**Bending and Adaptability to Proteins of the cAMP DNA- Responsive Element: Molecular Dynamics Contrasted with NMR**

Sylvie Derreumaux and Serge Fermandjian
Département de Biologie et Pharmacologie Structurales, UMR 8532 Centre National de la Recherche Scientifique, Institut Gustave Roussy, 94800 Villejuif, France

**ABSTRACT** DNA bending is assumed to play a crucial role during recognition of the cAMP-responsive element (CRE) by transcription factors. However, diverging results have been obtained for the bending direction of the unbound double helix. The refined NMR structures present a bend directed toward the minor groove, while biochemical methods conclude that there is a bend toward the major groove. The present 10-ns molecular dynamics (MD) simulation of d(GAGATGACGTCTC)₂, which contains the octamer CRE in its center, was carried out with AMBER in explicit water and counterions. It shows that CRE is a flexible segment, although it is bent slightly toward the major groove (7°-8°) on the average. The MD structure agrees with both the biochemical results and unrefined NMR data. The divergence with the NMR refined structures suggests an improper electrostatic parameterization in the refinement software. The malleability of the central CpG is certainly the major contribution to the curving of the whole CRE segment in both the unbound and bound states. Comparison with the crystal structure of CRE bound to GCN4 shows that the deformation induced by the protein is concentrated mainly on the CpG step, rendering the bound structure of CRE closer to the structure of the 12–0 tetradecanoylphorbol-β-acetate-responsive element.

**INTRODUCTION**

Recognition of DNA by bZIP (basic leucine zipper) dimers is a delicate process, involving subtle structural and dynamical variations that are generally related to dramatic biological effects. Some bZIP dimers recognize the cAMP-responsive element (CRE) in promoters of ubiquitary genes and the 12–0 tetradecanoylphorbol-β-acetate (TPA)-responsive element (TRE) in protooncogene promoters. Proteins of the ATF-CREB subfamily bind to CRE and very weakly to TRE, whereas the yeast transcription factor GCN4 (general control protein 4) and the proteins belonging to the FOS-JUN subfamily bind to both CRE and TRE (Ryseck and Bravo, 1991; Hai and Curran, 1991; Hurst, 1994; Kim and Struhl, 1995). Such cross-bindings could be explained by the resemblances between CRE, d(TGACGTCA)₂, and TRE, d(TGACTCA)₂ (same TGA:TCA half-sites) and among the bZIP proteins (similar DNA binding sequences). Complex formation will further depend on the relative concentrations of the protein components, which can vary with the cell physiology. Some of them will be response productive and others nonproductive or even detrimental to cell life.

The specific recognition of DNA by proteins is highly dependent on the local and sublocal deformations occurring in the helix. According to Paolella et al. (1994), bending of DNA is the major structural factor influencing protein-CRE/TRE recognition. Actually, results of biochemical methods indicate a straight structure for TRE and a structure bent slightly (~10°) toward the major groove for CRE (Paolella et al., 1994, 1997; Sitlani and Crothers, 1998; Hockings et al., 1998; Sloan and Schepartz, 1998). Bending of CRE shortens both the distance and the difference in orientation between the half-sites TGA:TCA, seemingly reducing the structural difference between CRE and TRE and allowing the recognition of both CRE and TRE by the same bZIP dimers (Paolella et al., 1994; Keller et al., 1995). The DNA bending accompanying the complex formation varies as a function of the bZIP dimer. It could be responsible for a change in the dimer orientation in the transcription machinery, influencing the promotion of the gene transcription (Paolella et al., 1994; Leonard and Kerppola, 1998).

Our previous NMR refined molecular modeling argued in favor of a straight structure for TRE included in a 15-mer DNA and a ~30° bend toward the minor groove for CRE at the center of a 16mer DNA (Chaoui et al., 1999). A structural analysis of the NMR atomic coordinates of Conte et al. (1997) reported in the Protein Data Bank provided the same bend toward the minor groove.

The results from biochemical methods (Paolella et al., 1994, 1997; Sitlani and Crothers, 1998; Hockings et al., 1998; Sloan and Schepartz, 1998) and NMR restrained modeling, while they converge for TRE, are thus diverging for CRE. This motivates the present modeling study of CRE, in which no NMR constraints are applied. We believe this study to be more “realistic,” as it resorts to a molecular dynamics (MD) simulation and includes neutralizing Na⁺ and water explicitly. To our knowledge, it is the first MD study of the CRE DNA, allowing us to investigate its flexibility and its conformational substates. For the sake of comparison, the selected 16-mer DNA d(GAGATGACGT- CATCTC)₂ was kept the same as the one previously ana-
lyzed by NMR (Chaoui et al., 1999). A MD trajectory of 10 ns was carried out, as a previous study revealed that simulations longer than 5 ns were necessary to reach a full convergence of the trajectory for a system of that size (Young et al., 1997b). Calculations were performed with the AMBER software and the particle mesh Ewald implementation for the treatment of electrostatics. Previous studies have revealed a good agreement with experimental data, particularly regarding the sequence-specific structure and bending of DNA (Cheatham and Kollman, 1996, 1997; Young et al., 1997b; Young and Beveridge, 1998; de Souza and Ornstein, 1998; Sprouss et al., 1999). The recently modified version of the force field of Cornell et al., parm98, was used (Cheatham et al., 1999).

The results indicate that the CRE DNA, although flexible, is on the average slightly curved toward the major groove (7°–8°). The inherent flexibility of CRE, mostly concentrated at the central CpG step, must allow it to easily adjust its curvature in the different DNA-protein complexes. In the cocrystal with GCN4 (Keller et al., 1995), the largest helical deformation involves the central CpG, which exhibits a correlated increase in the roll and a decrease in the twist, bringing the TGA:TCA half-sites closer to each other (distance and orientation) and attenuating the structural differences between CRE and TRE. MD results fit our NMR experimental data but disagree with the refined NMR structures (Conte et al., 1997; Chaoui et al., 1999), highlighting the impact of the electrostatic parameterization on the refinement of the DNA structure and, more particularly, of the axis curvature.

MATERIALS AND METHODS

Choice of the DNA sequence

The MD simulation was performed on the 16-mer duplex d(GAGAT-GACGTCATCTC)2, which contains the CRE consensus sequence d(T-GACGTCAT), of interest at its center. Three factors led to the choice of this oligonucleotide: it has the minimum size needed to achieve stable interactions with the bZIP transcription factors (Talanian et al., 1992); the crystal oligonucleotide: it has the minimum size needed to achieve stable interaction. Previous studies have revealed a good agreement with experimental data, particularly regarding the sequence-specific structure and bending of DNA (Cheatham and Kollman, 1996, 1997; Young et al., 1997b; Young and Beveridge, 1998; de Souza and Ornstein, 1998; Sprouss et al., 1999). The recently modified version of the force field of Cornell et al., parm98, was used (Cheatham et al., 1999).

The results indicate that the CRE DNA, although flexible, is on the average slightly curved toward the major groove (7°–8°). The inherent flexibility of CRE, mostly concentrated at the central CpG step, must allow it to easily adjust its curvature in the different DNA-protein complexes. In the cocrystal with GCN4 (Keller et al., 1995), the largest helical deformation involves the central CpG, which exhibits a correlated increase in the roll and a decrease in the twist, bringing the TGA:TCA half-sites closer to each other (distance and orientation) and attenuating the structural differences between CRE and TRE. MD results fit our NMR experimental data but disagree with the refined NMR structures (Conte et al., 1997; Chaoui et al., 1999), highlighting the impact of the electrostatic parameterization on the refinement of the DNA structure and, more particularly, of the axis curvature.

Structural analysis

CURVES version 5.2 was used to analyze the DNA structures stored every picosecond of MD (Lavery and Sklenar, 1988). CURVES fits a curved or a straight global helix axis that follows the double helix. It allows both global and local geometrical analysis, providing the values of the helical parameters and of the backbone dihedral angles. This work refers to the local parameters output by CURVES. To avoid end effects, only the 14 central base pairs were taken into account for the analysis. For the MD simulation (10 ns), each parameter was measured 10,000 times.

MD results were compared to the experimental NMR and x-ray data on unbound and bound CRE, respectively. For the 16-mer duplex d(GAGAT-GACGTCATCTC)2, one NMR refined structure (Chaoui et al., 1999) and one crystal structure of its complex with the GCN4 transcription factor (2.2 Å) (Keller et al., 1995) (PDB ID: 2dgc) are available. Another NMR study concerns a different oligomer with CRE at its center, d(CAT-GTGGACGCATATG)2 (Conte et al., 1997) (PDB ID: 1saa). Yet the loose NMR constraints used by these authors relative to the helical parameters (internucleotide distances) and phosphate positions (ε dihedral) result in a poor experimental assessment of these DNA morphological parameters. Thus comparisons were made with the NMR experimental structure of Chaoui et al. (1999), CRE_NMR; with the crystal structure (Keller et al., 1995), CRE_GCN4; and with the NMR data of Conte et al. (1997) for the sugar phases. Concerning the DNA bending, data of reference were provided by biochemical, NMR, and crystallography experiments. We calculated the local parameters and global bending describing the experimental structures, using the published PDB coordinates and CURVES 5.2. The canonical B-DNA and A-DNA conformations cited in the text are the ones generated by the AMBER software (Arnott et al., 1972).
RESULTS AND DISCUSSION

Molecular dynamics

A MD trajectory of 10 ns was simulated with the 16-mer duplex d(GAGATGACGTCATCTC)_2, using the AMBER software and the recent parm98 force field (Cheatham et al., 1999). According to Young et al. (1997b), simulations longer than 5 ns are necessary to reach a full convergence of the trajectory for a system with DNA of the above-mentioned size and containing explicit monovalent cations. The atom coordinates were saved and analyzed every picosecond. The results take into account the trajectory after 500 ps from its starting point (i.e., data collected through 9.5 ns).

In Fig. 1a are plotted the root mean square deviations (rmsds) as a function of time, calculated for the coordinates of heavy atoms in the 14 central residues of the DNA, between the instantaneous structures extracted from the molecular dynamics trajectory (MD), CRE_MD, and the canonical B DNA conformation and A DNA conformation. One can see that after 500 ps, the DNA has globally relaxed from the starting conformation and follows its own dynamics.

The DNA conformation calculated with the parm98 force field looks like more the canonical B form than the canonical A form, with rmsd values of nearly 2.5 Å and 5.5 Å, respectively. In comparison, calculations performed with the parm95 force field (unpublished results) provide a DNA conformation intermediate between the canonical B and A DNAs, with rmsd values of nearly 4.5 Å.

In the Appendix are reported the averaged values and the standard deviations over the MD trajectory (0.5–10 ns) of the interpair, pair-axis, base-base, and backbone structural parameters of the CRE oligomer. The DNA is globally stabilized in a B form, with averaged twist, rise, inclination, x displacement, and sugar phase values of 33.7°, 3.4 Å, 0.0°, −1.0 Å, and 147°, respectively, confirming that the new parm98 force field leads to DNA conformations that are more consistent with the experimental data compared to the parm95 force field (Young et al., 1997b; de Souza et al., 1998; Cheatham et al., 1997, 1999; Cieplak et al., 1997).

Backbone conformation

Sugar pucker. The average sugar phase angle in the CRE_MD structure is 147°. This value fits the B-DNA canonical value (162°) (Saenger, 1984) better than the value provided by the elder parm95 force field does (~120°-125°) (Cheatham et al., 1999). The phase angle as a function of the nucleotide (Fig. 2) follows the well-known sequence dependence, with values higher at purines (~160°) than at pyrimidines (~135°) (Poncin et al., 1992). The transitions toward the Northern state (C3'-endo, ~18°) are not numerous, and when they occur, they merely last more than 200 ps. Exceptions are the sugars of A12 (in the half-site TCA) and of C14 (outside the consensus sequence), which display transitions of nearly 500 ps between 1200 ps and 1700 ps and of 700 ps between 9000 ps and 9700 ps,
respectively. These two nucleotides are the only ones surrounded by pyrimidines.

**Phosphate conformations.** Numerous coupled variations of the dihedral angles \( \epsilon \) and \( \zeta \) associated with BI/BII transitions in the phosphate groups are observed during the dynamics, all over the DNA sequence (BI: \( \epsilon \) in *trans* and \( \zeta \) in *gauche*; BII: \( \epsilon \) in *gauche* and \( \zeta \) in *trans*) (Grzeskowiak et al., 1991; El antri et al., 1993; Gorenstein, 1994). The angle values are \( \sim 180^\circ \) (280°) for \( \epsilon \) and \( \sim 270^\circ \) (150°) for \( \zeta \) in the BI (BII) state. The difference \( \epsilon - \zeta \) is nearly \( -90^\circ \) for BI and \( 130^\circ \) for BII and is convenient for following the transitions. The value of \( \epsilon - \zeta \) averaged over time is proportional to the BII population percentage (Gorenstein, 1994). The BII population percentage averaged over the CRE_MD structure (21%) is in very good agreement with the percentage provided by NMR for solution DNAs and by x-ray for crystal DNAs (20–30%) (Gorenstein, 1994; Winger et al., 1998). Again the results are significantly improved compared to the results obtained with the parm95 force field (\( \sim 10\% \) of BII) (Cheatham et al., 1999).

The histogram of the BII population as a function of the nucleotide (Fig. 3 *a*) indicates that a period of 10 ns is sufficient to render the sequence dependency (similar variations along the two palindromic strands). Nevertheless, a significant difference (up to 20%) is found between the BII population of some symmetrical phosphates (for example, between G6pA7 and G22pA23). A longer MD simulation is needed to sample the phosphate conformations compared with the helical parameters (\( \sim 10\% \) versus \( \sim 2\% \) ns). This is due to the potential energy barrier between the BI and the BII states and to the dependence of the phosphate conformations on the diffusion of water molecules (Winger et al., 1998).

The largest BII percentages are found within the CRE consensus segment, that is, the half-sites TGA:TCA and the central CpG linker: 70% at G6pA7/G22pA23, 45% at C8pG9/C22pA23 and C11pA12/C27pA28. With exception of G6pA7/G22pA23 and T5pG6/T21pG22, where BII rates are unexpectedly very high and low, respectively, the hierarchy of BII percentages (Fig. 3 *a*) is similar to YpR > RpR > YpY > RpY (where Y and R stay for pyrimidine and purine, respectively), obtained from statistical analysis of NMR and crystallographic data (El antri et al., 1993; Gorenstein, 1994; Winger et al., 1998; Bertrand et al., 1998).

**Helical parameters**

The twist angle of CRE_MD averaged over the whole sequence, 33.7°, is that expected for B-DNAs in solution (34°) (Ulyanov and James, 1995). The averaged rise ranges between 3.2 Å (A4pT5) and 3.8 Å (C6pG7), and the averaged twist varies from 27.2° (A4pT5) to 44.4° (C6pG7) (Appendix). The averaged inclination and \( x \) displacement vary slightly, with values ranging from \( -1.8^\circ \) and 2.6°, and from \( -1.2 \) Å and \( -0.7 \) Å, respectively. The base-base buckle and

![FIGURE 3](https://example.com/figure3.jpg)
propeller twist parameters are submitted to important variations, with averaged values between $-13.9^\circ$ ($T_{15}$-$A_{18}$) and $14.1^\circ$ ($A_{2}$-$T_{31}$), and between $-16.0^\circ$ ($C_{14}$-$G_{19}$) and $-3.7^\circ$ ($C_{11}$-$G_{22}$), respectively. No disruption of hydrogen bonds is observed. In the phosphodiester backbone, most variable dihedral angles are $\zeta$ ($\Delta = 100^\circ$), $\epsilon$ ($\Delta = 70^\circ$), sugar phase ($\Delta = 50^\circ$), and $\chi$ ($\Delta = 30^\circ$). In contrast, the $\alpha$, $\beta$, and $\gamma$ angles are roughly constant, staying in $g^-$, $t$, and $g^+$ conformations, respectively, along the sequence. At last, both the minor and the major groove widths are typical of B-DNAs, with averaged values over CRE of 6.3 Å and 12.1 Å, respectively. Yet the grooves are narrower at the C$_8$G$_9$ step in the center of CRE, with width values of nearly 5 Å (minor) and 11 Å (major).

**DNA bending, roll, tilt**

There is a sampling of crescent-shaped DNA structures with helix curvatures mostly concentrated on the central CpG. The DNA global curvature is defined by CURVES 5.2 as the angle between the local helical axes of the penultimate base pairs $A_2$-$T_{31}$ and $T_{15}$-$A_{18}$ (Lavery and Sklenar, 1988) (Fig. 4a, left). The global curvature (Fig. 4b) varies from 0° to 51° (14 bp), with an averaged value of 16.1° and a SD of 8.5°. The bending direction is given by the Atip component of the kink angle (Fig. 4a, right). The kink is defined as the angle between the two optimal linear axes for the DNA half-sides, fragments $A_2$-$T_{31}$ to $C_8$-$G_{25}$ and $G_9$-$C_{24}$ to $T_{15}$-$A_{18}$, and the Atip is the component of the kink angle in the plane perpendicular to the long base-pair axis of the central base-pair step. The Atip component is negative or positive for a bending of the oligomer toward the minor or major groove, respectively. The variation of the Atip as a function of time (Fig. 4c) shows that the CRE curvature oscillates between the major and the minor grooves, with an intermediate straight conformation, revealing no apparent anisotropy (examples of structures are shown in Fig. 5, left and right). The averaged Atip value reported as a function of the simulation length (Fig. 4d) reveals that a trajectory of more than 5 ns is necessary to get a convergence of the DNA global bending. The distribution of the Atip values calculated over the entire trajectory (0.5–10 ns) is roughly Gaussian (data not shown), indicating that a trajectory of 10 ns is sufficient to sample the bending of a DNA 16-mer.

An understanding of why and how bending of CRE occurs requires examination of local perturbations in the double helix. Typical parameters responsible for DNA bending include the roll and, to a lesser extent, the tilt (EMBO Workshop, 1989; Young et al., 1995; Suzuki and Yagi, 1995; Dickerson, 1998). The roll is positive or negative for pinching toward the major or the minor groove of the DNA, respectively. The tilt is positive or negative for a pinching toward the second or the first strand backbone of the DNA, respectively. The MD structure shows loci of curvature, mainly in the central consensus sequence. The pyrimidine-purine steps $T_5$-$G_6$, $C_6$-$G_9$, and $C_{11}$-$pA_{12}$ display averaged positive roll values of 8.9°, 5.7°, and 8.6°, respectively, and high standard deviation values of 6.9°, 7.3°, and 6.8°, respectively (Fig. 6a). The major tilt deformations are found at the $T_5$-$G_6$ and $G_9$-$pA_3$ steps (2.4° ± 5.5° and $-3.2^\circ$ ± 4.5°), at the symmetrical $T_{10}$-$pC_{11}$ and $C_{11}$-$pA_{12}$ steps (4.4° ± 4.7° and $-3.4^\circ$ ± 5.6°), and at $C_8$-$G_9$ (1.0° ± 5.7°) (Fig. 6b). The CpG and TpG/CpA steps appear as the most malleable steps, which conforms to the crystallographic (Young et al., 1995; Suzuki and Yagi, 1995; Dickerson, 1998) and NMR (Ulyanov and James, 1995; El antri et al., 1993; Lefebvre et al., 1995) data; the effect is related to a lower stacking energy in these steps (Bertrand et al., 1998).

The small tilts at $T_5$-$G_6$ + $G_9$-$pA_3$ and $T_{10}$-$pC_{11}$ + $C_{11}$-$pA_{12}$ are in phase and additive (Dickerson, 1998) with the roll at $C_8$-$G_9$, resulting in a total kink angle of 7.5° for

![Figure 4](image-url)
the CRE consensus segment. Thus the CRE DNA segment is globally curved on the average by 7°–8° toward the major groove at its center, i.e., toward the dimer bZIP protein partner, with a global “C” form. The positive rolls (4.2°) at the A2pG3 and C14pT15 steps ending the DNA duplex are roughly in antiphase with the positive roll at the central CpG9, as these steps are separated by 6 bp. As a consequence the Atip is reduced, explaining the apparent isotropy of the Atip curve (Fig. 4c). In the binding half-sites, the high positive rolls at the T5pG6 and C11pA12 steps create local bends toward the major groove. Globally, the molecule is S-shaped in the plane perpendicular to the Atip one, with each half-site precurved toward one monomer of the bZIP protein.

The plot of rolls averaged over the MD structures and classified as a function of their Atip value (Atip > 10°, < −10°, or ≈ 0°) reveals a strong correlation between the roll of the central CpG and the Atip (Fig. 7). The structure of CpG modulates the overall bending: an increase in roll changes a global bending toward the minor groove (Atip < −10°) to a global bending toward the major groove (Atip > 10°); a null roll is related to a global bending toward the minor groove because of the presence of positive rolls at the A2pG3 and C14pT15 steps 6 bp from CpG. Regarding the tilts, no clear correlation with the Atip was found.

Comparison with experimental data

CRE bending: MD versus biochemical methods

The consensus CRE segment in CRE_MD (7–8°) presents an averaged curvature toward the major groove very similar to that obtained (nearly 10°) by phasing analysis (Paolella et al., 1994), cyclization kinetics (Hockings et al., 1998), and ligation ladder (Sloan and Schepartz, 1998). It can be noted, however, that the DNA bending is not due to the sole rolls...
The current MD data and the NMR data of Conte et al. (1997) related to the CRE segment is also good. Sugars at C\textsubscript{14}/C\textsubscript{30} show a preference for the Eastern state (O4'-endo, ~90°) in CRE\textsubscript{NMR} and for the C1'-exo state (~126°) in CRE\textsubscript{MD}. Actually, during the MD trajectory, these two sugars keep the C2'-endo conformation most of the time, with phase angles of nearly 150°. As they transit from time to time into the Northern conformation, their averaged conformation lies in the range of C1'-exo puckers. The Eastern state provided by NMR could better correspond to an average of Southern and Northern states. Whatever the real conformation adopted by residues C\textsubscript{14}/C\textsubscript{30}, there is a disagreement between the MD calculations and the NMR measurements.

Regarding the phosphate conformations, the BII populations predicted by MD simulation are close to those estimated from NMR \textsuperscript{31}P chemical shifts, using the Gorenstein (1994) relationship (Fig. 3 b). For the average over the sequence, both NMR and MD provide 21% of BII population, whereas for the BII population at each step, the absolute difference does not exceed 10%. Exceptions are G\textsubscript{6}pA\textsubscript{7}/G\textsubscript{26}pC\textsubscript{27} and their complementary steps T\textsubscript{10}pC\textsubscript{11}/T\textsubscript{26}pC\textsubscript{27}. For the first steps, relative BII populations of 70% and only 25% are predicted by MD and NMR, respectively, while for the second steps MD predicts a smaller BII population (~20%). Globally, the sequence dependence of the BII population is found to be more pronounced by MD calculations than by NMR measurements (1 SD = 21% versus 12%).

Helical parameters and bending propensity. The current MD simulation disagrees with the recently reported NMR restrained molecular modelings of both Conte et al. (1997) and Chaoui et al. (1999). These show a curvature of 30° (16-mer) toward the minor groove and rolls negative at CpG (~19°/−10°) and null or negative at TpG:CpA (0°/−12°). Actually, the plot of rolls as a function of dinucleotide steps for CRE\textsubscript{NMR} (Chaoui et al., 1999) is the opposite of that of CRE\textsubscript{MD} (Fig. 6 a), but this also proves to be true for the tilts at G\textsubscript{6}pA\textsubscript{7} and T\textsubscript{10}pC\textsubscript{11}. It seems that tilts adapt themselves to smooth the bending caused by rolls. In addition, the averaged twist in CRE\textsubscript{MD} is smaller compared with that found in CRE\textsubscript{NMR} (33.7° versus 35.4°). Yet the profiles of the twist sequence dependences are very similar (Fig. 6 c), with the greatest differences being observed at CpG and TpG, and C\textsubscript{14}pA\textsubscript{12}. The values of CRE\textsubscript{NMR} are within the averaged values ± 1 SD of CRE\textsubscript{MD}.

Influence of software and protocol on NMR structure determination. The curvature of CRE predicted by MD is similar to that derived from biochemical data but differs from that provided by NMR refinement methods.

DNA conformations resolved by NMR refinement depend on different factors such as measurements and methodology. The latter includes the modeling software and the refinement protocol. The helical parameters are determined by the internucleotide distances obtained from NOE, data which are refined through back-calculations of theoretical

![FIGURE 7 Rolls in (AGATGACGTCATCT)\textsubscript{2} as a function of the sequence, averaged over the MD conformations, classified as a function of their Atip value: Atip > 10° (solid line), Atip < 10° and > −10° (long dashed line), and Atip < −10° (short dashed line).](image-url)
NOESY spectra, using the NOE Simulate routine of INSIGHT II (MSI) (Boelens et al., 1989; Lefebvre et al., 1995, 1996; Chaoui et al., 1999). Moreover, the bending can be strongly influenced by the force field. Its impact can be tested by comparing, for instance, the rolls in the minimal energy structures provided by JUMNA, using Flex and parm95. Conditions are the same as the ones used for the CRE refinement by Chaoui et al. (1999) (dielectric function and screening of the phosphate charges), but the NMR constraints are excluded. One finds that rolls are modulated by the force field (Fig. 8 a), especially at the CpG and TpG/CpA steps, where JUMNA-Flex provides negative rolls and JUMNA-parm95 provides positive ones. Replacing parm95 with parm98 (force-field parameters modified for three dihedral angles, no change of the partial charges) in JUMNA does not significantly alter the rolls: they differ by less than 2° at each step of the CRE sequence. As in most softwares dedicated to the NMR refinement of molecular structures, the DNA environment in JUMNA is simulated in an implicit way: the dielectric effect of water is modeled by a sigmoid function with a slope value of 0.16 and a plateau of 78, and the screening effects of counterions on phosphates are taken into account by reducing the charge of the phosphate groups to a value of $-0.5e$ (Lavery et al., 1995).

Using different slope values of the dielectric function (0.16 and 0.356) and varying the apparent charge of the phosphates ($-0.5e$, $0.0e$, and $-1.0e$) for the minima provided by JUMNA-parm95 shows that the roll values increase parallel to the dielectric function and when the phosphate charge tends to zero. A MD simulation with the same parm95 force field (unpublished results) with an explicit consideration of the solvent provides a similar curve, but with more positive values of rolls (Fig. 8, b and c). However, we cannot conclude that it is possible to find an implicit electrostatic parameterization that would correctly simulate, whatever the DNA sequence, the influence of the solvent on the DNA structure. Actually, structure alterations caused by specific localization of water molecules or hydrated ions (Dickmann, 1987; Rouzina and Bloomfield, 1998; Lyubartsev and Laaksonen, 1998; Winger et al., 1998; McFail-Isom et al., 1998; Shui et al., 1998) cannot be simulated by varying the apparent phosphate charge or by using a dielectric function.

From the above results one can conclude that the electrostatic parameterization of the software exerts a significant effect on DNA bending, through both the force field
(higher polarity of the atomic charges for parm95/parm98 than for Flex) and the solvent model. As far as CRE is concerned, a major factor could be the electronegativity of the major groove along the consensus d(TGACGTCA)$_2$ helix. Actually, CRE displays a succession of intrastand electronegative pockets, at 5’-T$_2$pG$_3$-3’, 3’-T$_2$pG$_3$-5’, 5’-G$_9$pT$_{10}$-3’, and 3’-G$_{25}$pT$_{21}$-5’, connected one to another by interstrand electronegative pockets 5’-G$_9$pA$_{1}$-3’, 5’-C$_{23}$pG$_{23}$-3’, and 5’-T$_{110}$pC$_{11}$-3’ (Young et al., 1997a). The succession of thymine O4 and guanine O6 and N7 atoms creates a continuous negative channel along the major groove, which could push away subsequent base pairs and induce negative (or less positive) rolls. The repelling of base pairs would in this case decrease as the dielectric effect increases.

The NMR restraints introduced successively on the sugar phases and on the $\varepsilon - \zeta$ differences did not significantly modify the roll values. In contrast, the restraints on $^1$H-$^1$H distances, particularly the H6/8-H6/8 distances, have both “pulled” the rolls of CpG and TpG/CpA toward strong negative values and bent CRE toward the minor groove. This occurs whatever the force field used, Flex or parm95 (Fig. 9) (Chaoui et al., 1999). Although the interbase H6/8-H6/8 distance between subsequent bases is related to the roll value, it is not sufficient to determine precisely the roll because, for the same H6/8-H6/8 distance, different positive or negative rolls are possible, depending on the twist value. Actually, both the H6/8-H6/8 distances and the H3’-H6/8 distances (which are not always amenable to 2D NMR experiments for signal overlapping reasons) are needed to determine the rolls univocally (Lefebvre et al., 1997). For instance, in the CRE NMR restrained structures obtained with JUMNA-Flex and JUMNA-parm95 by Chaoui et al. (1999), the roll values of the central CpG are different, although these structures present similar H6/8-H6/8 distances (Fig. 10). On the other hand, the roll values in the CRE structures determined by Conte et al. (1997) and Chaoui et al. (1999) (with JUMNA-parm95) are similar, although the H6/8-H6/8 distances are very different. In these structures the restraints on the H3’-H6/8 distances strongly related to the rolls are severely missed (Lefebvre et al., 1997). Perhaps, whatever the starting structure, the software reacts to the adjunction of the H6/8-H6/8 distance restraints by “pushing” the CRE conformation toward a region of the conformational space that is “more convenient” for it (i.e., abnormally negative rolls but correct twists).

Another obvious fault of the refinement procedure that could influence the rolls concerns the internal molecular motions, so that the refinement procedure provides erroneous internucleotide $^1$H-$^1$H distances, especially at the most flexible steps (TpG:CpA and CpG).

The disagreement on the bending of CRE_NMR and CRE_Md could also result from the solvent conditions, which are different in the NMR and MD experiments, with regard to both the ionic strength (MD $= 0.5$; NMR $= 0.1$) and the nature of the ions (MD: Na$^+$; NMR: buffer + purification + impurities $\rightarrow$ Na$^+$, Li$^+$, K$^+$, and perhaps also traces of divalent cations like Mn$^{2+}$, Mg$^{2+}$, Zn$^{2+}$). Different monovalent ions could induce different local DNA deformations, as crystallographic (Tereshko et al., 1999) and MD (Lyubartsev and Laaksonen, 1998) studies performed with various monovalent cations argue that they bind to different specific DNA sites (for instance, Na$^+$ preferentially binds to the minor groove and Li$^+$ binds to the phosphate oxygens). In a recent MD study, Young and Beveridge (1998) have shown that the nature and the concentration of monovalent cations can have a significant impact on the DNA curvature: replacing neutralizing Na$^+$ with K$^+$ and KCl doubles the bending of a 25-bp DNA segment containing an A-tract. Moreover, gel retardation (Diekmann, 1987), theoretical investigations (Rouzina and...
conditions diverge from the MD conditions. For instance, the measurements of cyclization kinetics (Hockings et al., 1998) have been made in the presence of KCl (50 mM) and MgCl$_2$ (5 mM), and the ligation ladder experiments could have been performed in the presence of Mg$^{2+}$ impurities (Sloan and Schepartz, 1998).

In the end, the influence of the phosphorus NMR restraints has also to be considered. Such restraints could force the steps involved in a BI/BII equilibrium to adopt time-averaged $\epsilon - \zeta$ values devoid of physical sense. A forced backbone may strongly alter the shape of the helix because of the relation that exists between $\epsilon - \zeta$ and the helical parameters, including the rolls that are strongly involved in the DNA curvature (El antri et al., 1993; Hartmann et al., 1993; Gorenstein, 1994; Lefebvre et al., 1996; Bertrand et al., 1998). Actually, one could generate different initial structures with each phosphate either in the BI or in the BII conformation, making a combinatorial search with all of the steps susceptible to adopting both BI and BII states. One could then perform an energy minimization without constraint on the $\epsilon$ and $\zeta$ angles on the different initial structures and obtain final structures with different patterns of BI/BII phosphates. Finally, one could compare the averaged $\epsilon - \zeta$ values to the NMR data (Lefebvre et al., 1996; Tisné et al., 1998).

**CRE in the complex with GCN4**

The rmsd between $\langle$CRE_MD$\rangle$ (averaged MD structure without further minimization) and CRE_GCN4 is only 1.3 Å (Table 1 and Fig. 5, center). This is lower compared to the averaged rmsd of the instantaneous MD structures (~2 Å) (Fig. 1b), probably because, at a first approximation, it does not take into account the vibrational motions of the DNA molecule. The good agreement between the simulated CRE structure and the crystal structure of CRE bound to GCN4 indicates that the binding of the protein does not significantly affect the DNA, in agreement with the biochemical results (Paolella et al., 1994, 1997; Hockings et al., 1998; Bloomfield, 1998), and crystallographic data (McFail-Isom et al., 1998) suggest that divalent cations undergo sequence-specific interactions with DNA that have a particular impact on the DNA bending. Thus the nonequivalent ionic conditions between the NMR and MD studies could in principle contribute to the appearance of diverging results regarding the bending of CRE. As the MD results agree with the NMR data regarding the conformation of the phosphodiester backbone (sugar phases and phosphate conformations), this would mean that the diverging ionic conditions would only influence the helical parameters, including the rolls. But it is interesting to note that the biochemical data on the bending of CRE support the MD results, although the experimental

![FIGURE 10 H6/8-H6/8 distances (a) and rolls (b) in the CRE consensus sequence (TGACGTCA)$_2$, as a function of the sequence, for the NMR refined structures obtained by Conte et al. (1997) (solid line) and by Chaoui et al. (1999) with the Flex (long dashed line) and the parm95 (short dashed line) force fields.](image)

**TABLE 1** Rmsds (in Å) between the averaged MD, experimental, and canonical structures

<table>
<thead>
<tr>
<th></th>
<th>(CRE_MD)</th>
<th>B DNA</th>
<th>A DNA</th>
<th>CRE_NMR</th>
<th>CRE_GCN4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(CRE_MD)</td>
<td>1.5</td>
<td>4.8</td>
<td>2.2</td>
<td>1.4</td>
<td></td>
</tr>
<tr>
<td>B-DNA</td>
<td>2.2</td>
<td>5.4</td>
<td>1.6</td>
<td>2.1</td>
<td></td>
</tr>
<tr>
<td>A-DNA</td>
<td>6.1</td>
<td>7.5</td>
<td>6.0</td>
<td>4.3</td>
<td></td>
</tr>
<tr>
<td>CRE_NMR</td>
<td>3.1</td>
<td>2.1</td>
<td>8.4</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>CRE_GCN4</td>
<td>1.3</td>
<td>2.7</td>
<td>5.8</td>
<td>3.6</td>
<td></td>
</tr>
</tbody>
</table>

Root mean square deviations (rmsd) between the structure of (AGAT-GACGTCACTC), averaged over the MD trajectory (9.5 ns), (CRE_MD); the canonical B DNA and A DNA conformations; the NMR refined structure by Chaoui et al. (1999), CRE_NMR; and the conformation of the DNA in the cocrystal with GCN4 (Keller et al., 1995), CRE_GCN4. The upper right part of the table gives the rmsd values calculated for the CRE consensus sequence, i.e., the central octamer (TGACGTCA)$_2$; and the lower left part of the table gives the rmsd values for the whole 14-mer.
CpG stays within the MD averaged value reached with little consumption of energy, as a null roll at the complex with BP1 (Paolella et al., 1994) could also be mal energy. The straight conformation adopted by CRE in deformation induced by GCN4 is accessible through ther- MD values almost null in CRE_MD (7.1° versus 1.8° and 0.8°). How- ever, a higher value for CpG (12.3° versus 5.7°) and a slightly smaller value for TpG/CpA (7.3° versus 8.9° and 8.6°) (Fig. 6). A significant difference consists of rolls at G6pA7 and T10pC11 that are positive in the crystal and almost null in CRE_MD (7.1° versus 1.8° and 0.8°). However, the roll values of CRE_GCN4 stay within the averaged MD values ± 1 SD, suggesting that the small DNA bending deformation induced by GCN4 is accessible through thermal energy. The straight conformation adopted by CRE in the complex with BP1 (Paolella et al., 1994) could also be reached with little consumption of energy, as a null roll at CpG stays within the MD averaged value ± 1 SD. Rolls with averaged values and SD higher at the CpG and TpG/ CpA steps reflect the malleability of the helix at both the linker and the half-site positions. The roll differences between CRE_MD and CRE_GCN4 are caused by a drift of the kinks from TpG/CpA to CpG induced by the protein, thus confirming the gel results (Sloan and Schepartz, 1998).

Differences in tilts are observed only for G3pA4 and the T10pC11 steps (Fig. 6 b). The twist angle of CRE_MD averaged over the whole sequence (33.7°) is similar to that found in CRE_GCN4 (33.8°). However, although similar for the residues lying outside the CRE consensus octamer, the twists are very different inside CRE, especially at the CpG step (29.9° in CRE_GCN4 versus 44.4° in CRE_MD). Such a twist variation at CpG (correlated to the roll) is not accessible through thermal energy (Fig. 6 c) and must be induced by work done during complexation.

The comparison of CRE with TRE seemed interesting, as both are stable partners of GCN4. As far as the distance between the half-sites TGA/TCA and their relative orientations are concerned, the structural difference between CRE and TRE is reduced upon protein binding, as TRE keeps a straight conformation (Ellenberger et al., 1992; Keller et al., 1995). The fact that a greater positive roll is mechanically correlated to a smaller twist (Hartmann et al., 1993; Bertrand et al., 1998) suggests that the coupled change in the central CpG roll and twist during the formation of the complex of CRE with GCN4 follows a general mechanism in DNA that allows a reduction of the structural difference between promoters with half-sites that diverge by the spacer length.

CONCLUSION

A long molecular dynamics simulation of 10 ns was made on a 16mer DNA sequence containing the CRE consensus sequence. Results were compared with those provided by NMR and biochemical methods. The B-DNA conformation presented by CRE oscillates between a helix curved toward the major groove and a helix curved toward the minor groove, through an intermediate straight form. However, as a final result, the averaged helix is bent slightly toward the major groove. Although this finding is consistent with the biochemical results, it disagrees with the refined NMR structures, where the bending is directed toward the minor groove. This reversion is imputed to an improper electro- static parameterization in the NMR structure refinement softwares, inasmuch as the NMR data agree with the MD simulation results. In contrast, the MD structure appears to be very similar to the crystal structure of CRE bound to GCN4. In particular, the rolls (often involved in the DNA curvature) are roughly conserved in the complex. These parameters could be the structural determinants of the CRE recognition by GCN4. A large part of the helix deformation induced by GCN4 is concentrated at the central CpG step that links the TGA:TCA half-sites. CpG undergoes a concerted increase in roll and a decrease in twist that reduces both the distance between the TGA:TCA half-sites and their difference in orientation. As a result, the structure of CRE bound to GCN4 is more like the structure of TRE, which remains straight in its complex with GCN4.

Thus the current MD simulation reveals the high flexi- bility of the CRE consensus sequence, reflecting its ability to accommodate itself to various bZIP proteins. The impact of the central CpG dinucleotide on the conformation of CRE is clearly identified: the inherent bending and malleability of CpG are the major contributions to the curving and flexibility of CRE. The absence of CpG at the center of CRE could be the structural explanation for the observation by biochemical methods that TRE is unable to bind to such varied bZIP partners and in such numbers as CRE does.
APPENDIX

REFERENCES


Sprous, D., M. A. Young, and D. L. Beveridge. 1999. Molecular dynamics studies of axis bending in d(G5-(GA4T4C)2-C5) and d(G5-(GT4A4C)2-C5): effects of sequence polarity on DNA curvature. *J. Mol. Biol.* 285:1623–1632.


