Does TATA Matter? A Structural Exploration of the Selectivity Determinants in Its Complexes with TATA Box-Binding Protein

Nina Pastor,* Leonardo Pardo,*# and Harel Weinstein*

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

ABSTRACT The binding of the TATA box-binding protein (TBP) to a TATA sequence in DNA is essential for eukaryotic basal transcription. TBP binds in the minor groove of DNA, causing a large distortion of the DNA helix. Given the apparent stereochemical equivalence of AT and TA basepairs in the minor groove, DNA deformability must play a significant role in binding site selection, because not all AT-rich sequences are bound effectively by TBP. To gain insight into the precise role that the properties of the TATA sequence have in determining the specificity of the DNA substrates of TBP, the solution structure and dynamics of seven DNA dodecamers have been studied by using molecular dynamics simulations. The analysis of the structural properties of basepair steps in these TATA sequences suggests a reason for the preference for alternating pyrimidine-purine (YR) sequences, but indicates that these properties cannot be the sole determinant of the sequence specificity of TBP. Rather, recognition depends on the interplay between the inherent deformability of the DNA and steric complementarity at the molecular interface.

INTRODUCTION

The formation of a stable protein/DNA complex depends on multiple factors (Drapper, 1993; Matthews, 1988; Seeman et al., 1976; von Hippel, 1994; von Hippel and Berg, 1986). The analysis of protein/DNA complexes has revealed some of the structural determinants for the recognition of specific DNA sequences, reflected in the surface complementarity between the protein and the DNA. This complementarity makes possible the formation of van der Waals contacts, hydrogen bonds, salt bridges, and water-mediated contacts (Otwinowski et al., 1988). Several examples exist where the analysis of structural information has produced an understanding of the interactions that form the "direct readout" (Desjarlais and Berg, 1992, 1993; Suzuki, 1993, 1994; Zilliacus et al., 1995). This detailed understanding has made possible the design of DNA-binding proteins with specificity for a desired target sequence (Huang et al., 1994; Kim and Berg, 1996; Park et al., 1993) and the development of an "effective potential" for the interaction of specific zinc-finger protein side chains with particular bases (Lustig and Jernigan, 1995). In addition to the direct readout between the protein and the DNA, basepairs inside and outside of the protein-DNA interface can also confer specificity by allowing an induced fit to the recognition surface. For example, in the 434 repressor/DNA complex (Aggarwal et al., 1988), the DNA basepair step between the two sites recognized by the helix-turn-helix motifs must overtwist, so that the major grooves align for successive readout by α-helices. In this case, the ease of deformation is a major determinant for the specificity of interaction, and is referred to here as a "dynamic determinant" for recognition. The quest for similar insights for the interaction of the TATA box-binding protein (TBP) with DNA is motivated by the key importance of this complex for transcription in eukaryotes (Burley and Roeder, 1996). Thus TBP is a basal transcription factor that is absolutely required for transcription by the three nuclear RNA polymerases (Cormack and Struhl, 1992). In the case of RNA polymerase II transcription, TBP is responsible for promoter recognition and binds directly to DNA in the minor groove. The structural characteristics of the complexes between TBPs and various TATA sequences elucidated recently from x-ray crystallography (Geiger et al., 1996; Juo et al., 1996; Kim and Burley, 1994; Kim et al., 1993; Nikolov et al., 1995, 1996; Tan et al., 1996) suggest that both direct readout mechanisms and dynamic determinants may be functional in the formation of these constructs (Pastor and Weinstein, 1995).

TBP binds to DNA in the minor groove of an AT-rich sequence. Although AT and TA steps are very similar stereochemically in the minor groove, thus precluding sequence discrimination based entirely on "direct readout" (White et al., 1996), TBP has a clear preference for a subset of all possible AT-rich sequences (consensus TATA T/A A T/A X) (Bernues et al., 1996; Chen and Struhl, 1988; Wobbe and Struhl, 1990; Wong and Bateman, 1994). TBP binding causes unwinding, bending, and compression of the major groove without disrupting the hydrogen bonds between basepairs (Guzikevich-Guerstein and Shakked, 1996). In the crystal structures of the complexes, the DNA conformation returns to B-DNA abruptly, immediately outside the 8 bp of contact with TBP. The structural characteristics of these complexes seem to make them an extreme

Received for publication 27 February 1997 and in final form 30 April 1997.

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 10029-6574. Tel.: 212-241-7018; Fax: 212-860-3369; E-mail: hweinstein@msvax.mssm.edu.

0006-3495/97/08/640/13 $2.00

example of induced fit. The large difference in conformation of the bound DNA compared to B-DNA suggests that the energy penalty for the conformational change might be a selectivity determinant (Juo et al., 1996; Kim and Burley, 1994). DNA sequences favoring complexation with TBP might have average geometries biased toward the properties of the final complex. Alternatively, but not exclusively, such favorable sequences might have weaker stacking energies, making them more amenable to the type of distortions found in the complexes.

The idea that the sequence-dependent local conformation and dynamics of DNA contribute to its recognition by ligands is not new (Calladine, 1982; Calladine and Drew, 1986; Hagerman, 1990, 1992; Harrington and Winicov, 1994; Travers, 1991, 1992; Zhurkin et al., 1979). Consequently, there have been various attempts to rationalize preferential directions for DNA bending and flexibility, based on the assumption that the structure and dynamics of a DNA molecule can be understood from the properties of its constituent basepair steps. These studies (Goodsell and Dickerson, 1994; Gorin et al., 1995; Suzuki and Yagi, 1995; Suzuki et al., 1996; Ulyanov and James, 1995; Yanagi et al., 1991; Young et al., 1995) rest on the analysis of structures obtained from x-ray crystallography and NMR that are available in the Nucleic Acid Database (NDB) (Berman et al., 1992) or Protein Data Bank (PDB) (Bernstein et al., 1977), and on inferences from a variety of solution techniques, such as anomalous migration in gel electrophoresis (Crothers and Drak, 1992; Haran et al., 1994), differential reactivity to agents that break DNA (Price and Tullius, 1993), and cyclization efficiency assays (Kahn et al., 1994; Lyubchenko et al., 1993). Computational studies of sequence-dependent DNA properties include Monte Carlo simulations at various levels of detail (Olson et al., 1995; Sarai et al., 1988, 1989; Srinivasan et al., 1987; Zhurkin, 1985; Zhurkin et al., 1979, 1991), adiabatic mapping of the properties of basepair steps and DNA oligomers (Hunter, 1993; Poncin et al., 1992a,b; Sanghani et al., 1996; Zakrzewska, 1992), and mechanical models of DNA (Calladine, 1982; Calladine and Drew, 1986; Calladine et al., 1988; Goodsell and Dickerson, 1994). Detailed analyses of the structures of DNA bound to TBP have been carried out by Juo et al. (1996) and Suzuki et al. (1996), but no such studies have specifically addressed the question of TBP binding to the TATA sequence in a dynamic framework.

Here we report on the exploration of structural characteristics that TBP exploits in its DNA substrates, and aim to discern whether there are DNA sequences that are dynamically predisposed to attaining the particular geometric attributes related to TBP binding. In the absence of high-resolution structures of DNA oligomers containing TATA box sequences in solution, molecular dynamics simulations were carried out for seven DNA dodecamers whose sequences (Table 1) include three functional TATA boxes (**mlp**, **6t**, and **at**) (Wong and Bateman, 1994), two sequences recognized only by mutant TBPs (**2c** and **7g**) (Arndt et al., 1992, 1994), a reversed TATA box (**r28**), and a negative control (**gc**). The simulations were done with the CHARMM23 potential (MacKerell et al., 1995), with explicit water molecules (TIP3P) and sodium ions, using periodic boundary conditions and a spherical cutoff for the nonbonded interactions. To assess the convergence of the results and the dependence of the inferences on the force field used for the MD simulations, the **mlp** sequence was also simulated with the AMBER 4.1 potential (Cornell et al., 1995), using Ewald sums (Darden et al., 1993) for the electrostatic interactions, and with the CHARMM23 potential, using a different integrator for the equations of motion and a different set of initial velocities. The results of the methodological comparisons show that the simulations are robust, and that they adequately describe the behavior of general-sequence DNA, regardless of the force field used. Consequently, the specific findings that emerged from the present analysis relating conformational preferences of the DNA sequences to TBP binding can be considered to represent generalizable properties of DNA components that can combine dynamically in various ways to support different selectivity determinants.

TABLE 1 Sequences studied

TATA sequences crystallized in complexes with TBP		
Promoter	Sequence	NDB accession numbers
mlp	TATAAAAG	PDT009, PDT032, PDT034
CYC-1	TATATAAA	PDT012
CYC-1	TATAAAAC	PDT036
E4	TATATATA	PDT024
Consensus	TATA@A@X	
Simulated systems		
Name	Sequence	
mlp	C TATAAAAG	GGC
2c	C CATAAAAG	GGC
6t	C TATATAAG	GGC
7g	C TATAAGAG	GGC
r28	C TTTTATAG	GGC
at	A TATATATA	TAT
gc	G CGCGCGCG	CGC
bp step	1234567	
bp	1 23456789	10 11 12

@: A or T; sequences are shown for the coding strand only. bp: basepair

MATERIALS AND METHODS

Materials

Atomic coordinates of TBP/DNA complexes

The atomic coordinates of two copies of *A. thaliana* TBP2 (*ath*) (Kim and Burley, 1994), one copy of *H. sapiens* TBP (*hsa*) (Nikolov et al., 1996), and one copy of a ternary complex between *H. sapiens* TFIIB and *ath* bound to the adenovirus 2 major late promoter (Nikolov et al., 1995) were kindly provided by Drs. S. K. Burley and D. B. Nikolov. The atomic coordinates for two copies of *S. cerevisiae* TBP (*sce*) (Kim et al., 1993) and one copy of the ternary complex between *S. cerevisiae* TFIIA and *sce* bound to the -52 CYC1 promoter (Tan et al., 1996), and one copy of *hsa*

bound to the E4 promoter (Juo et al., 1996) were obtained from the NDB (see Table 1 for NDB accession numbers).

NMR-derived DNA oligomer structures

The atomic coordinates for the DNA oligomers whose basepair step geometry is summarized in Table 2 (see line labeled NMR) were obtained from the Brookhaven PDB, with accession numbers 142D, 1D18, 1D19, 1D20, 1D42, and 1D70. These structures were selected by Ulyanov and James (1995) as representative structures of DNA oligomers in solution.

High-resolution A- and B-DNA oligomer structures

The atomic coordinates for the DNA oligomers whose basepair step geometry is summarized in Table 2 (lines labeled NDB-A and NDB-B) were obtained from the NDB (accession numbers ADH008, ADH010, ADH026, ADH038, ADH047, ADH070, ADJ049, ADJ050, ADJ067, ADL025, BDJ017, BDJ019, BDJ025, BDJ031, BDJ036, BDJ051, BDJ052, BDJ060, and BDL020). All have a resolution better than 2 Å, and lack mismatches, nucleotide modifications, or ligands.

Methods

The molecular systems

The seven DNA dodecamers were built with B-DNA conformation (see the entry for fiber B-DNA in Table 2) using QUANTA (Molecular Simulations, 1992); the 5′ phosphate groups at the end of the strands were removed. The Na^+ counterions were placed at a distance of 5 Å from the P atom along the O-P-O bisector; one sodium ion was added per phosphate, for a total of 22 Na^+. The DNA and the sodium ions were solvated in InsightII (Biosym Technologies, 1993). The final simulation system, including the DNA dodecamer, 22 Na^+ ions, and >3400 TIP3 water molecules, was enclosed in a hexagonal prism of 72 Å length with a 24 Å side.

For the replica of **mlp** run in AMBER 4.1 (Pearlman et al., 1995), the dodecamer was built in standard B-DNA conformation, sodium ions positioned at a distance of 5 Å from the P atom along the O-P-O bisector, and solvated in AMBER 4.1 with an 11 Å shell (>4000 TIP3P water molecules). The final simulation system was enclosed in a square prism, the dimensions of which were adjusted by running at constant pressure and temperature to ensure the appropriate density (63.1 Å × 45.7 Å × 44.7 Å).

The simulation protocol

The molecular dynamics simulations were run with the CHARMM (Brooks et al., 1983) program in the NVE ensemble, using the CHARMM23 (MacKerell et al., 1995) all-atom potential, the Verlet integrator, and periodic boundary conditions. SHAKE was applied to all hydrogen-containing bonds. A cutoff value of 13 Å was used, with the shift and switch functions for the electrostatic and van der Waals interactions, respectively.

Keeping the DNA and sodium ions fixed, the water was equilibrated for 36 ps. Subsequently, the whole system was energy minimized, and then it was heated from 0K to 300K in 10 ps. Equilibration was carried out for 30 ps with a time step of 2 fs. For the production run of 510 ps (2080 ps for **mlp**), the time step was reduced to 1.5 fs. An independent **mlp** run was started in parallel using a different seed ($\sqrt{3}$, as opposed to the default $\sqrt{2}$) for assigning velocities for the heating of the whole system. Heating (10.5 ps), equilibration (31.5 ps), and production (510 ps) were done with a time step of 1.5 fs, using the Leapfrog Verlet integrator. Structures from the trajectories were saved every 0.075 ps.

The molecular dynamics simulation of **mlp** run in AMBER 4.1 was carried out in the NVT ensemble, using the AMBER 4.1 (Cornell et al., 1995) all-atom potential, the Verlet integrator, and periodic boundary conditions. SHAKE was applied to all hydrogen-containing bonds. A cutoff value of 9 Å was used for the van der Waals interactions, and the electrostatic interactions were treated with the particle mesh Ewald algorithm (Darden et al., 1993). In this run, water was heated for 15 ps and equilibrated for 85 ps; then the system was energy minimized, heated from 0K to 300K in 15 ps, and equilibrated for 50 ps. The production run extended to a total of 1 ns; all of the phases were carried out with a time step of 2 fs. Structures from this trajectory were saved every 0.1 ps.

Validation of the simulations

Conformational stability and independence of force field used. A major concern in simulations of oligoelectrolytes is the manner in which the electrostatic interactions are represented. Spherical truncation is known to

TABLE 2 Basepair step parameters for general-sequence DNA

Source	Shift	Slide	Rise	Tilt	Roll	Twist
CHARMM	0.1 ± 0.8	−1.5 ± 0.8	3.3 ± 0.5	0.8 ± 6.6	2.5 ± 12.2	32.1 ± 5.1
mlp[a]	0.0 ± 0.7	−1.7 ± 0.7	3.3 ± 0.4	0.4 ± 6.3	2.1 ± 10.9	32.0 ± 4.6
mlp[b]	0.0 ± 0.7	−1.4 ± 0.8	3.3 ± 0.4	0.6 ± 6.3	2.2 ± 11.6	32.3 ± 4.8
AMBER	0.1 ± 0.7	−1.6 ± 0.6	3.4 ± 0.4	0.8 ± 6.4	1.9 ± 8.1	30.5 ± 4.4
NMR	0.2 ± 0.4	−0.9 ± 0.4	3.1 ± 0.2	2.8 ± 8.3	3.9 ± 6.9	33.1 ± 2.3
NDB-A	0.1 ± 0.6	−1.9 ± 0.4	3.4 ± 0.4	0.7 ± 5.4	6.9 ± 6.2	30.4 ± 4.0
NDB-B	0.1 ± 0.5	0.2 ± 1.0	3.3 ± 0.2	1.1 ± 5.7	0.4 ± 6.4	36.7 ± 6.9
Fiber A	0.0	−2.0	3.2	0.0	10.5	30.7
Fiber B	0.0	−0.6	3.3	0.0	−2.6	35.9

Mean ± standard deviation.
CHARMM: all simulations done with the CHARMM23 potential, including mlp[a] and mlp[b].
mlp[a]: 2.0-ns simulation with CHARMM23, Verlet algorithm for integration, $\sqrt{2}$ used as seed for initial distribution of velocities.
mlp[b]: 0.5-ns simulation with CHARMM24, Leapfrog algorithm for integration, $\sqrt{3}$ used as seed for initial distribution of velocities.
AMBER: Simulation done with the AMBER 4.1 potential.
NMR: Structures determined by NMR.
NDB-A: Structures determined by x-ray crystallography and classified as A-DNA by the NDB.
NDB-B: Structures determined by x-ray crystallography and classified as B-DNA by the NDB.
Fibers A and B: A- and B-DNA conformations as defined by QUANTA.

produce an artefactual accumulation of structures where ionic species of the same charge are separated by a distance just outside of the cutoff distance. Fig. 1 shows the radial distribution functions (rdfs) for the three pairs of ionic species present in the simulations: Na^+-Na^+, P-Na^+, and P-P, calculated for **mlp** run with CHARMM23 (Verlet integrator, spherical truncation with a shifting function that makes the interaction zero beyond 12 Å) and AMBER 4.1 (particle mesh Ewald algorithm). None of the rdfs displays the pathological accumulation of pairs at the cutoff distance, indicated by the vertical line in the plots (Fig. 1), suggesting that the spherical cutoff with the shifting function at 12 Å is an acceptable approximation. Overall, the rdf from the CHARMM23 run shows structure at greater radii than the AMBER 4.1 run, and the distances between the peaks correspond roughly to a water diameter. For distances up to ~8 Å, there is good agreement in the position of the peaks between the two simulations, except for the first peak in the P-Na^+ rdf: the AMBER 4.1 potential appears to drive the Na^+ ions closer to the phosphates, to a contact distance. The P-P rdf indicates that AMBER 4.1 is more effective at preserving the B-DNA distances than CHARMM23.

Structural stability of the DNA molecules was monitored with a two-dimensional root mean square difference (rmsd) plot, in which the coordinates of the heavy atoms of each saved conformation were superimposed to get the best fit to all of the other conformations in each run. The plot for **mlp** is shown in Fig. 2; the left triangle contains the data from the comparison of all heavy atoms, and the right triangle has the data for the heavy atoms of the 8 bp of the TATA box (TATAAAAG). Three areas of low rmsd are obvious at 140–500 ps, 550-1630 ps (*square marked with a black line*), and 1630–2080 ps. There is an instance of recurrence of the structures sampled, ~1.2 ns apart (see the area comparing the structures at 240 ps and at 1240 ps). The corresponding plots of the other simulations are very similar to the area between 40 and 550 ps in Fig. 2. The recurrence of structures in the mlp simulation indicates that the runs carried out for 550 ps are representative of longer simulations.

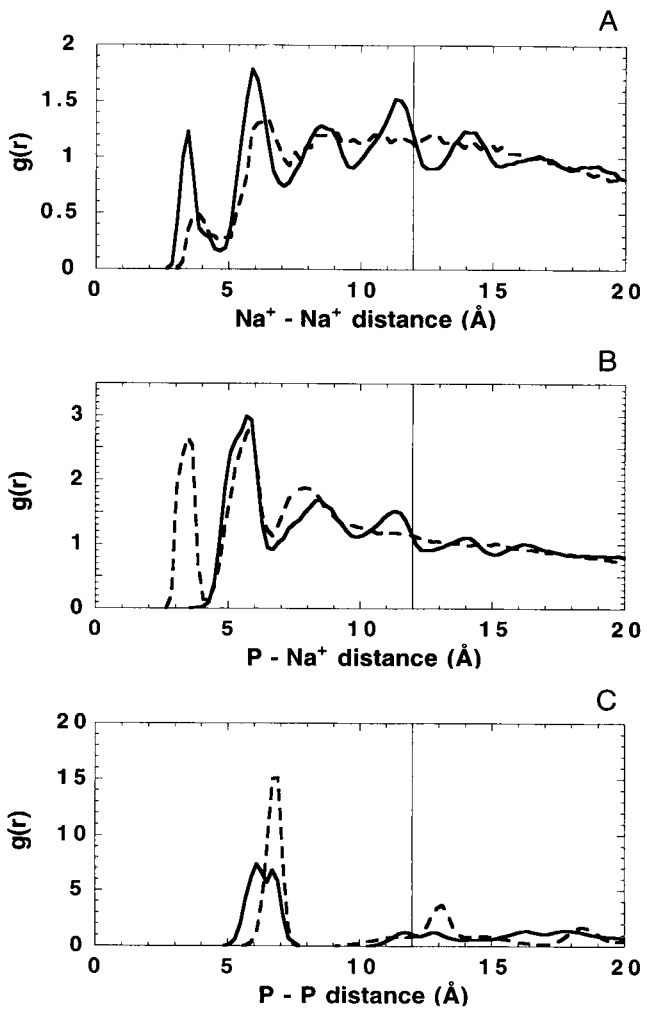


FIGURE 1 Comparison of the radial distribution functions [$g(r)$] for (*A*) Na^+-Na^+, (*B*) P-Na^+, and (*C*) P-P atoms for **mlp** run with the CHARMM23 potential (——) and with the AMBER 4.1 potential (– – –). The vertical line at 12 Å indicates the radius at which all electrostatic interactions are set to zero in CHARMM23.

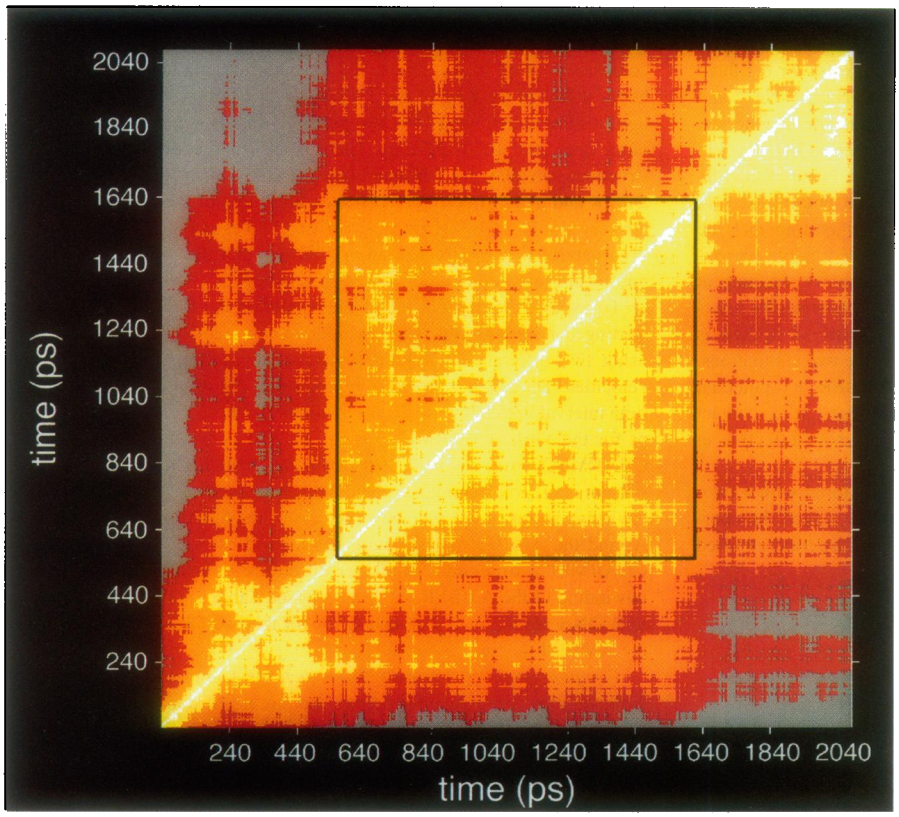
To explore the dependence of the results on the force field and on the starting conditions, **mlp** was run three times, with two force fields and using two different integration schemes and seeds for assigning initial velocities. The average structure representing the structures in the longest time interval with a low rmsd was obtained for these three runs (550–1630 ps for the Verlet CHARMM23 run, 220–550 ps for the Leapfrog Verlet CHARMM23 run, and 300–950 ps for the AMBER 4.1 run). The three structures are very similar in their global appearance. The rmsd between the heavy atoms of the average structures of the CHARMM23 runs is 0.74 Å, and the AMBER 4.1 structure has a rmsd of ~2.7 Å to each of the CHARMM23 structures. This latter difference is mainly due to a greater unwinding of the AMBER 4.1 structure (see Table 2). The rms differences to B-DNA are 4.2 Å, 3.9 Å, and 4.3 Å for the two CHARMM23 runs and the AMBER 4.1 run, respectively.

Comparison to crystal and NMR structures. Geometrical parameters of basepair steps were calculated from the simulated dodecamers, for comparison to results from oligomers whose structure was determined by NMR, high-resolution DNA oligomer structures deposited in the NDB, and fiber diffraction models. The results presented in Table 2 show that all of the simulations yield very similar values for the geometries of general-sequence DNA. The similarity persists for the various potentials used in the method of simulation, the different protocols for handling the electrostatic interactions, and the variations in the integrator for the equations of motion and the initial distribution of velocities. Nevertheless, the run with AMBER 4.1 consistently yields smaller standard deviations than the CHARMM23 runs, suggesting a more rigid DNA structure, but it remains uncertain whether this is due to the potential or to the formulation of the electrostatic interactions. Notably, the average geometry of the simulated DNA oligomers is more similar to that derived for structures from NMR studies and to A-DNA crystals (see, for example, the negative slide and low twist) than to B-DNA crystal structures. As a measure of the general reliability of the simulations, these results indicate that the simulations are appropriately reproducing the structure of general-sequence DNA in solution. Moreover, the simulations also reproduce adequately known features of the dynamics of general-sequence DNA, such as the anisotropy of motion expressed in the observation that the fluctuations in roll are larger than those of tilt or twist (Olson et al., 1995; Zhurkin et al., 1979).

Conformational analysis and statistics

Conformational analysis was carried out with the CURVES algorithm (Lavery and Sklenar, 1988, 1989), implemented in Dials and Windows of the MD Toolchest (Ravishanker et al., 1989). Because this algorithm performs a global fit to the DNA axis, the reported angles and displacements depend on the DNA length. To allow for a comparison of the local basepair step geometry between the simulated DNA dodecamers and the NMR and crystal structures of DNA oligomers, which have different lengths, all of the DNA oligomers analyzed in this work were disassembled into their constitutive basepair steps. Data were collected for all basepair steps, except those at the ends of the oligomers. Non-self-complementary steps were considered in both orientations. The values for A- and B-DNA (Table 2) were obtained from single basepair steps built in the corresponding canonical fiber conformations.

The DNA axes shown in Fig. 3 were calculated with CURVES from the average structures representing the longest time interval in the simulations with a <1.8-Å rms difference between the instantaneous structures in that



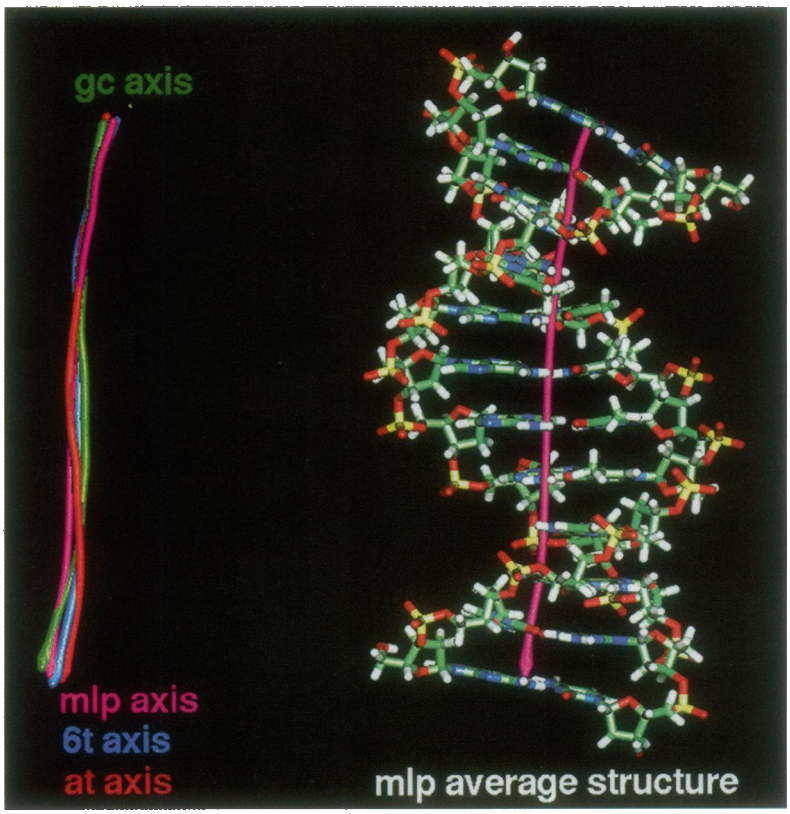


FIGURE 2 Two-dimensional rmsd plot for the 2-ns mlp run with the CHARMM23 potential. *Upper triangle*: rmsd for all heavy atoms. *Lower triangle*: rmsd for the heavy atoms of the TATA box (TATAAAAG). *White*: rmsd [0–0.7 Å]; *yellow*: rmsd [0.7 Å-1.2 Å]; *orange*: rmsd [1.2 Å-1.8 Å]; *red*: rmsd [1.8 Å-2.5 Å]; *gray*: rmsd > 2.5 Å.

FIGURE 3 (*Left*) Helix axes calculated with Curves (Lavery and Sklenar, 1988, 1989) for the average structures of **mlp** (*magenta*), **6t** (*blue*), **at** (*red*), and **gc** (*green*). (*Right*) The all-atom depiction of the average structure of **mlp** from 550 to 1630 ps of the 2-ns run. The helix axis is shown in magenta.

time interval: **mlp** 550-1630 ps (*boxed area* in Fig. 2), **2c** 265–550 ps, **6t** 280–550 ps, **7g** 210–550 ps, **r28** 250–550 ps, **at** 120–490 ps, **gc** 220–550 ps. All of the structures were aligned by superimposing the coordinates of all DNA atoms on the starting structure of each simulation.

Statistical analyses were made with SAS (SAS Institute, Cary, NC). Averages and standard deviations were calculated from all of the geometries in the production phase of the simulations. The same data set was used to calculate the frequency of occurrence of basepair step geometries within the 99% confidence interval defined by the DNA oligomers found in the TBP/DNA crystal structures. These frequencies were further scored by a χ^2 test to determine the basepair steps with the highest and the lowest frequency of occurrence.

RESULTS AND DISCUSSION

Conformational variability in the TBP/TATA box complexes

To determine the structural requirements for successful complex formation, we used the existing structures of TBP/TATA box complexes to characterize the geometrical parameters for the seven basepair steps in contact with TBP. Four different sequences have been crystallized with TBP (see Table 1).

To separate direct readout effects from the conformational variability, we first searched the protein-DNA interface for differences in these four sequences. These differences were sought in the eight crystal structures available for the four different DNA sequences complexed with TBP; they include two ternary complexes, one with TFIIB (Nikolov et al., 1995), and another with TFIIA (Tan et al., 1996). Examination of these high-resolution structures reveals that the interface between the protein and the DNA is almost invariant. The two kinks (steps 1 and 7 in Tables 1 and 3) are caused by the partial insertion of Phe residues, and the only place where TBP makes hydrogen bonds to the bases is at the center of the TATA box (step 4 in Tables 1 and 3), ignoring most hydrogen bond acceptors in the bottom of the minor groove. The protein/DNA interface is mostly hydrophobic, with Val and Leu residues making contact with the minor groove edges of the bases, and probably being responsible for the selectivity against GC basepairs (Juo et al., 1996; Kim and Burley, 1994; Kim et al., 1993; Nikolov et al., 1996). Inspection of the structures shows that the only apparent difference among the TBP/DNA complexes concerns the rotamer of the Thr residue (124 in *sce*, 82 in *ath*, and 218 in *hsa*) that interacts with basepair 6. The difference is significant for recognition, because it pertains to the discrimination of AA versus AT steps at step 4. If step 4 is AA, the Thr residue does not hydrogen bond to the T in the complementary strand; if this step is AT, the Thr residue orients the OH group toward the N3 of the A in the complementary strand. This suggests, first, that there is a preference of interaction between the OH of the Thr and the N3 of an A versus the interaction with the O_2 of a T; and second, that despite the closely packed interface between TBP and the DNA, there is some flexibility in the direct readout, similar to the situation found in the complexes of the estrogen receptor with different DNA sequences (Schwabe et al., 1995). Furthermore, the loss of a hydrogen bond between TBP and the DNA should be destabilizing, and this is borne out by the finding that the best TBP binding sequence has an AT step at this position (step 4) (Wong and Bateman, 1994). It should be noted that this destabilization is modulatory, and not a determinant of complex formation, which points again to the sequence-dependent deformability of the DNA as a major component of selectivity.

To characterize the sequence-dependent deformability, the conformational variability of the seven basepair steps from the TBP/TATA box complexes was evaluated from the average geometrical parameters presented in Table 3. By definition, these parameters must be consistent with the formation of a complex between TBP and a DNA fragment. Overall, the parameters are consistent with unwound DNA, most severely at step 4, which corresponds to the dyad of the TATA box, and with a direction of roll that causes the minor groove to open up. As noted previously, steps 1 and 7 constitute the sites of insertion of Phe residues and are characterized by high rise and roll.

To identify the special characteristics of the geometrical parameters in DNA sequences involved in known complexes with TBP, the geometrical parameters in Table 3 were compared to those of free, general-sequence DNA (see the *charmm* entry in Table 2). Following Olson and coworkers (Olson et al., 1995), the thermally accessible range of conformations for general-sequence DNA is defined as

TABLE 3 Conformational ranges consistent with the formation of a TBP/DNA complex: basepair step parameters for the TATA boxes complexed with TBP

Step	Shift	Slide	Rise	Tilt	Roll	Twist
1	0.2 ± 0.5	−1.9 ± 0.4	4.9 ± 0.4	−0.6 ± 3.3	40.9 ± 4.8	19.8 ± 2.5
2	−1.3 ± 0.5	−1.5 ± 0.1	3.4 ± 0.2	1.7 ± 2.0	16.5 ± 2.8	16.4 ± 1.3
3	0.2 ± 0.3	1.3 ± 0.1	3.3 ± 0.2	3.4 ± 1.8	7.6 ± 2.9	25.5 ± 1.7
4	0.3 ± 0.4	1.2 ± 0.3	3.4 ± 0.3	−1.8 ± 1.1	26.2 ± 3.2	8.1 ± 6.3
5	−0.4 ± 0.3	2.0 ± 0.3	3.5 ± 0.4	1.0 ± 2.5	25.9 ± 4.3	22.4 ± 3.1
6	0.4 ± 0.3	1.2 ± 0.5	3.3 ± 0.3	0.8 ± 4.5	24.2 ± 4.1	22.7 ± 3.7
7	−0.1 ± 0.7	0.8 ± 1.0	5.4 ± 0.6	1.7 ± 4.4	44.6 ± 6.3	22.5 ± 6.1
All	−0.1 ± 0.7	0.4 ± 1.5	3.9 ± 0.9	0.9 ± 2.9	26.6 ± 12.5	19.6 ± 6.3

Mean ± 99% confidence interval; the steps are defined in Table 1. All: mean ± standard deviation over the eight structures and the seven steps.

the interval contained between the mean ± 1 standard deviation. The comparison yields a range of geometrical parameters representing conformations that deviate significantly from those of general-sequence DNA. Within this definition, the DNA complexed with TBP shows the largest distortions in the parameters slide, rise, roll, and twist. The appearance of such special characteristics in the simulated structures, analyzed below, can serve as an indication of their predisposition toward adopting geometries supporting the binding of TBP.

Global structural analysis of the simulated structures

The global structural properties of the seven simulated dodecamers were probed first for differences that could be relevant to their binding to TBP. The average geometrical parameters of these structures appear indistinguishable, in that all of the dodecamers are slightly unwound (**2c** is the most unwound, with an average twist of 31.3°), and all have small positive roll (**gc** has the largest average roll at 6.9°) and negative slide. The average structures of the dodecamers, calculated from the longest equilibrated time interval of the simulations, also show almost indistinguishable features, as exemplified by the helical axes shown in Fig. 3. It is impossible to distinguish **gc**, a negative control that does not bind TBP, from **mlp**, **at**, or **6t**, which are among the best TBP binding sequences (Wong and Bateman, 1994). The observed curvature corresponds to a slight compression of the major groove, shown for **mlp** in Fig. 3, as expected for DNA in solution (Calladine et al., 1988; Ulyanov and James, 1995). These results suggest either that DNA shows no global sequence-dependent features at the dodecamer level, or that the simulations are incapable of revealing such a sequence-dependent behavior at the level of global structural properties, so that the differences must be sought from a more local analysis.

Local structural analysis and the properties of a TATA sequence

To characterize the behavior of individual basepair steps from the simulated dodecamers, the geometrical parameters were calculated separately along the dynamics trajectory for each basepair step, in the absence of the rest of the dodecamer. The data for all basepair steps with identical flanking sequences (i.e., identical tetrads) were pooled together, yielding a total of over 670,000 data points for averaging. The pooling of data for identical tetrads in the various dodecamers has the advantage of considerably improving the sampling for each basepair step. The average values for the six geometrical parameters for all of the unique tetrads found in the simulated sequences are gathered in Table 4. Highlighted in boldface and underlined are those values lying outside the thermally accessible range (defined in Methods) for general-sequence DNA (see Table 2). The geometric properties thus identified are akin to the "wedges" of Calladine and Drew (1986), representing persistent deviations from canonical, straight DNA. If the results of the analysis show them to be in a direction consistent with TBP binding, it is possible that these intrinsic propensities of the structure revealed in the dynamics could serve as preformed incipient TBP binding sites.

The results in Table 4 and the distributions shown in Fig. 4 reveal a sequence-dependent behavior in DNA. RR steps are essentially straight, as indicated by their low roll and tilt, and display an intermediate geometry between purine-pyrimidine (RY) and pyrimidine-purine (YR) steps in all parameters except slide, in which they are extreme. YR steps have a high twist and high rise profile, whereas RY steps display a low twist, low rise and high positive roll profile. Some of the average values do not agree with the data obtained from crystal analysis of B-DNA (Yanagi et al., 1991), especially the persistent negative slide, but this is not surprising, given that the simulations were found here to better reproduce the structures of DNA determined in solution.

The sequence-dependent conformational properties of basepairs are evident from the average values in Table 4 and from the distributions shown in Fig. 4. They suggest a mechanism for the sequence preferences displayed by TBP. Thus YR steps are seen to have high average values of rise, which is needed at steps 1 and 7 of the TBP/DNA complex. Notably, CG and TA have the highest rise values, and they are found at these positions in the complexes; AG also has a high rise value, accounting for its presence at the last step of some TATA boxes. The RY steps are clearly distributed toward the high positive values of roll (Fig. 4) required throughout the TBP recognition site, and they have the requisite low twist values (Table 4).

Taken together, these sequence-dependent characteristics suggest that the special alternating construction of the TATA site reflects the fact that the only way to maximize the number of RY steps and to have YR steps at the kink sites is to have an alternating RY sequence. However, TBP does not bind alternating GC sequences, suggesting that the additional selection between AT and GC basepairs for the consensus TATA sequences does not occur at the level of basepair step properties. Rather, it is likely that steric interactions, such as the putative clash of Val and Leu side chains of TBP with the exocyclic amino group of G bases (Juo et al., 1996; Kim and Burley, 1994; Kim et al., 1993), determine the ultimate preference for AT over GC basepairs.

Basepair step properties essential for TBP recognition

Because not all of the basepair steps in the complexed TATA box have the same geometric requirements (Table 3), additional specificity determinants must operate locally. For example, only steps 1 and 2 have negative slide, whereas only steps 1 and 7 have high rise, etc. Such discriminating

TABLE 4 Average sequence-dependent basepair step parameters

Tetrad	Shift	Slide	Rise	Tilt	Roll	Twist
aAAa	0.0	−1.5	3.2	0.7	4.9	31.6
aAAg	0.0	−1.4	3.2	−1.6	−1.8	33.0
tAAa	0.2	−1.5	3.3	1.7	5.0	31.4
tAAg	0.1	−1.3	2.9	−3.2	10.0	28.4
aAGa	−0.4	−1.1	3.7	0.7	−8.8	34.2
aAGg	0.1	−1.8	3.3	1.2	2.6	31.6
gAGg	−0.4	−0.8	3.3	**−6.2**	6.2	34.0
tAGg	−0.5	−0.8	3.4	−5.1	1.8	34.7
aGAg	0.0	−1.2	3.0	0.6	7.5	30.1
aGGg	0.2	−2.2	3.5	3.0	1.6	32.6
gGGc	0.3	−2.1	3.3	4.2	−5.8	32.4
xRRx	0.1	−1.7	3.3	1.2	1.2	32.1
cCAt	**1.1**	−1.2	3.6	4.3	−1.7	32.7
gCGc	0.3	−0.5	**3.9**	1.8	−4.8	35.1
aTAa	0.1	−1.0	3.8	3.4	4.6	32.9
aTAt	0.8	−0.8	3.5	5.5	−0.7	34.4
cTAt	0.4	−1.0	3.6	2.5	2.6	33.5
xYRx	0.3	−0.9	3.7	3.2	1.0	33.7
cATa	−0.4	−0.9	**2.7**	−3.1	**18.4**	**24.4**
tATa	0.1	−1.0	3.0	2.0	9.3	30.9
cGCg	0.0	−0.7	**2.7**	0.1	**21.5**	27.4
xRYx	0.0	−1.0	2.9	1.3	12.5	29.7

x: any base; R: purine; Y: pyrimidine. The basepair step for which the geometry was calculated is indicated in uppercase in each tetrad. Underlined and boldface entries have average values outside the thermal range defined from the CHARMM entry in Table 2. xRRx: average values for RR steps; xYRx: average values for YR steps; xRYx: average values for RY steps.

parameters were sought by exploring the relation between the specific step requirements and the properties of individual basepair steps. To this end, we searched the TBP/TATA box complexes for those basepair step parameters that are significantly different from those of the average DNA conformation (Table 2). A 99% confidence interval was obtained for each of the geometric parameters and for each of the steps, and these intervals were compared with the thermally accessible intervals for general-sequence DNA (Table 2). The results show that tilt is never a discriminating parameter, as the interval of its values in the complexed TATA boxes is always contained in the thermal fluctuations of free DNA. The list of properties that exhibit intervals completely outside the thermal fluctuations of the general-sequence DNA, and hence represent discriminating parameters, is presented in Table 5.

The importance of slide, roll, and twist is evident from this analysis. Because extreme values of these properties cause large excursions from the mean structure, they should occur infrequently during the simulations. However, they are likely to be key to selectivity, because such sequence-dependent properties relate to the actual deformation of the DNA fragment bound by TBP. As pointed out by Suzuki et al. (1996), the large shift present in step 2 is responsible for directing the curvature, and the slide at steps 3 and 5 is necessary to align the groove for recognition by the side chains of TBP.

Identification of likely discriminant parameters makes it possible to search for the probability of occurrence of an extreme geometry that favors TBP binding in the simulation trajectories. The 99% confidence intervals defined by the crystal structures of TBP/DNA complexes for the discriminant properties selected above were used as a filter in evaluating all of the conformations generated in the simulations. The search recorded the number of times that a particular geometrical parameter, for each tetrad, falls in the range corresponding to the DNA in the complexes with TBP. The frequency with which each tetrad visits the TBP/TATA box complex conformation was rated with a χ^2 test. These results are also summarized in Table 5, where the sequence identified as "best tetrad" is the one that appears with the highest frequency, and the "worst tetrad" is the one appearing with the lowest frequency, in the range of values corresponding to the TBP/DNA complex. There are no instances of a particular tetrad simultaneously satisfying all of the relevant geometry requirements for a given step in the complex. For example, the requirements for step 1 are high rise, low twist, and high roll, which are three characteristics that are not shared by any one of the tetrads. This finding shows that the simulation of any DNA sequence (in the

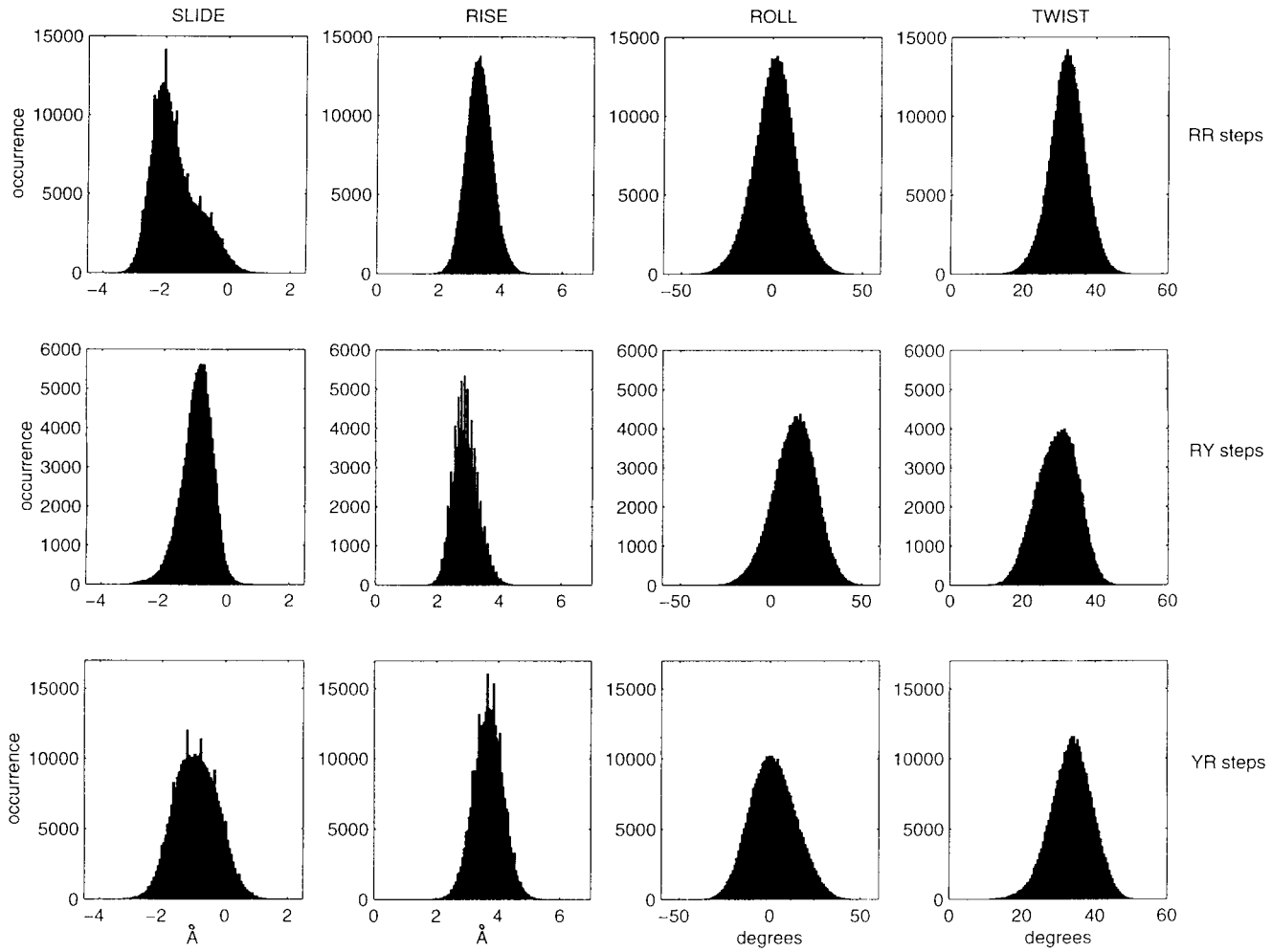


FIGURE 4 The distributions of basepair step parameters obtained from the populations of RR (*top*), RY (*middle*), and YR (*bottom*) basepair steps in the simulations of the sequences shown in Table 1. Occurrences are shown as histograms for the populated ranges of values. Only the parameters found to be most relevant to TBP binding are included.

absence of TBP) never produces a conformation that corresponds entirely to the one found in complex with TBP. However, several tetrads clearly show the propensity to adopt the essential geometrical attributes of the complexed DNA (Table 5), and they acquire in the dynamics values of the geometrical parameters that lie outside the range for average DNA and are close to those in TBP/DNA complexes.

A striking observation from Table 5 is the absence of RR steps from the best tetrad column. This can be explained here by the finding that these steps are straight on average, and hence are not predisposed to bend. Notably, sequences that are not bound efficiently by TBP do appear as best tetrads in steps of the TATA box, suggesting that aside from the general rule found above for alternating YR sequences, TBP selectivity is not reflected solely in the average geometry of the different basepair steps. This is particularly clear for step 4 in the TATA sequence, for which our search failed to pick out any tetrad found in the available crystallized complexes. It is interesting, however, that this step is the only one in which hydrogen bonds are formed between TBP and the minor groove of the DNA, suggesting that these hydrogen bonds might pay for the energy cost of unwinding, opening the minor groove and sliding the basepairs to induce the complex geometry. A rough estimate of the free energy penalty for this structural transition in the basepair step can be obtained from a statistical analysis of the simulation results, from

$$n_i/n_{\text{ground}} = \exp(\Delta G/RT)$$

where n_i is the number of conformations appropriate for this interaction, n_{ground} is the number of conformations inside the thermally accessible range, and ΔG is the free energy difference between the conformation found in the complexes with TBP and the conformation for the most populated state. The calculation for the two tetrads found in crystals with TBP indicate that tATa always has a lower penalty than tAAa (4.5 versus 5.1 kcal/mol for slide; 1.3 versus 2.0 kcal/mol for roll; 3.6 versus 4.2 kcal/mol for twist). Furthermore, tATa can form six hydrogen bonds to TBP, whereas tAAa can form only five. Assuming additiv-

TABLE 5 Basepair step properties contributing to selectivity

Step	Parameter	Best tetrad: % occurrence	Worst tetrad: % occurrence
1	Rise	ataa : 6.65	tata : 0.03
	Roll	cgcg : 6.36	gggc : 0.00
	Twist	cata : 26.25	aggg : 0.95
2	Shift	atag : 30.33	atat : 2.85
	Twist	cata : 3.85	aaaa : 0.02
3	**SLIDE**	aTAt : 0.23	aggg, gggc : 0.00
	Twist	cgcg : 20.94	gcgc : 2.34
4	Slide	atat : 1.02	aggg : 0.00
	Roll	cgcg : 23.65	gggc : 0.42
	Twist	atat : 0.58	aaaa, aggg : 0.00
5	**SLIDE**	aTAt : 0.01	tata : 0.00
	Roll	cgcg : 31.63	gggc : 0.63
	Twist	cgcg : 33.10	gcgc : 1.35
6	Slide	atat : 1.55	aggg : 0.00
	Roll	cgcg : 32.29	gggc : 0.88
	Twist	cgcg : 38.71	gcgc : 2.16
7	Slide	gcgc : 29.96	aggg : 0.02
	Rise	ataa : 1.67	tata : 0.00
	Roll	cgcg : 4.06	gggc : 0.00

Steps are defined in Table 1. Best tetrad: tetrad that visits the conformational range defined by the crystal structures of TBP/DNA complexes (Table 3) with the highest frequency, scored by a χ^2 test. Worst tetrad: Tetrad that visits the conformational range defined by the crystal structures of TBP/DNA complexes (Table 3) with the lowest frequency, scored by a χ^2 test (double entries shared the lowest frequency and χ^2 score). The parameters in uppercase are those that selected as the best tetrad one that has actually been crystallized in a complex with TBP.

ity and a 1.5 kcal/mol contribution from each hydrogen bond (Jen-Jacobson, 1995), the formation of six hydrogen bonds would balance the nearly 9.5 kcal/mol energy penalty for the distortion required in the tATa steps.

The entries highlighted in uppercase boldface in Table 5 identify the basepair step properties that are most likely to be used as selectivity determinants. These are the steps that satisfy both the "best tetrad" criterion, and actually occur in a crystal with TBP. Inspection shows that these steps have the most positive slide and are responsible for aligning the groove for recognition (Suzuki et al., 1996). It is of interest that the simulations yielded an extremely low population of steps with positive slide. If the geometrical rearrangement produced by the positive slide at these positions is indeed a limiting step for binding, then the low probability of such a feature in the trajectory may indicate a reason for the very slow binding of TBP (Coleman and Pugh, 1995; Coleman et al., 1995; Hoopes et al., 1992; Lieberman et al., 1991; Parkhurst et al., 1996; Perez-Howard et al., 1995; Petri et al., 1995; Starr et al., 1995).

Global binding properties of DNA oligomers

Having identified both average and dynamic geometrical properties of DNA that are pertinent to TBP binding, we can build a consensus sequence to optimize the propensity for TBP binding. By these criteria, the optimal binding sequence should conform to YRTATAYR, where the YR is required because of the tendency for high rise (see Table 4), whereas the TATA in the center is required because TA steps have the positive slide required at steps 3 and 5 (from Table 5). The expectation that affinity for TBP should increase the closer a sequence is to the proposed consensus is verified by the results in Table 6, which relate the sequences of the simulated dodecamers to the equilibrium binding constants determined by Wong and Bateman (1994). Thus, **at** and **6t**, which are the closest matches to the consensus sequence, are also the strongest binders. The sequences of **mlp**, **2c**, and **7g** match the consensus for the first half of the TATA box only, and **mlp** binds 3–4 times worse than **at** or **6t**, in keeping with the prediction that a reduction in sequence matches to the consensus (Table 6) should result in loss of affinity. The equilibrium binding constants of the TBP mutants that bind to **2c** and **7g** (Arndt et al., 1992, 1994) have not been measured, but the prediction from the relations in Table 6 is that the binding constants will be in the same range as for **mlp**.

CONCLUDING REMARKS

The successful formation of a TBP/DNA complex depends on multiple factors, including the generation of the appropriate contact surfaces between the protein and the DNA, the ease of deformation of the DNA, the dehydration of the interface, and the release of counterions from the DNA interface. The simulations described here have revealed conformational preferences of different DNA sequences in relation to TBP binding, identifying why certain sequence-dependent properties of DNA are essential for discrimination of TBP binding preferences. Thus the results suggest that alternating YR sequences are preferred for TBP binding because YR steps achieve on average the high rise required for the kink sites at steps 1 and 7, whereas the average conformational parameters for RY steps display the low twist and high positive roll needed throughout the recognition element. These inferences agree with the experimen-

TABLE 6 Comparison to experiment: equilibrium binding constants

Sequences	**Average** and *transient* specificity determinants*	K_{eq} (10^{-9}) (Wong and Bateman, 1994)
Consensus	**YR***TATA***YR**	
mlp	**TA***TAAAAG*	3.7
2c	**CA***TAAAAG*	NA
7g	**TA***TAAGAG*	NA
6t	**TA***TATAAG*	1.1
at	**TA***TATA***TA**	1.4
gc	**CG**CGCG**CG**	>10,000

*Average specificity determinants are in boldface; transient determinants are in italics.
NA, Not available.

tally determined preferred binding sites for TBP, and the analyses carried out by Suzuki et al. (1996) and Juo et al. (1996); they provide specific mechanistic insights into the binding site preferences.

Other criteria, developed from the analysis of basepair step properties in the context of experimentally known structures of the complexes, made possible the discriminant analysis of local structural properties required for TBP binding. Applied to the results of the simulations, these criteria revealed both "best tetrads" and "worst tetrads" and a set of basepair step properties (Table 5: positive slide, positive roll, low twist, and high rise) that are likely candidates for being discriminant of poor TBP binding sequences, on the basis that they are less likely to access conformations consonant with TBP binding. The results point to the details of the mechanisms involved in the recognition of specific DNA sequences by TBP that require further elucidation of the underlying molecular interactions responsible for the special properties of these steps, which are both beneficial and detrimental to specific binding. These are the subject of continuing investigations.

We are grateful to two anonymous reviewers for very valuable comments on the manuscript.

This work was supported in part by a grant from the Association for International Cancer Research (HW); grants from DGICYT: PB95-0624 and PR95-275 (LP); and a Fulbright/CONACyT scholarship (NP). Some of the simulations were run at the Cornell National Supercomputer Facility (sponsored by the National Science Foundation and IBM).

REFERENCES

Aggarwal, A. K., D. W. Rodgers, M. Drottar, M. Ptashne, and S. C. Harrison. 1988. Recognition of a DNA operator by the repressor of phage 434: a view at high resolution. *Science.* 242:899–907.

Arndt, K. M., S. L. Ricupero, D. M. Eisenmann, and F. Winston. 1992. Biochemical and genetic characterization of a yeast TFIID mutant that alters transcription in vivo and DNA binding in vitro. *Mol. Cell. Biol.* 12:2372–2382.

Arndt, K. M., C. R. Wobbe, H. S. Ricupero, K. Struhl, and F. Winston. 1994. Equivalent mutations in the two repeats of yeast TATA-binding protein confer distinct TATA recognition specificities. *Mol. Cell. Biol.* 14:3719–3728.

Berman, H. M., W. K. Olson, D. L. Beveridge, J. Westbrook, A. Gelbin, T. Demeny, S.-H. Hsieh, A. R. Srinivasan, and B. Schneider. 1992. The nucleic acid database: a comprehensive relational database of three-dimensional structures of nucleic acids. *Biophys. J.* 63:751–759.

Bernstein, F. C., T. F. Koetzle, G. J. Williams, E. E. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542.

Bernues, J., P. Carrera, and F. Azorin. 1996. TBP binds the transcriptionally inactive TA5 sequence but the resulting complex is not efficiently recognised by TFIIB and TFIIA. *Nucleic Acids Res.* 24:2950–2958.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 4:187–217.

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

Calladine, C. R. 1982. Mechanics of sequence-dependent stacking of bases in B-DNA. *J. Mol. Biol.* 161:343–352.

Calladine, C. R., and H. R. Drew. 1986. Principles of sequence-dependent flexure of DNA. *J. Mol. Biol.* 192:907–918.

Calladine, C. R., H. R. Drew, and M. J. McCall. 1988. The intrinsic curvature of DNA in solution. *J. Mol. Biol.* 201:127–137.

Chen, W., and K. Struhl. 1988. Saturation mutagenesis of a yeast his3 "TATA element": genetic evidence for a specific TATA-binding protein. *Proc. Natl. Acad. Sci. USA.* 85:2691–2695.

Coleman, R. A., and B. F. Pugh. 1995. Evidence for functional binding and stable sliding of the TATA binding protein on nonspecific DNA. *J. Biol. Chem.* 270:13850–13859.

Coleman, R. A., A. K. Taggart, L. R. Benjamin, and B. F. Pugh. 1995. Dimerization of the TATA binding protein. *J. Biol. Chem.* 270: 13842–13849.

Cormack, B. P., and K. Struhl. 1992. The TATA-binding protein is required for transcription by all three nuclear RNA polymerases in yeast cells. *Cell.* 69:685–696.

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.

Crothers, D. M., and J. Drak. 1992. Global features of DNA structure by comparative gel electrophoresis. *Methods Enzymol.* 212:46–71.

Darden, T. A., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98: 10089–10092.

Desjarlais, J. R., and J. M. Berg. 1992. Toward rules relating zinc finger protein sequences and DNA binding site preferences. *Proc. Natl. Acad. Sci. USA.* 89:7345–7349.

Desjarlais, J. R., and J. M. Berg. 1993. Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. *Proc. Natl. Acad. Sci. USA.* 90:2256–2260.

Drapper, D. E. 1993. Protein-DNA complexes: the cost of recognition. *Proc. Natl. Acad. Sci. USA.* 90:7429–7430.

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.

Goodsell, D. S., and R. E. Dickerson. 1994. Bending and curvature calculations in B-DNA. *Nucleic Acids Res.* 22:5497–5503.

Gorin, A. A., V. B. Zhurkin, and W. K. Olson. 1995. B-DNA twisting correlates with base-pair morphology. *J. Mol. Biol.* 247:34–48.

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. *Nature Struct. Biol.* 3:32–7.

Hagerman, P. J. 1990. Sequence-directed curvature of DNA. *Annu. Rev. Biochem.* 59:755–81.

Hagerman, P. J. 1992. Straightening out the bends in curved DNA. *Biochim. Biophys. Acta.* 1131:125–32.

Haran, T. E., J. D. Kahn, and D. M. Crothers. 1994. Sequence elements responsible for DNA curvature. *J. Mol. Biol.* 244:135–143.

Harrington, R. E., and I. Winicov. 1994. New concepts in protein-DNA recognition: sequence-directed DNA bending and flexibility. *Prog. Nucleic Acids Res. Mol. Biol.* 47:195–270.

Hoopes, B. C., J. F. LeBlanc, and D. K. Hawley. 1992. Kinetic analysis of yeast TFIID-TATA box complex formation suggests a multi-step pathway. *J. Biol. Chem.* 267:11539–11547.

Huang, L., T. Sera, and P. G. Schultz. 1994. A permutational approach toward protein-DNA recognition. *Proc. Natl. Acad. Sci. USA.* 91: 3969–3973.

Hunter, C. A. 1993. Sequence-dependent DNA structure. The role of base stacking interactions. *J. Mol. Biol.* 230:1025–1054.

Jen-Jacobson, L. 1995. Structural-perturbation approaches to thermodynamics of site-specific protein-DNA interactions. *Methods Enzymol.* 259:305–344.

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.

Kahn, J. D., E. Yun, and D. M. Crothers. 1994. Detection of localized DNA flexibility. *Nature.* 368:163–166.

Kim, C. A., and J. M. Berg. 1996. A 2.2 Å resolution crystal structure of a designed zinc finger protein bound to DNA. *Nature Struct. Biol.* 3:940–945.

Kim, J. L., and S. K. Burley. 1994. 1.9 Å resolution refined structure of TBP recognizing the minor groove of TATAAAAG. *Nature Struct. Biol.* 1:638–653.

Kim, Y., J. H. Geiger, S. Hahn, and P. B. Sigler. 1993. Crystal structure of a yeast TBP/TATA-box complex. *Nature.* 365:512–520.

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.

Lieberman, P. M., M. C. Schmidt, C. C. Kao, and A. J. Berk. 1991. Two distinct domains in the yeast transcription factor IID and evidence for a TATA box-induced conformational change. *Mol. Cell Biol.* 11:63–74.

Lustig, B., and R. L. Jernigan. 1995. Consistencies of individual DNA base-amino acid interactions in structures and sequences. *Nucleic Acids Res.* 23:4707–4711.

Lyubchenko, Y. L., L. S. Shlyakhtenko, E. Appella, and R. E. Harrington. 1993. CA runs increase DNA flexibility in the complex of lambda Cro protein with the OR3 site. *Biochemistry.* 32:4121–4127.

MacKerell, A. D., Jr., J. Wiorkiewicz-Kuczera, and M. Karplus. 1995. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* 117:11946–11975.

Matthews, B. W. 1988. Protein-DNA interaction. No code for recognition. *Nature.* 335:294–295.

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.

Olson, W. K., M. S. Babcock, A. Gorin, G. Liu, N. L. Marky, J. A. Martino, S. C. Pedersen, A. R. Srinivasan, I. Tobias, T. P. Westcott, and P. Zhang. 1995. Flexing and folding double helical DNA. *Biophys. Chem.* 55:7–29.

Otwinowski, Z., R. W. Schevitz, R. G. Zhang, C. L. Lawson, A. Joachimiak, R. Q. Marmorstein, B. F. Luisi, and P. B. Sigler. 1988. Crystal structure of trp repressor/operator complex at atomic resolution. *Nature.* 335:321–329.

Park, C., J. L. Campbell, and W. A. D. Goddard. 1993. Design superiority of palindromic DNA sites for site-specific recognition of proteins: tests using protein stitchery. *Proc. Natl. Acad. Sci. USA.* 90:4892–4896.

Parkhurst, K. M., M. Brenowitz, and L. J. Parkhurst. 1996. Simultaneous binding and bending of promoter DNA by the TATA binding protein: real time kinetic measurements. *Biochemistry.* 35:7459–7465.

Pastor, N., and H. Weinstein. 1995. Electrostatic analysis of DNA binding properties in lysine to leucine mutants of TATA-box binding proteins. *Protein Eng.* 8:543–549.

Pearlman, D. A., D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, III, 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.

Perez-Howard, G. M., P. A. Weil, and J. M. Beechem. 1995. Yeast TATA binding protein interaction with DNA: fluorescence determination of oligomeric state, equilibrium binding, on-rate, and dissociation kinetics. *Biochemistry.* 34:8005–8017.

Petri, V., M. Hsieh, and M. Brenowitz. 1995. Thermodynamic and kinetic characterization of the binding of the TATA binding protein to the adenovirus E4 promoter. *Biochemistry.* 34:9977–9984.

Poncin, M., B. Hartmann, and R. Lavery. 1992a. Conformational sub-states in B-DNA. *J. Mol. Biol.* 226:775–794.

Poncin, M., D. Piazzola, and R. Lavery. 1992b. DNA flexibility as a function of allomorphic conformation and of base sequence. *Biopolymers.* 32:1077–1103.

Price, M. A., and T. D. Tullius. 1993. How the structure of an adenine tract depends on sequence context: a new model for the structure of TnAn DNA sequences. *Biochemistry.* 32:127–136.

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(CGCGAATTCCGCG) double helix: "Curves," Dials and Windows. *J. Biomol. Struct. Dyn.* 6:669–699.

Sanghani, S. R., K. Zakrzewska, S. C. Harvey, and R. Lavery. 1996. Molecular modelling of (A4T4NN)n and (T4A4NN)n: sequence elements responsible for curvature. *Nucleic Acids Res.* 24:1632–1637.

Sarai, A., J. Mazur, R. Nussinov, and R. L. Jernigan. 1988. Origin of DNA helical structure and its sequence dependence. *Biochemistry.* 27: 8498–8502.

Sarai, A., J. Mazur, R. Nussinov, and R. L. Jernigan. 1989. Sequence dependence of DNA conformational flexibility. *Biochemistry.* 28: 7842–7849.

Schwabe, J. W. R., L. Chapman, and D. Rhodes. 1995. The oestrogen receptor recognizes an imperfectly palindromic response element through an alternative side-chain conformation. *Structure.* 3:201–213.

Seeman, N. C., J. M. Rosenberg, and A. Rich. 1976. Sequence-specific recognition of double helical nucleic acids by proteins. *Proc. Natl. Acad. Sci. USA.* 73:804–808.

Srinivasan, A. R., R. Torres, W. Clark, and W. K. Olson. 1987. Base sequence effects in double helical DNA. I. Potential energy estimates of local base morphology. *J. Biomol. Struct. Dyn.* 5:459–496.

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. 1993. Common features in DNA recognition helices of eukaryotic transcription factors. *EMBO J.* 12:3221–3226.

Suzuki, M. 1994. A framework for the DNA-protein recognition code of the probe helix in transcription factors: the chemical and stereochemical rules. *Structure.* 2:317–326.

Suzuki, M., M. D. Allen, N. Yagi, and J. T. Finch. 1996. Analysis of co-crystal structures to identify the stereochemical determinants of the orientation of TBP on the TATA box. *Nucleic Acids Res.* 24:2767–2773.

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.* 381:127–151.

Travers, A. A. 1991. DNA bending and kinking: sequence dependence and function. *Curr. Opin. Struct. Biol.* 1:114–122.

Travers, A. A. 1992. DNA conformation and configuration in protein-DNA complexes. *Curr. Opin. Struct. Biol.* 2:71–77.

Ulyanov, N. B., and T. L. James. 1995. Statistical analysis of DNA duplex structural features. *Methods Enzymol.* 261:90–120.

von Hippel, P. H. 1994. Protein-DNA recognition: new perspectives and underlying themes. *Science.* 263:769–770.

von Hippel, P. H., and O. G. Berg. 1986. On the specificity of DNA-protein interactions. *Proc. Natl. Acad. Sci. USA.* 83:1608–1612.

White, S., E. E. Baird, and P. B. Dervan. 1996. Effects of the A.T/T.A degeneracy of pyrrole-imidazole polyamide recognition in the minor groove of DNA. *Biochemistry.* 1996:12532–12537.

Wobbe, C. R., and K. Struhl. 1990. Yeast and human TATA-binding proteins have nearly identical DNA sequence requirements for transcription in vitro. *Mol. Cell. Biol.* 10:3859–3867.

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.

Yanagi, K., G. G. Prive, and R. E. Dickerson. 1991. Analysis of local helix geometry in three B-DNA decamers and eight dodecamers. *J. Mol. Biol.* 217:201–214.

Young, M. A., G. Ravishanker, and D. L. Beveridge. 1995. Analysis of local helix bending in crystal structures of DNA oligonucleotides and DNA-protein complexes. *Biophys. J.* 68:2454–2468.

Zakrzewska, K. 1992. Static and dynamic conformational properties of AT sequences in B-DNA. *J. Biomol. Struct. Dyn.* 9:681–693.

Zhurkin, V. B. 1985. Sequence-dependent bending of DNA and phasing of nucleosomes. *J. Biomol. Struct. Dyn.* 2:785–804.

Zhurkin, V. B., Y. P. Lysov, and V. I. Ivanov. 1979. Anisotropic flexibility of DNA and the nucleosomal structure. *Nucleic Acids Res.* 6:1081–1096.

Zhurkin, V. B., N. B. Ulyanov, A. A. Gorin, and R. L. Jernigan. 1991. Static and statistical bending of DNA evaluated by Monte Carlo simulations. *Proc. Natl. Acad. Sci. USA.* 88:7046–7050.

Zilliacus, J., A. P. Wright, J. Carlstedt Duke, L. Nilsson, and J. A. Gustafsson. 1995. Modulation of DNA-binding specificity within the nuclear receptor family by substitutions at a single amino acid position. *Proteins.* 21:57–67.