Binding Mechanisms of TATA Box-Binding Proteins: DNA Kinking is Stabilized by Specific Hydrogen Bonds

Leonardo Pardo,* Mercedes Campillo,* David Bosch,* Nina Pastor,† and Harel Weinstein‡

*Laboratori de Medicina Computacional, Unitat de Bioestadística, Facultat de Medicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain; †Facultad de Ciencias, Universidad Autónoma del Edo. de Morelos, 62210 Cuernavaca, Morelos, México; and ‡Department of Physiology and Biophysics, Mount Sinai School of Medicine, New York, New York 10029 USA

ABSTRACT One of the common mechanisms of DNA bending by minor groove-binding proteins is the insertion of protein side chains between basepair steps, exemplified in TBP (TATA box-binding protein)/DNA complexes. At the central basepair step of the TATA box TBP produces a noticeable decrease in *twist* and an increase in *roll*, while engaging in hydrogen bonds with the bases and sugars. This suggests a mechanism for the stabilization of DNA kinks that was explored here with ab initio quantum mechanical calculations and molecular dynamics/potential of mean force calculations. The hydrogen bonds are found to contribute the energy necessary to drive the conformational transition at the central basepair step. The Asn, Thr, and Gly residues involved in hydrogen bonding to the DNA bases and sugar oxygens form a relatively rigid motif in TBP. The interaction of this motif with DNA is found to be responsible for inducing the un*twist*ing and *roll*ing of the central basepair step. Notably, direct readout is shown not to be capable of discriminating between AA and AT steps, as the strength of the hydrogen bonds between TBP and the DNA are the same for both sequences. Rather, the calculated free energy cost for an equivalent conformational transition is found to be sequence-dependent, and is calculated to be higher for AA steps than for AT steps.

INTRODUCTION

The bending of DNA associated with protein-DNA interaction is an important element of biological activity (Dickerson, 1998 and references therein). Recently determined three-dimensional structures of minor groove-inserting DNA-bending proteins implicated in the control and regulation of transcription have revealed a new mechanism to produce the bend in DNA (for reviews see Werner and Burley, 1997; Werner et al., 1996)). This common mechanism involves the insertion of one or several side chains into the minor groove of DNA, and the unstacking of two contiguous basepairs, to produce noticeable kinks. As a consequence of the side chain insertion the DNA structure is severely bent toward the major groove. Known three-dimensional structures in this group are the complexes with TATA-box binding protein (TBP) (Geiger et al., 1996; Juo et al., 1996; Kim et al., 1993a, b; Kosa et al., 1997; Nikolov et al., 1995; Tan et al., 1996), with the purine PurR (Schumacher et al., 1994) and lactose operon LacI (Lewis et al., 1996) repressors, with the high-mobility group (HMG) domain of LEF-1 (Love et al., 1995) and SRY (Werner et al., 1995) proteins, with the integration host factor IHF (Rice et al., 1996), and with the hyperthermophile chromosomal protein Sac7d (Robinson et al., 1998). Other examples exist where bending of DNA is induced without the insertion of protein side chains (Clark et al., 1993; Escalante et al., 1998; Li et al., 1995; Schultz et al., 1991).

TBP specifically binds to the consensus sequence TATAt/aAt/aX (TATA box). The TATA box is usually located 30 basepairs upstream from the transcriptional start site of the core promoter (Burley and Roeder, 1996; Pugh, 1996). Structural comparisons (Guzikevich Guerstein and Shakked, 1996) and molecular dynamics simulations (Pardo et al., 1998a) led to the suggestion that the bent DNA can be characterized as an A-DNA structure in which the torsion angle of the glycosyl bond (χ), of the 16 nucleotides comprising the TATA box was modified by ~45°. Classification of known protein-DNA complexes (Dickerson, 1998) has suggested that the TATA box exhibits three-dimensional *writhe* resulting from positive *roll* at the successive basepairs steps along the TATA box. The insertion of side chains from TBP (two pairs of conserved Phe residues), at the first two (TA) and last two (t/aX) basepair steps of the TATA box, may have a key role in achieving the desirable change in χ and the positive *roll* at these steps. However, the mechanism by which the central base step (At/a) of the TATA box can undergo such a change in χ and *roll* remains unknown. In a previous study, we suggested from molecular dynamics simulations of TATA box-containing DNA oligomers, that the direct hydrogen bonds between two Asn and two Thr side chains with this central basepair step may provide the driving force for the kink in this step (Pastor et al., 1997). Here we explore the elements of the structural and energetic basis for this hypothesis in the context of the smooth transition from A-DNA to TA-DNA described earlier (Pardo et al., 1998a). We report results from 1) molecular dynamics/potential of mean force (MD/PMF) simulations used to calculate the energy required to kink the

Received for publication 24 August 1999 and in final form 29 December 1999.

Address reprint requests to Harel Weinstein, Dept. of Physiology and Biophysics, Mount Sinai School of Medicine, One Gustave Levy Place, Box 1218, New York, NY 10029-6574. Tel.: 212-241-7018; Fax: 212-860-3369; E-mail: hweinstein@inka.mssm.edu.

0006-3495/00/04/1988/09 $2.00

central basepair step of the DNA tetramers TATA and TAAA; 2) ab initio quantum mechanical calculations of the hydrogen bond interactions identified in two TBP complexes (Kim et al., 1993a, b), between the two Asn and two Thr residues of TBP and the TA and AA steps of DNA; and 3) analysis of hydrogen bond interactions during the conformational transition from A-DNA into the TA-DNA form (Guzikevich Guerstein and Shakked, 1996) found in these TBP/DNA complexes. The choice of A-DNA as the starting conformation reflects the fact that unrestrained molecular dynamics simulations suggest a predisposition of TATA sequences toward attaining a conformation close to A-DNA (Flatters et al., 1997; Pastor et al., 1998), a behavior that was not observed in other DNA oligomers without the TATA box (Flatters and Lavery, 1998).

METHODS

Free energy penalty for kinking the central basepair step of the TATA box

The model systems used in the free energy calculations consist of the DNA tetramers TATA and TAAA, built in canonical A-DNA conformation (Arnott and Hukins, 1972) defined in AMBER 4.1 (Pearlman et al., 1995). These sequences are found in the central basepair step of the TATA box, in complexes with *Saccharomyces cerevisiae* TBP (SCE) (Kim et al., 1993b) and *Arabidopsis thaliana* TBP2 (ATH) (Kim et al., 1993a), respectively. To achieve electroneutrality of the systems, counterions (a total of 6 sodium ions) positioned initially at a distance of 3.5 Å from each P atom along the O-P-O bisector were included in the simulations using the EDIT module in AMBER 4.1. The DNA and the counterions were placed in a rectangular box containing Monte Carlo-equilibrated TIP3P water (1327 and 1351 water molecules for TATA and TAAA, respectively) with the EDIT module in AMBER 4.1. Initial equilibration was carried out with the DNA atoms fixed, the sodium ions and water molecules minimized (100 steepest descent steps followed by 400 conjugate gradient steps), heated (from 0 to 300 K in 15 ps), and equilibrated (from 15 to 100 ps) at constant pressure with isotropic scaling (Berendsen coupling). The final box sizes after equilibration were 36.8 × 36.7 × 31.0 Å for TATA, and 37.1 × 37.0 × 31.5 Å for TAAA, resulting in a final density of ~1.04 g cm^{-3} in both cases. Subsequently, all elements of the systems were energy-minimized (100 steepest descent steps followed by 400 conjugate gradient 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 α, β, γ, δ, ϵ, ζ, and χ, and the Watson-Crick basepairing, were maintained close to the initial A-DNA conformation (±5° for torsional angles and ±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 results in similar equilibrated initial conformations for the DNA tetramers compared in the calculation, thus allowing for direct comparisons of free energy differences between them.

The transitions from the initial A-DNA structures to the kinked structures were carried out with an MD/PMF procedure by changing in 51 windows the torsional angles α, β, γ, δ, ϵ, ζ, and χ, and the Watson-Crick hydrogen bond distances from the initial A-DNA values to those found in complexes with SCE or ATH. The contribution of these constraints to the free energy [E(van der Waals) + E(electrostatic) + E(internal energies)] was evaluated by the thermodynamic integration/constraint forces method (Pearlman, 1993). Following a procedure described before (Pardo et al., 1998b), the simulation times were chosen from convergence tests. A total simulation time of 1530 ps, 51 windows × (5 ps of equilibration + 25 ps of data collection), was used. The Helmholtz free energy differences (ΔF) are reported, as the conformational transitions were conducted at constant volume.

The MD/PMF simulations were run with the Sander and Gibbs modules of AMBER 4.1, the all-atom force field (Cornell et al., 1995), SHAKE bond constraints in all bonds, a 2 fs integration time step, and constant temperature of 300 K coupled to a heat bath (Berendsen coupling). 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). Conformational analysis of the DNA structures was performed with CURVES 5.1 (Lavery and Sklenar, 1988, 1989). Because this algorithm performs a global fit to the DNA axis, the global basepair step parameters depend on the DNA length. Thus, the reported parameters of both simulated and experimental structures are calculated on DNA tetramers.

A model of the interaction between the central basepair step of the TATA box and selected protein side chains of TBP

Direct hydrogen bond interactions between the central basepair step of the TATA box and TBP have been identified in the crystal structures of TBP/DNA complexes (Kim et al., 1993a, b): 1) Ade-5:Thy-25 and Thy-6:Ade-24 of DNA (AT central basepair step) with Asn-69, Thr-124, Gly-125, Asn-159, Thr-215, and Gly-216 of SCE; or 2) Ade-6:Thy-23 and Ade-7:Thy-22 of DNA (AA central basepair step) with Asn-27, Thr-82, Gly-83, Asn-117, Thr-173, and Gly-174 of ATH. To characterize the properties of these interactions with ab initio quantum mechanical calculations, the model system shown in Figs. 1 and 2 was constructed to include the following molecular fragments: the free bases Ade or Thy (the sugar moieties were replaced by a hydrogen) and the side chains of Asn (the C_α atom of the backbone is included), Thr (the C_α and CO moieties of the backbone are included), or Gly (the C_α and NH moieties of the backbone are included). All free valences were capped with hydrogen atoms. Thr-82 in ATH orients the $O_\gamma H$ group toward O_1' of Thy (Kim et al., 1993a). Thus, the C_1', O_1', and C_4' groups of the sugar moiety were attached to the bases that are involved in the hydrogen bond to Thr: Ade-5 and Ade-24 in the complex with SCE and Ade-6 and Thy-22 in the complex with ATH. During the energy optimization of the system (each strand of DNA with the corresponding protein side chains were optimized separately) the atoms of Thy and Ade, and the backbone atoms of Asn, Thr, and Gly were kept fixed at the positions originally determined by x-ray crystallography (Kim et al., 1993a, b). Becke's three-parameter hybrid functional using the LYP correlation functional (B3LYP) (Becke, 1993) and the 6-31G basis set were used for the optimization of the released atoms. The energy of interaction, E_{int}, evaluated with the 6-31G** basis set at the B3LYP level of density functional theory (Parr and Yang, 1989), represents the stabilization of the complex. This energy is defined as the difference in energy between the optimized complex and the sum of the energies of the DNA bases and TBP side chains, calculated in the conformation obtained in the complex. All the quantum mechanical calculations were performed with the GAUSSIAN94 system of programs (Frisch et al., 1995) using the default criteria for convergence.

Evolution of the interaction between the protein side chains of TBP and the central basepair step of the TATA box during the conformational transition

To characterize the evolution of the hydrogen bond interactions between TBP side chains and DNA bases (see above) during the conformational transition from the A to the TA form of DNA observed in the complex with SCE, a model system was constructed consisting on the heavy atoms of the average structures computed from the MD/PMF calculations. Specifically, the stages were defined at windows 1 (A-DNA), 11 (20% of TA-DNA

character), 21 (40% of TA-DNA character), 31 (60% of TA-DNA character), 41 (80% of TA-DNA character), and 51 (TA-DNA) in the MD/PMF simulations (see above). The heavy atoms of the DNA bases, at positions obtained from energy optimization at the B3LYP level of theory with the 6–31G basis set, were superimposed on the central AT base step of the coding strand A (corresponding to Ade-5 and Thy-6 in Kim et al., 1993b). The protein side chains of Asn-159 (the C_α atom of the backbone is included), Thr-215 (the C_α and CO moieties of the backbone are included), and Gly-216 (the C_α and NH moieties of the backbone are included) were added to the AT base step in the orientation determined by x-ray crystallography (Kim et al., 1993b). Energy minimization of the systems, at the B3LYP level of theory and the 6–31G basis set, were performed as described above. The energy of the systems was evaluated with the 6–31G** basis set at the B3LYP level of theory for the optimized geometries.

RESULTS AND DISCUSSION

Free energy cost of kinking the central basepair step of the TATA box

The free energy for the conformational transition of the central base step (At/a) of the TATA box from the A to the TA form of DNA was evaluated with the MD/PMF simulations described in Methods, on the DNA tetramers TATA and TAAA, found in complexes with SCE and ATH, respectively. The choice of the A-DNA conformation as a starting point reflects the following findings: 1) the conformational parameters of the TATA box in the complex with TBP resemble those of the A-DNA structure, with the exception of the glycosyl-bond torsion angle χ (Guzikevich Guerstein and Shakked, 1996); 2) unrestrained molecular dynamics simulations reproduce the apparent predisposition of DNA dodecamers containing TATA sequences toward attaining a conformation close to A-DNA (Flatters et al., 1997; Pastor et al., 1998) that could not be observed in other DNA oligomers without the TATA box (Flatters and Lavery, 1998); and 3) the bending of the TATA box can be achieved by gradual changes in the torsion angle of the glycosyl-bond χ, within the constraints of an A-DNA backbone (Pardo et al., 1998a). Thus, the conformational transition was completed primarily by modifying the value of χ in 51 windows (see Methods) through a smooth change (see Fig. 3 for a display of the average structures computed from the data collection trajectory at each window for the TATA tetramer), in a similar manner to the previously reported conformational transition for the first, and most conserved, basepair step (TA) of the TATA box (Pardo et al., 1998b). Note that the dynamics simulations resulted in small adjustments of all other backbone dihedral angles as well, but these remained within the range of values corresponding to A-DNA.

Fig. 4 *a* shows the root mean square deviation (rmsd) between the average structures computed from the data collection trajectory in each window of the TATA and TAAA tetramers and the crystal structures of the complex with SCE and ATH, respectively. The rmsd profiles indicate that as the simulation progresses, the computed average structures at each window become more similar to the conformation found in the crystal structures. TBP binding to the TATA box significantly alters the helical parameters *rise*, *roll*, and *twist* (Werner et al., 1996). Fig. 5 depicts these inter-base parameters for the central basepair step of the average structures computed during the data collection trajectory in the first window (initial structure) and the last window (final structure) of the MD/PMF simulations, compared to the experimental values in the complexes with SCE (Kim et al., 1993b), ATH (Kim et al., 1993a), human TBP (Juo et al., 1996), yeast TFIIA/TBP (Tan et al., 1996), or human TFIIB/TBP (Nikolov et al., 1995). The conformational parameters of the simulated structures (Fig. 5) agree well with the corresponding data from experiment. This extent of agreement with the experimental data is noteworthy, because the conformational transitions were achieved by modifying only the backbone torsional angles. Differences between the DNA tetramers are also evident, as the central AT step of the TATA tetramer is found to have higher *rise* and *roll*, and lower *twist* than the central AA step of the TAAA tetramer. This indicates that TBP binding results in a more significant conformational change of the TATA than the TAAA tetramer.

Fig. 4 *b* shows the computed free energy changes during the conformational transition from A-DNA to the conformation found in the complex with SCE and ATH ($\lambda = 0 \rightarrow 1$, 0 representing A-DNA and 1 TA-DNA). These free energy values must be considered only relative to each

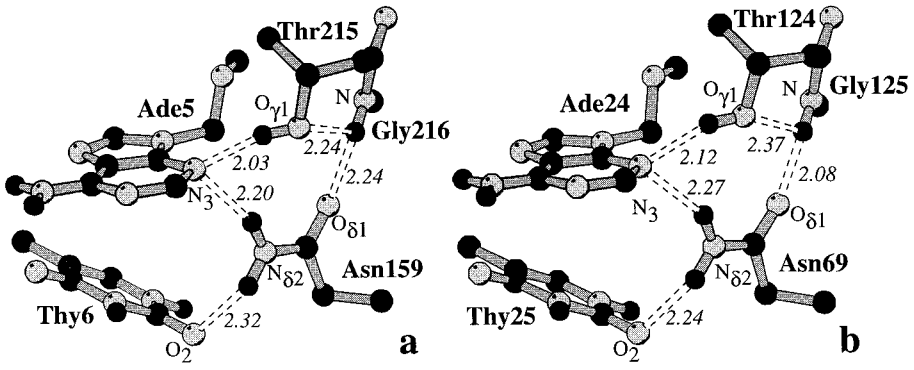

FIGURE 1 Molecular models used in the calculation of the hydrogen bond interactions between (a) the Asn-159, Thr-215, and Gly-216 side chains of SCE and Ade-5 and Thy-6 bases of the TATA box; and (*b*) the Asn-69, Thr-124, and Gly-125 side chains of SCE and Ade-24 and Thy-25 bases of the TATA box. The positions of the atoms of Thy and Ade, and the backbone atoms of Asn, Thr, and Gly were kept fixed at the positions originally determined from x-ray crystallography (Kim et al., 1993b).

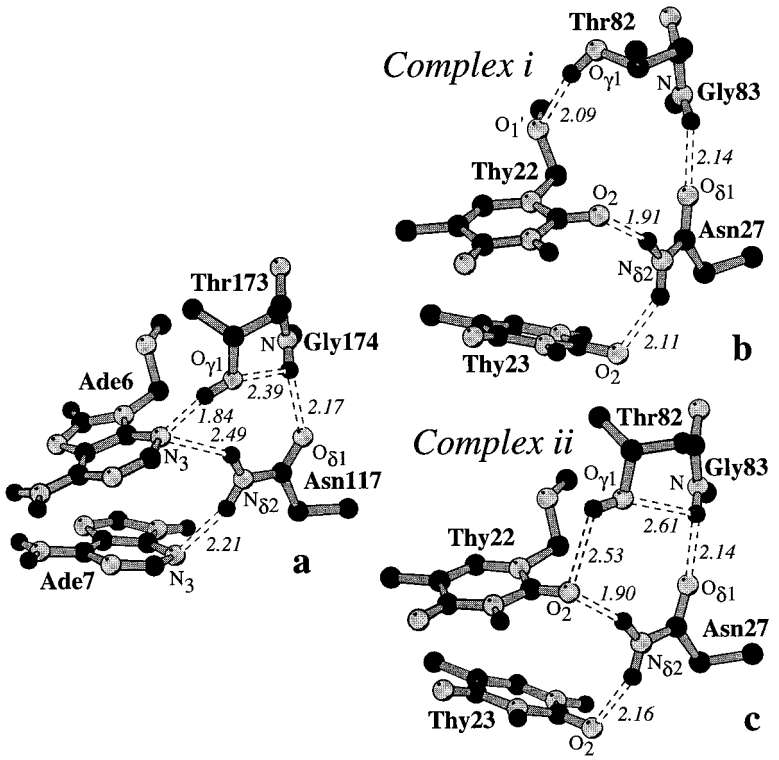

FIGURE 2 Molecular models used in the calculation of the hydrogen bond interactions between (*a*) the Asn-117, Thr-173, and Gly-174 side chains of ATH and Ade-6 and Ade-7 bases of the TATA box; and (*b*, *c*) the Asn-27, Thr-82, and Gly-83 side chains of ATH and Thy-22 and Thy-23 bases of the TATA box. *Complex i* denotes the mode of interaction, observed in the crystal structure (Kim et al., 1993a), in which Thr-82 orients the $O_\gamma H$ group toward O_1' of Thy-22 (*b*). *Complex ii* denotes the mode of interaction in which Thr-82 orients the $O_\gamma H$ group toward O_2 of Thy-22 (*c*) in a similar manner to the other Thr side chains of SCE and ATH.

other, as they cannot be expected to accurately describe the absolute energies of DNA kinking due to the following approximations: 1) the choice of the starting conformation; 2) the effects of the interaction with the protein have not been taken into account; and 3) the discrepancies between computed and crystal structures (e.g., see Pardo et al., 1998b for a more detailed consideration of these limitations). Because these approximations may be assumed to equally affect the compared structures, the difference in the values of free energy change, $\Delta\Delta F$, should yield a qualitative indication of the comparative energy differences. Thus, the energetic cost for the conformational change in the TATA tetramer (11.8 kcal/mol, see solid line in Fig. 4 *b*) is higher by 3.7 kcal/mol than for the TAAA tetramer (8.1 kcal/mol). The larger conformational change produced by SCE (see above) is reflected in the calculated energies.

We have previously reported that the energy cost to kink AA or TT steps is higher than TA or AT steps because of the steric clash between methyl groups of Thy in the kinked conformation (Pardo et al., 1998b). To confirm this observation we evaluated the free energy penalty for the conformational transition produced in the TAAA sequence by SCE. Thus, for a similar conformational change the in TAAA tetramer, the cost is 2.6 kcal/mol higher (14.4 kcal/mol, results not shown in Fig. 4 *b*) than the for TATA tetramer (11.8 kcal/mol, see Fig. 4 *b*). This is in good qualitative agreement with the previous observation (Pardo et al., 1998b). An estimate of the relative free energy penalty for this structural transition was obtained earlier from a statistical analysis on unrestrained molecular dynamics simulations of seven DNA dodecamers (Pastor et al., 1997). A comparison of the relative propensities of these two sequences to achieve the conformation found at the central basepair step in their complexes with TBP indicated that TATA requires less energy than TAAA (9.5 versus 11.3 kcal/mol) to achieve the average conformation in TBP/DNA complexes (see Pastor et al., 1997). A free energy difference of 1.8 kcal/mol between the two tetramers was obtained through Boltzmann weighing of the spontaneous population of the bound conformation, and this result is in good agreement with the energies evaluated by the MD/PMF simulations described here.

Notably, the energy required to kink the central basepair step is lower than the previously reported energy required to produce the kink in the first basepair step (TA) of the TATA box associated with the insertion of two Phe side chains (compare the results in Table 3 and Fig. 6 *B* of Pardo et al., 1998b with those of Fig. 4 *b*). The smaller kink produced by SCE in the central basepair step (see Table 3 in Pastor et al., 1997) is also reflected in the lower energy calculated here for this deformation.

The interaction between the central basepair step of the TATA box and the protein side chains of TBP

The direct hydrogen bond interactions between TBP and the central base step of the TATA box involve the side chains

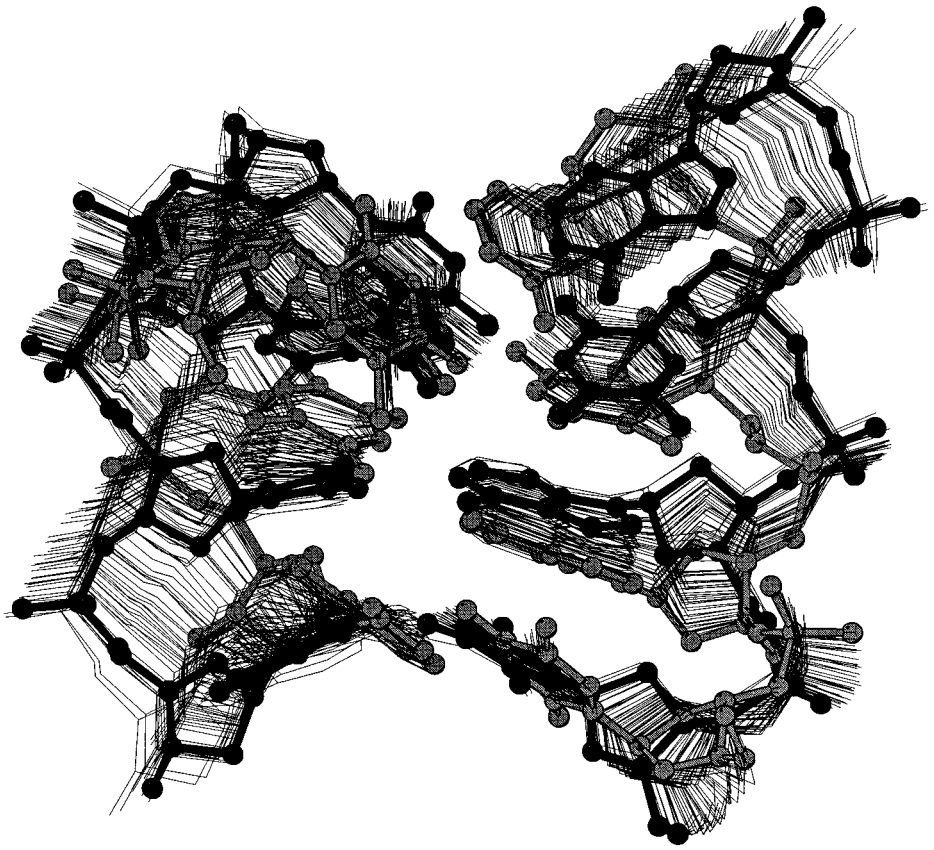

FIGURE 3 Average structures computed from the data collection trajectory at each window during the conformational transition for the TATA tetramer. The initial A-DNA (*black*) and final TA-DNA (*gray*) conformations are shown in the ball-and-stick rendering. The figure was created using MOLSCRIPT (Kraulis, 1991).

of two Asn (69 and 159 of SCE, and 27 and 117 of ATH) and two Thr residues (124 and 215 of SCE, and 82 and 173 of ATH) (Kim et al., 1993a, b). Figs. 1 and 2 depict the molecular models used in the ab initio calculation of the corresponding hydrogen bond interaction energies (see Methods for computational details). We assumed that the orientations of both the Asn/Thr side chains and the central Ade/Thy bases are determined by the protein secondary structure and the DNA sugar-phosphate backbone to which these moieties are bound. Consequently, the positions of the atoms of Thy and Ade, and the backbone atoms of Asn, Thr, and Gly were kept fixed at the positions originally determined by x-ray crystallography (Kim et al., 1993a, b). In a separate test of the role of the glycine residues in the protein-DNA interface region (Gly-125 and Gly-216 of SCE, and Gly-83 and Gly-174 of ATH), optimization of the molecular models in the absence of these residues yielded structures in which the side chains of Asn and Thr could not achieve the hydrogen bond interactions with the DNA bases (results not shown). The apparent reason for this result is the electrostatic repulsion between the $O_{\gamma 1}$ atom of Thr and the $O_{\delta 1}$ atom of Asn (see Figs. 1 and 2). Thus, Gly appears to stabilize the appropriate conformation to the Asn and Thr side chains through hydrogen bonds of the solvent-accessible backbone: N-H···$O_{\gamma 1}$(Thr) and N-H···$O_{\delta 1}$(Asn) (see Figs. 1 and 2).

Two possible modes of interaction between Asn-27, Thr-82, and Gly-83 of ATH and the noncoding strand Thy-22 and Thy-23 of the TATA box were considered in the models: *Complex i* denotes the mode of interaction observed in the crystal structure (Kim et al., 1993a), in which the $O_{\gamma 1}H$ group of Thr-82 is oriented toward O_1' of Thy-22 (see Fig. 2 *b*). *Complex ii* denotes the mode of interaction in which the $O_{\gamma 1}H$ group of Thr-82 is oriented toward O_2 of Thy-22 (see Fig. 2 *c*), in a manner similar to the other Thr side chains of SCE and ATH. Table 1 shows the interaction energies between the ATH side chains and the DNA bases in both conformations. The ATH side chains form stronger hydrogen bond interactions in *Complex i* (−17.9 kcal/mol) than in *Complex ii* (−14.6 kcal/mol). That the O_1'···$H_{\gamma 1}$-$O_{\gamma 1}$ hydrogen bond (*Complex i*) is stronger than the O_2···$H_{\gamma 1}$-$O_{\gamma 1}$ bond (*Complex ii*) is also corroborated by the shorter calculated length of the O_1'···$H_{\gamma 1}$ bond (2.09 Å, see Fig. 2 *b*) than the O_2···$H_{\gamma 1}$ distance (2.53 Å, see Fig. 2 *c*). The former bond O_1'···$H_{\gamma 1}$-$O_{\gamma 1}$ is also more linear (145°) than the O_2···$H_{\gamma 1}$-$O_{\gamma 1}$ bond (122°). It is important to note that the reported interaction energies do not include the difference in energy of the ATH side chains in both conformations (see Methods). This component may be significant because Thr-82 in *Complex i* cannot form the hydrogen bond between the NH backbone of Gly-83 and the $O_{\gamma 1}$ atom of Thr (see Fig. 2, *b* and *c*). A single point energy evaluation of the

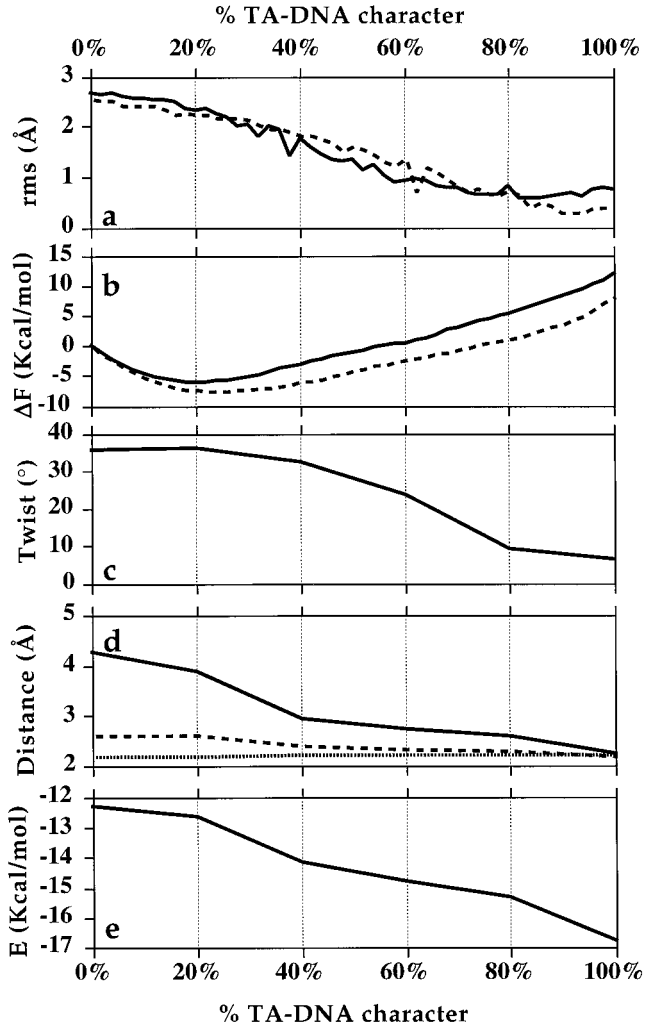

FIGURE 4 (*a*) Rmsd comparisons of the average structures computed from the data collection trajectory at each window for the TATA (*solid line*) and TAAA (*broken line*) tetramers, and the crystal structures of TA-DNA in the complex with SCE (Kim et al., 1993b) and ATH (Kim et al., 1993a), respectively. The rmsd values are shown for the heavy atoms of the entire sequence. (*b*) Computed free energy changes during the conformational transition of the TATA (*solid line*) and TAAA (*broken line*) tetramers. (*c*) Evolution of the basepair step parameter *twist* during the conformational transition of the TATA tetramer. (*d*) Evolution of the hydrogen bond interactions between the O_2 atom of Thy-6 and the hydrogen atom of the $H_{\delta 2}$-$N_{\delta 2}$ bond of Asn-159 (O_2···$H_{\delta 2}$-$N_{\delta 2}$, *solid line*) and the H atom of the NH bond of the Gly-216 backbone and the $O_{\gamma 1}$ atom of Thr-215 (N-H···$O_{\gamma 1}$, *dotted line*) and the $O_{\delta 1}$ atom of Asn-159 (N-H···$O_{\delta 1}$, *broken line*). (*e*) Evolution of the energy of interaction between the modeled residues of SCE (Asn-159/Thr-215/Gly-216) and the central base step (Ade-5/Thy-6) of the TATA box.

ATH side chains in both conformations (using B3LYP/6–31G**) shows that Asn-27, Thr-82, and Gly-83 are more stable in *Complex ii* than in *Complex i* by 3.2 kcal/mol. This difference in energy represents the cost of breaking the N-H···$O_{\gamma 1}$ hydrogen bond. However, the stronger interaction of the ATH side chains with the DNA bases in *Complex i* compensates for the breakage of the N-H···$O_{\gamma 1}$ hydrogen bond (a comparison of the total energies of both complexes shows that *Complex i* is more stable than *Complex ii* by only 0.1 kcal/mol). Given that the *complexes* are isoenergetic and that *Complex i* is also representative of the interactions found in the human TBP/DNA complex (Juo et al., 1996), we will consider *Complex i* as the model for the interaction.

Note that, as for the case of the first basepair step of the TATA box (Pardo et al., 1998b), the ability of the protein side chains to discriminate between AT or AA base steps at the central basepair step of the TATA box is not determined at the level of interaction energies. Thus, the E_{int} values in Table 1, which represent interaction energies between the modeled residues of SCE (see Fig. 1) or ATH (Fig. 2) and the coding strand (Figs. 1 *a* and 2 *a*), or the noncoding strand (Figs. 1 *b* and 2 *b*) of the central basepair step of the TATA box, are nearly equal for the *Complex i* arrangement.

These results are in qualitative agreement with experimental observations. Thus, electrophoretic mobility shift assays (Wong and Bateman, 1994) have been used to classify the sequences of the TATA box from a pool of random double-stranded oligonucleotides into four different classes according to their affinity for TBP: TATATAA (K_{eq} = 1.1×10^{-9} M), TATATATA (1.4×10^{-9} M), TATAAATA (1.6×10^{-9} M), and TATAAAA (3.7×10^{-9} M). It seems reasonable to assume that the TATATATA sequence adopts, at the central basepair step, the conformation observed in the complex with SCE, and that the TATAAATA sequence adopts the conformation observed in the complex with ATH. On this basis we conclude that the binding affinity of SCE or ATH for the central sequence TATA or TAAA, respectively, is similar. In agreement with this experimental result, the calculated hydrogen bond energies between TBP and the central base step of the TATA box are of the same magnitude for SCE and ATH (see Table 1 and above). However, we found the conformational change of the central TATA tetramer, driven by SCE, to require 3.7 kcal/mol more than the conformational change of the TAAA tetramer, driven by ATH (see Fig. 1 *b*). This larger free energy cost of distorting the TATA sequence by SCE is compensated in the complex with ATH by the energy cost (3.2 kcal/mol) evaluated with ab initio quantum mechanical calculations (see above) for the conformational change of the Asn-27, Thr-82, and Gly-83 side chains of ATH (breaking the N-H···$O_{\gamma 1}$ hydrogen bond; see Fig. 2 *b*). This "compensation" is reflected in the good agreement between the theoretical analysis and the experimental results showing nearly identical complexation energies for the two cases. The clear importance of hydrogen bonding as a mechanism to compensate for DNA bending is also identifiable in the hyperthermophile TBP/DNA complex, where the interactions extend to three bases on each side of the center of the TATA box (Kosa et al., 1997).

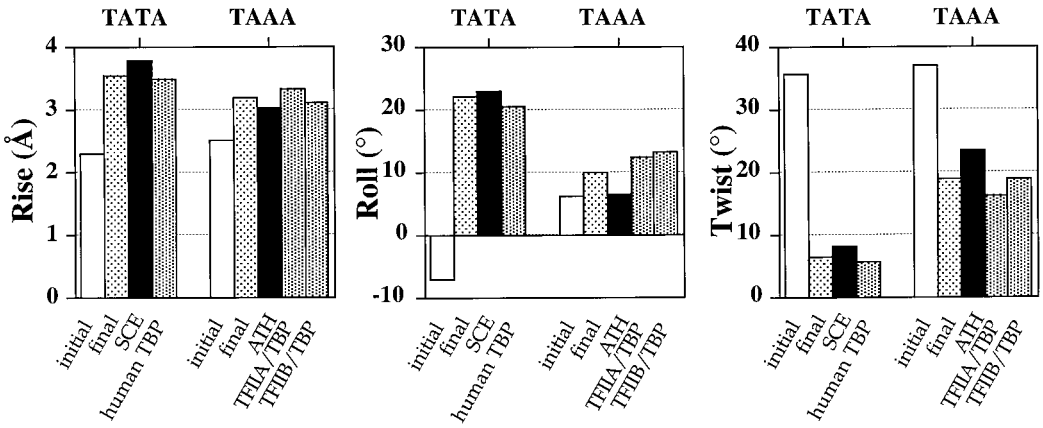

FIGURE 5 Helical global parameters *rise*, *roll*, and *twist*, calculated with CURVES 5.1 (Lavery and Sklenar, 1988, 1989), of the average structures computed during the data collection trajectory at the first (initial structure, *white bars*) and the last (final structure, *white dotted bars*) windows of the MD/PMF simulations, and of the experimental complexes with SCE (Kim et al., 1993b; *solid bars*), ATH (Kim et al., 1993a; *solid bars*), human TBP (Juo et al., 1996; *black dotted bars*), yeast TFIIA/TBP (Tan et al., 1996; *black dotted bars*), or human TFIIB/TBP (Nikolov et al., 1995; *black dotted bars*).

The interaction between TBP and the central basepair step of the TATA box during the conformational transition

The direct hydrogen bonds between the Asn/Thr/Gly motif of TBP and the central basepair step of the TATA box are likely to be essential in the formation of the kink in this basepair step (Pastor et al., 1997). The mechanistic role of these hydrogen bond interactions in the kinking of DNA was evaluated from the conformational transition of A-DNA into the TA-DNA form observed in the complex with SCE. The DNA/SCE complex was chosen for this analysis over the DNA/ATH complex for the following reasons: 1) the conformational change produced by SCE in the central basepair step of the TATA box is larger than in the ATH complex (see above); 2) the interaction of the Asn, Thr, and Gly side chains of SCE with the AT base step is symmetrical in the coding and the noncoding strand (see Table 1, Fig. 1, and above), so that only one side needs to be calculated; and 3) the Asn, Thr, and Gly side chains of SCE interact directly with the AT bases and not with the sugars.

The molecular models used in these calculations consisted of the free bases Ade-5 and Thy-6 of DNA, and the residues of Asn-159, Thr-215, and Gly-216 of SCE (see Methods). The simplified representation of the protein reflects the assumption that the backbone of TBP does not change, in position and conformation, during the process of DNA bending for the two following reasons. The first reason is that DNA recognition involves the contact between a series of Arg/Lys residues with the negatively charged phosphate moieties of the DNA backbone (Yamamoto et al., 1992). These strong interactions set and maintain the position of the TBP residues that insert (the Phe residues), and of those forming hydrogen-bond interactions (the Asn/Thr/Gly motif) in the minor groove of the TATA box during the process of protein-assisted DNA bending. The second reason is that TBP undergoes only a small conformational change upon DNA binding (Kim et al., 1993a, b). Consequently, the protein backbone atoms can be considered fixed during the energy optimization of the systems at the positions determined in the crystal structure (Kim et al., 1993b).

TABLE 1 Energies of interaction between the modeled residues of SCE (Fig. 1) or ATH (Fig. 2) and the coding strand (Figs. 1 *a* and 2 *a*) or the noncoding strand (Figs. 1 *b* and 2 *b*) of the central basepair step of the TATA box

	Protein Side Chain	DNA Step	E_{int}* (kcal/mol)
SCE	Asn-159/Thr-215/Gly-216	Ade-5/Thy-6 (coding)	−17.9
	Asn-69/Thr-124/Gly-125	Ade-24/Thy-25	−17.8
ATH	Asn-117/Thr-173/Gly-174	Ade-5/Ade-6 (coding)	−17.9
	Asn-27/Thr-82/Gly-83	Thy-24/Thy-25 *Complex i*	−17.9
		Complex ii	−14.6

*Calculated at the B3LYP/6-31G**//B3LYP/6-31G level of theory.

Fig. 6 depicts the structures of the central basepair step (Ade-5 and Thy-6) at the conformations calculated at windows 1 (A-DNA), 11 (20% of TA-DNA character), 21 (40% of TA-DNA character), 31 (60% of TA-DNA character), 41 (80% of TA-DNA character), and 51 (TA-DNA) of the MD/PMF simulations, and the optimized side chains of Asn-159, Thr-215, and Gly-216 for each of the conformations. The comparison of the evolving structures brings into evidence the repositioning of Thy-6 and Asn-159 to achieve the hydrogen bond interaction between the O_2 atom of Thy-6 and the $H_{\delta 2}$-$N_{\delta 2}$ atoms of Asn-159. Thus, the central basepair step of the TATA box must undergo a significant decrease of the *twist* parameter (Fig. 4 *c* shows the changes in *twist* during the conformational transition) to accomplish the interaction with Asn-159. To further illustrate the formation of this hydrogen bond, Fig. 4 *d* shows the evolution

of the O_2···$H_{\delta 2}$-$N_{\delta 2}$ distance during the conformational transition. The profile indicates that as the simulation progresses the hydrogen bond distance is shortened to a final O_2···$H_{\delta 2}$ value of 2.25 Å. Fig. 4 *d* also shows the evolution of the other hydrogen bond distances between the NH atoms of the Gly backbone and both the $O_{\gamma 1}$ atom of Thr (N-H···$O_{\gamma 1}$) and the $O_{\delta 1}$ atom of Asn (N-H···$O_{\delta 1}$) during the A-DNA to TA-DNA transition. It is remarkable that these two hydrogen bond distances are invariant during the conformational transition. These findings suggest that the rigidity of the Asn/Thr/Gly motif of TBP, supported by the hydrogen bonds of the NH atoms of the Gly backbone (see above), drives the distortion of the central basepair step of the TATA box to facilitate the formation of the hydrogen bond interaction between the O_2 atom of Thy-6 and the $H_{\delta 2}$-$N_{\delta 2}$ atoms of Asn-159. These interactions could not be realized for a TATA box in the A-DNA conformation. The significant rigidity of this motif in TBP has also been observed in unrestrained MD simulations of the ATH and SCE proteins, both free and in complex with DNA (N. Pastor, unpublished observations).

Finally, Fig. 4 *e* shows the evolution of the energy of interaction between the modeled residues of SCE and the central base step of the TATA box. The gradual transformation from the A-DNA to the TA-DNA form produces a significant increase in the magnitude of the hydrogen bond interaction from an initial value of −12.3 kcal/mol (A-DNA) to a final value of −16.8 kcal/mol (TA-DNA), consonant with the gradual formation of the hydrogen bond interactions observed in the crystal structure.

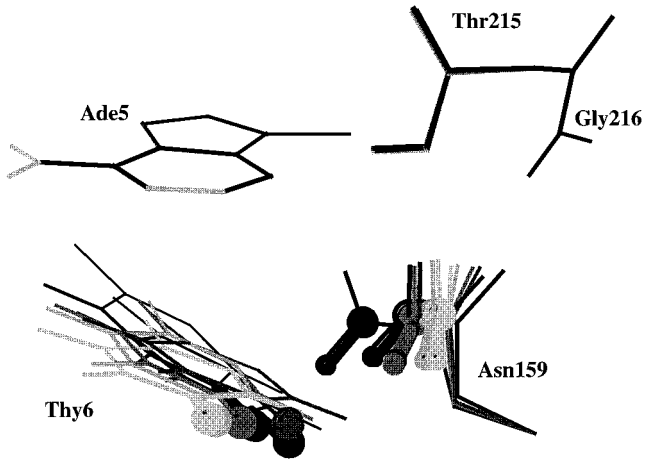

FIGURE 6 Structures of the central basepair step Ade-5 and Thy-6 of the TATA box at the conformations calculated at windows 1 (A-DNA, the *lightest gray*), 11 (20% of TA-DNA character), 21 (40% of TA-DNA character), 31 (60% of TA-DNA character), 41 (80% of TA-DNA character), and 51 (TA-DNA, the *darkest gray*) of the MD/PMF simulations, and the optimized side chains of Asn-159, Thr-215, and Gly-216 at each conformation (Kim et al., 1993b). Only polar hydrogens are shown. The O_2 atom of Thy-6 and the $H_{\delta 2}$-$N_{\delta 2}$ atoms of Asn-159 are shown in the ball-and-stick representation.

CONCLUSIONS

We have recently proposed (Pardo et al., 1998a) that the mechanism of DNA bending from an A-DNA-like conformation to the TA-DNA structure observed in the complex with TBP is achieved by gradual changes in the torsion angle of the glycosyl bond of the nucleotides comprising the TATA box. This putative conformational transition had been identified earlier from crystallographic data (Guzikevich Guerstein and Shakked, 1996). The present study suggests that the energy required for this conformational transition differs for TATA compared to TAAA tetramers at the central basepair step of the TATA box: 11.8 kcal/mol vs. 8.1 kcal/mol, respectively. However, this energetic cost is partially compensated by a significant increase in the energy of the hydrogen bond interactions between the Asn/Thr/Gly motif of SCE and the AT central base step of the TATA box (−4.5 kcal/mol per strand, as was evaluated by ab initio quantum mechanical calculations). Thus, the conformational change at the central base step facilitates the formation of the hydrogen bond interaction between the O_2 atom of Thy and the $H_{\delta 2}$-$N_{\delta 2}$ atoms of Asn, which cannot be formed when the TATA box is in the A-DNA conformation. Consequently, the mechanism of protein-assisted DNA bending by changing the glycosyl bond extends beyond the insertion of two pairs of conserved Phe residues, at the first two (TA) and last two (t/aX) basepairs of the TATA box, to the formation of direct hydrogen bonds between two Asn, two Gly, and two Thr side chains with the central two basepairs (At/a) of the TATA box. To account for the entire effect of the protein, the calculations will need to consider as well the overall electrostatic effects of the protein on the energetics of bending. The low dielectric interior of proteins (Elcock and McCammon, 1996) and phosphate neutralization (Lebrun et al., 1997) are likely to favor the energetics of bending.

The present findings, and the differences in structure and energetics calculated for the different TATA box sequences, point to a complex interplay of direct readout and induced fit as the likely source of selectivity differences in the various DNA/TBP complexes, both wild-type and mutant. These energetic contributions are superimposed on the effect of dehydration of the interaction surfaces on the total energy of formation of the protein-DNA complex. The sequence-dependent insights into the mechanism of complex formation, combined with the relative energetic costs evaluated here for the various sequence-dependent steps, should be useful in guiding novel experiments to probe the function of specific TATA boxes and the design of novel TBP constructs with predetermined specificity.

This work was supported in part by a grant from the Association for International Cancer Research (to H.W.), a Fulbright/CONACYT/IIE scholarship (to N.P.), and DGES Grant PB98−0907 (to L.P.). Some of the simulations were run at the Cornell National Supercomputer Facility (spon-

sored 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.

Becke, A. D. 1993. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* 98:5648.

Burley, S. K., and R. G. Roeder. 1996. Biochemistry and structural biology of transcription factor IID (TFIID). *Annu. Rev. Biochem.* 65:769–799.

Clark, K. L., E. D. Halay, E. Lai, and S. K. Burley. 1993. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature.* 364:412–420.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, Jr., 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. 1998. DNA bending: the prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res.* 26:1906–1926.

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.

Escalante, C. R., J. Yie, D. Thanos, and A. K. Aggarwal. 1998. Structure of IRF-1 with bound DNA reveals determinants of interferon regulation. *Nature.* 391:103–106.

Flatters, D., and R. Lavery. 1998. Sequence-dependent dynamics of TATA-box binding sites. *Biophys. J.* 75:372–381.

Flatters, D., M. Young, D. L. Beveridge, and R. Lavery. 1997. Conformational properties of the TATA-box binding sequence of DNA. *J. Biomol. Struct. Dyn.* 14:757–765.

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.

Geiger, J. H., S. Hahn, S. Lee, and P. B. Sigler. 1996. Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science.* 272:830–836.

Guzikevich Guerstein, G., and Z. Shakked. 1996. A novel form of the DNA double helix imposed on the TATA-box by the TATA-binding protein. *Nat. Struct. Biol.* 3:32–37.

Juo, Z. S., T. K. Chiu, P. M. Leiberman, I. Baikalov, A. J. Berk, and R. E. Dickerson. 1996. How proteins recognize the TATA box. *J. Mol. Biol.* 261:239–254.

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.

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.

Kosa, P. F., G. Ghosh, B. S. DeDecker, and P. B. Sigler. 1997. The 2.1-angstrom 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.

Lavery, R., and H. Sklenar. 1988. The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* 6:63–91.

Lavery, R., and H. Sklenar. 1989. Defining the structure of irregular nucleic acids: conventions and principles. *J. Biomol. Struct. Dyn.* 6:655–667.

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.

Lewis, M., G. Chang, N. C. Horton, M. A. Kercher, H. C. Pace, M. A. Schumacher, R. G. Brennan, and P. Lu. 1996. Crystal structure of the lactose operon repressor and its complexes with DNA and inducer. *Science.* 271:1247–1254.

Li, T., M. R. Stark, A. D. Johnson, and C. Wolberger. 1995. Crystal structure of the MATa1/MAT alpha 2 homeodomain heterodimer bound to DNA. *Science.* 270:262–269.

Love, J. J., X. Li, D. A. Case, K. Giese, R. Grosschedl, and P. E. Wright. 1995. Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature.* 376:791–795.

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.

Pardo, L., H. Pastor, and H. Weinstein. 1998a. Progressive DNA bending is made possible by gradual changes in the torsion angle of the glycosyl-bond. *Biophys. J.* 74:2191–2198.

Pardo, L., H. Pastor, and H. Weinstein. 1998b. Selective binding of the TATA box-binding protein to the TATA box-containing promoter: analysis of structural and energetic factors. *Biophys. J.* 75:2411–2421.

Parr, R. G., and W. Yang. 1989. Density-Functional Theory of Atoms and Molecules. Oxford University. Press, Oxford.

Pastor, N., L. Pardo, and H. Weinstein. 1997. Does TATA matter? A structural exploration of the selectivity determinants in its complexes with TBP. *Biophys. J.* 73:640–652.

Pastor, N., L. Pardo, and H. Weinstein. 1998. How the TATA box selects its protein partner. *In* Molecular Modeling of Nucleic Acids. N. B. Leontis and J. J. SantaLucia, editors. Ame. Chem. Soc., Washington, D.C. 329–345.

Pearlman, D. A. 1993. Determining the contributions of constraints in free energy calculations: development, characterization, and recommendations. *J. Chem. Phys.* 98:8946–8957.

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.

Rice, P. A., S. W. Yang, K. Mizuuchi, and H. A. Nash. 1996. Crystal structure of an IHF-DNA complex: a protein-induced DNA U-turn. *Cell.* 87:1295–1306.

Robinson, H., Y. G. Gao, B. S. McCrary, S. P. Edmondson, J. W. Shriver, and A. H. J. Wang. 1998. The hyperthermophile chromosomal protein Sac7d sharply kinks DNA. *Nature.* 392:202–205.

Schultz, S. C., G. C. Shields, and T. A. Steitz. 1991. Crystal structure of a CAP-DNA complex: the DNA is bent by 90 degrees. *Science.* 253: 1001–1007.

Schumacher, M. A., K. Y. Choi, H. Zalkin, and R. G. Brennan. 1994. Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. *Science.* 266:763–770.

Tan, S., Y. Hunziker, D. F. Sargent, and T. J. Richmond. 1996. Crystal structure of a yeast TFIIA/TBP/DNA complex. *Nature.* 381:127–151.

Werner, M. H., and S. K. Burley. 1997. Architectural transcription factors: proteins that remodel DNA. *Cell.* 88:733–736.

Werner, M. H., A. M. Gronenborn, and G. M. Clore. 1996. Intercalation, DNA kinking, and the control of transcription. *Science.* 271:778–784.

Werner, M. H., J. R. Huth, A. M. Gronenborn, and G. M. Clore. 1995. Molecular basis of human 46X, Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell.* 81:705–714.

Wong, J. M., and E. Bateman. 1994. TBP-DNA interactions in the minor groove discriminate between A:T and T:A base pairs. *Nucleic Acids Res.* 22:1890–1896.

Yamamoto, T., M. Horikoshi, J. Wang, S. Hasegawa, P. A. Weil, and R. G. Roeder. 1992. A bipartite DNA binding domain composed of direct repeats in the TATA box binding factor TFIID. *Proc. Natl. Acad. Sci. USA.* 89:2844–2848.