doi:10.1016/j.jmb.2004.03.078 J. Mol. Biol. (2004) 339, 785–796

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

The Role of DNA Structure and Dynamics in the Recognition of Bovine Papillomavirus E2 Protein Target Sequences

D. Djuranovic[1]*, C. Oguey[2] and B. Hartmann[1]

[1]*Laboratoire de Biochimie Théorique, CNRS UPR 9080 Institut de Biologie Physico-chimique, 13 rue P. et M. Curie, Paris 75005, France*

[2]*LPTM, Université de Cergy-Pontoise, 95031 Cergy-Pontoise, France*

The papillomavirus E2 transcription and replication factors bind to the DNA consensus ACCGN_4CGGT sequence (E2-BS), through both direct and indirect readout mechanisms. The two symmetric half-sites ACCG·CGGT are highly conserved in the genomes and are hydrogen bound with E2. Although E2 does not contact the N4 spacer, the affinities are modulated by the base composition of this DNA part. Nevertheless, the origin of either the global recognition mechanism or the spacer effect remains unclear, particularly in the case of the bovine papillomavirus type 1 E2 (BPV-1-E2) system, used as model to study the papillomaviruses. We present, herein, studies carried out on oligomers differently recognized by the BPV-1-E2 protein and based on molecular dynamic simulations including counterions and water. The sequences contain the conserved half-sites but three different spacers (CCAT, ACGT and AAAC), resulting in very high, high and low affinity targets for BPV-1-E2. In order to estimate how much the free DNAs resemble the bound conformations, comparisons are made with two DNAs extracted from E2-BS-BPV-1 crystallographic complexes, representative of high and moderate affinity structures. The analysis of 15 ns trajectories reveals that the ACCG/CGGT half-sites, whatever the spacer, have the same behavior and adopt average stable base-pair parameters very close to those of the bound conformations. In contrast, the three different free spacers strongly differ in their BI ↔ BII backbone dynamics. The low affinity AAAC spacer exhibits stable BI backbone conformations, the high affinity ACGT spacer is characterized by a dramatic instability of the CpG phosphate groups, and the CpA and GpG backbones in the very high affinity CCAT·ATGG spacer are trapped in BII conformations. All resemble more of the moderate affinity complex DNA than the high affinity one. Nevertheless, the particular behavior of the CCAT and ACGT backbones allows the emergence of BII-rich spacers, a configuration reproducing both local and global helical features of the bound DNA conformation of the high affinity complex and favoring the minor groove curvature required in the complex. In particular, the CCAT-containing site spends almost half of the time in this form that well mimics the bound one. Thus, we propose that the E2 protein could take advantage of the invariant favorable structures of the half-sites to form a pre-complex, but would require a specific spacer intrinsic malleability to lock the interaction. Finally, the backbone conformational states, by their ability to translate information coded in the sequence into structural properties, provide insight into the mechanisms that contribute to fine binding site selection and specific nucleic acid ligand recognition.

Keywords: E2-BS; BPV-1-E2; BI and BII phosphate groups; specific recognition; molecular dynamics simulation

**Corresponding author*

Abbreviations used: BPV, bovine papillomavirus; HPV, human papillomavirus; MD, molecular dynamics; RMSD, root-mean-square deviation.

E-mail address of the corresponding author: djuranov@ibpc.fr

Introduction

The papillomaviruses are a family of DNA viruses that infect a variety of mammalian species, including human.[1] They infect skin and mucosa tissues where they induce benign warts or papillomas that can, for some members of this family, evolve into malignant carcinoma (HPV-16, HPV-18). The proteins E2 are a key factor for the transcriptional regulation and for the replication of the bovine papillomavirus-1 (BPV-1) and the human papillomavirus HPV-16 and HPV-18 strains. The proteins E2 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 protein E1.[2] Multiple copies of the dodecameric $ACCGN_4CGGT$ sequence are found in the viral genomes, and this sequence is the consensus E2 binding site (E2-BS).[3,4] If the presence of the highly conserved ACCG·CGGT half-sites is decisive for recognition of proteins E2, the affinities of the recognition sites are modulated by the four base-pair spacer sequences (N_4).[5] The affinities that the E2 proteins have for the naturally occurring binding sites on DNA vary by a factor of 11 for BPV-1 and by a factor of 25 for HPV-16 and HPV-18.[6–8] The resulting hierarchies of binding site affinities allow the fine regulation of transcription and replication. By exploring the range of E2 affinities for DNA sites in the BPV-1 and HPV genomes,[6,9] it was shown that HPV-E2 binds with significantly greater affinities to sequences containing A·T-rich spacers, while no clear spacer sequence signature has been detected for the BPV-1-E2 high affinity sites. In order to explore the relationship between DNA flexibility and E2 binding, other binding measurements[5,10,11] have been made on series of binding sites that have intact or backbone-nicked spacers. This study showed that, generally, breaks in the central step backbone can weakly enhance the affinity for BPV-1-E2 (at best by a factor of 2), but strongly diminish the affinity for HPV-E2. Despite the weakness of this effect on BPV-1-E2, as compared to the severe decrease in HPV-E2 binding, it was concluded that BPV-1-E2 is sensitive to flexible spacers.

The importance of the half-sites for recognition is confirmed by the crystal structures of two DNA sequences (N4: ACGT and AATT) bound to either the HPV-18 or the BPV-1-E2 proteins.[11,12] The proteins E2, folded into an unusual dimeric antiparallel β-barrel, insert α-helices into successive major grooves of *B*-DNA. The specificity is attributed to direct interactions between conserved amino acids in each recognition helix and the two ACCG·CGGT half-sites of the DNA. Either bases or phosphate groups of these target parts are contacted by often more than one amino acid side-chain. The four base-pair of the spacers are not contacted by E2: this observation eliminates any explanation of the spacer effect in terms of direct contacts. Regarding the bound DNA conformations, the roll patterns are similar along the whole 12 base-pair target sites, inducing the same bend. The positive rolls located at the CpG steps within the two symmetric half-sites and a block of negative rolls observed in all the spacers results in a 30° curvature toward the minor groove. Similarly, the twist and rise values are retrieved in the half-sites of all the co-crystals. The spacer conformations appear more variable; indeed our own analysis of the co-crystal structures shows that the tilt, the rise and the twist may have opposite values (positive/negative tilt, high/low rise/twist) when the ACGT or the AATT spacers are bound to BPV-1 or HPV-18-E2. The spacer backbone conformations also differ: in the complex of BPV-1-E2 with the best DNA sequence in terms of affinity for BPV-1-E2, there is, in the ACGT spacer, a distortion of four phosphate groups adopting the unusual BII conformation (CpG and GpT on the two strands).

The protein E2/E2-BS system is of particular interest because the recognition clearly involves both direct and indirect readouts of DNA. The selection of the specific target by the protein may proceed along several non-exclusive strategies: (i) internal sequence characteristics make the free DNA structure close to the bound conformation; (ii) free DNA exhibits sequence-dependent intrinsic flexibilities that lead to favorable features for the protein. To study the role of the DNA, in addition to the co-crystal structures, three DNA sequences have been crystallized without protein. These sequences differ in their spacer: ACGT (four structures per crystal unit),[13] GTAC[13] and AATT (three structures per crystal unit).[14] In the first stage, the analysis reveals a strong variability of the helical values of these free DNA structures, perhaps reflecting the different resolutions (from 1.6 Å to 2.8 Å). Contrasting the co-crystal data, the structures of the symmetric half-sites are different: for example, according to our analysis, the twist values of the CpG step in this region range from 28° to 38°. Furthermore, recent simulations, carried out on the sequence containing the AATT spacer, showed that numerous free DNA crystal features were not retrieved in solution.[15] Despite this limitation, the data suggest that the free and bound structures are more similar for the half-sites than for the spacers. The final information comes from the temperature factors, that are unusually large in the ACGT free spacer (Rozenberg, personal communication), in line with the hypothesis of a pre-existing deformability of high affinity BPV-1 spacers.

The overall results, so far, allow a better understanding of the E2/E2-BS recognition, in particular for the HPV-18 strain.[5,14] For all the E2 proteins, the resemblance between the free and bound half-sites, although poorly indicated by the crystal data, suggests that the cost in this DNA part deforming will be reduced upon binding. Concerning the spacers, the preference of HPV-18-E2 for A·T-rich sequences is attributed to a large accumulation of positive charges in the central

region of the protein that is juxtaposed to the enhanced negative potentials in the A·T minor groove.[5] The explanation of the recognition between HPV-16-E2 and its target is more hypothetical since the experimental structure of the complex is still lacking. Nevertheless, considering that electrostatic forces can be exerted upon the DNA phosphate backbone to facilitate bending, this protein, poorly charged all along the DNA interaction surface, could prefer rigid AT-rich spacers that could appropriately pre-bent the DNA molecule.[5]

In contrast, no conclusive explanation accounts for the recognition mechanism of BPV-1-E2/E2-BS, in particular for the spacer sequence discrimination. The absence, in the protein BPV-1-E2, of any positive charge close to the interaction surface explains its inability to favor A·T-rich spacer, but it does not account for the requirement of a dynamic or flexible structure. Furthermore, HPV-18-E2 and BPV-1-E2 make the same number of contacts with the half-sites phosphate groups and, thus, should have the same ability to force the spacer structure. Thus, we have carried out molecular dynamics simulations 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 (AAAC, $K_{rel} = 0.3$) affinity BPV-1-E2 targets.[6,10] The free DNA structures are compared to the DNA structures extracted from the two BPV-1-E2/E2-BS crystallographic complexes, one containing the ACGT spacer,[12] the other the moderate affinity ($K_{rel} = 0.8$) AATT spacer.[11] Our results confirm that, whatever the spacer, the ACCG·CGGT half-sites remain in a stable conformation very close to the bound conformations. Furthermore, the three studied sequences reproduce several features of the crystallographic DNA structure in the moderate affinity complex. However, only the best target sequences, thanks to the unusual backbone dynamics of their spacers, spontaneously reach characteristics of the high affinity bound conformation. In other words, our simulations provide a new insight into the mechanisms of BPV-1-E2/E2-BS recognition and outline the relevance of the phosphate dynamics to the recognition process.

Results

Here, we recapitulate the theoretical and experimental sequences used in this work, their origins and their denominations in Table 1.

General characteristics

During the MDs, the sequences clearly stay in *B* form, with an average root-mean-square deviation (RMSD) of 2.6(0.5) Å for 12 bp, the RMSD with respect to the *A*-DNA form being 5.0(0.5) Å.

All the sequences contain the two symmetric ACCG·CGGT sites; recall that the CG and CGb sequences have the same recognition site but the simulations start from different structures (initial RMSD = 2.4 Å, on 12 bp), and the ending sequences differ. So, several intra and intermolecular comparisons can be made to test the convergence of the simulations. The ACCG·CGGT half-sites all converge to the same structure (RMSD: 0.4(0.3) Å), indicating furthermore insensitivity to the ending steps (both in terms of sequence and length) and to the spacer sequences. The CG and CGb structures converge rapidly to the same average state (RMSD = 0.7 Å for 12 bp), supporting again an effectively good sampling of the conformational space by our simulations. Because of their identical dynamic behavior and for clarity, we will present only the data referring to the CG sequence. Finally, the fact that CGb sequence moves rapidly away from its starting point to reach an average RMSD of 3.5 Å, led to the conclusion that the bound DNA conformation is unstable for a free DNA in solution.

Backbone dynamic

ε ($C_{4'}$–$C_{3'}$–$O_{3'}$–P) and ζ ($C_{3'}$–$O_{3'}$–P–$O_{5'}$) are the most variable phosphodiester backbone angles in the free DNA. They define two different phosphate conformations: the BI conformation, characterized by ε: *t*, ζ: g^- and (ε–ζ) around −90° and the BII conformation characterized by ε: g^-, ζ: *t*, (ε–ζ) around +90°.[16–18] The percentage of BII conformation for each dinucleotide linkage is shown in Figure 1 for the recognition sites of CA, CG and AA sequences. In the three simulations, the proportions of BII conformers, similar for the comparable steps (same dinucleotides at same or symmetric positions), confirm the good sampling of the phosphate dynamics. Thus, whatever the sequence, BII-rich conformers (>30%) are found in GpA and GpG steps, in agreement with previous

Table 1. Sequences of the studied oligomers

Sequence	Designation	Origin	
5′-CGCACCGCCATCGGTGCG-3′	CA	MD	Free DNA
5′-CGCACCGACGTCGGTGCG-3′	CG	MD	Free DNA
5′-CGCACCGAAACCGGTGCG-3′	AA	MD	Free DNA
5′-CGACCGACGTCGGTCG-3′	CGb	MD	Free DNA
5′-CGACCGACGTCGGTCG-3′	CGc	Crystal	Bound DNA
5′-CAACCGAATTCGGTTG-3′	ATc	Crystal	Bound DNA

The bases directly contacted by the cognate protein are underlined.

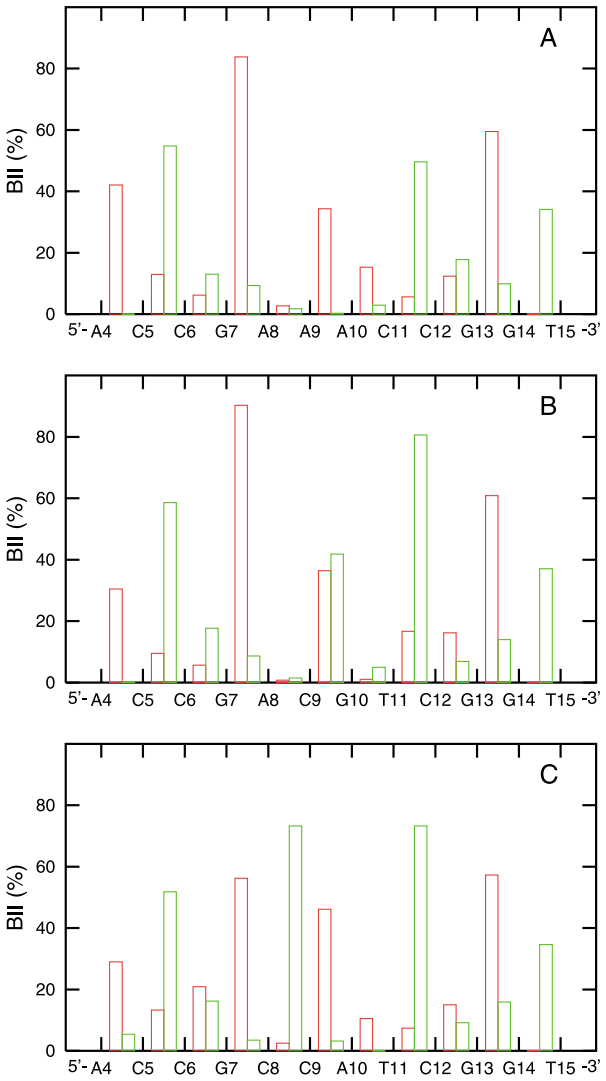

Figure 1. Percentage population of the backbone BII state (A, CA sequence; B, CG sequence; C, AA sequence) for each linkage along the DNA sequences. Strand 1 is red and strand 2 is green.

NMR studies[19,20] and high resolution crystallographic structure analysis.[21] Similarly, the central CpA, CpG·CpG and ApA backbones in the spacers of CA, CG and AA sequences exhibit a significant propensity for BII conformers. They clearly present higher percentages of BII population than their counterparts: the central CpA (Figure 3A), CpG·CpG (Figure 2A) and ApA (Figure 2B) backbones show 46%, 39% and 34% of BII conformers, respectively, against 21% for the CpA steps in 3′- and 5′- of the target site (Figure 3B), at best ~12% for any outer CpG (Figure 2C) and ~2% for the A8pA9 step in the AAAC spacer (Figure 2D). In addition, despite the fact that all the GpG in the half-sites (Figure 3C) exhibit a high percentage of BII conformers (in average 57%), this propensity is once more enhanced in the GpG step in the CCAT spacer (73%, Figure 3A). Finally, we note that only a few facing phosphate groups are almost totally devoid of BII conformers and that they are located within the spacers: a stable BI·BI configuration characterizes the ApT·ApT, ApC·GpT and $A_8pA_9{\cdot}T_{28}pT_{29}$ steps in the CCAT, ACGT and AAAC spacers, respectively. Overall, the observation of different backbone dynamical behaviors for a given dinucleotide is in line with earlier theoretical studies,[22,23] reinforced by several NMR[19,20,24–26] that have shown that the dynamic of YpR backbones depends on the neighboring bases.

The behavior of the BII-rich steps within the CA and CG spacers requires a further comment. Along the trajectories, in most cases, only one of the backbones undergoes BII transitions while the other steadily remains in BI. The backbone of the central ApA·TpT in the AAAC spacer (Figure 2B) well illustrates this behavior. The CpA and GpG steps in the CCAT spacer (Figure 3A) are distinguishable by both the particularly high percentages and the large lifetimes of BII conformers. In contrast, the central CpG·CpG in the ACGT spacer undergoes non-correlated BI/BII transitions at high frequency (Figure 2A); this chronic instability is a spectacular characteristic of these phosphate groups. In these two spacers, the

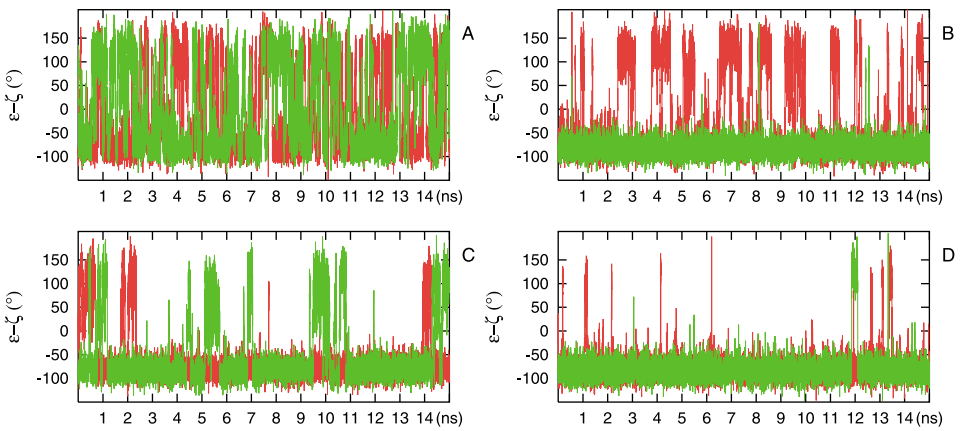

Figure 2. (ε−ζ) Fluctuations during the simulations for (A) central CpG·CpG backbones in the A**CG**T spacer, (B) central ApA (red) and TpT (green) backbones in the A**AA**C spacer, (C) C6pG7.C30pG31 backbones in the AC**CG** half-site of the CG sequence, (D) A8pA9 (red) and T28pT29 (green) backbones in the **AA**AC spacer.

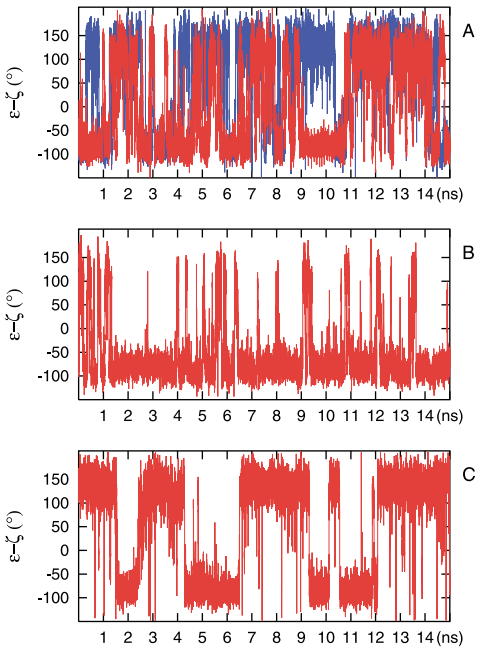

Figure 3. (ε–ζ) Fluctuations during the simulations for (A) CpA (red) backbone in the C**CA**T spacer and GpG (blue) backbone in the CCAT·AT**GG**, spacer, (B) C3pA4 backbone in the CG sequence, (C) G13pG14 (red) backbone in the C**GG**T half-site of the CG sequence.

backbone dynamic is unusual: in the CCAT spacer, some BII steps are stabilized while, in the ACGT spacer, the CpG phosphate flexibility is increased.

Rather than the simple BI/BII conformers, let us now consider the possible combinations of the conformations of the six phosphate groups in the spacers. Three combinations emerge from the trajectories, as shown in Table 2. The first case, that we will call C0, common to the three spacers, consists of only BI phosphate groups. In the second case (C1), one phosphate in each sequence adopts the BII conformation. The remaining case (C2) is characterized by the simultaneous presence of two BII phosphate groups, one on the first strand, the other located on the facing strand. Only the CA and the CG spacers can adopt this BII-rich configuration, where they spend 39% and 12% of time, respectively. As the BII-rich spacers characterize the CGc high affinity bound DNA (four BII phosphate groups), it seems to us that the impact of this backbone state on the free MD structures should be tested. Considering the percentage of each possible phosphate configuration *versus* the RMSD between the free MD structures and the bound DNA in the CGc crystal, we find that the RMSD of the AA sequence is insensitive to the C0 and C1 percentages, whereas, for the CA and CG sequences, the C1 and the C2 configuration percentages increase when the RMSD decreases (Figure 4). Thus, the closest structures (RMSD ≤ 2.5 Å) populate either almost equally the C1 (44%) and C2 (40%) states, in the case of the ACGT containing site, or only the C2 (87%) configuration for the CA sequence. This result strongly suggests that the backbone states in the CCAT and the ACGT spacers affect the conformation of the whole oligomers and play a role in the recognition mechanism.

Table 2. The different spacer phosphate configurations and their percentages, as observed in the MD trajectories

MD			RX
C0 (no BII)	C1 (one BII)	C2 (two BII)	
	C•C•A•T G•G•T•A 34%		
C•C•A•T G•G•T•A 20%	C•C•A•T G•G•T•A 7%	C•C•A•T G•G•T•A 39%	
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%	A•C•G•T T•G•C•A CGc
A•A•A•C T•T•T•G 66%	A•A•A•C T•T•T•G 34%	– –	
			A•A•T•T T•T•A•A ATc

The last column (RX) shows the crystallographic conformations of the bound CGc and ATc spacers. •: BI phosphate; •: BII phosphate.

Helical parameters

Let us now analyze some helical parameters, critical for the protein–DNA interaction process. The rise and the twist govern the overall form of the fragment. The roll is the major source of bending, which is also modulated in a second hand by

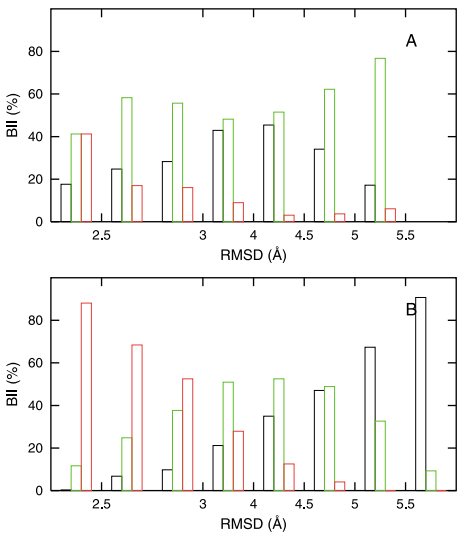

Figure 4. Population distribution of the C0 (black), C1 (green) and C2 (red) configurations found in the ACGT (A) and the CCAT (B) spacers *versus* RMSD calculated between the CGc bound sequence and the CA and the CG free sequences for 12 bp.

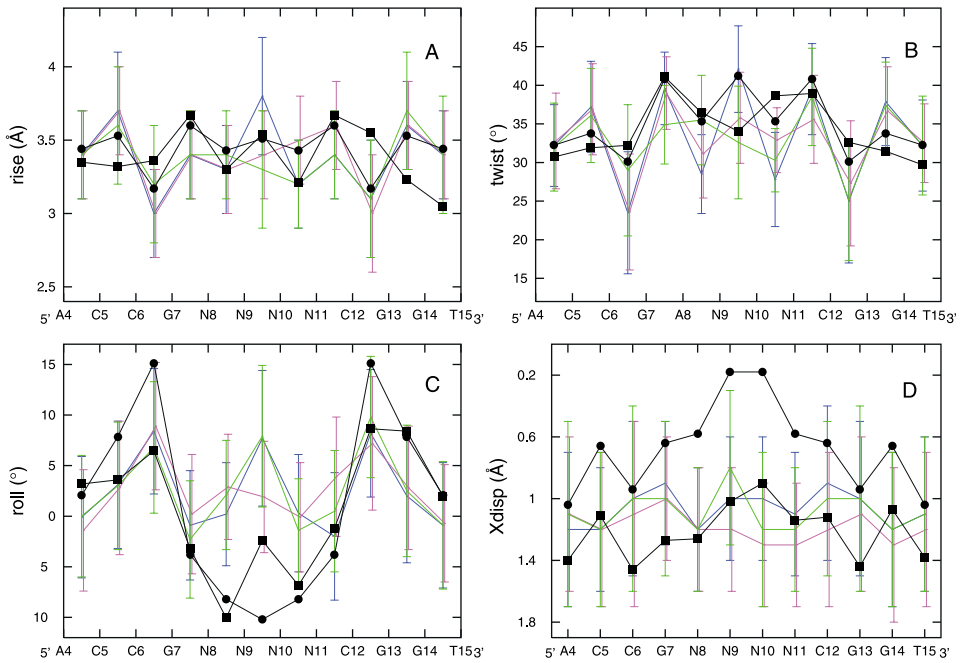

Figure 5. Average and standard deviation values of the A, rise, B, twist, C, roll and D, *X*-disp along the free MD sequences. Free MD DNAs (CA sequence, green; CG sequence, blue; AA sequence, pink) are compared to the bound CGc (●) and ATc (■) structures.

the tilt. *X*-disp quantifies the base displacement towards the major (*X*-disp > 0) or the minor groove (*X*-disp < 0), and is related to the groove's depth and the base atom's accessibility.

Rise, twist, roll and *X*-disp average values of CA, CG, AA, CGc and ATc sequences are plotted in Figure 5. Let us first discuss the two DNA crystal structures. These two bound DNAs are similar in terms of RMSD (0.8 Å for the ten common base-pairs). The two sequences contain ACGT and AATT spacer sequences and thus are palindromic, although the structure of the ATc sequence is not rigorously symmetric. The alternating sequence of positive and negative rolls is preserved (Figure 5(C)); nevertheless, the extreme roll values are more pronounced in the CGc sequence than in the ATc sequence: the positive rolls of the CpG half-sites reach 16° in CGc while only 6° in ATc; similarly, the negative roll of the central spacer step is −8° in the CGc sequence and only −2° in the ATc sequence. The central twists are correlated to the central rolls and thus they show different values: a high twist is associated to the strong negative roll, while a moderate twist is found with the moderate negative roll. However, the major discrepancy between these two structures concerns the *X*-disp (Figure 5D). In the ATc sequence, the bases are located in the minor groove, whereas they are pushed towards the center in the CGc helix. A re-examination of our data-set of high resolution protein/DNA crystal structures (90 sequences) recently published[21] shows that the major peak of the *X*-disp distribution is centered around −0.8 Å for the DNA major groove binding complexes, and around −1.5 Å for the DNA minor groove binding complexes, the same value as we found for the free *B*-DNA. From this point of view, the ATc *X*-disp (−1.25 Å) appears unusual for a major groove complex, whereas the CGc *X*-disp value (−0.7 Å) is completely in line with the statistical analysis.

Concerning the MD structures, in all cases but the central CpG step in the ACGT spacer, the average values of the inter-base parameters are found to depend only on the dinucleotide type in the three studied sequences.

As observed above for the backbone dynamics, the helical characteristics of the ACCG·CGGT half-sites are identical whatever the sequence and the starting point of the simulations. Moreover, along the conserved fragments extended to the adjacent G7pN8 and N11pC12 dinucleotides, a good fit is observed step-by-step with the experimental data for the rise (Figure 5A), the twist (Figure 5B), the roll (Figure 5C) and the tilt (not shown). Alternating low and high values are preserved and the values of the bound structure parameters are always within thermal fluctuations of the free DNA parameters. These similarities are reflected through the weak RMSD values (1.3 Å in average) measured between the MD and the crystallographic half-sites.

The rise and the twist values encountered in the free spacer sequences are also acceptable in regard to those of the bound structures. However, for the three studied spacers in the free DNAs, major discrepancies appear in the roll values that are far away from the negative rolls observed in the complexes (Figure 5C). Despite a large standard deviation, the roll value of the central CpA and

Table 3. Average value and standard deviation of RMSA (deg.) made on four base-pair fragments (the half-sites and the spacers) between crystallographic (CGc and ATc) and MD structures, in function of selections made on the spacer backbone configurations

	CGc				ATc			
Spacer selection	No selection	C0	C1	C2	No selection	C0	C1	C2
Half-sites CG and AA	28(9)	28(9)	28(9)	28(7)	26(9)	26(9)	26(9)	26(8)
Half-sites CA	28(9)	28(9)	28(9)	28(7)	34(8)	32(8)	35(8)	34(7)
Spacer CCAT	42(6)	47(4)	44(4)	37(2)	31(6)	25(5)	31(5)	33(6)
Spacer ACGT	41(6)	47(5)	40(5)	30(4)	25(5)	23(5)	26(5)	33(6)
Spacer AAAC	45(4)	47(5)	41(5)	–	23(5)	22(5)	26(5)	–

CpG steps in the CCAT and ACGT spacers remain positive. The corresponding roll in the AAAC spacer is able to adopt moderate negative values, as found for the ATc bound sequence. Nevertheless, the values of the three successive rolls in the three free spacers do not reach those of the high affinity CGc bound sequence. Considering the base-pair *X*-disp parameter, the values in free DNAs are similar to those of the ATc sequence, whereas markedly different from the CGc sequence. Table 3 summarizes these results in terms of RMSA, which allows the comparison of different base sequences. These RMSA values are calculated between the crystallographic structures of the four base-pair fragments (the half-sites and the spacers) and the corresponding MD structures. As a reference, the RMSA between the CGc and ATc structures are 23° for the half-sites, and 34° for the spacers. Comparing the simulated DNAs with the ATc DNA, we find similar weak RMSA values for the half-sites and the spacers, revealing that the free structures are close to this bound conformation, even if the CA half-sites are somewhat further than the two other sequences. In contrast, only the free half-sites agree with the CGc ones (RMSA of 28° in average), while free and bound spacers adopt different conformations.

Interesting couplings are observed for some successive rolls. In the three studied sequences, the rolls of the CpG steps in the ACCG·CGGT half-sites and the rolls of the steps that form the junction between the half-sites and the spacers (AC**CGA**NN**NCG**GT) evolve conjointly: when the CpG rolls increase, the junction rolls decrease. This coupling is more pronounced for the CA and CG sequences than for the AA sequence. As an example, in the palindromic CG sequence, GpA·TpC junction roll values of $< -5°$ correspond to CpG roll values of $+11°$ in average, almost the same values as encountered in the bound CGc structure ($-4°$ and $+15°$, respectively). Another coupling is found for rolls in the AAAC spacer. The two successive ApA·TpT rolls are anti-correlated: negative rolls in the central ApA·TpT step are associated with 5′- neighboring positive rolls (in average $+8°$). In the bound structures, both are negative, and thus, in this case, the intrinsic roll properties of the AAAC spacer and the effect of the protein are contra-variant.

The major discrepancies between the free and the high affinity bound DNA are found for *X*-disp and for the spacer roll values. Keeping in mind that the difference between the two spacers CCAT and ACGT on one hand and the AAAC spacer on the other hand, is mostly manifest in their backbone dynamics and that the different phosphate conformations are known to be associated with large changes in DNA local (ex: roll) and global (ex: *X*-disp) structural properties,[21,22,27] we select the structures according to the three phosphate combinations found in our trajectories (Table 2). Following the phosphate configurations, we have three families for the CA and CG sequences but only two for the AA sequence (configuration C2 does not show up in the AA spacer). The RMSA values (Table 3) show that, whatever the selection, the free ACCG·CGGT half-sites remain at the same distance from the bound structures. The selections slightly affect the RMSA between the free and bound ATc spacers, the good agreement already noticed being reinforced by the BI-rich spacers. The same comparisons with the CGc sequence are, in contrast, strongly dependent on the selections. The RMSA values for the MD spacers, clearly higher than for the half-sites without any selection, diminish with the number of BII conformers in the spacers, until reaching 30° for the C2 structure group of the ACGT spacers. The better agreement between the free CG and the bound CGc spacer conformations is essentially due to the fact that BII conformers, especially two facing BIIs, disfavor positive rolls (Figure 6A) and improve the tilt pattern (Figure 6B), without damaging the fit with the other inter base parameters. Nevertheless, the most crucial impact is noted for the *X*-disp base-pair parameter. The BII·BII configuration strongly affects the global *X*-disp of the 12 base-pair fragments, and, as shown in Figure 6(C), the presence of two facing BII central phosphate groups pushes the bases towards the major groove, as found in the CGc high affinity complex. This effect is characteristic of the BII·BII configuration. As an example, if we select the structures on the basis of spacer negative rolls, the *X*-disp is not affected.

The curvature

The two crystal bound DNAs are bent towards the minor groove with a curvature of 30°. In our

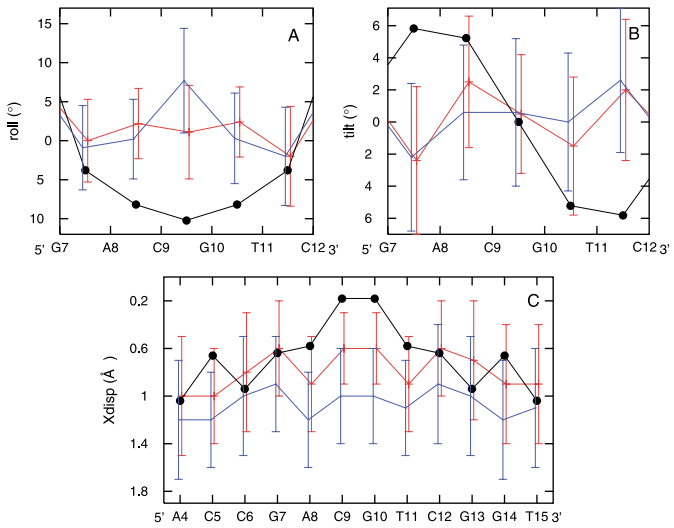

Figure 6. Average and standard deviation values of the A, roll, B, tilt and C, *X*-disp calculated on the selected pool of CG structures containing the central CpG·CpG phosphate groups in BII·BII configuration (red), compared to the structures without selection (blue) and to the bound CGc DNA structure (●).

MD structures, we observe a slight anisotropic curvature directed towards the major groove, with an average curvature magnitude of 15(9)°. Restricting considerations to the spacers, on average, the CCAT and ACGT parts are bent towards the major groove, while the AAAC spacer curvature is directed towards the minor groove, as expected for an A·T-rich sequence. Nevertheless, the presence of the strong positive rolls on the CpG half-sites dominates the spacer sequence effects, so that the overall curvature ends up being towards the major groove, for all oligomers. However, in the three trajectories, 10% of the snapshots display a curvature with a significant magnitude (>15°, the average value) and directed towards the minor groove. These groups of structures, in the case of the CA and CG sequences, contain many BII-rich spacers (C2 family): 42% for CA and 30% for CG. This observation should be confronted to Figure 4 showing the effect of the C2 containing sites on the RMS between the whole free and bound states of DNA. The roll values and the standard deviations of the structures that fit the minor groove curvature are plotted in Figure 7A (AA sequence), B (CG sequence) and C (CA sequence), in comparison with those of the non-selected free structures and the bound DNAs. From these roll patterns, it becomes clear that the curvature direction is not linked to marked negative values of the spacer rolls, since the free oligomers can accommodate a minor groove curvature with central roll values around zero or even slightly positive. The block of five consecutive negative rolls in the G7N$_4$C12 part of the oligomers seems to act on the curvature magnitude, since, when it occurs, the bent amplitude increases. Thus, from the curvature point of view, the three studied oligomers are similarly capable of adopting temporarily the complex curvature. Nevertheless, we observe that the roll pattern of the CCAT and ACGT sequences becomes strictly parallel and very close to the ATc bound sequence one, notably reducing the RMSD between the ten common base-pairs of ATc and CA or CG sequences (2.4 Å against 3.0 Å without selection). For comparison, the corresponding RMSDs (also calculated for the same ten base-pairs) between the AA and ATc structures are on average 2.8 Å for the selected curved structures, and 3.0 Å without selection.

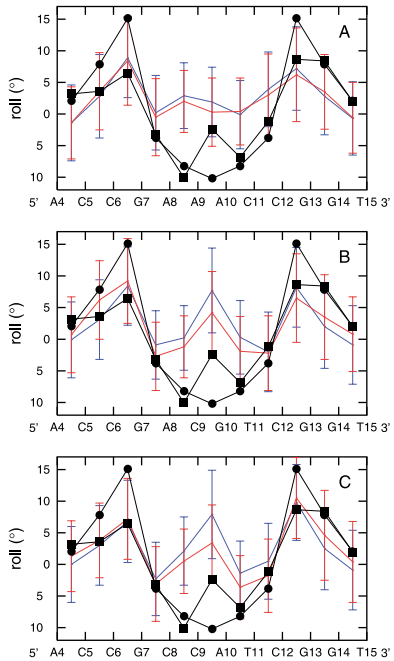

Figure 7. Average and standard deviation values of the rolls calculated on the selected pool of AA (A), CG (B) and CA (C) structures adopting a minor groove curvature with a magnitude of >15° (red), compared to the structures without selection (blue) and to the bound CGc (●) and ATc (■) structures.

Conclusion and Discussion

The affinity of the BPV-1-E2 protein depends critically upon the DNA spacer sequence, a region of the target not in direct contact with the protein. Since the early stage of recognition depends on key features of the structure of the uncomplexed

DNA in solution, we have performed a series of MD studies of duplex sequences with the E2 recognition motif ACCG·CGGT and three different spacers, i.e. CCAT·ATGG (in the CA sequence), ACGT·ACGT (in the CG sequence) and AAAC·GTTT (in the AA sequence), that correspond, respectively, to very high, high and low affinity targets for BPV-1-E2. Here, we examine in detail how the free MD and the bound crystallographic DNAs compare. These experimental data encompass the BPV-1-E2 protein complexed with (i) the high affinity target used here (the CGc structure[12]); (ii) a moderate affinity site, containing the AATT spacer (the ATc structure[11]).

The first result concerns the structure of the conserved half-sites extended to the adjacent dinucleotide steps making the junction with the spacer (ACCGN·NCGGT). Whatever the sequence and the conformation of the four base-pair central spacers, they adopt a stable conformation in which the base-pair parameter values are similar to those of the two bound conformations, which are themselves very close to each other. These parts of the free targets reproduce the alternation of weak and high values observed for the bound base-pair parameters and they achieve all the bound base-pair parameter values within the thermal fluctuations. Furthermore, in the CA and the CG sequences, a coupling is found between the two CGA·TCG adjacent rolls, so that they can easily evolve towards the marked positive/negative roll pattern encountered in the high affinity bound DNA.

The second series of results focuses on the free spacers. They are not very far from the ATc spacer structure, although they behave differently. The major difference comes from the backbone dynamics. The AAAC spacer shows a classical backbone dynamic, with the phosphate groups in majority in BI conformation. In contrast, the high affinity ACGT spacer is characterized by intense conformational fluctuations in the central CpG·CpG backbones that are not observed elsewhere along the CA, CG or AA oligomers. These high frequency non-correlated BI/BII transitions allow a relatively low but significant percentage of the unusual BII·BII face-to-face phosphate configuration. Similarly, the unusual stabilization of BII CpA and GpG steps on the facing strands of the CCAT spacer leads to a high percentage of BII-rich spacer structures. The occurrence of these BII-rich regions is significant. On one hand, the CGc co-crystal DNA structure contains four BII phosphate groups (two per strand) in the spacer. On the other hand, the free CA and CG structures show several semi-local and global characteristics (null or negative rolls, tilts, and especially in the case of CG sequence, weak base displacement values) that are turned to those of the high affinity bound CGc conformation.

Considering curvature, which is an important feature of the bound DNA structures, the 30° minor groove curvature found in the complexes is not retrieved, since, on average, the three considered oligomers behave similarly and show a slight bending preference towards the major groove. Nevertheless, some MD snapshots in all the trajectories exhibit a significant (magnitude $>15°$) curvature towards the minor groove, and, among these structures, we note, again, an advantage for the CA and CG sequences containing a BII-rich spacer. In addition to the curvature direction, these selected structures agree with the ATc bound DNA parameters, in particular the rolls. Nevertheless, the protein E2 is required to stabilize the curvature, in terms of both magnitude and direction.

From these considerations, it appears that the different affinities of these three BPV-1-E2 targets could be directly related to their ability to more or less adopt pre-distorted structures that obviously favor the achievement of the interaction without implying a large cost in DNA deforming.

Now, we can understand better the affinity measurements made on intact and nicked spacer sequences.[10] The backbone instability that we observe for the central step of the ACGT spacer seems to have the same effect as breaks in the backbone.[28] A nick introduced in such an intrinsically instable backbone should not change anything, apart from prohibiting the BII·BII configuration found in the CGc structure. This could be the reason for the decrease in affinity observed for the nicked ACGT spacer, while a nick located on less malleable steps, such as ApT, increases the affinity. Although no such measurement is yet available for the CCAT containing site, we may suppose that spacer nicks, removing the BII conformers possibility, would give the same type of results as for the ACGT spacer.

Concerning the recognition mechanism, and keeping in mind that DNA deformation is an important component of the driving forces for protein/DNA association, our simulations show that the AA, CG and CA sequences exhibit "extended" half-sites structures closely adapted to BPV-1-E2. However, along the trajectories, the conformation of the whole AA fragment never fits to the bound structures as well as the CG and CA sequences. The AA sequence conformational space is not very far from the ATc structure, but solely the CA and the CG sequences, thanks to the special spacer behavior, explore a conformational space that clearly overlaps the two conformations observed in the ATc and the CGc bound DNAs. Furthermore, we have seen that it is tempting to consider the ATc structure as an instance of moderate affinity complex, i.e. less stable than the CGc one which is more representative of a high affinity complex. In this context, the formation of a moderate affinity complex with the AA sequence seems yet possible, although presumably with a lower efficiency than for the two other sequences, whereas adopting the high affinity complex form appears easily accessible only to the CA and the CG sequences. The CG sequence does not spend

much time in a CGc-like conformation and, although the CA sequence exhibits a predistorted structure nearly the half-time of the trajectory, it seems improbable that the protein E2 directly recognizes this transient state conformation. Instead, the role of E2 could be to impose the curvature, taking advantage of a stabilization of the BII-rich spacer configuration in the free CA and CG sequences, that in turn facilitates both the required pronounced negative rolls within the spacer and the shift of the whole base-pairs toward the major groove, thereby definitively locking the complex. Whatever the detailed process is, the energetic cost of these three sequences' deformation certainly becomes minimal upon complexation. However, as underlined by Hegde's group,[11] reducing the energy expended in deforming DNA could be counterbalanced by an entropic decrease due to a loss of flexibility in the complex, especially if the free DNA is particularly flexible. Nevertheless, in the studied sequences, speaking of global enhanced flexibility is inappropriate to qualify the CA and the CG sequences compared to the AA sequence: the helical parameter standard deviations are similar for the three sequences; only the high frequency of BI/BII transitions on the two facing central backbone of the ACGT spacer and the stabilization of BII conformers in the CCAT spacer are unusual. The major difference between, on one part, the CA and the CG sequences, and, on the other part, the AA sequence, is not in their global flexibility but in the fact that they explore different extreme conformational regions.

Let us now compare the recognition mechanisms of the BPV and HPV systems. The HPV-18 E2 DNA binding specificity seems clearly established:[5] this protein has an accumulation of positive charge in the center of its DNA interaction surface, discriminating the A·T-rich spacers that exhibit a complementary negative minor groove. This model of selectivity is similar to ones highlighted on other systems; furthermore, it is convincingly supported by protein mutation experiments.[29] In contrast, until now, the recognition mechanism of HPV-16 E2 remains hypothetical. Only the X-ray structure of BPV-16 E2 is known,[30] while the E2 protein bound to its DNA target seems to resist crystallization attempts. The potential interaction surface of HPV-16 E2 is less charged than the BPV-1 one, the most charged surface being HPV-18 E2. In addition, HPV-16 E2 makes fewer contacts with the DNA backbone than BPV-1 E2.[9] Because of these two observations, Hedge[5] postulated that this protein could not exert a strong force to bend the DNA, and thus required a pre-bent DNA, sterically complementary to the protein surface. This conclusion raises questions in the perspective of our results. Indeed, the AAAC containing site that we found only slightly pre-bent towards the minor groove, is recognized by HPV-16 E2 with a maximal efficacy.[10] However, the prerequisite for HPV-16 E2 could be not limited to the curvature. Comparing the DBD regions of the three dimeric proteins E2 of interest (HPV-16, -18 and BPV-1) shows that the relative orientation of the two symmetric recognition helices α1 are identical for HPV-18 and BPV-1, whereas the HPV-16 ones do not agree with them, resulting in a displacement of 4 Å and a tilt of 25°.[5] Thus, the relative orientations of the two DNA half-sites should have an important impact on the recognition by HPV-16 E2. In this line, we observe that the angle between the two axes of the half-sites base planes is in line with the CGc structure for the CG and the CA sequences but increased by almost 10° for the AA sequence, the distances between the half-sites mass centers in the three free DNA being all close to the bound CGc value. However, a complete understanding of the molecular mechanism of the recognition of DNA by HPV-16 E2 at the atomic level depends on obtaining the complex structure. Similarly, in order to confirm our hypothesis for the BPV-1 E2 recognition mechanism, the next step would be to carry out MD of the complexes themselves, to probe other contributions to the binding such as the relative stability of each DNA conformer in the presence of E2.

Finally, it appears that the phosphate conformations could be decisive actors in the inducing curvature in complexes. This idea is strongly supported by the analysis of the high resolution structure of the nucleosome core particle containing 147 DNA base-pairs.[31] This structure reveals an alternation of major and minor groove curvatures, which forms the super-helix. The minor groove blocks are composed of 3 bp or 4 bp having strong negative roll angles, and our own analysis shows that all these 13 blocks are associated to either BI·BII (six cases) or BII·BII (seven cases) phosphate configurations. This work, consistent with the nucleosome observations and added to previous results[27,32] on transcription factor targets that account for curvature-dependent backbone motion, shows that they should be taken into consideration for understanding the origin of DNA curvature at the molecular level.

Methods

Molecular dynamics

Model building and simulations were performed using the AMBER 6.0 program[33] and the Parm98 force field,[34] employing the procedure successfully used in earlier studies.[21,22] Three MD of 15 ns were carried out on the 18 bp 5′-CGCACCGCCATCGGTG CG-3′, 5′-CGCACCGACGTCGGTGCG-3′ and 5′-CGC ACCGAAACCGGTGCG-3′ sequences, constructed with a canonical *B*-DNA conformation (numbering: 5′-C_1G_2 $C_3A_4C_5C_6G_7N_8N_9N_{10}N_{11}C_{12}G_{13}G_{14}T_{15}G_{16}C_{17}G_{18}$-3′). The underlined bases are the recognition sites. Examining the flanking sequences around the BPV-1 genome recognition sites of E2 protein, we found that Pyr-G-C was the most often encountered motif, and thus, we selected CGC·GCG as terminal bases. A fourth

MD of 15 ns was made on the 16 bp 5′-CGACCGACGTCGGTCG-3′ sequence, built using the DNA sequence and coordinates of the high resolution crystal structure of the BPV-1-E2/E2-BS complex[12] (NDB code: pdv001).

The oligomers were neutralized with Na^+ counterions (one Na^+ for one phosphate group) and explicitly solvated by a 10-Å water shell in all directions (roughly 11,000 and 9000 TIP3P water molecules for the 18 bp and 16 bp oligomers, respectively) in a truncated octahedral box.

After 1000 cycles of energy minimization, each minimized system was heated to 300 K, re-scaling the velocities as necessary, and coupling the system to a heat bath using the Berendsen algorithm.[35] During each of these phases, harmonic restraints were imposed on the atomic positions of the oligomers and then slowly relaxed over several periods of 50 ps until a free system was achieved. The simulations were then performed at constant temperature and pressure (NTP) using the Berendsen algorithm. Bond lengths involving hydrogen atoms were constrained using SHAKE,[36] which enabled an integration step of 2 fs. Long range electrostatic interactions were treated using the Particle Mesh Ewald (PME) approach[37,38] with a 9 Å direct space cutoff, a direct sum tolerance criteria of 10^{-5} and a reciprocal space charge grid spacing of roughly 1 Å. During the production phase, translations and rotations of the DNA oligomer were removed every 100 steps.

Structural analysis

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

The pairwise comparisons of DNA structures are quantified by root mean square deviations: RMSD of heavy atom Cartesian coordinates is used in parallel with RMSA of torsion angles (α, β, γ, ε, ζ, χ, sugar phase and amplitude), which allows comparison of DNAs composed of different bases.

Analyses of DNA structures were carried out using CURVES,[39] which calculates the optimal helical axis and a complete set of conformational and helicoidal parameters. Analyses were made in terms of both local and global parameters; however, since the two analyses give quasi-identical values, only the global values are reported here. The helical parameters, which obey the Cambridge convention for DNA conformation,[40] are divided into two categories: the inter-base 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 parentheses next to the corresponding mean values. In order to avoid end effects, the first and the last base-pairs are dropped out of the statistics. Generally, we focus on the data on the 12 central base-pairs, i.e. the recognition sites.

Finally, our structures are be compared with the crystallographic DNA structures bound to the protein E2 of BPV-1. One corresponds to the high affinity sequence ACCGACGTCGGT and is available on the PDB or NDB site (NDB code: pdv001[12]). The second has been resolved by the group of R. Hedge[11] with the moderate affinity sequence ACCGAATTCGGT. This crystal contains two complexes per unit cell, and the two DNAs slightly differ (RMSD = 0.8 Å on 12 bp). For clarity, we present only one of these two structures, choosing the conformation that mostly differs from pdv001.

Acknowledgements

The authors thank Dave Beveridge and Suzie Byun for very productive discussions and Brahim Heddi for helpful advice. Rashmi Hedge is gratefully acknowledged for providing the crystallographic coordinates of the ATc complex.

References

1. Sousa, R., Dostatni, N. & Yaniv, M. (1990). Control of papillomavirus gene expression. *Biochim. Biophys. Acta*, **1032**, 19–37.
2. Steger, G., Ham, J. & Yaniv, M. (1996). E2 proteins: modulators of papillomavirus transcription and replication. *Methods Enzymol.* **274**, 173–185.
3. Dostatni, N., Thierry, F. & Yaniv, M. (1988). A dimer of BPV-1 E2 containing a protease resistant core interacts with its DNA target. *EMBO J.* **7**, 3807–3816.
4. McBride, A. A., Byrne, J. C. & Howley, P. M. (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.
5. Hegde, R. S. (2002). The papillomavirus E2 proteins: structure, function, and biology. *Annu. Rev. Biophys. Biomol. Struct.* **31**, 343–360.
6. Li, R., Knight, J., Bream, G., Stenlund, A. & Botchan, M. (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.
7. Sanders, C. M. & Maitland, N. J. (1994). Kinetic and equilibrium binding studies of the human papillomavirus type-16 transcription regulatory protein E2 interacting with core enhancer elements. *Nucl. Acids Res.* **22**, 4890–4897.
8. Steger, G. & Corbach, S. (1997). Dose-dependent regulation of the early promoter of human papillomavirus type 18 by the viral E2 protein. *J. Virol.* **71**, 50–58.
9. Bedrosian, C. L. & Bastia, D. (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.
10. Hines, C. S., Meghoo, C., Shetty, S., Biburger, M., Brenowitz, M. & Hegde, R. S. (1998). DNA structure and flexibility in the sequence-specific binding of papillomavirus E2 proteins. *J. Mol. Biol.* **276**, 809–818.
11. Kim, S. S., Tam, J. K., Wang, A. F. & Hegde, R. S. (2000). The structural basis of DNA target discrimination by papillomavirus E2 proteins. *J. Biol. Chem.* **275**, 31245–31254.
12. Hegde, R. S., Grossman, S. R., Laimins, L. A. & Sigler, P. B. (1992). Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature*, **359**, 505–512.

13. Rozenberg, H., Rabinovich, D., Frolow, F., Hegde, R. & Shakked, Z. (1998). Structural code for DNA recognition revealed in crystal structures of papillomavirus E2-DNA targets. *Proc. Natl Acad. Sci. USA*, **95**, 15194–15199.
14. Hizver, J., Rozenberg, H., Frolow, F., Rabinovich, D. & Shakked, Z. (2001). DNA bending by an adenine–thymine tract and its role in gene regulation. *Proc. Natl Acad. Sci. USA*, **98**, 8490–8495.
15. Byun, K. S. & Beveridge, D. L. (2003). Molecular dynamics simulations of papilloma virus E2 DNA sequences: dynamical models for oligonucleotide structures in solution. *Biopolymers*, **73**, 369–379.
16. Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. (1982). Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. *J. Biol. Chem.* **257**, 14686–14707.
17. Karslake, C., Schroeder, S., Wang, P. L. & Gorenstein, D. G. (1990). 31P NMR spectra of an oligodeoxyribonucleotide duplex lac operator–repressor headpiece complex. *Biochemistry*, **29**, 6578–6584.
18. Grzeskowiak, K., Yanagi, K., Prive, G. G. & Dickerson, R. E. (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.
19. Tisne, C., Hantz, E., Hartmann, B. & Delepierre, M. (1998). Solution structure of a non-palindromic 16 base-pair DNA related to the HIV-1 kappa B site: evidence for BI-BII equilibrium inducing a global dynamic curvature of the duplex. *J. Mol. Biol.* **279**, 127–142.
20. Tisne, C., Hartmann, B. & Delepierre, M. (1999). NF-kappa B binding mechanism: a nuclear magnetic resonance and modeling study of a GGG → CTC mutation. *Biochemistry*, **38**, 3883–3894.
21. Djuranovic, D. & Hartmann, B. (2003). Conformational characteristics and correlations in crystal structures of nucleic acid oligonucleotides: evidence for sub-states. *J. Biomol. Struct. Dyn.* **20**, 771–788.
22. Djuranovic, D. & Hartmann, B. (2004). DNA fine structure and dynamics in crystals and in solution: the impact of BI/BII backbone conformations. *Biopolymers*, **73**, 356–368.
23. Bertrand, H., Ha-Duong, T., Fermandjian, S. & Hartmann, B. (1998). Flexibility of the B-DNA backbone: effects of local and neighbouring sequences on pyrimidine-purine steps. *Nucl. Acids Res.* **26**, 1261–1267.
24. Mauffret, O., Hartmann, B., Convert, O., Lavery, R. & Fermandjian, S. (1992). The fine structure of two DNA dodecamers containing the cAMP responsive element sequence and its inverse. Nuclear magnetic resonance and molecular simulation studies. *J. Mol. Biol.* **227**, 852–875.
25. Lefebvre, A., Mauffret, O., Hartmann, B., Lescot, E. & Fermandjian, S. (1995). Structural behavior of the CpG step in two related oligonucleotides reflects its malleability in solution. *Biochemistry*, **34**, 12019–12028.
26. Lefebvre, A., Mauffret, O., Lescot, E., Hartmann, B. & Fermandjian, S. (1996). Solution structure of the CpG containing d(CTTCGAAG)2 oligonucleotide: NMR data and energy calculations are compatible with a BI/BII equilibrium at CpG. *Biochemistry*, **35**, 12560–12569.
27. Tisne, C., Delepierre, M. & Hartmann, B. (1999). How NF-kappaB can be attracted by its cognate DNA. *J. Mol. Biol.* **293**, 139–150.
28. Roll, C., Ketterle, C., Faibis, V., Fazakerley, G. V. & Boulard, Y. (1998). Conformations of nicked and gapped DNA structures by NMR and molecular dynamic simulations in water. *Biochemistry*, **37**, 4059–4070.
29. Koudelka, G. B., Harrison, S. C. & Ptashne, M. (1987). Effect of non-contacted bases on the affinity of 434 operator for 434 repressor and Cro. *Nature*, **326**, 886–888.
30. Hegde, R. S. & Androphy, E. J. (1998). Crystal structure of the E2 DNA-binding domain from human papillomavirus type 16: implications for its DNA binding-site selection mechanism. *J. Mol. Biol.* **284**, 1479–1489.
31. Richmond, T. J. & Davey, C. A. (2003). The structure of DNA in the nucleosome core. *Nature*, **423**, 145–150.
32. Hartmann, B., Sullivan, M. & Harris, L. (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.
33. Case, A., Pearlman, D. A., Caldwell, J. W., Cheatham, T. E. III, Ross, W. S., Simmerling, C. L., *et al.* (1999). *Amber V.6*, University of California, San Francisco.
34. Cheatham, T. E., III, Cieplak, P., Kollman, P. A. *et al.* (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.
35. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690.
36. Ryckaert, J. P., Ciccotti, G. & Berendsen, H. J. C. (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.
37. Darden, T., York, D. & Pedersen, L. (1993). Particle mesh Ewald: an N log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092.
38. Cheatham, T. E., III, Miller, J. L., Fox, T., Darden, T. A. & Kollman, P. A. (1995). Molecular dynamics simulations on solvated biomolecular systems: the particule mesh Ewald methods leads to stable trajectories of DNA, RNA, and proteins. *J. Am. Chem. Soc.* **117**, 4193–4194.
39. Lavery, R. & Sklenar, H. (1988). The definition of generalized helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dyn.* **6**, 63–91.
40. Dickerson, R. E., Bansal, M., Calladine, C. R., Diekmann, S., Hunter, W. N., Kennard, O. *et al.* (1989). Definition and nomenclature of nucleic acid structure parameters. *EMBO J.* **8**, 1–4.

Edited by M. Yaniv

(Received 14 November 2003; received in revised form 22 March 2004; accepted 30 March 2004)