Binding of Proteins to the Minor Groove of DNA: What Are the Structural and Energetic Determinants for Kinking a Basepair Step?

DAVID BOSCH, MERCEDES CAMPILLO, LEONARDO PARDO
Laboratori de Medicina Computacional, Unitat de Bioestadistica, Facultat de Medicina, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

Received 27 March 2002; Accepted 13 August 2002

Abstract: The structural and energetic determinants for kinking a basepair step by minor groove–insertion of the protein side chains of PurR, LacI, LEF–1, IHF, Sac7d, and Sso7d, have been calculated by molecular dynamics/potential of mean force simulations. The structural determinants of the kinked structures are: two contiguous furanose rings achieve different conformations, in the region of C3′endo (A–DNA) and C2′endo (B–DNA); the χ torsion angle always takes values characteristic of the C2′endo conformation of B–DNA, independently of sugar puckering; and protein side chain insertion increases *slide* (from negative to positive values), *rise*, and *roll*, and decreases *twist*. The energetic determinants of DNA kinking are: the conformational transition of the sugar–phosphate backbone is not energetically demanding; the relative importance of the interbase parameters in the free energy penalty is *slide*, followed by *twist* and *rise*, and concluding with *shift* and *roll*; and the characteristic increase of *roll* and decrease of *twist*, upon side chain insertion, tends to stabilize the process of DNA kinking.

© 2003 Wiley Periodicals, Inc. J Comput Chem 24: 682–691, 2003

Key words: protein–DNA interactions; minor groove-binding proteins; molecular dynamics; potential of mean force

Introduction

A growing class of proteins that recognizes DNA primarily through interaction with the minor groove has emerged by the recently determined three–dimensional structures of protein–DNA complexes implicated in the control and regulation of transcription.[1–3] To date, the available three–dimensional structures of this family are the TATA–box binding protein (TBP) complexed with a 12 basepair duplex containing the TATA box of the CYC1 promoter,[4] with the TATA element of the adenovirus major late promoter,[5] and with the symmetric TATATATA sequence from the adenovirus E4 promoter;[6] the purine repressor (PurR) bound to the corepressor, hypoxanthine, and the 16 basepair purF operator;[7] the lactose repressor (LacI) complexed with a 21 basepair symmetric operator DNA;[8,9] the high–mobility group (HMG) of mouse lymphoid enhancer–binding factor (LEF–1) complexed with a 15 basepair oligonucleotide duplex from the TCR–α gene enhancer;[10] the HMG domain of the human sex–determining region Y (SRY) complexed with its DNA target site in the promoter of the Müllerian inhibitory substance gene;[11] the integration host factor (IHF) complexed with a 35 basepair DNA fragment containing the H′ site of phage l;[12] the hyperthermophile chromosomal protein, Sac7d complexed with either the $d(GCGATCGC)_2$ or the $d(GTAATTTAC)_2$ DNA sequences;[13] and the hyperthermophilic archaebacteria Sso7d protein complexed with the sequence $d(GTAATTAC)_2$.[14] These proteins insert one or several hydrophobic side chains into the minor groove of DNA, unstacking two contiguous basepairs and thus producing noticeable kinks at one or more sites. The helical axis of the DNA structure is severely bent towards the major groove as a consequence of the side chain insertion.

The successful formation of the protein–DNA complex requires a negative binding free energy. It has recently been shown for 10 different site–specific DNA binding proteins, including minor groove–inserting DNA–bending proteins, that the binding free energy ranges only from -9 to -16 kcal/mol.[15] This small range in binding free energy suggested an isothermal enthalpy–entropy compensation: complexes with undistorted DNA have favorable enthalpy and unfavorable entropy, whereas complexes with distorted DNA have unfavorable enthalpy and favorable entropy. The formation of the protein–DNA interface, which includes van der Waals contacts, hydrogen bonds, and salt bridges, is expected to be enthalpically favorable in all type of complexes. Nevertheless, systems with distorted DNA complexes have an

Correspondence to: M. Campillo

Contract/grant sponsor: DGES; contract/grant number: PB98–0907

© 2003 Wiley Periodicals, Inc.

Table 1. Protein Data Bank Code, Notation of the Basepair Steps at the Place of Insertion, and the Intercalating Protein Side Chains of All Protein–DNA Complexes.

Protein	PDB code	Basepair steps		Protein side chains	Name
		Strand A	Strand B		
PurR	2pua	Cyt707[a]	Gua707′[d]	Leu54, Leu54′	CpG_{PurR}
		Gua708[b]	Cyt708′[c]		
LacI	1efa	Cyt11[a]	Gua12′[d]	Leu56	CpG_{LacI}
		Gua12[b]	Cyt11′[c]		
LEF-1	2lef	Thy7[a]	Ade24[d]	Met10	TpT_{LEF-1}
		Thy8[b]	Ade23[c]		
IHF	1ihf	Ade37[a]	Thy37′[d]	Pro65	ApA_{IHF}
		Ade38[b]	Thy38′[c]		
Sac7d	1azp	Cyt102[a]	Gua115′[d]	Val26, Met29	CpG_{Sac7d}
		Gua103[b]	Cyt114′[c]		
Sac7d	1azq	Ade103[a]	Thy114′[d]	Val26, Met29	ApA_{Sac7d}
		Ade104[b]	Thy113′[c]		
Sso7d	1bnz	Ade68[a]	Thy79′[d]	Val26, Met29	ApA_{Sso7d}
		Ade69[b]	Thy78′[c]		

These sugar-base moieties will be denoted throughout the manuscript as:
[a]StrandA:sugar1.
[b]StrandA:sugar2.
[c]StrandB:sugar1.
[d]StrandB:sugar2.

additional energetic cost, due to DNA kinking (denoted as "molecular strain" in Jen–Jacobson et al.[15]), which yields, in contrast to undistorted DNA, unfavorable binding enthalpies. We aim to explore, in this article, this energetic cost of DNA kinking and bending by minor groove–inserting DNA–bending proteins. We report the free energy penalty required for the conformational transition to the kinked conformation observed in the experimental protein–DNA complexes, calculated by molecular dynamics/potential mean force (MD/PMF) simulations. The structural and energetic determinants of base step kinks have been studied in view of the sugar–phosphate backbone and basepair step moieties.

Methods

Atomic Coordinates of Protein–DNA Complexes

The atomic coordinates and notation of the complexes between DNA and PurR [PDB[16] access number 2pua]; LacI (1lbg); LEF–1 (2lef); IHF (1ihf); Sac7d (1azp and 1azq); and Sso7d (1bnz) proteins are employed throughout the article. To achieve electroneutrality of the systems, counterions were included in the simulations using the CION module in AMBER 5. The protein–DNA complexes and the counterions were placed in a rectangular box containing Monte Carlo-equilibrated TIP3P[17] water with the EDIT module in AMBER 5. Initially, the counterions and water molecules were energy minimized (100 steepest descent steps followed by 1400 conjugate gradient steps). Subsequently, all elements of the systems were energy minimized (500 steepest descent steps followed by 5000 conjugate gradient steps). The resulting systems are used as reference in the article and are named CpG_{PurR}, CpG_{LacI}, TpT_{LEF-1}, ApA_{IHF}, CpG_{Sac7d}, ApA_{Sac7d}, or ApA_{Sso7d}. Table 1 shows, for all protein–DNA complexes, the DNA basepair steps at the place of insertion, the coordinates of which are employed in the calculations, the intercalating protein side chains, and the notation employed throughout the article. The TBP–DNA complexes were not included in the simulations because the TATA box sequence appears to be remarkably unstable in the B–DNA conformation (as seen in Pardo et al.[18]). In the case of the TBP–DNA interaction, the mechanism of DNA bending has been proposed to involve the recognition of an A–DNA–like conformation,[19–21] and the change of the torsion angle of the glycosyl–bond χ[22,23] by both the insertion of hydrophobic side chains and the formation of direct hydrogen bonds.[24] The SRY-DNA complex was not included in the simulations because the furanose rings of Ade5 and Ade6 possess in the experimental structure (PDB access number 1hry) particularly flat rings with low amplitudes of pseudorotation (24.1° and 4.0°, respectively). These values are in noticeable contrast to the mean value of 35.2° found in crystal structures of deoxynucleosides and deoxynucleotides or in helical DNA crystal structures.[25]

Molecular Model of the DNA Basepair Steps at the Place of Insertion

The double–stranded DNA basepair steps $d(CpG)_2$, observed in the PurR, LacI, and Sac7d complexes; $d(TpT)_2$, observed in the LEF–1 complex; and $d(ApA)_2$, observed in the IHF, Sac7d, and Sso7d complexes, were built in the canonical B–DNA conformation, a common starting conformation for simulations with the AMBER package.[26,27] Moreover, a single backbone formed by two deoxyribose sugars connected by a phosphate group was also

built in the B–DNA conformation. This single backbone coincides with a single–stranded DNA in which the bases have been replaced by hydrogens. Similar molecular models were previously employed to study the free energy of the stacking-unstacking process of all 16 naturally occurring DNA dimers.[28] To achieve electroneutrality of the systems, counterions were included, positioned initially at a distance of 3.5 Å from each P atom. The solute and the counterions were placed in a rectangular box containing ≈1450 Monte Carlo-equilibrated TIP3P water. Initially, the DNA was kept fixed whereas the sodium ions and water molecules 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 80 ps) at constant pressure with isotropic scaling (Berendsen coupling). The final box sizes after equilibration were ≈46 × 36 × 33 Å, resulting in a final density in the 1.00–1.01 g cm^{-3} range. Subsequently, the entire systems were energy minimized (500 steepest descent steps followed by 1000 conjugate gradient steps), heated (from 0 to 300 K in 15 ps), and equilibrated (from 15 to 1000 ps) at constant volume. During these 1000 ps of simulation the torsion angles α, β, γ, δ, ε, ζ, χ, and ν_1, and the Watson–Crick base pairing were maintained close to the initial B–DNA conformation with flat harmonic restraints (32 kcal mol^{-1} rad^{-2} for torsional angles and 32 kcal mol^{-1} $Å^{-2}$ for hydrogen bond distances). The restraint was a well with a square bottom (±5° for torsional angles and ±0.1 Å for hydrogen bond distances) with parabolic sides. It is important to note that the choice of the starting conformation will affect the values obtained in the free energy calculations (see below). Although the sequence-dependent DNA structure has been studied,[29–31] it is not possible to predict the most likely conformation in solution for these sequences. The procedure employed here produces similar equilibrated initial conformations for the sequences compared, thus allowing for direct comparisons of the free energy calculations. The MD simulations were run with the Sander module of AMBER5,[32] the all–atom force field,[33] SHAKE bond constraints,[34] 2 fs integration time step, and constant temperature of 300 K coupled to a heat bath (Berendsen coupling algorithm).

Free Energy of Basepair Step Kinking Determined by MD/PMF Simulations

The conformational transitions of the double–stranded DNA from the initial B–DNA (named CpG_B, TpT_B, or ApA_B) to the final kinked (named CpG_{PurR}, CpG_{LacI}, TpT_{LEF-1}, ApA_{IHF}, CpG_{Sac7d}, ApA_{Sac7d}, or ApA_{Sso7d}) structures were carried out with an MD/PMF procedure by changing in 101 windows the torsional angles α, β, γ, δ, ε, ζ, χ, and ν_1, and the Watson–Crick hydrogen bond distances from the initial B–DNA values to those found in the optimized protein–DNA complexes. Moreover, the conformational transition of the deoxyribose–phosphate–deoxyribose (RpR) backbone from B–DNA (RpR_B) to the conformation observed in strand A (named $RpR_{PurR/A}$, $RpR_{LacI/A}$, $RpR_{LEF-1/A}$, $RpR_{IHF/A}$, $RpR_{Sac7dCG/A}$, $RpR_{Sac7dAA/A}$, or $RpR_{Sso7d/A}$) and strand B (named $RpR_{PurR/B}$, $RpR_{LacI/B}$, $RpR_{LEF-1/B}$, $RpR_{IHF/B}$, $RpR_{Sac7dCG/B}$, $RpR_{Sac7dAA/B}$, or $RpR_{Sso7d/B}$) of the protein–DNA complexes was also accomplished by changing the torsional angles α, β, γ, δ, ε, ζ, χ, and ν_1 from the initial B–DNA values to those found in the protein–DNA complexes. It has recently been shown that the all atom force-field of AMBER[33] properly reproduces the empirical torsional energy profiles of these ε, β, γ, and χ torsions as compared with *ab initio* calculations.[35] 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.[36,37] The simulations were conducted at constant volume so the Helmholtz free energies are reported. The values of free energy for the conformational transitions of the double–stranded DNA are designated as ΔF_{dimer}, whereas the free energy values for single RpR backbones are designated as ΔF_A (strand A) and ΔF_B (strand B). The convergence of the simulation was checked by systematically increasing the number of steps of data collection in each window (see Fig. 3 in Pardo et al.[18] for the convergence test of similar conformational transitions). A total simulation time of 2020 ps, using 101 windows × (5 ps of equilibration + 15 ps of data collection), was used. The MD/PMF simulations were run with the Gibbs modules of AMBER 5. All solute–solute interactions were evaluated, and an 8 Å cutoff was applied to the solute–water and water–water interactions.

Conformational and Statistical Analysis of Basepair Step Kinking

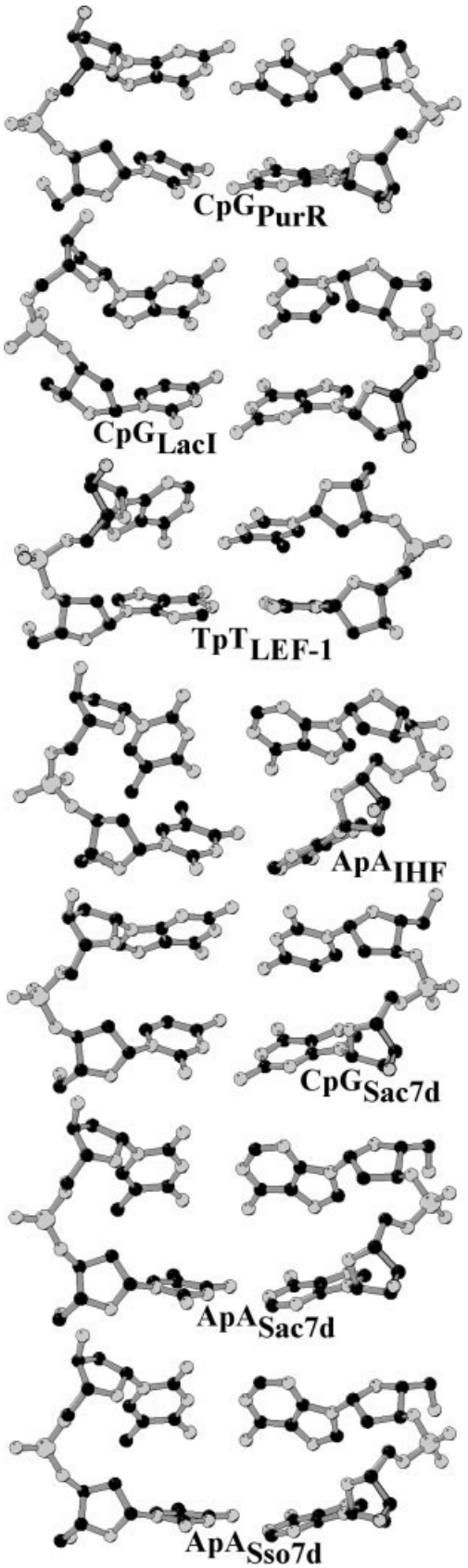
The average structures computed from the data collection trajectories during the conformational transitions at each window were analyzed with the CURVES 5.1 program.[38,39] The sugar–phosphate and the interbase step parameters were obtained for these average structures. The sugar–phosphate parameters included the torsional and pseudorotation angles $\gamma_{A:1}$, $\delta_{A:1}$, $\chi_{A:1}$, $\varepsilon_{A:1}$, $\zeta_{A:1}$, and $P_{A:1}$ of strandA:sugar1; $\alpha_{A:2}$, $\beta_{A:2}$, $\gamma_{A:2}$, $\delta_{A:2}$, $\chi_{A:2}$, and $P_{A:2}$ of strandA:sugar2; $\gamma_{B:1}$, $\delta_{B:1}$, $\chi_{B:1}$, $\varepsilon_{B:1}$, $\zeta_{B:1}$, and $P_{B:1}$ of strandB:sugar1; and $\alpha_{B:2}$, $\beta_{B:2}$, $\gamma_{B:2}$, $\delta_{B:2}$, $\chi_{B:2}$, and $P_{B:2}$ of strandB:sugar2. The interbase step parameters included local values of *shift* (D_x), *slide* (D_y), *rise* (D_z), *tilt* (τ), *roll* (ρ), and *twist* (Ω).

The statistical analysis comprised a principal component analysis with Varimax/Kaiser rotation of the sugar–phosphate and interbase step parameters, performed independently, and the regression between the calculated free energies required for kinking the basepair step (ΔF_{dimer}) and the interbase parameters (denoted Φ) through the $\Delta F_{dimer} = K_\Phi(\Phi - \Phi_0)^2$ equation, where Φ_0 is defined as the value of the conformational parameter at the first window of the simulation, in which $\Delta F_{dimer} = 0$. The statistical analyses were performed with the SPSS 9.0 program.[40]

Results and Discussion

Conformational Transition to the Kinked Structures

The conformational transitions from the initial structures (CpG_B, TpT_B, ApA_B) to the final kinked structures (CpG_{PurR}, CpG_{LacI}, TpT_{LEF-1}, ApA_{IHF}, CpG_{Sac7d}, ApA_{Sac7d}, ApA_{Sso7d}, see Fig. 1) were accomplished by changing the torsional angles α, β, γ, δ, ε, ζ, χ, and ν_1, and the Watson–Crick hydrogen bond distances from the initial values to those found in the protein–DNA complexes (see Table 2) in 101 windows (see Methods). Figure 2 shows the average structures computed from the data collection trajectory at



each window during the $ApA_B \rightarrow ApA_{Sso7d}$ transition. The initial ApA_B structure (white) and final ApA_{Sso7d} conformations (black) are shown in "ball and sticks" rendering. The evolution of these computed structures points to the possibility of finding many intermediates on the path between the initial (lightest gray) conformation and final (darkest gray) kinked conformation. Such a change is also observed for the other transitions (results not shown). Thus, the computed structures gradually become more similar to the conformation found in the protein–DNA complexes. This is also reflected in the all atom root mean square deviation (rmsd) between the average computed structures and the structures observed in the protein–DNA complexes (see gray lines in Fig. 3). The highest final rmsd value of 0.8 Å corresponds to the $ApA_B \rightarrow ApA_{IHF}$ conformational transition.

The conformational transition of DNA by minor groove insertion of protein side chains can be studied in view of two main structural factors: the sugar–phosphate backbone conformation, and the basepair step conformation.

Sugar–Phosphate Backbone Conformation

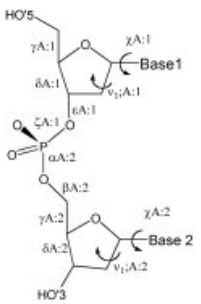
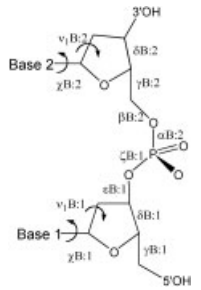
Figure 3 (right panels) depicts the evolution of the pseudorotation angle during the computed conformational transition (color lines) and the values of sugar puckering observed in the kinked structures of the protein–DNA complexes (color circles). The color code is: strand A:sugar 1 (black) sugar 2 (red); strand B:sugar 1 (blue) and sugar 2 (green). As the simulated conformational transition is driven by changing intrastrand and ν_1 dihedral angles and basepair H–bond lengths, sugar puckering becomes gradually similar to the values found in the experimental structures. It is important to note that sugar puckering (torsional angles ν_0, ν_1, ν_2, ν_3, and ν_4^{41}) is explicitly constrained in the computer simulations by ν_1 and δ (correlated with ν_3) torsional angles. The only minor deviation is found in the pucker of strand A:sugar 2 of LEF-1 (see red line and circle in Fig. 3, right panels).

Table 2 shows the results of a principal component analysis on the sugar–phosphate backbone torsions and pseudorotation angles (24 parameters × 101 windows, see Methods). The backbone parameters fall into three different categories.

The first component, which accounts for 32% of the total variance, is a combination of the torsions related to the pucker of sugar 2 in strands A and B ($\delta_{A:2}$, $P_{A:2}$, $\delta_{B:2}$, and $P_{B:2}$), and the torsion β of both strand A and B ($\beta_{A:2}$ and $\beta_{B:2}$). The second component (27%) of the principal component analysis includes the torsions of sugar 1 in strands A and B ($\delta_{A:1}$, $P_{A:1}$, $\delta_{B:1}$, and $P_{B:1}$). A striking feature of the conformational transition from B–DNA to the kinked structure is the change in the pucker of the furanose rings. The conformation of sugar 1 of both strand A (black) and strand B (blue) undergoes a transition from the C2′endo conformation, representative of B–DNA, to the vicinity of the C3′endo conformation, representative of A–DNA, during the $CpG_B \rightarrow CpG_{PurR}$, $TpT_B \rightarrow TpT_{LEF-1}$, $ApA_B \rightarrow ApA_{IHF}$, $CpG_B \rightarrow CpG_{Sac7d}$,

Figure 1. Detailed view of the experimental kinked structures of CpG_{PurR}, CpG_{LacI}, TpT_{LEF-1}, ApA_{IHF}, CpG_{Sac7d}, ApA_{Sac7d}, and ApA_{Sso7d}. Figures were created using MOLSCRIPT.[54]

Table 2. Backbone Dihedral Angles of the Kinked Basepair Step Found in the Experimental Protein–DNA Complexes.

									Principal components		
Strand	Torsion	CpG_{PurR}	CpG_{LacI}	TpT_{LEF-1}	ApA_{IHF}	CpG_{Sac7d}	ApA_{Sac7d}	ApA_{Sso7d}	1	2	3
	$\gamma_{A:1}$	10.0	61.7	52.4	188.4	187.3	60.2	53.9			0.96
	$\delta_{A:1}$	75.5	103.2	86.0	83.1	80.3	81.1	90.8		0.92	
	$\chi_{A:1}$	267.8	237.5	245.2	239.4	234.7	232.2	217.0			
	$\varepsilon_{A:1}$	190.4	226.2	198.4	208.2	207.7	223.4	244.3			
	$\zeta_{A:1}$	284.2	306.9	280.6	279.1	290.3	284.3	292.7			
	$\alpha_{A:2}$	297.1	292.7	291.3	292.5	292.6	292.1	288.8			
	$\beta_{A:2}$	175.2	171.4	174.3	194.7	197.0	194.3	180.1	0.84		
	$\gamma_{A:2}$	61.9	49.1	66.0	62.4	50.7	43.9	53.2			
	$\delta_{A:2}$	148.7	122.9	119.6	135.4	136.8	135.3	108.4	0.86		
	$\chi_{A:2}$	258.6	268.7	254.0	259.6	255.3	253.0	243.7			
	$\nu_{1;A:1}$	349.1	359.5	343.0	359.4	351.7	353.5	349.9			
	$\nu_{1;A:2}$	36.2	41.9	12.0	22.1	24.5	23.8	36.0			
	$P_{A:1}$	39.4	57.6	23.9	54.7	43.4	45.6	35.9		0.96	
	$P_{A:2}$	164.4	130.7	159.2	174.5	173.6	175.7	117.6	0.89		
	$\gamma_{B:1}$	22.4	70.9	32.8	53.2	45.1	50.2	49.8			
	$\delta_{B:1}$	72.9	114.2	69.2	93.6	76.9	77.4	87.8		0.94	
	$\chi_{B:1}$	262.5	240.1	241.1	260.5	259.0	243.3	246.5			
	$\varepsilon_{B:1}$	212.1	250.1	200.7	225.7	192.2	200.5	193.8			
	$\zeta_{B:1}$	294.0	293.2	279.4	292.9	284.5	275.2	279.2			
	$\alpha_{B:2}$	295.0	288.5	307.6	296.6	298.1	297.5	295.8			
	$\beta_{B:2}$	197.7	149.6	164.0	180.6	189.1	180.8	187.9	0.87		
	$\gamma_{B:2}$	57.5	60.4	69.5	55.3	56.0	59.0	62.7			
	$\delta_{B:2}$	146.7	119.7	139.4	130.8	140.9	135.3	140.1	0.83		
	$\chi_{B:2}$	268.1	264.7	245.2	251.3	254.9	242.8	242.3			
	$\nu_{1;B:1}$	344.7	20.0	348.1	335.7	346.2	0.7	349.7			
	$\nu_{1;B:2}$	28.4	43.2	31.0	24.6	24.5	27.4	32.1			
	$P_{B:1}$	34.2	114.4	40.7	359.3	36.3	56.5	35.6		0.97	
	$P_{B:2}$	176.0	128.4	167.8	158.1	178.3	163.0	164.0	0.89		
								% Variance	32.0	27.3	7.2

Principal components analysis of the sugar-phosphate backbone torsions and pseudorotation angles was obtained during the conformational transition from double-stranded B-DNA to the kinked structures (24 parameters × 101 windows). Factor coefficients with absolute value less than 0.8 are omitted for clarity.
The variance population was normalized following the Varimax/Kaiser method.

$ApA_B \rightarrow ApA_{Sac7d}$, and $ApA_B \rightarrow ApA_{Sso7d}$ transitions (Fig. 3, right panels). The only exception is strand A:sugar 1 (black) in the $CpG_B \rightarrow CpG_{LacI}$ transition. In contrast, the other two sugar moieties of the dimer, sugar 2 of strand A (red color) and sugar 2 of strand B (green color), remain in the vicinity of the C2′endo conformation during the $CpG_B \rightarrow CpG_{PurR}$, $CpG_B \rightarrow CpG_{LacI}$, $TpT_B \rightarrow TpT_{LEF-1}$, $ApA_B \rightarrow ApA_{IHF}$, $CpG_B \rightarrow CpG_{Sac7d}$, $ApA_B \rightarrow ApA_{Sac7d}$, and $ApA_B \rightarrow ApA_{Sso7d}$ conformational transitions (see Fig. 3, right panels). Thus, the conformation of the furanose rings in the final kinked basepair step structure follows similar trends in all transitions considered. Clearly, two contiguous sugar moieties adopt different conformations in the vicinity of either C2′endo (sugar 2, red and green) or C3′endo (sugar 1, black and blue) conformations in the kinked basepair step. The values of the torsion angle δ, which constrain the conformation of the sugar in the MD/PMF simulations, are always in concordance with the pucker of the sugar: in the vicinity of the standard value of 130°, characteristic of the C2′endo conformation,[42] for sugar 2, or 80°,

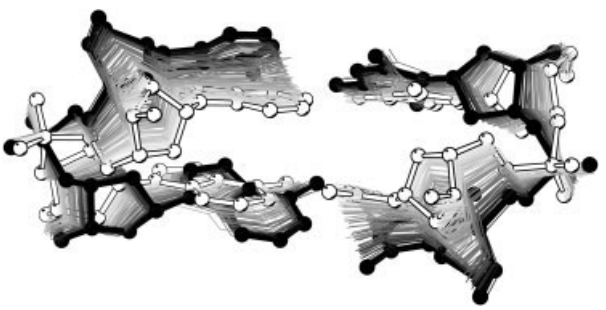


Figure 2. Average structures computed from the data collection trajectory at each window during the $ApA_B \rightarrow ApA_{Sso7d}$ transition. The initial ApA_B structure (white) and final ApA_{Sso7d} conformations (black) are shown in "ball and sticks" rendering. The evolving structures are depicted from the lightest gray (initial) to the darkest gray (final).

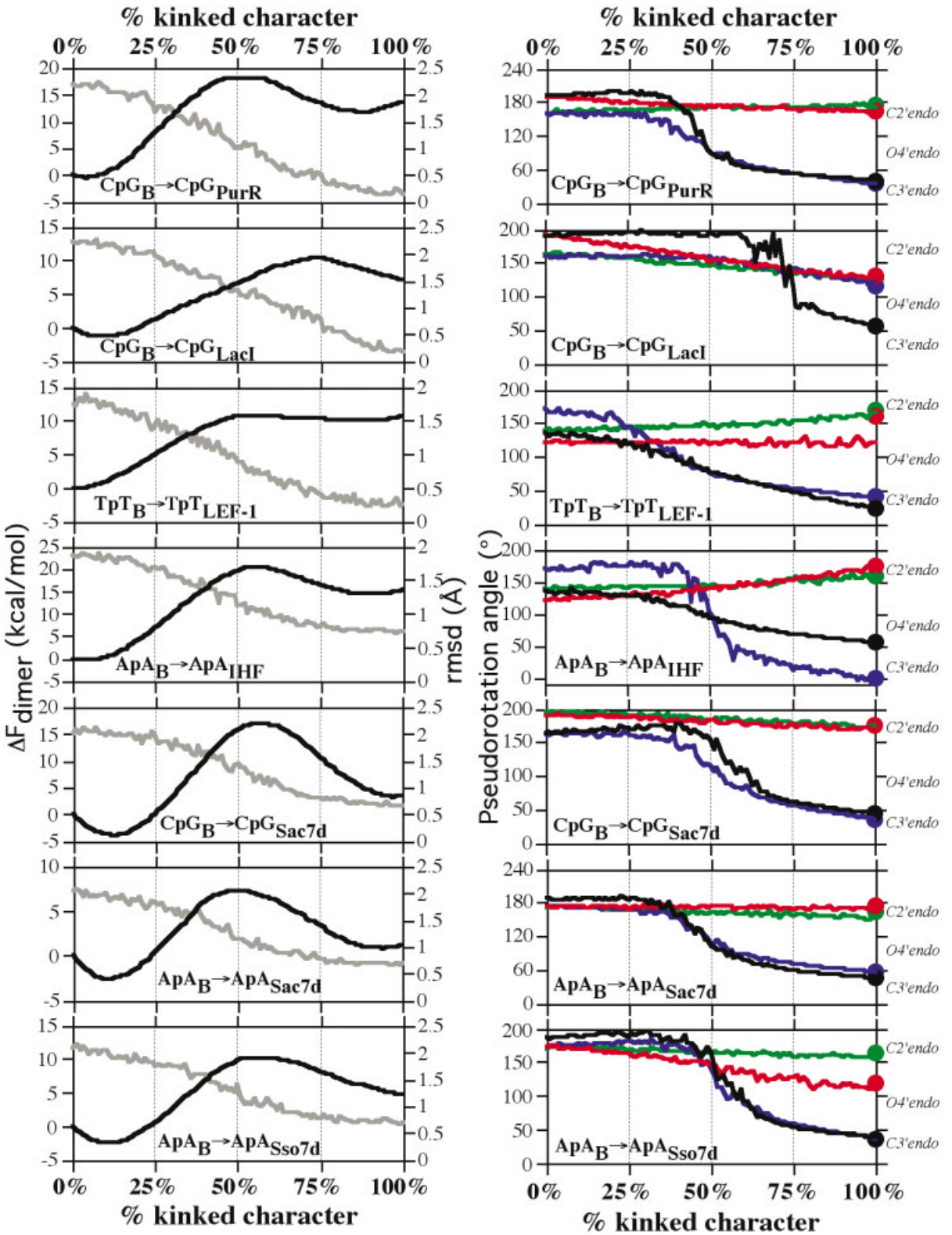


Figure 3. Computed free energy changes (black lines in left panels), rmsd comparisons of the average structures computed from the data collection trajectory at each window and the experimental structures (gray lines in left panels), and the evolution of the pseudorotation angle (color lines in right panels) during the conformational transition from double-stranded B–DNA to the kinked structures found in experimental protein–DNA complexes, and the values of sugar puckering observed in the kinked structures of the experimental protein–DNA complexes (color circles in right panels). The color code of sugar puckering is: strandA:sugar1 (black) sugar 2 (red); strandB:sugar1 (blue) and sugar 2 (green).

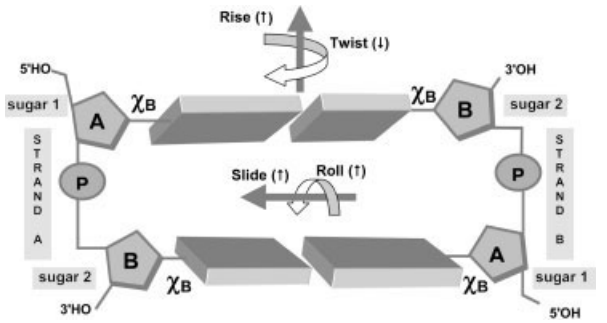


Figure 4. Structural determinants of DNA kinking by minor groove–inserting DNA-bending proteins. The view is from the minor groove side of the basepairs. Two contiguous furanose rings, in the final kinked basepair step structure, achieve different conformations, in the region of C3′endo (A–DNA) in sugar 1 and C2′endo (B–DNA) in sugar 2. However, the χ torsional angle always take values characteristic of the C2′endo conformation of B–DNA, independently of sugar puckering. Moreover, protein side chain insertion increases the basepair step parameters *slide* (from negative to positive values), *rise*, and *roll*, and decreases the *twist* parameter. *Shift* and its rotational related parameter *tilt* always take opposite signs in the kinked structures.

characteristic of the C3′endo conformation,[42] for sugar 1. In contrast, the χ torsional angle always takes the value in the vicinity of 250°, characteristic of the C2′endo conformation of B–DNA,[42] independently of sugar puckering (see Table 2). These structural determinants of DNA kinking are summarized in Figure 4. Moreover, β increases from the standard B–DNA value of 176°[42] to 197.7° (strand B in CpG_{PurR}), 194.7° (strand A in ApA_{IHF}), 197.0° or 189.1° (strands A and B in CpG_{Sac7d}), 194.3° (strand A in ApA_{Sac7d}), or 187.9° (strand B in ApA_{Sso7d}). This variation has facilitated the principal component analysis to also include in the first component the torsions $\beta_{A:2}$ and $\beta_{B:2}$ of strands A and B.

Finally, the third component (7%) includes $\gamma_{A:1}$, which varies considerably from the B–DNA value of 48°[42] to 10.0° in CpG_{PurR}, 188.3° in ApA_{IHF}, and 187.3° in CpG_{Sac7d} (see Table 2).

Basepair Step Conformation

Figure 5 depicts the interbase parameters *shift*, *slide*, *rise*, *tilt*, *roll*, and *twist* for canonical B–DNA conformation, for the average structures computed during the data collection trajectory at the first (initial structure) and the last (final structure) windows of the MD/PMF simulations, and for the kinked DNA basepair step obtained in the protein–DNA complexes (see Fig. 1). Clearly, there is a general coincidence of the basepair step parameters between the final simulated structures and experimental structures.

Table 3 shows the principal components analysis on the basepair step parameters (6 parameters × 101 windows, see Methods). The unique component of this analysis, which accounts for 52% of the total variance, includes *roll*, *slide*, and *rise* with positive factor

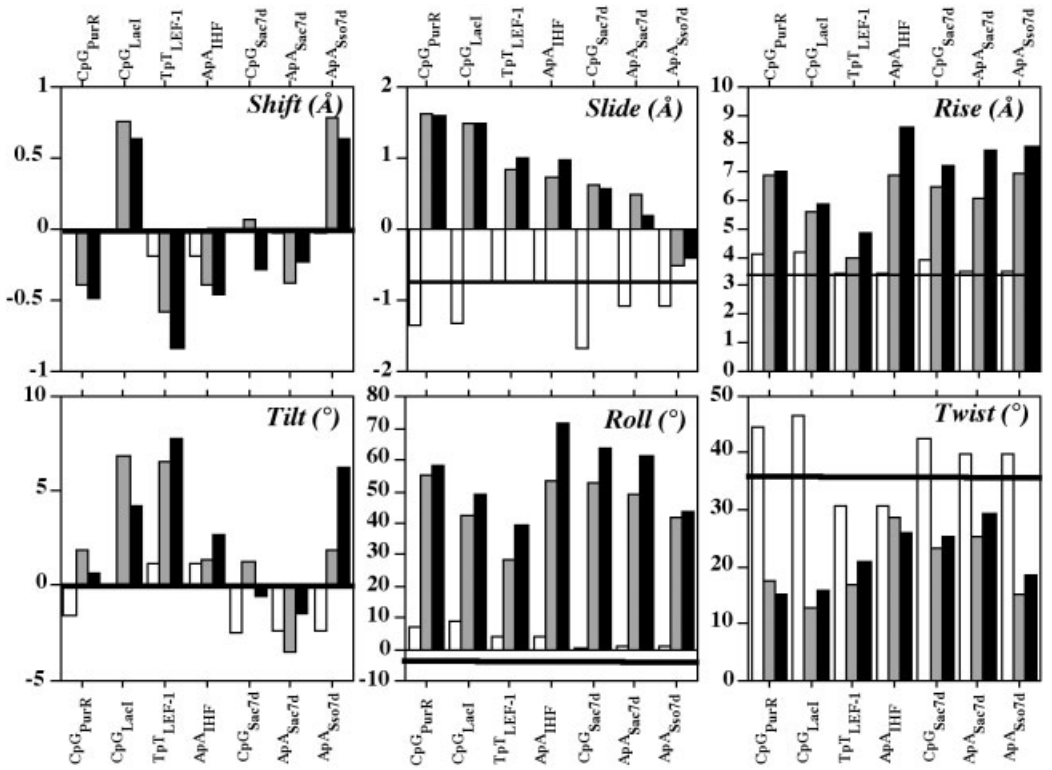


Figure 5. Helical local parameters *shift*, *slide*, *rise*, *tilt*, *roll*, and *twist*, calculated with CURVES 5.1[38,39] for B–DNA (solid lines) for the average structures computed during the data collection trajectory at the first (initial structure, white bars) and the last (final structure, gray bars) windows of the MD/PMF simulations, and of the experimental protein–DNA complexes (black bars).

Table 3. Local Basepair Step Geometric Parameters, Calculated with CURVES 5.1,[38,39] of the Kinked Basepair Step Found in the Experimental Protein–DNA Complexes.

Base step parameter	CpG_{PurR}	CpG_{LacI}	TpT_{LEF-1}	ApA_{IHF}	CpG_{Sac7d}	ApA_{Sac7d}	Apa_{Sso7d}	Principal components 1
Shift	−0.49	0.63	−0.84	−0.46	−0.29	−0.23	0.63	
Slide	1.60	1.48	1.00	0.98	0.55	0.18	−0.41	0.83
Rise	7.05	5.84	4.82	8.59	7.23	7.75	7.88	0.84
Tilt	3.75	4.19	7.77	2.69	−0.61	−1.44	6.19	
Roll	58.04	49.20	39.39	71.75	63.95	61.09	43.46	0.99
Twist	15.13	15.82	20.91	26.00	25.35	29.30	18.43	−0.70
							% Variance	51.5

Principal components analysis of the basepair step parameters was obtained during the conformational transition from double-stranded B-DNA to the kinked structures (6 parameters × 101 windows).
Factor coefficients with absolute value less than 0.6 are omitted for clarity and are normalized following the Varimax/Kaiser method.

coefficients, and *twist* with negative positive factor coefficients. Analysis of the evolution of the basepair step parameters depicted in Figure 5 shows that protein side chain insertion into a B–DNA step increases the *slide* (from negative to positive values), *rise*, and *roll* parameters, and decreases the *twist* parameter. This profile is common to all the considered conformational transitions, with the only exception being *slide* in the $ApA_B \rightarrow CpG_{Sso7d}$ transition. These common trends of DNA kinking by protein side chain insertion are summarized in Figure 4.

Factors Governing the Free Energy Cost of Basepair Step Kinking

Table 4 and Figure 3 (solid black lines, left panels) show the computed free energy changes, evaluated with the MD/PMF simulations described in Methods, for the conformational transition from double–stranded B–DNA to the kinked structures found in the experimental protein–DNA complexes. The insertion of the protein side chain into the basepair step causes a conformational change that demands (see ΔF_{dimer} in Table 4) 13.7 kcal/mol for $CpG_B \rightarrow CpG_{PurR}$, 7.1 kcal/mol for $CpG_B \rightarrow CpG_{LacI}$, 10.8 kcal/mol for $TpT_B \rightarrow TpT_{LEF-1}$, 14.1 kcal/mol for $ApA_B \rightarrow ApA_{IHF}$, 3.4 kcal/mol for $CpG_B \rightarrow CpG_{Sac7d}$, 1.3 kcal/mol for $ApA_B \rightarrow ApA_{Sac7d}$, and 4.9 kcal/mol for $ApA_B \rightarrow ApA_{Sso7d}$ transitions. It is important to note that the overall electrostatic effects of the protein on the energetics of kinking have not been considered in these calculations. It has been proposed that the low dielectric environment of the protein increases phosphate repulsion across the minor groove and facilitates the bending towards the major groove.[43,44] Moreover, asymmetrical neutralization of the phosphates across the major groove by positively charged residues of the protein is likely to favor the energetics of bending.[45–48] We have focused in this work on the influence of both the backbone and the basepair step conformations on the energetics of kinking (denoted as "molecular strain" in Jen–Jacobson et al.[15]). Thus, the reported energies are upper limits. A major goal of future studies will be to determine how the binding of proteins to the minor groove of DNA decreases these energies.

Table 4. Computed Free Energy Changes for the Conformational Transition of Double-Stranded B-DNA (ΔF_{dimer}), Backbone A (ΔF_A), and Backbone B (ΔF_B) from the Initial B-DNA Conformation to the Experimental Kinked Conformation.

	ΔF_{dimer}	ΔF_A	ΔF_B
$CpG_B \rightarrow CpG_{PurR}$	13.7	−2.3	−2.0
$CpG_B \rightarrow CpG_{LacI}$	7.1	−3.4	−2.2
$TpT_B \rightarrow TpT_{LEF-1}$	10.8	−5.2	−4.2
$ApA_B \rightarrow ApA_{IHF}$	14.1	−4.9	−5.0
$CpG_B \rightarrow CpG_{Sac7d}$	3.4	−5.5	−6.1
$ApA_B \rightarrow ApA_{Sac7d}$	1.3	−4.2	−6.4
$ApA_B \rightarrow ApA_{Sso7d}$	4.9	−5.9	−6.0

Energies are in kcal/mol.

The potential free energy profiles of Figure 3 (left panels) display two energy minima connected by an energy maximum during all conformational transitions. Clearly, the maxima of the potential free energy profiles are located when the 50–70% kinked character has been accomplished in the computed conformational transitions. Analysis of the evolution of sugar puckering, in the right panels of Figure 3, reveals that sugar 1 of both strand A (black) and strand B (blue) is adopting the O4′endo conformation at this stage of the simulation. The O4′endo region corresponds to the maximum in the lowest energy path between deoxyribose C2′endo and C3′endo conformations.[41,49,50]

Furthermore, we performed independent MD/PMF simulations of strand A and strand B, in which the bases were replaced by hydrogens (see Methods). Table 4 shows the energetic cost for these conformational transitions of strand A (ΔF_A) and strand B (ΔF_B). The conformational changes of the RpR backbones require free energies between −6.4 and −2.0 kcal/mol. Thus, the conformational transition of the sugar-phosphate backbone, in the absence of the bases, is not energetically demanding.

Table 5. Nonstandardized Coefficients K_{Dx}, K_{Dy}, K_{Dz}, K_τ, K_ρ, and K_Ω, Standard Error of the Coefficients, and Standardized Coefficients β_{Dx}, β_{Dy}, β_{Dz}, β_τ, β_ρ, and β_Ω.

	Shift (Dx)	Slide (Dy)	Rise (Dz)	Tilt (τ)	Roll (ρ)	Twist (Ω)
K	23.86*	4.07*	0.92*	-8.7×10^{-3}	-2.4×10^{-3}*	-3.8×10^{-2}*
Standard error	2.1	0.17	0.15	0.028	0.001	0.002
β	0.46	1.19	0.50	−0.02	−0.25	−0.92

*$p < 0.0001$

Data were obtained by fitting ΔF_{dimer} to the basepair step parameters.

The energy required for kinking the basepair step, from the initial B–DNA to the final experimental conformations, can be grouped into transitions with either relatively high (>7 kcal/mol) or low (<5 kcal/mol) free energy cost. Conformational transitions that induce the largest change (see Fig. 5 and Table 3) in *slide* (PurR, LacI) and *rise* (IHF) belong to the first group, whereas intermediate changes in these basepair step parameters belong to the second group. A direct relationship between (ΔF_{dimer}) and the local base–step parameters *shift* (Dx), *slide* (Dy), *rise* (Dz), *tilt* (τ), *roll* (ρ), and *twist* (Ω), at each window, can be obtained through the following quadratic expression:

$$\Delta F_{\text{dimer}} = K_{Dx}(Dx - Dx_0)^2 + K_{Dy}(Dy - Dy_0)^2 + K_{Dz}(Dz - Dz_0)^2 + K_\tau(\tau - \tau_0)^2 + K_\rho(\rho - \rho_0)^2 + K_\Omega(\Omega - \Omega_0)^2$$

this equation assumes that the base–step parameters behave as harmonic oscillators. $D\text{x}_0$, $D\text{y}_0$, Dz_0, τ_0, ρ_0, and Ω_0 are the values of the basepair step parameters at the first window of the MD/PMF simulations, in which $\Delta F_{\text{dimer}} = 0.0$ kcal/mol, and K_{Dx}, K_{Dy}, K_{Dz}, K_τ, K_ρ, and K_Ω the corresponding energy constants. A gradual change in the torsional angles α, β, γ, δ, ε, ζ, χ, and ν_1, as performed in the MD/PMF simulations, does not produce a gradual change in Dx, Dy, Dz, τ, ρ, and Ω. Despite the fact that there is a clear tendency towards the final basepair–step parameters, these vary considerably between two successive windows. Thus, the values of ΔF_{Step}, Dx, Dy, Dz, τ, ρ, and Ω were smoothed by averaging the value at each window with the previous one. Table 5 shows the obtained nonstandardized coefficients K, the standard error of the coefficients, and the standardized coefficients β. All the coefficients are significantly different from zero ($p < 0.0001$), except *tilt*. The relative importance of the basepair step parameters in the free energy of kinking can be determined by the magnitude of the standardized coefficients β that, in contrast to nonstandardized coefficients, do not take into account the magnitude of the independent variable. Consequently, the parameters that most determine the free energy cost of kinking the basepair step are *slide* (1.19), followed by *twist* (−0.92) and *rise* (0.50), and concluding with *shift* (0.46) and *roll* (−0.25). It is important to note the positive sign of the coefficients of all the translational parameters, which produces positive energies of basepair step kinking, and the negative sign of all the rotational parameters, which produces negative energies. Thus, the characteristic variation of the rotational parameters *roll* (increase) and *twist* (decrease), upon side chain insertion (increase *slide* and *rise*), tends to stabilize the process of DNA kinking.

Conclusions

The simulations presented in this work provide new insights into the structural and energetic determinants associated with the process of kinking a DNA basepair step. These determinants have been studied in view of the sugar–phosphate backbone and basepair step moieties. Two contiguous furanose rings, in the final kinked basepair step structure, achieve different conformations, in the region of C3′endo (A–DNA) and C2′endo (B–DNA). This conformational transition of the sugar-phosphate backbone is not energetically demanding in the absence of the bases. Analysis of the basepair step parameters, at the place of insertion, shows a significant increase of *slide* (from negative to positive values), *rise*, and *roll*, and decreases of *twist*. The energy required for modifying these basepair step parameters is in the 1.3–14.1 kcal/mol range as obtained by MD/PMF. The parameters that most determine the free energy cost of kinking the basepair step are *slide*, followed by *twist* and *rise*, and ending with *shift* and *roll*. The more pronounced the changes in *slide* and *rise*, the larger the energetic penalty of kinking. In contrast, the characteristic variation of the rotational parameters *roll* and *twist*, upon side chain insertion, tends to stabilize the process of DNA kinking.

These energetic contributions of DNA kinking and bending are superimposed on the effect of water and counterion release from the interaction surfaces,[51–53] protein–DNA recognition interactions, immobilization of the interaction surface,[15] the coupling between DNA binding and partial protein folding,[15] and the total energy of formation of the protein–DNA complex.

Acknowledgments

Some of the simulations were run at the Centre de Computació i Comunicacions de Catalunya.

References

1. Werner, M. H.; Gronenborn, A. M.; Clore, G. M. Science 1996, 271, 778.
2. Werner, M. H.; Burley, S. K. Cell 1997, 88, 733.
3. Bewley, C. A.; Gronenborn, A. M.; Clore, G. M. Annu Rev Biophys Biomol Struct 1998, 27, 105.
4. Kim, Y.; Geiger, J. H.; Hahn, S.; Sigler, P. B. Nature 1993, 365, 512.
5. Kim, J. L.; Nikolov, D. B.; Burley, S. K. Nature 1993, 365, 520.

6. Juo, Z. S.; Chiu, T. K.; Leiberman, P. M.; Baikalov, I.; Berk, A. J.; Dickerson, R. E. J Mol Biol 1996, 261, 239.
7. Schumacher, M. A.; Choi, K. Y.; Zalkin, H.; Brennan, R. G. Science 1994, 266, 763.
8. Lewis, M.; Chang, G.; Horton, N. C.; Kercher, M. A.; Pace, H. C.; Schumacher, M. A.; Brennan, R. G.; Lu, P. Science 1996, 271, 1247.
9. Bell, C. E.; Lewis, M. Nat Struct Biol 2000, 7, 209.
10. Love, J. J.; Li, X.; Case, D. A.; Giese, K.; Grosschedl, R.; Wright, P. E. Nature 1995, 376, 791.
11. Werner, M. H.; Huth, J. R.; Gronenborn, A. M.; Clore, G. M. Cell 1995, 81, 705.
12. Rice, P. A.; Yang, S.; Mizuuchi, K.; Nash, H. A. Cell 1996, 87, 1295.
13. Robinson, H.; Gao, Y. G.; McCrary, B. S.; Edmondson, S. P.; Shriver, J. W.; Wang, A. H. Nature 1998, 392, 202.
14. Gao, Y. G.; Su, S. Y.; Robinson, H.; Padmanabhan, S.; Lim, L.; McCrary, B. S.; Edmondson, S. P.; Shriver, J. W.; Wang, A. H. Nat Struct Biol 1998, 5, 782.
15. Jen-Jacobson, L.; Engler, L. E.; Jacobson, L. A. Structure Fold 2000, 8, 1015.
16. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. Nucleic Acids Res 2000, 28, 235.
17. Jorgensen, W. L.; Chandreskhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. J Chem Phys 1982, 79, 926.
18. Pardo, L.; Pastor, N.; Weinstein, H. Biophys J 1998, 75, 2411.
19. Flatters, D.; Young, M.; Beveridge, D. L.; Lavery, R. J Biomol Struct Dyn 1997, 14, 757.
20. Pastor, N.; Pardo, L.; Weinstein, H. Biophys J 1997, 73, 640.
21. Flatters, D.; Lavery, R. Biophys J 1998, 75, 372.
22. Guzikevich Guerstein, G.; Shakked, Z. Nat Struct Biol 1996, 3, 32.
23. Pardo, L.; Pastor, N.; Weinstein, H. Biophys J 1998, 74, 2191.
24. Pardo, L.; Campillo, M.; Bosch, D.; Pastor, N.; Weinstein, H. Biophys J 2000, 78, 1988.
25. Gelbin, A.; Schneider, B.; Clowney, L.; Hsieh, S.H.; Olson, W. K.; Berman, H. M. J Am Chem Soc 1996, 118, 519.
26. Cieplak, P.; Cheatham, T. E.; Kollman, P. A. J Am Chem Soc 1997, 119, 6722.
27. Cheatham, T. E.; Kollman, P. A. J Mol Biol 1996, 259, 434.
28. Norberg, J.; Nilsson, L. Biophys J 1995, 69, 2277.
29. Basham, B.; Schroth, G. P.; Ho, P. S. Proc Natl Acad Sci USA 1995, 92, 6464.
30. Dickerson, R. E. Methods Enzymol 1992, 211, 67.
31. Hunter, C. A. J Mol Biol 1993, 230, 1025.
32. Case, D. A.; Pearlman, D. A.; Caldwell, J. W.; Cheatham, T. E.; Ross, W. S.; Simmerling, C.; Darden, T. A.; Merz, K. M.; Stanton, R. B.; Cheng, A.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P.; Kollman, P. A. AMBER 5.0; University of California: San Francisco, 1995.
33. Cornell, W. D.; Cieplak, P.; Bayly, C.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179.
34. Rickaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. J Comput Phys 1977, 23, 327.
35. Bosch, D.; Foloppe, N.; Pastor, N.; Pardo, L.; Campillo, M. J Mol Struct (THEOCHEM) 2001, 537, 283.
36. Pearlman, D. A. J Chem Phys 1993, 98, 8946.
37. Pearlman, D. A. J Phys Chem 1994, 98, 1487.
38. Lavery, R.; Sklenar, H. J Biomol Struct Dyn 1989, 6, 655.
39. Lavery, R.; Sklenar, H. J Biomol Struct Dyn 1988, 6, 63.
40. SPSS 9.0; SPSS Inc.: Chicago, 2000.
41. Saenger, W. Principles of Nucleic Acid Structure; Springer Verlag: New York, 1984.
42. Schneider, B.; Neidle, S.; Berman, H. M. Biopolymers 1997, 42, 113.
43. Suzuki, M.; Yagi, N. Nucleic Acids Res 1995, 23, 2083.
44. Elcock, A. H.; McCammon, J. A. J Am Chem Soc 1996, 118, 3787.
45. Strauss, J. K.; Maher, L. J. Science 1994, 266, 1829.
46. Lebrun, A.; Shakked, Z.; Lavery, R. Proc Natl Acad Sci USA 1997, 94, 2993.
47. Travers, A. A. Nat Struct Biol 1995, 2, 615.
48. Lebrun, A.; Lavery, R. Biopolymers 1999, 49, 341.
49. Foloppe, N.; MacKerell, A. D. J. Biophys J 1999, 76, 3206.
50. Foloppe, N.; MacKerell, A. D. J. J Phys Chem B 2000, 102, 6669.
51. Spolar, R. S.; Record Jr, M. T. Science 1994, 263, 777.
52. Sharp, K. A. Biopolymers 1995, 36, 227.
53. Record Jr, M. T.; Zhang, W.; Anderson, C. F. Adv Protein Chem 1998, 51, 281.
54. Kraulis, J. J Appl Cryst 1991, 24, 946.