 1 (bottom) and optimized structures in Methods). The largest change, compared to the reference structures [Table 1 (top)], is in the van der Waals contact between Leu-205 and Gua-3:Cyt-27 of the CG step (32.9 vs. 1.7 kcal/mol). It is significant that although both TBP and DNA can rearrange to reduce the steric clash between Leu and the  $-\text{NH}_2$  group of Gua, the resulting energy of interaction remains repulsive, and the effect on the rest of the interaction surface between the protein and the DNA as a result of this local rearrangement still remains to be appraised. The energy of interaction of the other steric clash between Cyt-2:Gua-28 and Pro-191 becomes negative (changing from 1.5 to  $-0.4$  kcal/mol), but is not as favorable as the values obtained for the TA step ( $-1.6$  kcal/mol). Remarkably, the energies of interaction of Pro-191 or Leu-205 with the other AT, AA, or TT steps remain as favorable as for the TA step, indicating that these bulky and hydrophobic Pro and Leu residues cannot selectively discriminate among TA, AT, AA, or TT steps. However, the results confirm earlier observations (Juo et al., 1996; Kim et al., 1993a, b) that the steric contact of these residues with the protruding  $-\text{NH}_2$  group of guanine is likely to be responsible for the discrimination against any Gua-containing base step.

### Free energy penalty for the conformational transition into the TA-DNA form of DNA observed in the complex with TBP

The energy required to kink the TA, AT, AA, and TT steps, associated with the insertion of the Phe residues, was cal-

culated for the DNA tetramers GTAT, GATT, GAAT, and GTTT. The free energy penalty required for the conformational transition was evaluated from MD/PMF simulations (see Methods) starting from both A-DNA and B-DNA and achieving the TA-DNA form found in a TBP/DNA complex (Kim et al., 1993b).

### Testing the convergence of the simulation

Fig. 3 shows the computed free energy changes during the B  $\rightarrow$  TA transition ( $\lambda = 0 \rightarrow 1$ ) for the GTAT sequence at different simulation lengths. The 1020 ps run (5 ps of equilibration + 15 ps of data collection per window, for 51 windows) reproduces a situation in which the TA-DNA conformation is more stable than the B-DNA conformation by  $-3.7$  kcal/mol. This free energy change between the two

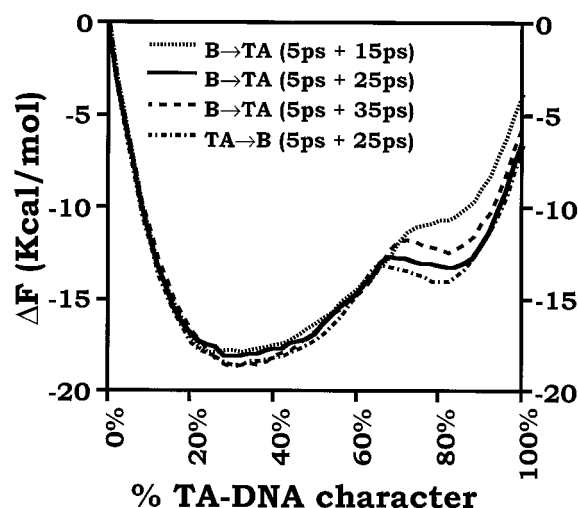


FIGURE 3 Computed free energy changes of the GTAT sequence during the B  $\rightarrow$  TA transition ( $\lambda = 0 \rightarrow 1$ ) at simulation lengths of 51 windows  $\times$  (5 ps of equilibration + 15 ps data collection) = 1020 ps ( $\Delta F = -3.7$  kcal/mol); 51  $\times$  (5 + 25) = 1530 ps ( $\Delta F = -6.5$  kcal/mol); and 51  $\times$  (5 + 35) = 2040 ps ( $\Delta F = -5.5$  kcal/mol); and for the TA  $\rightarrow$  B transition ( $\lambda = 1 \rightarrow 0$ ) at a simulation length of 51  $\times$  (5 + 25) = 1530 ps ( $\Delta F = 5.5$  kcal/mol).

conformations decreased to a value of  $-6.5$  kcal/mol when the simulation was run for 1530 ps (5 ps + 25 ps). Although the value of free energy obtained at 1530 ps is relatively close to the value obtained at 1020 ps, the free energy profiles are different (see Fig. 3). However, the free energy profile obtained for 2040 ps (5 ps + 35 ps) is analogous to that obtained for 1530 ps (see Fig. 3), and the free energy change obtained from these simulations is similar ( $-6.5$  vs.  $-5.5$  kcal/mol). The free energy profile for the TA  $\rightarrow$  B transition ( $\lambda = 1 \rightarrow 0$ ) at a simulation time of 1530 ps (5 ps + 25 ps) is also shown in Fig. 3. The small differences between the “forward” and “reverse” free energy profiles support the convergence of the results obtained with the 1530 ps runs. Therefore, the free energy changes during the A  $\rightarrow$  TA and B  $\rightarrow$  TA transitions were calculated with simulation times of 1530 ps for all other tetrads.

### The conformational transition

Statistical analysis of various conformational parameters of the crystal structure of DNA bound to TBP showed that the torsion angles of the TATA box resemble those of the A-DNA structure, with the exception of the glycosyl bond torsion angle ( $\chi$ ) which is close to the B-DNA structure (Guzikevich Guerstein and Shakked, 1996). Thus, the A  $\rightarrow$  TA transition was shown to proceed smoothly to completion primarily by modifying the value of  $\chi$  (Pardo et al., 1998). In contrast, the B  $\rightarrow$  TA transition involved changes in the values of  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ ,  $\epsilon$ , and  $\zeta$ , while maintaining the value of  $\chi$  almost invariant. It is important to note that in the complex with TBP the local conformation of the DNA structure

outside the TATA box is close to the B-DNA conformation (Guzikevich Guerstein and Shakked, 1996; Kim et al., 1993b). Thus, the flanking G:C basepair of the DNA tetramers adopts torsion angles close to the B-DNA conformation observed in the crystal structure (Kim et al., 1993b). Fig. 4 shows the average structures computed from the data collection trajectory at each window during the B  $\rightarrow$  TA transition for the GTAT tetramer. The initial B-DNA (*black*) and final TA-DNA (*gray*) conformations are shown in “ball and sticks” rendering. Clearly, the process of DNA bending occurred through a smooth change as reflected by the proximity between subsequent structures in the display. Such a smooth change is also found for the A  $\rightarrow$  TA transition, and for the other DNA tetramers (results not shown). The root mean square (rms) best fit of the crystal structure of TA-DNA (*gray*) and the average structure computed from the data collection trajectory of the last window (*black*) are shown in Fig. 5 for the B  $\rightarrow$  TA transition (Fig. 5 *a*) and the A  $\rightarrow$  TA transition (Fig. 5 *b*). The minimal and the average all atoms rms deviation (rmsd) of the structures computed during the last window of the PMF simulations from the optimized structures (see Methods) are given in Table 2. Note that initial PMF calculation of the GTTT conformational transitions in 51 windows produced an rmsd  $>1.6$  Å, i.e., structures quite dissimilar to the TA-DNA conformation, but this problem was overcome by running the conformational transitions of GTTT in 81 windows, for a simulation time of 2025 ps (see Methods). Typically, the inner basepair step shows lower rmsd values than the entire tetramer sequence, which includes the more mobile four bases at the ends of the DNA helix.

FIGURE 4 Average structures computed from the data collection trajectory at each window during the B  $\rightarrow$  TA transition for the GTAT tetramer. The initial B-DNA (*black*) and final TA-DNA (*gray*) conformations are shown in the “ball and sticks” rendering.

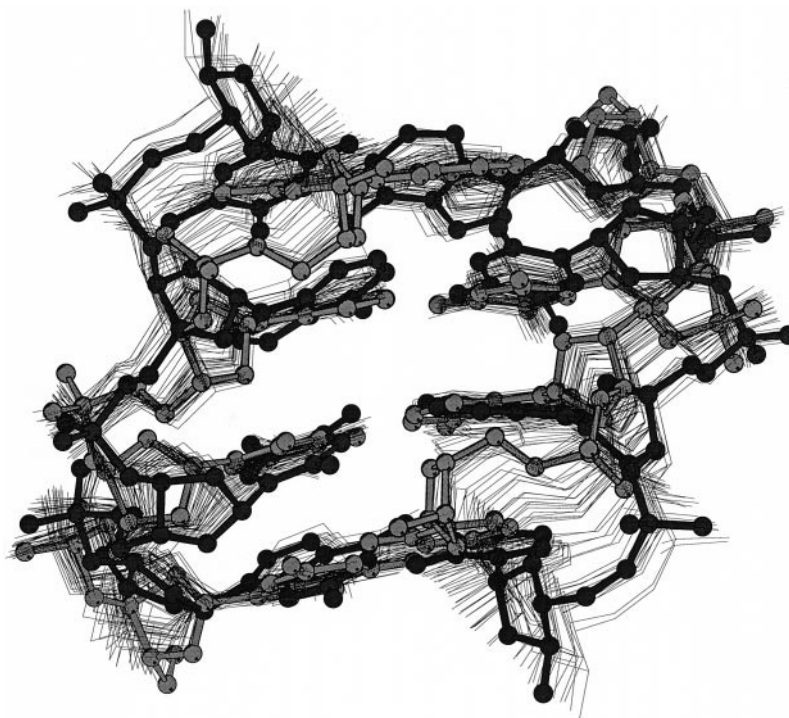
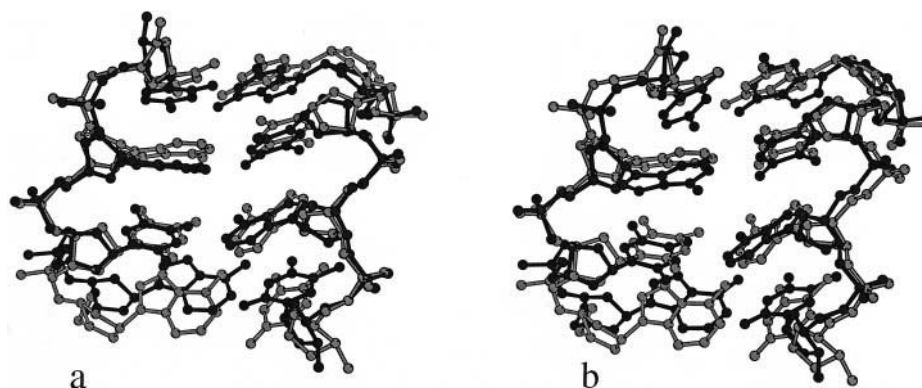


FIGURE 5 The root mean square best fit of the GTAT tetrad of the crystal structure (Kim et al., 1993b), in gray, and the average structure computed from the data collection trajectory of the last window, in black, from the B  $\rightarrow$  TA transition (a), and the A  $\rightarrow$  TA transitions (b).



### Free energy penalty

Fig. 6 and Table 3 show the computed free energy changes for the B  $\rightarrow$  TA transition (Fig. 6 a) and the A  $\rightarrow$  TA transition (Fig. 6 b) calculated for the GTAT, GATT, GAAT, and GTTT sequences. The B  $\rightarrow$  TA simulations reveal that the TA-DNA conformation is more stable than the B-DNA conformation for the GTAT (−6.5 kcal/mol), GATT (−4.2 kcal/mol), and GAAT tetramers (−1.8 kcal/mol). In contrast, the GTTT sequence prefers the B-DNA conformation by 0.4 kcal/mol. The order of stability is reversed for the transition from the A-DNA conformation to TA-DNA: energies of 25.0 kcal/mol for GTAT, 27.2 kcal/mol for GATT, 36.4 kcal/mol for GAAT, and 29.1 kcal/mol for GTTT are required to accomplish this transition. Notably, the computed free energy of 25.0 kcal/mol for the A  $\rightarrow$  TA transition for the TA step is in good agreement with the energy of deformation of 22 kcal/mol obtained with the JUMNA program by stretching the TATATAAA segment in the A-DNA conformation to mimic the distortion caused by TBP (Lebrun et al., 1997). Note that having calculated the B  $\rightarrow$  TA and A  $\rightarrow$  TA free energy penalties, one might feel inclined to calculate the free energy difference between the initial canonical B- and A-DNA conformations (B  $\rightarrow$  A) simply by adding the reported values. However, this addi-

tion is not appropriate because the TA-DNA conformations obtained in the A  $\rightarrow$  TA and in the B  $\rightarrow$  TA transitions were not the same. Thus, the differences among the three TA-DNA structures (from the crystal, from the endpoint of the B  $\rightarrow$  TA transition, and from the endpoint of the A  $\rightarrow$  TA transition) preclude the estimation of the B  $\rightarrow$  A free energy from a thermodynamic cycle approach.

The free energy values obtained in the simulations may not accurately describe the energetics of DNA bending in

**TABLE 2** All atoms root mean square deviation between the structures computed during the last window of the potential mean force for the A  $\rightarrow$  TA and B  $\rightarrow$  TA transitions of the GTAT, GATT, GAAT, and GTTT tetramers, and the TA<sub>ref</sub>, AT<sub>opt</sub>, AA<sub>opt</sub>, and TT<sub>opt</sub> structures (see Methods)

Transition	Sequence	Entire Sequence (Å)		Inner 2:7 and 3:6 Basepairs (Å)	
		Minimal	Avg.	Minimal	Avg.
A $\rightarrow$ TA	GTAT*	0.7	1.0	0.5	0.8
A $\rightarrow$ TA	GATT*	0.7	0.9	0.5	0.7
A $\rightarrow$ TA	GAAT*	0.8	1.0	0.6	0.8
A $\rightarrow$ TA	GT <sup>#</sup>	0.9	1.1	0.8	0.9
B $\rightarrow$ TA	GTAT*	1.0	1.2	0.9	1.0
B $\rightarrow$ TA	GATT*	1.1	1.2	0.9	1.0
B $\rightarrow$ TA	GAAT*	1.1	1.2	1.0	1.1
B $\rightarrow$ TA	GT <sup>#</sup>	1.1	1.3	1.1	1.2

\*51 windows  $\times$  (5 ps of equilibration + 25 ps data collection) = 1530 ps.

<sup>#</sup>81 windows  $\times$  (3 ps of equilibration + 22 ps data collection) = 2025 ps.

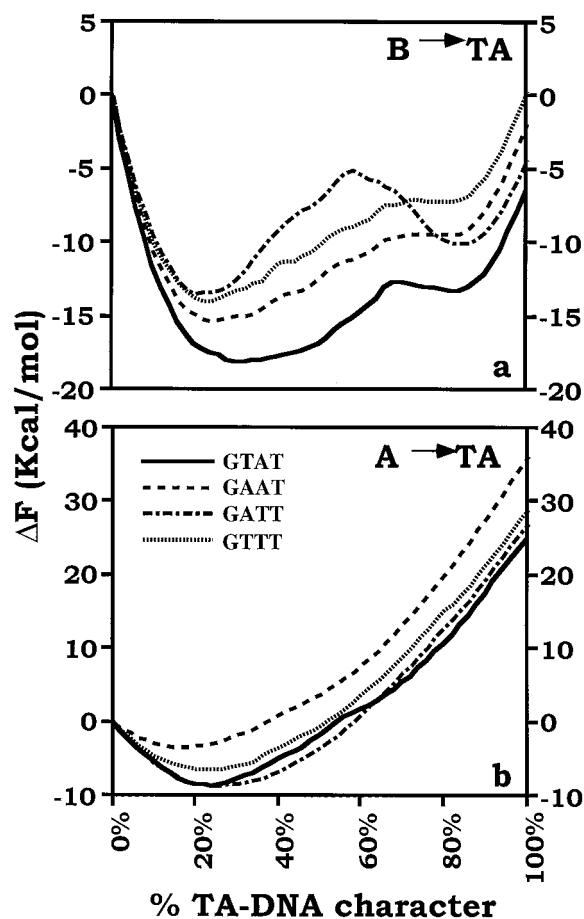


FIGURE 6 Computed free energy changes during (a) the B  $\rightarrow$  TA transition and (b) the A  $\rightarrow$  TA transition for the GTAT, GATT, GAAT, and GTTT sequences.



TABLE 3

Method	Sequence	$\Delta F$	$\Delta\Delta F$	$\Delta\Delta E$	$E_{\text{int}}(2/3)$	$E_{\text{int}}(27/28)$
MD/PMF	B $\rightarrow$ TA					
	GTAT <sup>#</sup>	-6.5	0.0			
	GATT <sup>#</sup>	-4.2	2.3			
	GAAT <sup>#</sup>	-1.8	4.7			
	GTTT <sup>§</sup>	0.4	6.9			
MD/PMF	A $\rightarrow$ TA					
	GTAT <sup>#</sup>	25.0	0.0			
	GATT <sup>#</sup>	27.2	2.2			
	GAAT <sup>#</sup>	36.4	11.4			
	GTTT <sup>§</sup>	29.1	4.1			
MP2/6-31G*	TA			0.0	-7.4 (Thy/Ade)	-5.0 (Thy/Ade)
	AT			6.1	-2.2 (Ade/Thy)	-4.0 (Ade/Thy)
	AA			6.6	-4.2 (Ade/Ade)	-1.3 (Thy/Thy)
	TT			7.3	-0.3 (Thy/Thy)	-4.8 (Ade/Ade)

$\Delta F$ , potential mean force for the A  $\rightarrow$  TA and B  $\rightarrow$  TA transitions of the GTAT, GATT, GAAT, and GTTT tetramers;  $\Delta\Delta F$ , the difference in  $\Delta F$  relative to the GTAT tetramer;  $\Delta\Delta E$ , stability of the AT, AA, and TT steps, relative to the TA step;  $E_{\text{int}}(2/3)$ , energy of interaction between bases 2 and 3; and  $E_{\text{int}}(27/28)$ , energy of interaction between bases 27 and 28 (see Methods). Units are in kcal/mol.

<sup>#</sup>51 windows  $\times$  (5 ps of equilibration + 25 ps data collection) = 1530 ps.

<sup>§</sup>81 windows  $\times$  (3 ps of equilibration + 22 ps data collection) = 2025 ps.

the process of complex formation, for several reasons. First, the choice of the starting conformation significantly affects the values obtained from the free energy calculation, and the most likely torsion angles of the GTAT, GATT, GAAT, or GTTT tetramers in solution might not correspond to the considered values of either the A-DNA or B-DNA conformation. Although the sequence-dependent DNA structure has been studied extensively (Basham et al., 1995; Dickerson, 1992; Hunter, 1993) it is not possible to predict the most likely conformation in solution for these different sequences. Second, the overall electrostatic effects of the protein on the energetics of bending has not been considered in these calculations, while it has been proposed that the low dielectric interior of proteins, such as TBP, is sufficient to cause major structural changes in the DNA molecule (Elcock and McCammon, 1996; Travers, 1995). Third, there are some discrepancies between the resulting computer simulated structure in the PMF simulations and the crystal or modeled structures (see Fig. 5 and Table 2). Given the stress imposed on DNA structure in the TA-DNA conformation, a small change in structure might represent a large change in free energy. Nevertheless, the difference in the values of free energy change between a given DNA tetramer and GTAT,  $\Delta\Delta F$  (see Table 3), might be useful to understand the basis for the selectivity toward the TA step, if the above-mentioned approximations are assumed to affect all tetramers equally. This assumption is reasonable because no major differences appeared in the process of DNA bending during the explored conformational transitions.

In addition to the MD/PMF simulations, the relative stabilities of the TA step in the TA<sub>ref</sub> structure, the AT step in the AT<sub>opt</sub> structure, the AA step in the AA<sub>opt</sub> structure, and the TT step in the TT<sub>opt</sub> structure were evaluated with ab initio quantum mechanical calculations (see Methods). The molecular models employed in the calculation consist

of the four free bases of the 2:28 and 3:27 basepairs (depicted as enlarged ball and stick in Fig. 7). The differences in energy ( $\Delta\Delta E$ ) relative to the TA step are given in Table 3.

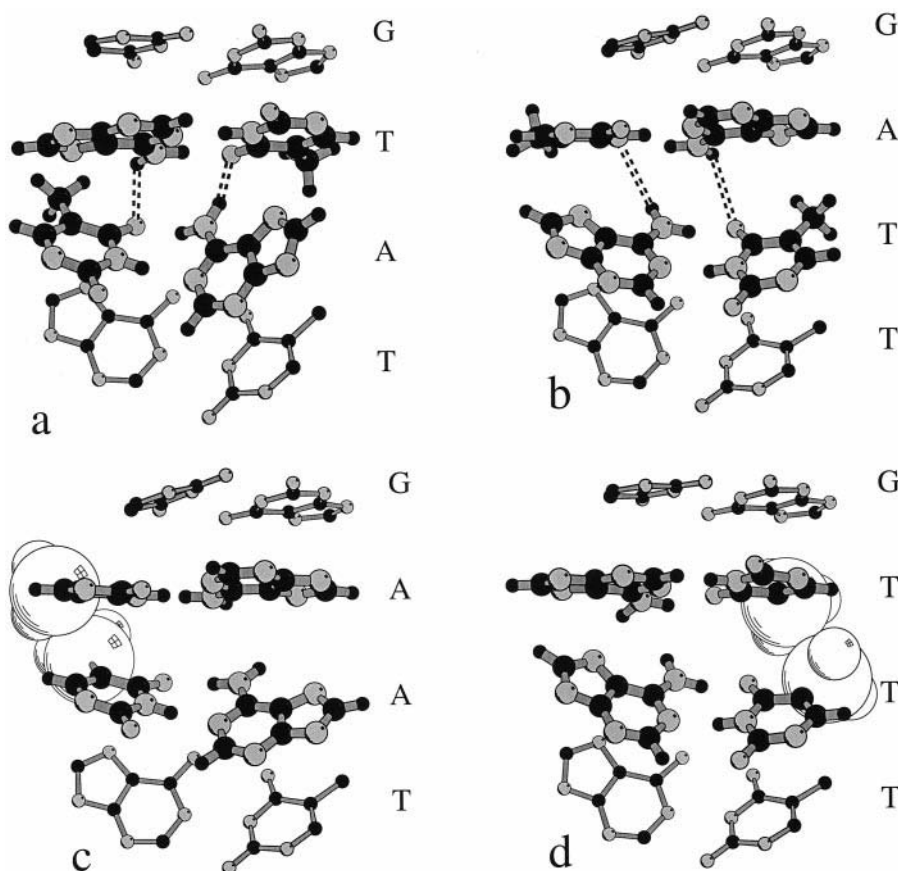
#### Comparison between TA and AT steps

Fig. 7 *a* shows the TA-DNA conformation adopted by the TA step. The noticeable kink produced by the insertion of both Phe side chains promotes the formation of two intermolecular hydrogen bonds between Thy-2 and Ade-3, and Thy-27 and Ade-28 (see *dashed bonds* in Fig. 3). The O<sub>4</sub> atom of Thy-2 is hydrogen bonded to the N<sub>6</sub> atom of Ade-3 with an O<sub>4</sub>-N<sub>6</sub> distance of 2.82 Å and an O<sub>4</sub>-H  $\cdot$  N<sub>6</sub> angle of 146°, close to the optimal of 180°. The energy of interaction between Thy-2 and Ade-3, evaluated by ab initio methods (see Methods and Table 3), is -7.4 kcal/mol. The interaction between Thy-27 and Ade-28 in the complementary strand is weaker (-7.4 versus -5.0 kcal/mol, see Table 3). This is probably due to a longer O<sub>4</sub>-N<sub>6</sub> distance (3.5 Å) and a less linear O<sub>4</sub>-H  $\cdot$  N<sub>6</sub> angle (106°).

This hydrogen bond network cannot be achieved as efficiently in the AT step (see Fig. 3). The N<sub>6</sub>-O<sub>4</sub> distances between Ade-2 and Thy-3 and between Ade-27 and Thy-28 are 3.6 Å and 3.7 Å, respectively. Moreover, the N<sub>6</sub>  $\cdot$  H-O<sub>4</sub> angles are almost perpendicular: 84° for Ade-2 and Thy-3, and 117° for Ade-27 and Thy-28. Therefore, the energies of interaction between bases cannot achieve the values obtained for the bases of the TA base step (-7.4 or -5.0 kcal/mol versus -2.2 or -4.0 kcal/mol, see Table 3). These findings are reflected in the values of  $\Delta\Delta F$  or  $\Delta\Delta E$  obtained in the MD/PMF or ab initio calculations. Thus, the insertion of the Phe side chains in the AT step demands 2.3 kcal/mol (B  $\rightarrow$  TA), or 2.2 kcal/mol (A  $\rightarrow$  TA), or 6.1 kcal/mol (ab initio) more than in the TA step. We can conclude that the efficiency of the interbase hydrogen bond network of the



FIGURE 7 Detailed view of the (a) GTAT tetrad of the TA<sub>ref</sub> structure, (b) the GATT tetrad of the AT<sub>opt</sub> structure, (c) the GAAT tetrad of the AA<sub>opt</sub> structure, and (d) the GTTT tetrad of the TT<sub>opt</sub> structure. The interbase hydrogen bonds between Thy-2 and Ade-3, and Thy-27 and Ade-28 of TA<sub>ref</sub>, and between Ade-2 and Thy-3, and Ade-27 and Thy-28 of AT<sub>ref</sub> are shown as dashed bonds. The steric clash between the major groove methyl groups of Thy-2 and Thy-3 of TT<sub>opt</sub>, and of Thy-27 and Thy-28 of AA<sub>opt</sub> is identified with van der Waals spheres.



TA base step constitutes a significant component of the discrimination between TA and AT base steps by TBP.

### Comparison between AA and TT base steps

Fig. 7, *c* and *d* show the conformation adopted by the AA and TT base steps. Both AA and TT base steps possess two contiguous Thy bases, which cause a significant steric clash between methyl groups in the kinked conformation (see van der Waals spheres in Fig. 7, *c* and *d*). The importance of the methyl group of Thy in the overall conformation of the DNA structure has been discussed (Suzuki and Yagi, 1995; Suzuki et al., 1996). Thus, in contrast to the TA base step,

the interaction energies between Thy-2 and Thy-3 of the TT step, and between Thy-27 and Thy-28 of the AA step, are very small ( $-0.3$  kcal/mol for TT and  $-1.3$  kcal/mol for AA, see Table 3). Hyperthermophiles bind efficiently to TT steps [the consensus TATA box sequence for these organisms is YTTAT/aANN, where Y is a pyrimidine (Kosa et al., 1997)], albeit at temperatures close to  $100^{\circ}\text{C}$ , and inducing a slightly different conformation at this basepair step. As shown by the values compiled in Table 4, a TT step apparently requires a higher *rise* and a less negative *slide* than the equivalent TA steps in order to avoid the methyl-methyl clash. The energies of interaction between Ade-2 and Ade-3 of the AA step or between Ade-27 and Ade-28 of the TT

TABLE 4 Basepair step geometric parameters for the first step of the TATA box in the available crystal structures of TBP/DNA complexes

Step	Shift	Slide	Rise	Tilt	Roll	Twist	Reference
TA	0.52	-1.31	5.20	-4.22	44.11	18.10	(Kim <i>et al.</i> , 1993a)
TA	0.18	-1.91	4.75	0.53	37.24	18.50	(Kim <i>et al.</i> , 1993a)
TA	0.12	-1.99	4.93	-0.59	47.72	17.18	(Nikolov <i>et al.</i> , 1996)
TA	0.34	-2.20	4.41	-2.58	37.34	18.56	(Nikolov <i>et al.</i> , 1995)
TA	0.20	-1.98	4.63	0.25	37.22	19.80	(Kim <i>et al.</i> , 1993b)
TA	0.82	-2.24	4.74	0.89	41.70	22.75	(Kim <i>et al.</i> , 1993b)
TA	-0.56	-2.28	5.22	-3.46	44.01	22.25	(Tan <i>et al.</i> , 1996)
TA	-0.01	-1.43	5.54	4.42	38.20	21.58	(Juo <i>et al.</i> , 1996)
TT	0.94	-0.79	5.82	-0.84	37.62	22.64	(Kosa <i>et al.</i> , 1997)

Parameters were calculated with CURVES in Dials and Windows (Ravishanker *et al.*, 1989).

step are  $-4.2$  or  $-4.8$  kcal/mol, respectively, but are not as attractive as those of the TA base step (see Table 3), due to the lack of the hydrogen bond network. Notably, these two base steps are found to be the most difficult to bend in the  $B \rightarrow TA$  and  $A \rightarrow TA$  simulations, and in the evaluation of the relative stability of the base steps by ab initio methods (see Table 3).

## CONCLUSIONS

First, the three separate simulations carried out to study the difference in energy required to kink the AT, AA, and TT steps—which are not bound efficiently by mesophile TBPs—relative to the best bound TA basepair step, provide the same rank order of preferred DNA sequences for TBP binding:  $GTAT > GATT > GAAT \approx GTTT$  (see  $\Delta\Delta F$  and  $\Delta\Delta E$  in Table 3). The only apparent mismatch is the GTTT tetramer in the A-DNA  $\rightarrow$  TA-DNA transition. However, the low value of  $\Delta\Delta F$  found in the A-DNA  $\rightarrow$  TA-DNA transition for this GTTT tetramer might be due to the instability of the TTT triplet in the A-DNA conformation (Basham et al., 1995).

Second, from the analysis of the interaction energies that may contribute to the selectivity among structural components of the TBP/DNA complexes, the results indicate that *the energy required to achieve the TA-DNA conformation is a key determinant of the observed selectivity for TA steps relative to AT, AA, and TT base steps.*

Third, the selectivity determinants for TBP discrimination against AA or TT base steps *reside in the presence of two contiguous Thy bases that oppose kinking because of the a steric clash between their methyl groups in the kinked conformation.* Furthermore, these basepairs lack the ability to form the interbase hydrogen bond network that was shown to be stabilizing in the TA step.

Taken together, the selectivity determinants resulting from the local interaction energies described here complement and enhance the effects of the dynamic properties shown earlier (Pastor et al., 1997) to support TBP binding preferences for the various forms of TATA box sequences.

This work was supported in part by a grant from the Association for International Cancer Research (to H.W.); DGICYT Grants PB95-0624 and PR95-275 (to L.P.); and a Fulbright/CONACyT scholarship (to N.P.). Some of the simulations were run at the Cornell National Supercomputer Facility (sponsored by the National Science Foundation and IBM) and at the Centre de Computació i Comunicacions de Catalunya.

## REFERENCES

- Arnott, S., and D. W. L. Hukins. 1972. Optimized parameters for A-DNA and B-DNA. *Biochem. Biophys. Res. Commun.* 47:1504–1509.
- Basham, B., G. P. Schroth, and P. S. Ho. 1995. An A-DNA triplet code: thermodynamic rules for predicting A- and B-DNA. *Proc. Natl. Acad. Sci. USA.* 92:6464–6468.
- Bernues, J., P. Carrera, and F. Azorin. 1996. TBP binds the transcriptionally inactive TA<sub>5</sub> sequence but the resulting complex is not efficiently recognized by TFIIB and TFIIA. *Nucleic Acids Res.* 24:2950–2958.
- Burley, S. K., and R. G. Roeder. 1996. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.* 65:769–799.
- Cheatham III, T. E., and P. A. Kollman. 1996. Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.* 259:434–444.
- Cieplak, P., T. E. Cheatham III, and P. A. Kollman. 1997. Molecular dynamics simulations find that 3' phosphoramidate modified DNA duplexes undergo a B to A transition and normal DNA duplexes an A to B transition. *J. Am. Chem. Soc.* 119:6722–6730.
- 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.
- Dickerson, R. E. 1992. DNA structure from A to Z. *Methods Enzymol.* 211:67–111.
- Elcock, A. H., and J. A. McCammon. 1996. The low dielectric interior of proteins is sufficient to cause major structural changes in DNA on association. *J. Am. Chem. Soc.* 118:3787–3788.
- Frisch, M. J., G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. A. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. J. P. Stewart, M. Head-Gordon, C. Gonzalez, and J. A. Pople. 1995. Gaussian 94. Gaussian Inc., Pittsburgh, PA.
- Geiger, J. H., S. Hahn, S. Lee, and P. B. Sigler. 1996. Crystal structure of the yeast TFIIB/TBP/DNA complex. *Science.* 272:830–836.
- Guzikevich Guershtein, G., and Z. Shakked. 1996. A novel form of the DNA double helix imposed on the TATA-box by the TATA-binding protein. *Nature Struct. Biol.* 3:32–37.
- Hunter, C. A. 1993. Sequence-dependent DNA structure. The role of base stacking interactions. *J. Mol. Biol.* 230:1025–1054.
- Juo, Z. S., T. K. Chiu, P. M. Leiber, I. Baikalov, A. J. Berk, and R. E. Dickerson. 1996. How proteins recognize the TATA box. *J. Mol. Biol.* 261:239–254.
- Kim, Y., J. H. Geiger, S. Hahn, and P. B. Sigler. 1993b. Crystal structure of a yeast TBP/TATA-box complex. *Nature.* 365:512–520.
- Kim, J. L., D. B. Nikolov, and S. K. Burley. 1993a. Co-crystal structure of TBP recognizing the minor groove of a TATA element. *Nature.* 365:520–527.
- Kosa, P. F., G. Ghosh, B. S. DeDecker, and P. B. Sigler. 1997. The 2.1-Å crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box. *Proc. Natl. Acad. Sci. USA.* 94:6042–6047.
- Kraulis, J. 1991. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structure. *J. Appl. Crystallogr.* 24:946–950.
- Lebrun, A., Z. Shakked, and R. Lavery. 1997. Local DNA stretching mimics the distortion caused by the TATA box-binding protein. *Proc. Natl. Acad. Sci. USA.* 94:2993–2998.
- 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.
- Nikolov, D. B., H. Chen, E. D. Halay, A. A. Usheva, K. Hisatake, D. K. Lee, R. G. Roeder, and S. K. Burley. 1995. Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature.* 377:119–128.
- Norberg, J., and L. Nilsson. 1995. Potential of mean force calculations of the stacking-unstacking process in single-stranded deoxyribonucleoside monophosphates. *Biophys. J.* 69:2277–2285.
- Pardo, L., N. Pastor, and H. Weinstein. 1998. Progressive DNA bending is made possible by gradual changes in the torsion angle of the glycosyl-bond. *Biophys. J.* 74:2191–2198.
- Pastor, N., L. Pardo, and H. Weinstein. 1997. Does TATA matter? A structural exploration of the selectivity determinants in its complexes with TATA box-binding protein. *Biophys. J.* 73:640–652.
- Pearlman, D. A., D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, D. M. Ferguson, G. L. Seibel, U. C. Singh, P. K. Weiner, and P. A. Kollman. 1995. AMBER 4.1. University of California, San Francisco.

- Pugh, B. F. 1996. Mechanisms of transcription complex assembly. *Curr. Opin. Cell Biol.* 8:303–311.
- Ravishanker, G., S. Swaminathan, D. L. Beveridge, R. Lavery, and H. Sklenar. 1989. Conformational and helicoidal analysis of 30 PS of molecular dynamics on the d(CGCGAATTCGCG) double helix: “curves,” dials and windows. *J. Biomol. Struct. & Dyn.* 6:669–699.
- Starr, D. B., B. C. Hoopes, and D. K. Hawley. 1995. DNA bending is an important component of site-specific recognition by the TATA binding protein. *J. Mol. Biol.* 250:434–446.
- Suzuki, M., and N. Yagi. 1995. Stereochemical basis of DNA bending by transcription factors. *Nucleic Acids Res.* 23:2083–2091.
- Suzuki, M., N. Yagi, and J. T. Finch. 1996. Role of base-backbone and base-base interactions in alternating DNA conformations. *FEBS Lett.* 379:148–152.
- Tan, S., Y. Hunziker, D. F. Sargent, and T. J. Richmond. 1996. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature.* 381:127–134.
- Travers, A. A. 1995. Reading the minor groove. *Nature Struct. Biol.* 2:615–618.
- Wong, J. M., and E. Bateman. 1994. TBP-DNA interactions in the minor groove discriminate between A:T and T:A basepairs. *Nucleic Acids Res.* 22:1890–1896.