Molecular Dynamics Studies on Free and Bound Targets of the Bovine Papillomavirus Type I E2 Protein: The Protein Binding Effect on DNA and the Recognition Mechanism

D. Djuranovic and B. Hartmann
Laboratoire de Biochimie Théorique, CNRS UPR 9080, Institut de Biologie Physico-chimique, Paris, France

ABSTRACT Molecular dynamics simulations of a total duration of 30 ns in explicit solvent were carried out on the BPV-1-E2 protein complexed to a high-affinity DNA target containing the two hydrogen-bonded ACCG.CGGT half-sites separated by the noncontacted ACGT sequence. The analysis of the trajectories focuses on the DNA structure and on the dynamics. The data are compared to those issued from recent simulations made on three free targets that recognize E2 with different affinities. E2 does not drastically perturb the mechanic properties of the free DNA: the structural relationships between the BI/BII backbone substates and some helical parameters are preserved in the complex despite a severe slowing down of the phosphate group motions. The structures of both free and bound half-sites are very close to each other although the conformational space explored by these regions is narrowed when they are contacted by the protein. The enhanced plasticity found in the best free target spacers, mainly manifested through the backbone motions, allows a clear overlap between several free and bound global DNA features such as the base displacement. Furthermore, this flexibility is preserved in the complex. Our results support the hypothesis that E2 takes advantage of free predistorted structures that may minimize the DNA deformation cost. In addition, we observe that E2 is far from totally stiffening the DNA, suggesting that the entropic penalty inherent in the complex formation could be limited.

INTRODUCTION

E2 proteins are key factors for the transcriptional regulation and replication of the bovine papillomavirus-1 (BPV-1). The E2 proteins are involved in the regulation of transcription initiation from all the viral promoters and are required for the initiation of viral DNA replication through their interaction with the E1 protein (1). The DNA targets of E2 (E2BS) found in the viral genomes are the $ACCGN_4CGGT$ sequences (2,3). The highly conserved ACCG.CGGT half-sites are decisive for the recognition of E2 proteins. The crystal structures of two DNA sequences (N_4: ACGT and AATT) bound to the BPV-1-E2 protein (4,5) show that E2, folded into an unusual dimeric antiparallel β-barrel, inserts α-helices into the successive major grooves of the half-sites. The DNA is moderately distorted by E2: a 30° curvature toward the minor groove is observed, originating in negative rolls in the spacer combined with the positive rolls located at the CpG steps within the two symmetric half-sites. The specificity is attributed to direct interactions between conserved amino acids in each recognition helix and the half-sites, in which either bases or phosphate groups are contacted by often more than one amino acid side chain. The affinities of the recognition sites, varying by a factor of 11 for the naturally occurring DNA binding sites, are modulated by the four basepair spacer sequences (N_4) (6). The four basepairs of the spacers are not contacted by E2: this observation eliminates any explanation of the spacer effect in terms of direct contacts. Moreover, by exploring the range of E2 affinities for DNA sites in the BPV-1 (7,8), no clear spacer sequence signature was detected for the BPV-1-E2 high-affinity sites. To explore the relationship between DNA flexibility and affinity, binding measurements (9) were made on a series of targets that had intact or backbone-nicked spacers. Despite the fact that breaks in the central step backbone at best enhance the affinity for BPV-1-E2 by a factor of 2, it was concluded that BPV-1-E2 was sensitive to flexible spacers.

Recently, we carried out molecular dynamics simulations of 15 ns in explicit solvent on oligomers containing the CCAT, ACGT, and AAAC spacers, to contrast the behavior of very high- (CCAT, $K_{rel} = 3$), high- (ACGT, $K_{rel} = 1$), and low-affinity (AAAC, $K_{rel} = 0.3$) BPV-1-E2 targets (10). Our results showed that, whatever the sequence and the conformation of the spacers, the structure of the half-sites was close to those of the bound conformations extracted from the two BPV-1-E2/E2BS crystallographic complexes (4,5). The major difference between the three studied sequences comes from their spacer backbone dynamics. The AAAC spacer shows most phosphates in the BI conformation. In contrast, the high-affinity ACGT spacer is characterized by intense conformational BI/BII fluctuations in the central CpG.CpG backbones. Similarly, we observed an unusual stabilization of BII conformers on the CpA and GpG facing steps in the CCAT spacer. The occurrence of these BII-rich regions is relevant, since they are accompanied by several semilocal and global characteristics (null or negative spacer rolls, weak base displacement values, and a preference for bending toward the minor groove) that are turned to those of the bound conformation. It was thus concluded that the formation of a complex with the three sequences seemed yet possible, although with different efficiencies directly dependent on the plasticity

Submitted November 26, 2004, and accepted for publication June 8, 2005.

Address reprint requests to D. Djuranovic, Laboratoire de Biochimie Théorique, CNRS UPR 9080, Institut de Biologie Physico-chimique, Paris 75005, France. E-mail: djuranov@ibpc.fr.

0006-3495/05/10/2542/10 $2.00

doi: 10.1529/biophysj.104.057109

of the spacers. This malleability permits overlapping of the conformational spaces covered by the bound and free DNAs and could favor the achievement of binding without implying a large cost in DNA deforming. However, as underlined by Hegde's group (5,6), reducing the energy spent in deforming DNA could be counterbalanced by an entropic decrease if the flexibility totally disappears when E2 surrounds E2BS. Examining DNA flexibility in the complex, and, in particular, in the spacer one, is thus important to strengthen our hypothesis.

Overall, at this time, the mechanical and dynamical consequences of the protein binding on the DNA properties remain underexplored. In free B-DNA, the mechanical properties are reflected by marked structural relationships between the sugar puckers, the backbone conformations, and several helical parameters. An analysis of crystallographic data showed that the same trends exist in bound DNA (11), but data in solution are still lacking. Concerning the DNA dynamics, the common idea is that the DNA is stiffened by a surrounding protein. Nevertheless, some experimental data, such as the range of chemical shifts of the DNA phosphate group in NMR, enlarged in the presence of a protein (12,13) could be interpreted as an enhancement of the flexibility of the bound DNA backbone. Although several molecular dynamics (MD) studies have been devoted to protein/DNA systems, they often focused on the behavior of the protein and on the interaction interface (14–17). Only one article (18) on a covalent complex between a DNA and the human topoisomerase I reported low base fluctuations in the regions where direct DNA protein contacts were present. Moreover, to investigate the change in DNA flexibility induced by a protein, both free and bound DNA should obviously be studied in parallel. Some MD studies (19–21) consider the two forms of the relevant DNA, but, again, they do not explicitly deal with the effect of the protein on the DNA motion. Finally, the DNA behavior was examined by means of a 50-bp fragment containing the OR1/OR2 sites, free and complexed with the phage 434 operators (22), and, although the size of the system limited the trajectory duration to 1 ns, it was concluded that the OR1 dynamics were enhanced by the operator whereas the OR2 plasticity slightly decreased. Thus, the DNA freezing within a complex is clearly not a trivial problem.

Here, we perform MD simulations in explicit solvent on the complex composed of BPV-1-E2 and the high-affinity DNA target CGACCGACGTCGGTCG, the sequence used in one (4) of the two x-ray complexes (4,5). The total duration of the simulations is 30 ns, i.e., much longer than what is usually presented ($\leq$5 ns, except in the simulations by Liedl's group (23), where the duration was 10 ns). Indeed, we expect a slowdown of DNA motion in regard to those in free oligomers, correctly sampled by 15-ns trajectories (10,24). The MD structure of the complex is at first compared with the crystallographic one, and then with the free DNA structures previously studied under the same conditions (10). Even if the simulated complex remains on average very close to the crystallographic structure, the MD approach provides new insight into the bound DNA dynamics. Our results highlight that E2 does not perturb the relationship between the backbone substates and the twist-and-roll helical parameters. In contrast, E2 modifies the dynamical behavior of its target with regard to the free state, freezing or preserving various motions according to the type of parameter involved and the part of the oligomer. These findings strengthen the recognition mechanism we proposed in our previous studies, and, furthermore, suggest which DNA structural elements could be the most important in the recognition of the high-affinity sites by major groove binding proteins.

METHODS

Molecular dynamics

We performed two simulations of 15 ns using the program AMBER 7.0 (25), with the Parm98 force field (26) on the E2BS/E2 complex. The DNA-protein complex was built using the DNA/protein sequences and coordinates of the high-resolution crystal structure of the complex (4) (Protein Data Bank code 2BOP; Nucleic Acid Database code pdv001), consisting of the 16-bp DNA high-affinity target fragment 5′-CGACCGACGTCGGTCG-3′ bound to the DNA binding domain of BPV-1-E2.

The DNA and the protein carry a total charge of -30 e and $+10$ e, respectively; 30 Na^+ (one Na^+ for each phosphate) and 10 Cl^- neutralized the system. The complexes were solvated in a 12-Å water box shell in all directions (9603 TIP3P water molecules) in a truncated octahedral box. The water molecules close to the DNA/protein interface within the crystallographic structures were retained on our starting point.

After 1000 cycles of energy minimization, each minimized system was heated to 300 K, rescaling the velocities as necessary and coupling the system to a heat bath using the Berendsen algorithm (27). During each of these phases, harmonic restraints were imposed on the atomic positions of the complex and then slowly relaxed over periods of 50 ps until a free system was achieved. In the first MD simulation, the harmonic restraints were classically relaxed after 300 ps. In the second simulation, this phase was completed by a period of 1 ns in which the DNA backbone angles ε and ζ were restrained around their crystallographic values, to ensure the best stabilization of these conformations. The simulations were then performed at constant temperature and pressure using the Berendsen algorithm. The bond lengths involving hydrogen atoms were constrained using SHAKE (28), which enabled an integration step of 2 fs. Long-range electrostatic interactions were treated using the particle mesh Ewald approach (29,30) with a 9-Å direct space cutoff, a direct sum tolerance criterion of 10^{-5}, and a reciprocal space charge grid spacing of $\sim$1 Å. During the production phase, translations and rotations of the complex were removed every 100 steps.

DNA analysis

In all trajectory analyses, to ensure a good equilibration the first nanosecond of the production phase was not taken into account,. The results are presented for structures extracted from the trajectories at the frequency of one snapshot every picosecond.

The pairwise comparison of DNA and protein structures is quantified by root mean-square deviations (RMSD) of heavy-atom Cartesian coordinates. For the protein, the RMSD is performed on the backbone and the amino acid side chains. The MD snapshots were compared with the 2BOP (4) crystallographic complex and the x-ray complex solved by Hegde et al. (5). In addition, both the simulated protein and DNA were compared to their free counterparts: the crystallographic structure of E2 (31) (Protein Data Bank code 1JJH), and three free DNA targets previously studied (10) by molecular dynamics carried out under the same conditions of temperature and Na^+

concentration. These three targets consist of oligomers containing the half-sites ACCG.CGGA and different spacers: 1), CCAT (the ''CA'' sequence); 2), ACGT (the ''CG'' sequence, identical to the one we used in the complex); and 3), AAAC (the ''AA'' sequence).

To describe sugar conformations, the pseudorotational phase angle P is divided into two categories, ''north'' ($-60° < P < 50°$) and ''southeast'' ($50° < P < 220°$). The BI conformation is characterized by ε: t, ζ: $g-$, and (ε-ζ) around $-90°$ and the BII conformation is characterized by ε: $g-$, ζ: t, and (ε-ζ) around $+120°$ (12,32,33). The distribution of (ε-ζ) values shows that 50° is the point that best differentiates the BI and BII populations. Thus, we define the BI state by (ε-ζ) $< +50°$ and the BII state by (ε-ζ) $> +50°$.

Analyses of DNA structures were carried out using CURVES (34,35), which calculates the optimal helical axis and a complete set of conformational and helicoidal parameters. Our analyses were made in terms of both local and global parameters; however, since the two analyses gave quasiidentical values, only the global values are reported. The helical parameters, which obey the Cambridge convention for DNA conformation (36,37), are divided into two categories: the interbase parameters defined between one nucleotide and its nearest neighbor, and the base parameters that refer to the nucleotide position with respect to the helix axis. The standard deviations are given in brackets next to the corresponding mean values.

RESULTS AND DISCUSSION

The protein amino acids are numbered following the crystallographic sequence (two identical chains, residues 326–410). The numbering of the oligomer bases is

$$5'\text{-}C_1\ G_2\ \underline{A_3\,C_4\ C_5\ G_6}\,N_7\ N_8\ N_9\ N_{10}\underline{C_{11}G_{12}G_{13}T_{14}}C_{15}G_{16}\text{-}3'$$
$$3'\text{-}G_{32}C_{31}\underline{T_{30}G_{29}G_{28}C_{27}}N_{26}N_{25}N_{24}N_{23}\underline{G_{22}C_{21}C_{20}A_{19}}G_{18}C_{17}\text{-}5'.$$

The spacer corresponds to the N_4 part, and the conserved half-sites are underlined.

General characteristics

We performed two simulations, the starting point being the crystallographic structure and the difference lying in the way the constraints were applied to the DNA during the equilibrium phase. For the two MD simulations, the RMSDs presented below are all very stable after ~100 ps of the production phase, indicating the convergence of the trajectories. We first analyzed the two trajectories separately. As the results (in terms of either RMSD or standard deviations) were identical, from now on, we present the analysis made on the two simulations taken together.

The DNA-protein complex structures extracted from the trajectories remain always close to the structure of the 2BOP crystallographic complex, with an average RMSD value of 1.5 ± 0.2 Å. Keeping in mind that these RMSDs are calculated on all the heavy atoms, comprising those of the side chains, and taking into account the large size of the simulated system, the values we obtain are very small.

Considering the protein, comparisons are made between the MD snapshots and the two crystallographic structures of E2, without the DNA (1JJH) or extracted from the 2BOP complex (Table 1). The MD bound E2 is almost as close to the x-ray form in the E2/E2BS complex (average RMSD value of 1.4 ± 0.2 Å) as the free E2 x-ray conformation (average RMSD value of 1.7 ± 0.2 Å), suggesting that the DNA target binding induces only weak perturbations on the protein. Furthermore, the RMSD is perfectly stable during the trajectories, as indicated by the small values of the standard deviations, and thus, any motion of subdomains can be excluded. This result is in line with the comparison made between the free and bound E2 crystallographic structures, which gives an RMSD value of 1.2 Å. Analyzing E2 by regions, we find that the part of E2 containing most of the amino acids involved in hydrogen bonds with DNA (amino acids 336–344) are all around the average (1.7 Å), whereas the largest RMSD values (2.55 Å) are found for the β-turn (amino acids 363–370).

TABLE 1 RMSD values (Å) calculated on the heavy atoms of the backbone and side chains of E2 and on the heavy atoms of the whole oligomers

	X-ray free E2 (1JJH)	X-ray bound E2 (2BOP)	
MD bound E2	1.7 (0.2)	1.4 (0.2)	
	Arnott B-DNA	Arnott A-DNA	X-ray bound DNA (2BOP)
MD free CG	3.1 (0.5)	5.7 (0.5)	3.5 (0.7)
MD bound CG	3.05 (0.2)	7.2 (0.2)	1.4 (0.2)

Table 1 shows the RMSD computed on the CG sequence. The bound and free MD DNA structures are clearly in B-form (average RMSD value ~3 Å), the RMSD with respect to the A-form lying between 6 and 7 Å. The bound DNA structures in crystal and MD are close to each other (average RMSD of 1.4 Å). In contrast, the free CG structures are notably different from the bound DNA form (average RMSDs of 3.5 Å with the 2BOP DNA and 3.2 Å with the MD bound DNA), showing that the bound DNA conformation is unstable for a free DNA in solution. Comparing the RMSD values between the free and bound protein on one hand and the free and bound DNA on the other, we conclude that, upon DNA-E2 binding, E2 causes larger deformations to the DNA than the DNA does to E2.

Finally, the atomic fluctuations are calculated for the nucleotides in free and bound DNA. The bound DNA exhibits weaker fluctuation values (1.1 Å) than the free DNA (1.5 Å). Thus, the presence of E2 seems to reduce the global flexibility of the DNA molecule after protein binding. Nevertheless, the decrease in fluctuations induced by E2 is more pronounced for the half-sites (0.5 Å) than for the spacer (0.3 Å).

Interaction between E2 and DNA

The intermolecular hydrogen bonds observed in the MD simulations and in the 2BOP crystal structure are summarized in Table 2 for one half-site and one chain of E2, both the experimental and simulated complexes being symmetric. Four DNA base atoms and six oxygens belonging to the phosphate groups are contacted by E2 in both the MD and the crystallographic structures. The hydrogen bonds are stable along the entire simulations, apart from one of the bidentate

TABLE 2 Hydrogen bonds observed between one monomer of BPV-1-E2 and one half-site of DNA, in the crystals and the MD structures

BPV1-E2		DNA			*N*
X-ray and MD			X-ray	MD	MD
ASN 336	ND2	A_3	N7	N7	100
ASN 336	OD1	C_4	N4	N4	100
LYS 339	NZ	G_{28}	N7	**O6**	70
LYS 339	NZ	G_{29}	O6	O6	100
ARG 370	NH2	C_{27}	O3′	O3′	100
GLN 337	NE2	G_2pA_3	O2P	O2P	85
ARG 342	NH2	$C_{27}pG_{28}$	O1P	O1P	100
ARG 344	NH2	C_1pG_2	O1P	**O3′**	100
THR 359	N and OG1	$T_{26}pC_{27}$	O2P and O1P	O2P and O1P	100
ARG 370	NH1	$C_{27}pG_{28}$	O2P	O2P	100

The DNA atoms that differ between the x-ray and the MD structures are indicated in bold type. *N* is the occurrence (%) of hydrogen bonds in simulations.

contacts between lysine 339 and the adjacent guanines 28 and 29 (and their symmetric 12 and 13). The hydrogen bond with guanine 29, occurring at 100% of the trajectory, is more stable than the one with guanine 28, present at 70%. Furthermore, these contacts occur with the O6 atom of the guanine and not with the N7 atom, as observed in the crystal. Similarly, the crystallographic contact between arginine 344 and O1P of the CpG outermost steps is broken at the beginning of the equilibration in favor of O3′ of the same step. Despite these minor modifications, the interaction pattern of the MD complex remains strikingly consistent with the experimental data.

Sugar puckering

In the presence of E2, a large majority of sugars (20 out of a total of 24) appears almost locked in S-form, the percentage of N-form for each sugar being 2% in the complex, against 5% in the free DNAs. Both the pucker standard deviations and the sugar atomic fluctuations decrease when E2 surrounds the DNA. Fig. 1 *A* illustrates the behavior of the C8 cytosine sugar, representative of the other 18 sugars in S-form: observed in a normal S-form range in the free oligomer, the pucker remains in a restricted C2′-endo conformation in the bound states. As in the crystallographic complex, N-form sugars are observed on the symmetric cytosines located within the contacted half-sites in our trajectories. The sugars of cytosines 5 (Fig. 1 *B*) and 21 oscillate between the N-form (25%) and the S-form; those of cytosines 11 (Fig. 1 *C*) and 27, wedged in the 3′ and 5′ sides by two phosphate group/arginine side chain salt bridges, are trapped during all the trajectories in the C3′-endo conformation. In the free MD oligomer, these four sugars either populate an exceptionally large S-form range (Fig. 1 *B*) or show a significant tendency to adopt the N-form (~15%; Fig. 1 *C*). Thus, the emergence or the stabilization of the N-form in the complex clearly takes advantage of the sugar intrinsic flexibility noticed in the free oligomer. Finally, it should be noted that these north sugars

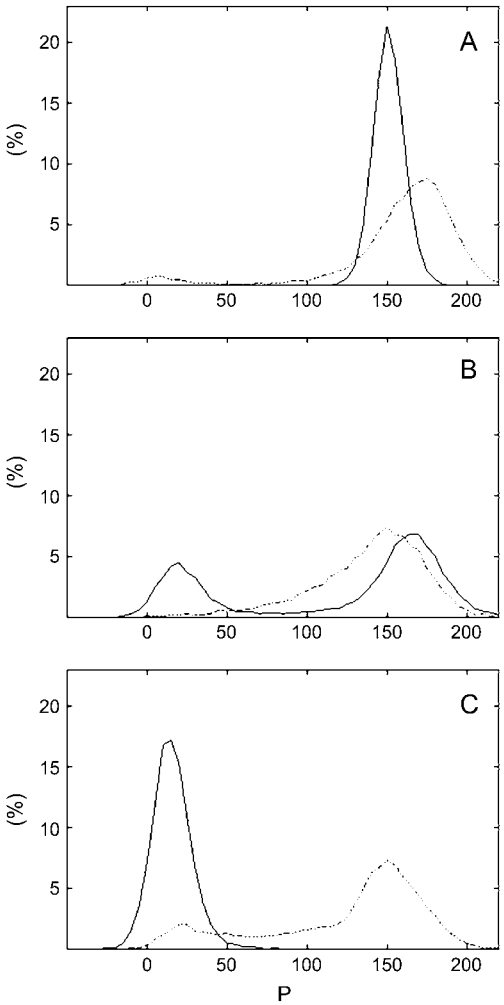

FIGURE 1 Sugar pucker distribution (%) by steps of 5° in the MD free (*dashed line*) and bound (*solid line*) DNA calculated for cytosines 8 (*A*), 5 (*B*), and 11 (*C*).

belong to the 5′-$Y_{\text{N-sugar}}$-$R_{\text{S-sugar}}$-3′ pattern, the so-called ''sugar switch'' found in severely kinked steps of DNA bound to proteins (46) and, in our case, associated to marked positive rolls, as we will see.

Backbone structure and dynamics

Although some unusual $\alpha/\beta/\gamma$ conformations are found in DNA surrounded by a protein (11), neither in the x-ray nor in the MD complexes of BPV-1-E2/E2BS do the phosphate groups exhibit such alternative states. Consequently, we will focus on the backbone angles ε ($C_{4'}$-$C_{3'}$-$O_{3'}$-*P*) and ζ ($C_{3'}$-$O_{3'}$-*P*-$O_{5'}$), defining the two phosphate conformations BI and BII. The effect of E2 on the BI/BII dynamics differs according to the parts of the DNA we consider: in regard to their free counterparts, the phosphate groups of the spacer and the GpA junction lose 0.4 Å in atomic fluctuations, whereas in the other regions they decrease by 1.2 Å. In each CGACCGA. TCGGTCG symmetric region, the eight phosphate groups

that are observed in BI conformation in the free MD DNA remain in this state when E2 is present. On the BII-rich phosphate groups in the free DNA (**CpGpA**CC**GpA**.TC**GpG**TCG, with BII-rich groups indicated in bold print), the percentage of BII conformers decreases; it even totally disappears in the case of the GpG steps (Fig. 2 *A–C*). In parallel, we note that some bound steps within or next to the half-sites (Figs. 2, *B* and *C*, and 3, *A* and *B*) accommodate larger BI and BII ranges than in the free DNA. Finally, the typical BII state lifetime of ~100 ps in the free DNA notably increases, sometimes reaching several nanoseconds (2–6 ns). Thus, E2 tends to maintain or stabilize the BI state, in accordance with the emergence of north sugars, known to destabilize the BII phosphates. Moreover, E2 is able to modify the backbone dynamics in terms of both substate lifetimes and the (ε-ζ) distribution. This effect is not specific to E2, as it is in line with earlier observations from MD simulations (22) and from analysis of crystallographic DNA/protein structures (11).

In the ACGT spacer, the situation is somewhat different. Step by step, the proportion of the BII conformer in the bound spacer is equal to or greater than in the free spacer. The ApC steps are in BI conformation in all cases. The propensity to adopt BII conformers remains identical for free and bound CpG backbones, whereas it notably increases for the bound GpT phosphates. Furthermore, the range of (ε-ζ) values for both conformers is not enlarged in the bound steps, as shown in Fig. 3 *C* for the CpG central step. Thus, the E2 effect seems to be inverted with respect to what is observed on the extended half-sites. However, as for the phosphates located in the CGACCGA.TCGGTCG parts of the target, the BII state lifetime increases. This effect is particularly evident in the case of the central CpG.CpG phosphates which, in the free state, are characterized by noncorrelated BI/BII transitions occurring at high frequency and by BII lifetimes shorter than usual (Fig. 2 *D*), whereas in the bound DNA the BII state can be continuously populated for 6 ns (Fig. 2, *E* and *F*).

Thus, the bound spacers appear to be constituted of BII-rich steps. Only three conformational combinations of the six phosphates emerge from the trajectories, as shown in Table 3. The first one (C1) consists of one of the two GpT phosphates in the BII conformation. The other two, occurring 73% of the time, correspond to the simultaneous presence of two facing-strand BII conformers (C2), either on one GpT and one CpG step (C2a) or on the two symmetric GpT steps (C2b). To ensure a good sampling of the BII-rich spacers, we performed dynamics in which all the other possible BII combinations were constrained for 1 ns, then relaxed. None of the combinations is stable, and all of them immediately relax toward one of the three observed patterns. In particular, two adjacent BII phosphates on the same strand found in the 2BOP spacer are clearly unstable in solution, whatever their locations.

Let us now compare these spacer substates with the ones found for the three E2BS previously studied. We recall that these sequences differ by the base composition in the spacer and their affinities for E2: apart from the free CG sequence already presented here, the CCAT-containing site (CA sequence) is a very high-affinity target, whereas the AAAC (AA sequence) is one of the lowest. As shown in Table 3, the AAAC spacer is lacking in BII conformers, whereas the spacers of CG and CA sequences spend 66% and 73% of the trajectories, respectively, in C1 or C2 substates. Furthermore, the CA spacer exactly mimics two (C1b and C2a) among the three backbone combinations reported for the bound DNA. Thus, a qualitative relationship between the DNA backbone behavior and the affinities seems to emerge from this analysis.

Finally, it should be noted that the classification based on the spacer backbone substates is as relevant as any clustering made on the whole oligomer, as the other parts of the sequence do not exhibit such variability in phosphate groups during the trajectories, either in the bound or the free DNAs. Furthermore, keeping in mind that the various phosphate conformations are known to be associated with distinct DNA

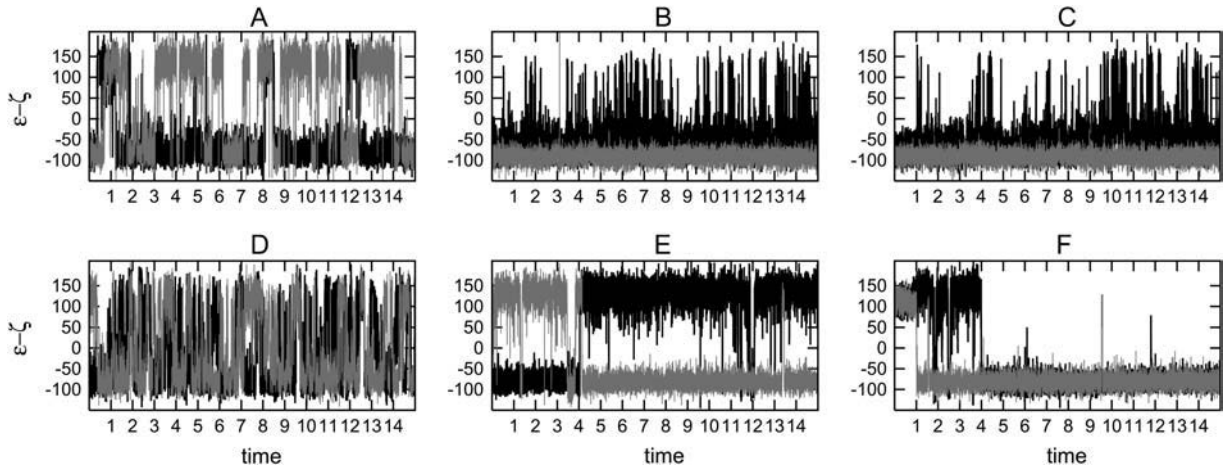

FIGURE 2 (ε-ζ) (°) fluctuations during the simulations (time in ns) for 1), the CpC (*black*). GpG (*shaded*) backbones, in the free (*A*) and bound (*B* and *C*) oligomer, and 2), the facing CpG.CpG backbones (strand 1, *black*; strand 2, *shaded*) in the free (*D*) and bound (*E* and *F*) oligomer.

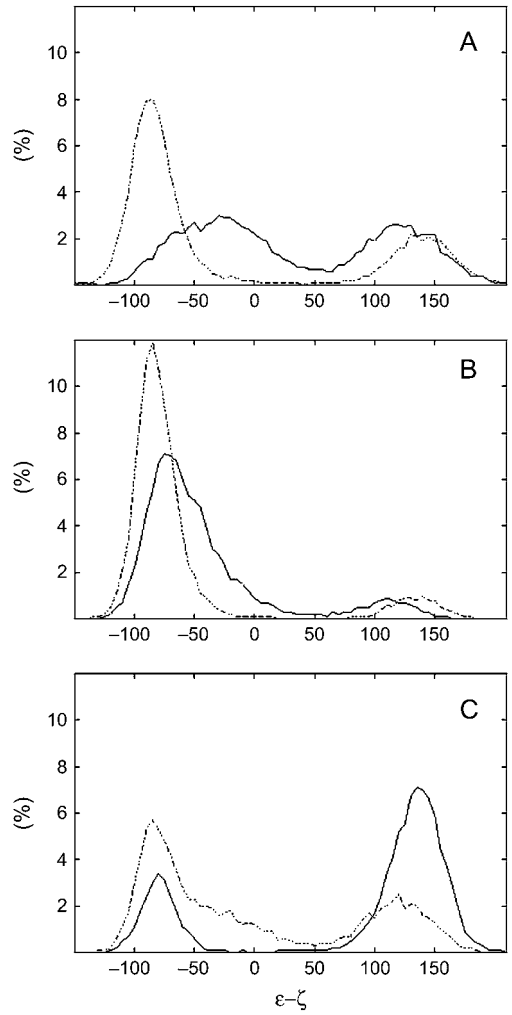

FIGURE 3 (ε-ζ) (°) distribution (%) by steps of 5° in the MD free (*dashed line*) and bound (*solid line*) DNA calculated on the GpA (*A*), CpC half-site phosphate groups (*B*), and on the central CpG spacer steps (*C*).

helical characteristics, we expect that this classification will contribute to refining our knowledge of the conformational space explored by the DNA sequences.

Helical parameters and curvature

For all the parameters describing the double helix, a very good fit is observed step by step with the experimental data, as shown in Fig. 4, *A* and *B*, for the roll and twist parameters. Nevertheless, a fine comparison shows that the X-displacement (X-disp) (Fig. 4 *C*) and the spacer rolls and twists are closer to the structure of Hegde et al. (5) than to the 2BOP structure.

The three spacer roll standard deviations are larger in the spacer (6.7°) than in the half-sites (4.7°). A similar observation holds for the CpG twist (6.3° vs. 3.6° on average for the other steps), whereas the standard deviations of the other helical parameters are invariant along the whole sequence. This twist/roll spacer variability is in correspondence with the backbone flexibility. Both experimental and theoretical data have highlighted that, on free B-DNA, a BII phosphate group is accompanied by local disturbances (high twist and negative roll) and by semilocal variations: the bases enclosing the BII phosphate and their neighbors are displaced toward the major groove (11,38–40). Selecting the structures according to the spacer phosphate combinations defined in Table 3, we find that the spacers belonging to different substates are characterized by different roll/twist patterns, whereas the helical interbasepair parameters of the CGACC-GA.TCGGTCG parts are totally insensitive to selection. The largest differences are observed for the central CpG step, which exhibits twist and roll values of 26° and +5° or 37.8° and −3.5°, according to the BI or BII phosphate substate, respectively. As a consequence, the MD simulation shows that the 30° curvature accommodates either alternated (negative/positive/negative) or uniformly moderate (−5°) spacer rolls. Finally, a BII-rich spacer should help E2 to stabilize the base location in the major groove (X-disp standard deviations of 0.3 Å in the bound state against 0.5 Å in the free state).

The analysis of the standard deviations confirms the flexibility of the bound spacer, in which both twist and roll standard deviations increase by 2.5 and 1.1°, respectively, compared to their free counterparts. In the rest of the bound oligomer, the twist and roll standard deviations are reduced by 2.5° and 1.3°, respectively. To take into account the whole base motion, atomic fluctuations were measured on the base's

TABLE 3 Different spacer phosphate configurations and their percentages, as observed in MD trajectories and crystallographic bound CG spacer

	C0	C1	C2	C4
X-ray bound DNA				A•C○G○T T○G○C•A
MD bound DNA		A•C•G•T * T○G•C•A 27%	A•C○G•T * T○G•C•A a: 60% A•C•G○T T○G•C•A b: 13%	
MD free DNA (CG seq.)	A•C•G•T T•G•C•A 34%	A•C○G•T * T•G•C•A 54%	A•C○G•T T•G○C•A 12%	
MD free DNA (CA seq.)	C•C•A•T G•G•T•A 20%	C•C○A•T G•G•T•A a: 7% C•C•A•T G○G•T•A b: 34%	C•C○A•T G○G•T•A 39%	
MD free DNA (AA seq.)	A•A•A•C T•T•T•G 66%	A•A○A•C T•T•T•G 34%		

(•), BI phosphate; (○), BII phosphate. Note that, due to the palindromic nature of the ACGT sequence, the configurations noted with an asterisk correspond to two symmetric structures: A•C•G○T.A•C•G•T is obviously equivalent to T○G•C•A.T•G•C•A.

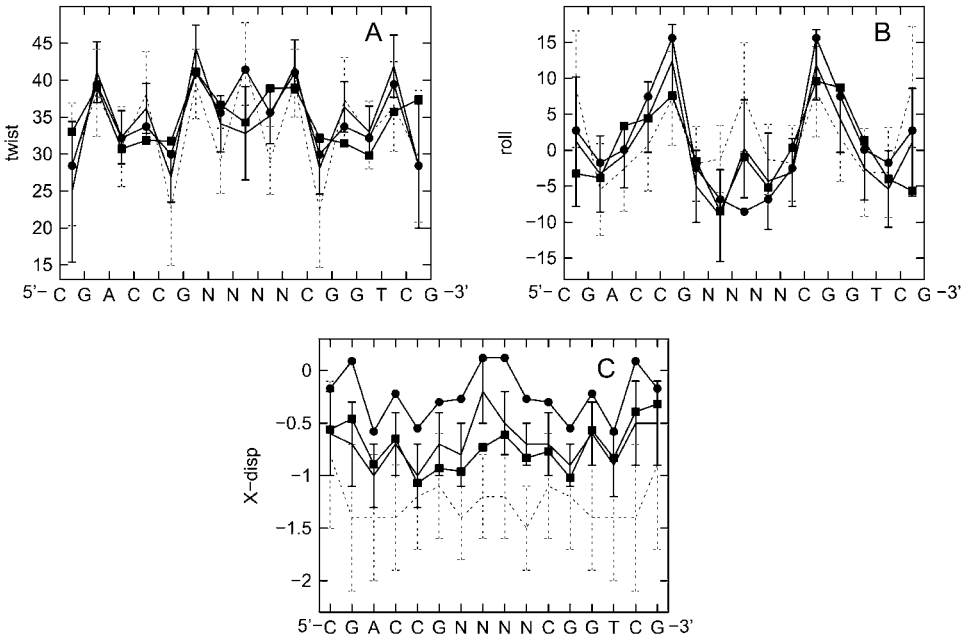

FIGURE 4 Average and standard deviation values of (*A*) the twist (°), (*B*) the roll (°), and (*C*) the X-disp (Å) along the MD free (*dashed line*) and bound (*solid line*) CG sequences (N4 = ACGT). The MD DNAs are compared to the crystallographic bound structures 2BOP: (●), N4 = ACGT, and the structure of Hegde's group: (■), N4 = AATT.

heavy atoms. Comparing the values obtained for the free and bound CG sequences (Table 4) shows that the presence of E2 reduces the fluctuations to ~0.6 Å on the half-site bases. In these DNA parts, only the two guanines in the 5′ of the spacer conserve a noticeable flexibility. E2 also acts on the spacer base fluctuations, which, however, maintain the greatest values in both free and bound states. The malleability of the spacer and the stability of the ACCG part in the presence of E2 are illustrated on a selection of structures in Fig. 5.

The free CG, CA, and AA MD snapshots were previously compared to the crystal DNA structures. The information gleaned from the MD of the complex, in particular the multiplicity of the spacer helical parameter patterns, justifies a new comparison between the bound and free structures extracted from the trajectories. We deduce that, whatever the spacer sequence, the half-site interbasepair parameters have almost identical average values in both the free and the bound forms, as shown for the twist and roll of the CG free sequence in Fig. 4, *A* and *B*. Taking into account the standard deviations, the free and the bound spacer interbase parameters overlap in all cases. Nevertheless, step by step along the whole oligomer, the rolls of the free CA and CG C2 substate structures are very close to the bound ones (Δ of 3.4° on average). Concerning the curvature, the analysis of the 16-bp oligomers shows that, even if the intensity of the bound DNA curvature is never reproduced in the free DNA, the three free sequences are able to adopt temporarily the complex curvature direction. In line with recent curvature measurements (41), the AA sequence presents the best propensity to bend toward the minor groove. The CG and CA sequences are not intrinsically curved on average, as experimentally shown for the CG sequence (41,42). However, the presence of BII-rich spacers (C2 substates) again helps these two sequences to temporally prebend toward the minor groove.

TABLE 4 Atomic fluctuations (Å) calculated and averaged on the heavy atoms of the bases in free and bound CG sequence trajectories

	Base	Free	Bound	Base	Free	Bound
Half-sites	$A_{3,19}$	1.3	0.5	$T_{30,14}$	1.1	0.6
	$C_{4,20}$	1.1	0.55	$G_{29,13}$	0.9	0.5
	$C_{5,21}$	1.2	0.6	$G_{28,12}$	1.2	0.55
	$G_{6,22}$	1.3	0.9	$C_{27,11}$	1.3	0.5
Spacer	$A_{7,23}$	1.4	0.7			
	$C_{8,24}$	1.55	0.9			
	$G_{9,25}$	1.45	0.8			
	$T_{10,26}$	1.3	0.6			

The ACGT spacer is a palindromic sequence; thus, the first and second strands have identical atomic fluctuations.

The X-disp appears to be the best parameter to discriminate the free sequences. Along the whole oligomer, the average base displacement is far from the values found in the crystallographic complexes, as shown for the CG free sequence in Fig. 4 *C*. Although the average X-disp is approximately the same (−1.6/−1.5 Å) for the three free sequences, the CG and the CA in their BII-rich substates exhibit base displacements around −1.15 Å, not so different from the bound value (−0.8 Å). This property is not found in the AA sequence, as it lacks any BII-rich spacer. This parameter is known to be related to the major groove depth (11,39). Here, in the MD structures, the average major groove depth is 4.3 Å for the bound CG sequence, 6.35 Å for the free AA sequence, and 5.1 and 4.9 Å for the BII-rich substates of the free CG and CA sequences, respectively (Fig. 6). The X-disp parameter was previously invoked as an essential component favoring the binding of the transcription factor NF-κB to its native DNA site, in regard to

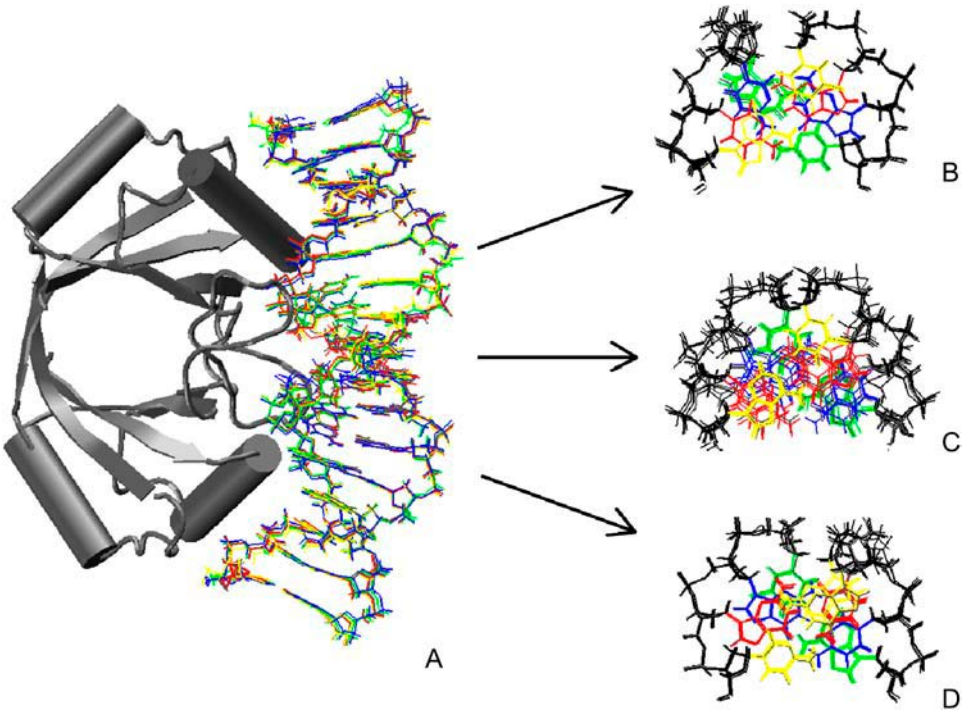

FIGURE 5 (*A*) Side view of the superimposed average structures of the complex coming from the four DNA substates (C0, C1a, C1b, and C2). For the same bound DNA structures, top views are shown of the half-sites (*B* and *D*) and of the spacer (*C*).

mutated sequences (43–45). A study of a DNA complexed to the even-skipped transcriptional repressor homeodomain (21) leads to the same conclusion. In these two cases, the proteins belong to the group of DNA major groove binding proteins, as E2. In this line of ideas, Tolstorukov et al (46) recently found that the complexes of DNA with minor groove binding proteins are characterized by interactions between hydrophobic amino acids and short clusters of A-like nucleotides (i.e., bases associated with north sugars). Although not mentioned by these authors, it is likely that a supplementary effect of these A-like clusters is to favor the typical A-DNA base displacement toward the minor groove. Thus, the propensity of the basepairs to shift toward either the major or the minor groove could be important at two levels of the recognition process involving the major and minor groove binding proteins, respectively: they could facilitate external access to the base sequence in the early stage of site discrimination and minimize the cost of DNA deformation during the locking mechanism.

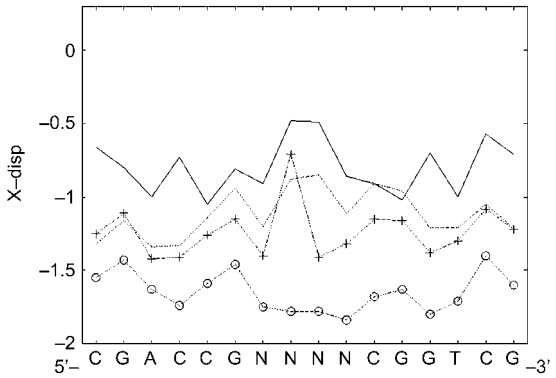

FIGURE 6 Average values of the X-disp (Å) extracted from the trajectories of the bound CG sequence (*solid line*, N4 = ACGT), the free C2 substates of the CG (*dashed line*, N4 = ACGT), the free AA (*dashed line* with *open circles*, N4 = AAAC), and the free C2 substates of the CA (*dashed line* with *crosses*, N4 = CCAT) sequences.

CONCLUSION

We recently presented modeling studies on three BPV-1-E2 targets that differed in their spacer sequences (the AA, CA, and CG sequences), which modulated the affinity between the DNA and E2. Here, to explore the effect of E2 on DNA and to better understand the role of DNA in the recognition mechanism, we have carried out molecular dynamics simulations of the complex constituted of E2 and one of the high-affinity sequences (the CG sequence). Our analysis focused on the DNA behavior, comparing the free and bound sequences.

Our first aim is to describe the effect of E2 on DNA. Overall, the general trends in the relationships between sugars, backbone angles, and selected helical parameters are found to be similar, little sensitive to the free and the bound DNA states: the main intrinsic mechanical properties of DNA are preserved in the complex. The effect of E2 is to diminish globally the DNA motion, to impose and lock the base displacements and the curvature. In the contacted half-sites, E2 favors some north sugar puckers; it freezes BI backbone substates and the twist/roll pattern. Within the noncontacted DNA part,

the numerous BI/BII transitions observed in the free CG and CA oligomers are preserved, although the BII lifetime strongly increases in the complex. In the same way, the twist/roll fluctuations of this region remain clearly higher than those of the half-sites. So we distinguish the same two regions in both the free and bound states, i.e., the relatively rigid cGACCga.TCGGTC region, overlapping the major part of the extended half-sites, and the malleable GACGtc.GACGtc region (bases in upper case), containing the last guanines of the half-sites and the first three bases of the noncontacted spacer.

The binding affinity of a protein for various binding sites reflects an energy balance between the favorable free energy release by the protein/DNA interaction and the unfavorable cost of altering the conformation of the binding site and decreasing the flexibility. To quantify the real energy balance of the E2/E2BS recognition, further investigations are required, in particular simulations of the complexes containing the CA and AA sequences. These trajectories, now under study, will supply the data necessary for the calculation of both relative deformation and entropic costs. Thus, we would like to underline that, at this stage, we can only make hypotheses on the driving forces that govern this interaction. Comparing the conformational spaces explored during the trajectories by three free E2 targets and the bound CG sequence, we find that all the free sequences contain well-defined prestructured half-sites. The best free E2BS contain ACGT and CCAT spacers proposed to enhance backbone dynamics and, thanks to this flexibility, temporarily but significantly adopt several shapes (in particular the base displacement) that approximate the bound DNA conformation. These CCAT and ACGT containing sequences do not spend much time in a bound-like conformation and it seems improbable that the protein E2 directly recognizes this transient state conformation. Nevertheless, these half-site and spacer properties mimicking the bound structure should limit the energy penalties for DNA deformation. Concerning the entropic cost that a severe decrease in flexibility would involve, the noticeable spacer malleability preserved in the complex suggests that this cost could be nonprohibitive.

The direct readout appears to be common to the three exhaustively studied E2 proteins of HPV-18, HPV-16, and BPV-1 strains. In contrast, the mechanisms of noncontacted sequence recognition differ (for a review, see Hegde (6)). In the HPV-18-E2 case, the predistorted spacers are preferred but not strictly essential: this protein is charged all along the DNA interaction surface, and can exert electrostatic forces upon the DNA phosphate backbone to facilitate bending. Nevertheless, HPV-18-E2 prefers an A:T-rich spacer, essentially because of the electrostatic complementarity between the E2 cationic charges and the enhanced negative potentials in the A:T minor groove. This favorable indirect interaction should play a role in both the selection of the target and the locking stage. Although no structural data are yet available on DNA complexes of HPV-16-E2, this protein, lacking any accumulation of positive charges, clearly requires appropriately prebent A:T-rich spacers (6,41). BPV-1-E2 is intermediate between HPV-18 and HPV-16 E2. On one hand, this protein is capable to actively bend the DNA via phosphate groups/positively charged side chain interactions. On the other hand, in the absence of any cationic zone close to the spacer, BPV-1-E2 does not meet the favorable electrostatic feature found between HPV-18-E2 and A:T-rich spacers. Our results suggest that an equivalent favorable energy could originate in an unusual spacer flexibility in the BPV-1 complex. In summary, the E2/E2BS systems illustrate the various and subtle strategies used to select the desired targets and to modulate the affinities among nonspecific DNA and multiple but similar binding site sequences.

The authors thank Dave Beveridge and Suzie Byun for very productive discussions, and Christophe Oguey for helpful advice. Rashmi Hegde is gratefully acknowledged for providing the crystallographic coordinates of the complex.

REFERENCES

1. Steger, G., and S. Corbach. 1997. Dose-dependent regulation of the early promoter of human papillomavirus type 18 by the viral E2 protein. *J. Virol.* 71:50–58.
2. Dostatni, N., F. Thierry, and M. Yaniv. 1988. A dimer of BPV-1 E2 containing a protease resistant core interacts with its DNA target. *EMBO J.* 7:3807–3816.
3. McBride, A. A., J. C. Byrne, and P. M. Howley. 1989. E2 polypeptides encoded by bovine papillomavirus type 1 form dimers through the common carboxyl-terminal domain: transactivation is mediated by the conserved amino-terminal domain. *Proc. Natl. Acad. Sci. USA.* 86: 510–514.
4. Hegde, R. S., S. R. Grossman, L. A. Laimins, and P. B. Sigler. 1992. Crystal structure at 1.7Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature.* 359:505–512.
5. Kim, S. S., J. K. Tam, A. F. Wang, and R. S. Hegde. 2000. The structural basis of DNA target discrimination by papillomavirus E2 proteins. *J. Biol. Chem.* 275:31245–31254.
6. Hegde, R. S. 2002. The papillomavirus E2 proteins: structure, function, and biology. *Annu. Rev. Biophys. Biomol. Struct.* 31:343–360.
7. Li, R., J. Knight, G. Bream, A. Stenlund, and M. Botchan. 1989. Specific recognition nucleotides and their DNA context determine the affinity of E2 protein for 17 binding sites in the BPV-1 genome. *Genes Dev.* 3:510–526.
8. Bedrosian, C. L., and D. Bastia. 1990. The DNA-binding domain of HPV-16 E2 protein interaction with the viral enhancer: protein-induced DNA bending and role of the nonconserved core sequence in binding site affinity. *Virology.* 174:557–575.
9. Hines, C. S., C. Meghoo, S. Shetty, M. Biburger, M. Brenowitz, and R. S. Hegde. 1998. DNA structure and flexibility in the sequence-specific binding of papillomavirus E2 proteins. *J. Mol. Biol.* 276:809–818.
10. Djuranovic, D., C. Oguey, and B. Hartmann. 2004. The role of DNA structure and dynamics in the recognition of bovine papillomavirus E2 protein target sequences. *J. Mol. Biol.* 339:785–796.
11. Djuranovic, D., and B. Hartmann. 2003. Conformational characteristics and correlations in crystal structures of nucleic acid oligonucleotides: evidence for sub-states. *J. Biomol. Struct. Dyn.* 20:771–788.
12. Karslake, C., S. Schroeder, P. L. Wang, and D. G. Gorenstein. 1990. 31P NMR spectra of an oligodeoxyribonucleotide duplex lac operator-repressor headpiece complex. *Biochemistry.* 29:6578–6584.

13. Castagne, C., E. C. Murphy, A. M. Gronenborn, and M. Delepierre. 2000. 31P NMR analysis of the DNA conformation induced by protein binding SRY/DNA complexes. *Eur. J. Biochem.* 267:1223–1229.

14. Yang, L., W. Beard, S. Wilson, B. Roux, S. Broyde, and T. Schlick. 2002. Local deformations revealed by dynamics simulations of DNA polymerase β with DNA mismatches at the primer terminus. *J. Mol. Biol.* 321:459–478.

15. Chillemi, G., P. Fiorani, P. Benedetti, and A. Desideri. 2003. Protein concerted motions in the DNA-human topoisomerase I complex. *Nucleic Acids Res.* 31:1525–1535.

16. Marco, E., R. Garcia-Nieto, and F. Gago. 2003. Assessment by molecular dynamics simulations of the structural determinants of DNA-binding specificity for transcription factor Sp1. *J. Mol. Biol.* 328:9–32.

17. Gorfe, A. A., A. Caflisch, and I. Jelesarov. 2004. The role of flexibility and hydration on the sequence-specific DNA recognition by the Tn916 integrase protein: a molecular dynamics analysis. *J. Mol. Recognit.* 17:120–131.

18. Chillemi, G., T. Castrignano, and A. Desideri. 2001. Structure and hydration of the DNA-human topoisomerase I covalent complex. *Biophys. J.* 81:490–500.

19. Rauch, C., M. Trieb, W. Flader, B. Wellenzohn, R. H. Winger, E. Mayer, A. Hallbrucker, and K. R. Liedl. 2002. PvuII-endonuclease induces structural alterations at the scissile phosphate group of its cognate DNA. *J. Mol. Biol.* 324:491–500.

20. Obika, S., S. Y. Reddy, and T. C. Bruice. 2003. Sequence specific DNA binding of Ets-1 transcription factor: molecular dynamics study on the Ets domain–DNA complexes. *J. Mol. Biol.* 331:345–359.

21. Flader, W., B. Wellenzohn, R. H. Winger, A. Hallbrucker, E. Mayer, and K. R. Liedl. 2003. Stepwise induced fit in the pico- to nano-second time scale governs the complexation of the even-skipped transcriptional repressor homeodomain to DNA. *Biopolymers.* 68: 139–149.

22. Hartmann, B., M. Sullivan, and L. Harris. 2003. Operator recognition by the phage 434 cI repressor: MD simulations of free and bound 50 bp DNA reveal important differences between the OR1 and OR2 sites. *Biopolymers.* 68:250–264.

23. Wibowo, F. R., C. Rauch, M. Trieb, B. Wellenzohn, and K. R. Liedl. 2004. Water-mediated contacts in the trp-repressor operator complex recognition process. *Biopolymers.* 73:668–681.

24. Beveridge, D. L., G. Barreiro, S. K. Byun, D. A. Case, T. E. Cheatham, S. B. Dixit, E. Giudice, F. Lankas, R. Lavery, J. H. Maddocks, R. Osman, E. Seibert, H. Sklenar, G. Stoll, K. M. Thayer, P. Varnai, and M. A. Young. 2004. Molecular dynamics simulations of the 136 unique tetranucleotide sequences of DNA oligonucleotides. *Biophys. J.* 87: 3799–3813.

25. Case, A., D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Metz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. K. Weiner, and P. A. Kollman. 1999. Amber 6. University of California, San Francisco.

26. Cheatham III, T. E., P. Cieplak, and P. A. Kollman. 1999. A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* 16:845–862.

27. Berendsen, H. J. C., J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690.

28. Ryckaert, J. P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23: 327–341.

29. Darden, T., 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.

30. Cheatham III, T. E., J. L. Miller, T. Fox, T. A. Darden, and P. A. Kollman. 1995. Molecular dynamics simulation on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA and proteins. *J. Am. Chem. Soc.* 117:4193–4194.

31. Hegde, R. S., A. F. Wang, S. S. Kim, and M. Schapira. 1998. Subunit rearrangement accompanies sequence-specific DNA binding by the bovine papillomavirus-1 E2 protein. *J. Mol. Biol.* 276:797–808.

32. Fratini, A. V., M. L. Kopka, H. R. Drew, and R. E. Dickerson. 1982. Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. *J. Biol. Chem.* 257:14686–14707.

33. Grzeskowiak, K., K. Yanagi, G. G. Prive, and R. E. Dickerson. 1991. The structure of B-helical C-G-A-T-C-G-A-T-C-G and comparison with C-C-A-A-C-G-T-T-G-G. The effect of base pair reversals. *J. Biol. Chem.* 266:8861–8883.

34. 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.

35. Lavery, R., and B. Hartmann. 1994. Modelling DNA conformational mechanics. *Biophys. Chem.* 50:33–45.

36. Dickerson, R. E., M. Bansal, C. R. Calladine, S. Diekmann, W. N. Hunter, O. Kennard, E. von Kitzing, R. Lavery, H. C. M. Nelson, W. K. Olson, W. Saenger, Z. Shakked, H. Sklenar, D. M. Soumpasis, C. S. Tung, A. H.-J. Wang, and V. B. Zhurkin. 1989. Definition and nomenclature of nucleic acid structure parameters. *EMBO J.* 8:1–4.

37. Olson, W. K., M. Bansal, S. K. Burley, R. E. Dickerson, M. Gerstein, S. C. Harvey, U. Heinemann, X. J. Lu, S. Neidle, Z. Shakked, H. Sklenar, M. Suzuki, C. S. Tung, E. Westhof, C. Wolberger, and H. M. Berman. 2001. A standard reference frame for the description of nucleic acid base-pair geometry. *J. Mol. Biol.* 313:229–237.

38. Hartmann, B., D. Piazzola, and R. Lavery. 1993. BI-BII transitions in B-DNA. *Nucleic Acids Res.* 21:561–568.

39. Tisne, C., M. Delepierre, and B. Hartmann. 1999. How NF-kappaB can be attracted by its cognate DNA. *J. Mol. Biol.* 293:139–150.

40. van Dam, L., and M. H. Levitt. 2000. BII nucleotides in the B and C forms of natural-sequence polymeric DNA: A new model for the C form of DNA. *J. Mol. Biol.* 304:541–561.

41. Zhang, Y., Z. Xi, R. S. Hegde, Z. Shakked, and D. M. Crothers. 2004. Predicting indirect readout effects in protein-DNA interactions. *Proc. Natl. Acad. Sci. USA.* 101:8337–8341.

42. Zimmerman, J. M., and L. J. Maher 3rd. 2003. Solution measurement of DNA curvature in papillomavirus E2 binding sites. *Nucleic Acids Res.* 31:5134–5139.

43. Tisne, C., B. Hartmann, and M. Delepierre. 1999. NF-kappa B binding mechanism: a nuclear magnetic resonance and modeling study of a GGG $\rightarrow$ CTC mutation. *Biochemistry.* 38:3883–3894.

44. Wecker, K., M. C. Bonnet, E. F. Meurs, and M. Delepierre. 2002. The role of the phosphorus BI-BII transition in protein-DNA recognition: the NF-kappaB complex. *Nucleic Acids Res.* 30:4452–4459.

45. Djuranovic, D., and B. Hartmann. 2004. DNA fine structure and dynamics in crystals and in solution: the impact of BI/BII backbone conformations. *Biopolymers.* 73:356–368.

46. Tolstorukov, M. Y., R. L. Jernigan, and V. B. Zhurkin. 2004. Protein-DNA hydrophobic recognition in the minor groove is facilitated by sugar switching. *J. Mol. Biol.* 337:65–76.