
RESEARCH ARTICLES

Structure Prediction of a Complex Between the Chromosomal Protein HMG-D and DNA

Alexander Balaeff,^{1,2} Mair E.A. Churchill,^{2,3} and Klaus Schulten^{1,2,4*}

¹Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, Illinois

²Center for Biophysics and Computational Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois

³Department of Cell and Structural Biology, University of Illinois at Urbana-Champaign, Chemistry and Life Sciences Laboratory, Urbana, Illinois

⁴Department of Physics, University of Illinois at Urbana-Champaign, Urbana, Illinois

ABSTRACT Non-histone chromosomal proteins are an important part of nuclear structure and function due to their ability to interact with DNA to form and modulate chromatin structure and regulate gene expression. However, the understanding of the function of chromosomal proteins at the molecular level has been hampered by the lack of structures of chromosomal protein–DNA complexes. We have carried out a molecular dynamics modeling study to provide insight into the mode of DNA binding to the chromosomal HMG-domain protein, HMG-D. Three models of a complex of HMG-D bound to DNA were derived through docking the protein to two different DNA fragments of known structure. Molecular dynamics simulations of the complexes provided data indicating the most favorable model. This model was further refined by molecular dynamics simulation and extensively analyzed. The structure of the corresponding HMG-D–DNA complex exhibits many features seen in the NMR structures of the sequence-specific HMG-domain–DNA complexes, lymphoid enhancer factor 1 (LEF-1) and testis determining factor (SRY). The model reveals differences from these known structures that suggest how chromosomal proteins bind to many different DNA sequences with comparable affinity. *Proteins* 30:113–135, 1998. © 1998 Wiley-Liss, Inc.

Key words: HMG proteins; protein–DNA complex; HMG-box; nonsequence-specificity; molecular dynamics

INTRODUCTION

Chromatin structure and gene regulation are orchestrated by numerous macromolecular interactions between proteins and nucleic acids that form different types of complexes. The best understood of

these protein–DNA interactions is the binding of the sequence-specific transcription factors to their cognate DNA sites.^{1–3} In contrast, the DNA-binding properties of abundant chromosomal proteins are rather poorly understood.⁴ Chromosomal proteins are essential for cellular regulation. They bind to DNA in a relatively sequence-independent manner, although they may have slight preferences for some sequences because of sequence-dependent structural and dynamic properties.^{3,5,6} Such nonsequence-specificity makes the structural studies of chromosomal proteins technically more challenging than those of their sequence-specific counterparts.^{6,7}

Several groups of homologous chromosomal proteins called High Mobility Group (HMG) proteins take part in general modulation of chromatin structure and gene activity (reviewed in ref. 8). Of these, the HMG1 chromosomal proteins are typified by vertebrate HMG1, its close relative, HMG2, HMG-D of *Drosophila melanogaster*, and NHP6 of *Saccharomyces cerevisiae*.^{8–13} These are abundant proteins with still somewhat poorly defined functions.⁸ They bend DNA upon binding and stabilize bent and supercoiled DNA structures and thus may facilitate the formation of higher order nucleoprotein complexes, DNA packaging, or interactions with other proteins in chromatin.^{5,14,15} HMG-D is thought to associate with a loosely condensed state of chromatin during early *Drosophila* embryogenesis.¹² Recent studies suggest that HMG1 may interact with transcription factors and possibly also nucleosomes¹⁶

Contract grant sponsor: Roy J. Carver Charitable Trust; Contract grant sponsor: National Institute of Health; Contract grant number: PHS 5 P41 RR05969-04; Contract grant sponsor: National Science Foundation (K.S., A.B.); Contract grant number: BIR-9318159; Contract grant sponsor: National Institute of Health, Shannon Award (M.E.A.C); Contract grant number: GM52100.

*Correspondence to: Klaus Schulten, Beckman Institute, 405 N. Mathews Ave., Urbana, IL 61801.

Received 2 July 1997; Accepted 9 September 1997

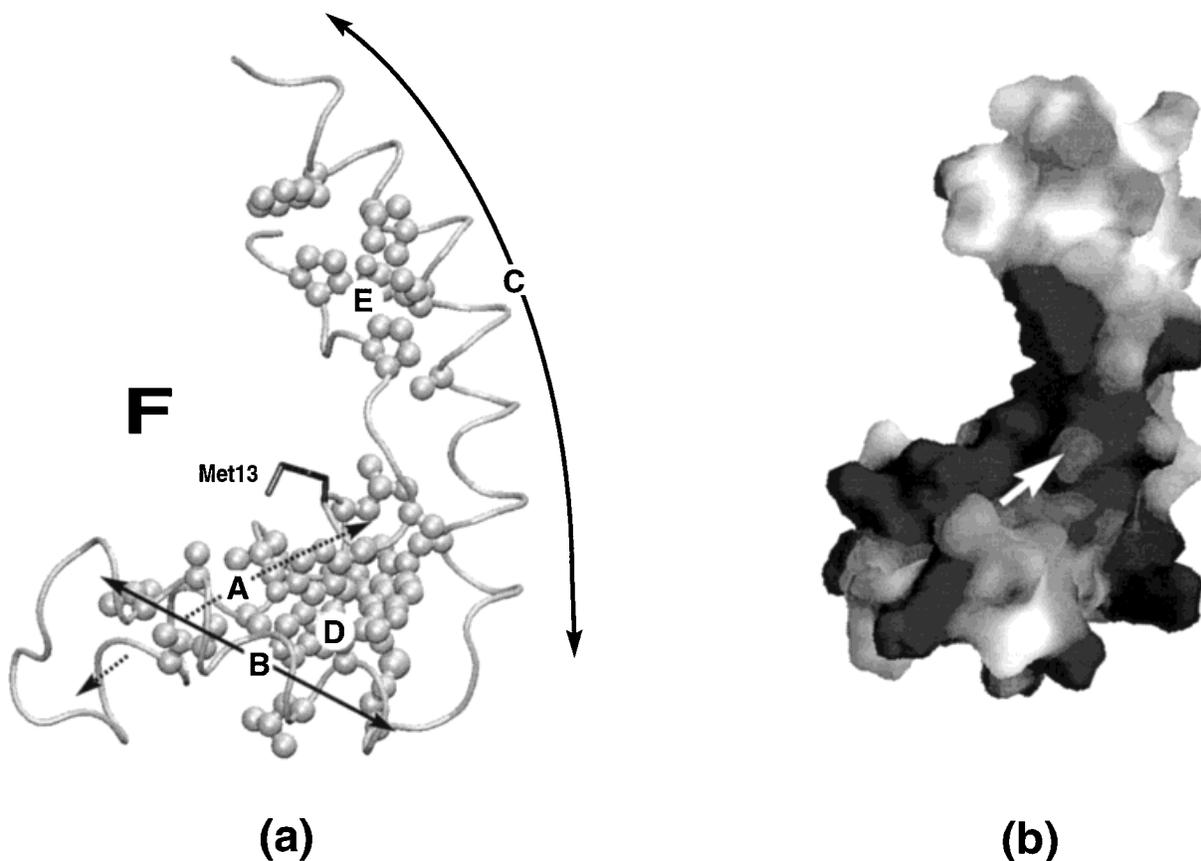


Fig. 1. The HMG-domain of HMG-D. (a) Solution structure of HMG-D obtained by NMR.³² Letters A, B, C refer to α -helices I, II, and III; D, E refer to hydrophobic cores 1 and 2; F indicates the concave face of the HMG-box. The diagram was prepared using

the program VMD.⁵⁷ (b) The electrostatic potential on the surface of the HMG-box, computed by GRASP;⁶⁷ black color indicates positively charged and light shading shows negatively charged regions. The arrow points to the location of Met13.

with functional consequences ranging from repressive¹⁷ to targeted positive effects on gene expression.¹⁸

The chromosomal HMG proteins belong to a larger family of HMG-domain-containing proteins.^{10,13,19} This protein family also includes several well-characterized sequence-specific transcription factors, such as the sex-reversal transcription factor SRY,²⁰ T-lymphocyte enhancer factor,²¹ and others (reviewed in ref. 13). Despite utilizing the same DNA-binding domain (known as the HMG-box), the two protein subfamilies have different specificities for preferred DNA sites, approximately 2–5-fold for the chromosomal proteins^{6,22} compared to approximately 50-fold for the transcription factors.^{21,23} The chromosomal proteins are much more abundant in the cell than the HMG-box containing transcription factors and characteristic amino acid differences in the HMG-box can be used to distinguish between the two subclasses of HMG-box domains.¹³ Both subfamilies possess “architecture specificity,”^{15,19} that is, an increased affinity for distorted DNA structures such as synthetic Holliday junction models²⁴ or DNA duplexes pre-bent either by circularization,⁵ cisplatin-DNA adducts,^{6,25,26} or interstrand disulfide cross-

links.²⁷ All of the HMG proteins bend DNA upon binding with a range of bend angles from 70° to 130°^{5,24,28,29} and the preference for pre-bent DNA is thought to arise from a reduced energetic penalty for this bending.^{23,27}

The DNA binding domain of the proteins, the HMG-box, defines the homology of the group.^{9,13} The HMG-box homology and structural domain consists of approximately 75 amino acids. The structures of several HMG-box domains have been determined by NMR methods.^{7,30–33} The domain has an L-shaped fold of three α -helices which are held together by two hydrophobic cores, as illustrated in the model of HMG-D³² in Figure 1a. The primary hydrophobic core (D) lies within the short wing of the “L” and at the junction between the two wings; the secondary hydrophobic core (E) lies within the long wing of the protein. The angle between the two wings is approximately 80°.^{30–32}

Recent solution structures of two sequence-specific HMG-box proteins, SRY and lymphoid enhancer factor 1 (LEF-1), in complex with DNA,^{34,35} confirm earlier observations that HMG-boxes bind to the minor groove of DNA at the outside of the DNA

bend.^{27,36,37} These structures provide valuable detailed insights into the interface of the proteins complexed with DNA,³⁸ revealing extensive hydrophobic interactions mediated by a ridge of aliphatic and aromatic residues contributing substantially to binding. Further, specific hydrogen bonds between the protein and particular DNA sequences can account for the specificity of this subfamily of HMG-boxes. However, these structures do not fully explain how the chromosomal proteins may form a complex with DNA in a nonsequence-specific fashion.

The HMG-box proteins are an ideal target for the study of chromosomal protein–DNA binding by modeling techniques, because of the opportunity they provide to predict a novel protein–DNA complex structure using the knowledge of structural and DNA binding properties of related sequence-specific proteins. In particular, the prediction may be achieved by docking the protein to DNA, guided by all available experimental data, then adjusting the resulting model in a series of molecular dynamics (MD) runs. Molecular dynamics is a classical powerful tool in theoretical biology^{39–41} that has been previously applied to study the protein–DNA contacts in the structures of *lac* repressor headpiece-operator complex⁴² and of nuclear hormone receptor–DNA complexes.^{43–47} Additionally, the role of water in protein–DNA interactions^{43,47,48} and the consequences of the modifications of experimental protein–DNA complex structures^{45–47} have been recently studied using MD. The methodology developed in these investigations has created a good foundation for the application of MD in structure prediction studies.⁴⁹ However, modeling of protein–DNA complexes is extremely demanding due to a need to include a large bath of water and ions in the simulations. The required large-scale simulations have been carried out so far only in the systems with known structures, but not in the context of structure prediction.

Here, we try to predict the structure of the complex of the HMG-box of HMG-D with DNA. Three trial structures of the protein–DNA complex were generated through docking and were subjected to 160 ps of free MD simulations within an environment of water and sodium ions (involving altogether 18,000 atoms). An analysis of the simulations revealed that one of the trial models was superior to the other two. This best model was subjected to a second, 60-ps long, MD simulation, which generated data for the structural analysis. A family of 60 snapshots from the second MD run constituted our prediction of the HMG-D–DNA complex structure. Comparison of the HMG-D–DNA model with the known models of SRY–DNA and LEF-1–DNA complexes, and inspection of the protein–DNA interface revealed many common details and also several features that may explain the difference in specificity between the two subfamilies of HMG-domain proteins.

MATERIALS AND METHODS

Initial Docking

The program “Quanta”⁵⁰ was employed to dock the protein to the DNA. For the docking, we used an NMR-determined HMG-D protein structure by Jones et al.,³² PDB entry 1HMA. Two DNA structures used were 1) an NMR structure of 12 bp DNA segment pre-bent by an interstrand disulfide cross-link²⁷ (the coordinates for this DNA model were kindly provided by Drs. Scot Wolfe and Greg Verdine), and 2) the DNA from the complex with the TATA-binding protein (TBP) reported by Kim et al.,⁵¹ PDB entry 1YTB. The latter DNA is a 29-nucleotide hairpin from which a 12 base pair double helical region lacking 3' and 5' phosphates was extracted for the modeling. The sequences of the DNA fragments are CGC-GAATTCGCG and GTATATAAAACG, respectively.

The concave surface of HMG-D was docked to the minor groove of the DNA. The putative intercalating Met13 was located as close as possible to the space between the 6th and 7th base pairs of the DNA. This is the site of the main kink in the structure of the cross-linked DNA and the site between the two intercalations of TBP, the third TA base step in the “TATA” DNA. We avoided as many steric clashes between the protein and DNA atoms as possible. A few clashes were neglected if it appeared that they could be eliminated by a minor change of an amino acid side-chain conformation, e.g., by a rotation around a C–C bond. When necessary, these changes were introduced manually. A short minimization procedure (50 to 100 steps of the Powell method) followed the docking. The resulting structures were used for further molecular dynamics simulations.

Solvation

The MD simulations of the protein–DNA complexes were carried out in an environment consisting of the solvent molecules (TIP3 water model) and sodium ions. A sphere of water molecules was centered at the center of mass of the complex at the beginning of the simulation. The radius of the sphere was 35 Å, which covers most of the complex by at least two layers of water molecules. To counterbalance the increase of the water density/pressure in the sphere due to surface tension, the oxygen atoms of the outer (3.2 Å thick) layer of the water were harmonically constrained to their initial positions. The spring constant was chosen such that the average displacement of the constrained water molecules from their initial positions would be the order of 1 Å.

Each model system contained the DNA carrying a charge of -22 units, and the protein with a charge of +4 units. Eighteen counterions were employed to maintain electrical neutrality and to approximate a solvent environment. The solvent environment around the fixed protein–DNA complex was con-

structed in three steps: 1) pure water was equilibrated for 15 ps; 2) 18 water molecules with largest electrostatic energy of the oxygen atoms were replaced by sodium ions—no two ions were positioned closer than 9 Å from each other, to allow each ion to have at least a single solvation shell; and 3) the water and the ions were equilibrated for 25 ps. Using the solvent was necessary to avoid possible simulation artifacts, but it significantly increased the size of the system to approximately 18,000 atoms, thus reducing the affordable simulation time.

Equilibration and MD Simulations

After the environment was built, the constraints on the protein-DNA complexes were released and the completed systems were equilibrated. In the initial stage of the equilibration, the protein and DNA heavy atoms were harmonically constrained to their initial positions. The temperature was controlled by coupling to a 300K heat bath using the Berendsen method.⁵² During ten steps of 1 ps dynamics, the harmonic constraints were gradually turned off. The second stage of equilibration consisted of five steps of 1 ps, during which the temperature coupling was also turned off. As a test, 5 ps of constraint-free dynamics was carried out. In all of the trial structures the temperature of the system remained steady at 300K during the test stage, proving the successful equilibration of the systems.

The equilibrated systems were simulated for up to 160 ps total time (equilibration included) without any temperature control or harmonic restraints (except for the outer water layer). One system was selected from the three trials (see Results), and for this system a second MD simulation of 60 ps was carried out. Snapshots from the MD trajectories were taken every picosecond.

The program X-PLOR⁵³ was used to carry out all the simulations. The interactions within the system were defined by the Charmm22^{54,55} force field. Van der Waals and Coulomb interactions between non-bonded atoms were computed within a cutoff distance of 11.35 Å, at which a local energy minimum is achieved (data not shown). At the cutoff boundary, the interactions were turned off smoothly by applying a switching function starting from 8.5 Å. Since the solvent was explicitly included in our simulations, the dielectric constant was set to 1.

Analysis of the Simulations

Two-dimensional root-mean-square deviation (r.m.s.d.) maps, similar to those developed by Beveridge et al.⁴⁰, were used to analyze the structural stability of the three protein-DNA complexes during the MD runs. Each cell of the map with coordinates (i, j) contains the r.m.s.d. between the i -th and j -th MD snapshots.

During the dynamics, various parameters were monitored using the library of X-PLOR scripts that process data extracted from the MD trajectories. The software package “MDToolchest”⁵⁶ was also used. To characterize the DNA unwinding, we used the difference between the average twist per helical step and the canonical twist of B-form DNA: $34.6^\circ - \text{Twist}/N_{\text{steps}}$. The DNA bend was defined as the angle between the vectors normal to the planes of the lateral base pairs of a DNA segment. If the planarity of a lateral base pair was not well maintained during the simulation, the adjacent base pair was used instead. Intercalation events were visually revealed by monitoring the simulation process using the program VMD⁵⁷ and characterized by the roll angle between the disturbed base pairs and the depth of intercalation. The latter was defined as the average distance between the carbon atom of an inserted methyl group and the geometric centers of the disturbed nucleotide bases.

More extensive analysis was performed on the chosen (see Results for selection criteria) model of the HMG-D-DNA complex. The data for the analysis were collected during the final (60 ps) round of MD simulations. Most parameters are averaged over the second MD round. Where the dynamics of a parameter is of interest, we present the complete time course. The analyzed properties of the complex included angles between α -helices, hydrophobic core sizes (gyration radii), the surface area of the protein-DNA contact region, and the hydrogen bonds between the protein and the DNA. A hydrogen bond was considered established when the donor-acceptor distance was lower than 3.5 Å, and the donor-hydrogen-acceptor angle was in the range 100° – 180° . To characterize the nonpolar (hydrophobic) interactions, the protein-DNA contact surface was computed as the difference between the accessible surface area of the protein alone and in complex with the DNA. The accessible area was computed using the algorithm by Lee and Richards⁵⁸ implemented in the X-PLOR package. The probe radius for the surface construction was set to 1.7 Å.

We compared the HMG-D-DNA complex model to the structures of the SRY-DNA and LEF-1-DNA complexes determined by NMR. Coordinates for SRY (current PDB entry 1HRY) were kindly provided to us by Dr. Marius Clore,³⁵ and those for LEF-1 were obtained from the PDB (entry 1LEF). Superposition and values of r.m.s.d. between the complexes were computed for sets of common atoms of protein or DNA backbone. Thus, residues 3–46 and 49–73 of the HMG-box were used and the HMG-D and LEF-1 residues that are absent in the SRY structure as well as the loop of variable length between α -helices II and III were excluded from the calculations. The 12 bp-DNA complexed with HMG-D best overlaps with base pairs 4–15 of the 15 bp-DNA from the LEF-1-DNA complex and was compared to that DNA frag-

ment. The 8 bp-DNA complexed with SRY was compared to base pairs 1–8 of the DNA from the HMG-D-DNA model and, correspondingly, to base pairs 4–11 of the DNA complexed with LEF-1.

After the structures were overlapped as described above, the following protein regions were compared: short wing—residues 11–50, long wing—residues 3–10 and 51–73, N-terminal stretch—residues 3–9, α -helix I—residues 10–26, α -helix II—residues 31–44, α -helix III—residues 50–73, residues at the HMG-box-DNA interface (DBI)—4–24 and 31–37. DNA bases 4–10 (HMG-D-DNA complex) were considered to be the protein binding site (PBS). The r.m.s.d. values were averaged over all (60) snapshots from the second MD round as well as the complete families of structures available from the PDB for LEF-1 (12 structures) and SRY (35 structures).

RESULTS

Docking and First Round of Molecular Dynamics Simulations

Trial models of the complex between the HMG-box of HMG-D and DNA were constructed using the experimentally determined structures of HMG-D and DNA fragments. The solution structure of the HMG-box of HMG-D was chosen as the initial protein model with the assumption that HMG-D does not significantly change its structure upon binding to DNA. Since the structures of the SRY- and LEF-1-DNA complexes had not been determined at that time, two DNA fragments, that we expected to resemble most closely the structure of a DNA complexed with a minor groove binding protein, were chosen for docking. One structure is the 12 base pair segment of DNA pre-bent by an interstrand disulfide cross-link.²⁷ HMG-D binds particularly well to such a cross-linked DNA site,²⁷ suggesting that this DNA fragment may better approximate the DNA structure in the complex with HMG-D than straight duplex DNA. The other structure, the DNA from the complex with the TATA-box binding protein (TBP) determined by X-ray crystallography,⁵¹ was chosen because of the great similarity between the DNA binding properties of TBP and those of the HMG-box proteins. TBP binds to the minor groove on the outside of a severely bent (by approximately 90°) DNA fragment and has two amino acids with side chains inserted between the DNA bases.⁵¹ The structures of the cross-linked and of the TATA-box DNA have bending angles, unwinding angles, and minor groove widths that are either larger or smaller than those expected for DNA bound by HMG-D by analogy with HMG1.⁵ Thus, it was anticipated that the DNA geometry of the two fragments may converge to an intermediate structure of DNA complexed with HMG-D during the molecular dynamics simulations.

The following considerations guided the docking of the protein to the DNA. The HMG-box of HMG-D is positively charged (+4 total charge), and the concave

face of the “L”-shaped domain (Fig. 1b) provides a favorable binding surface for negatively charged DNA. In NMR experiments on the SRY-DNA complex, King and Weiss suggested that an isoleucine side chain partially intercalates into the DNA minor groove.^{37,59} The intercalation of this residue was later confirmed by the structures of SRY and LEF-1 complexed with DNA.³⁴⁻³⁵ In HMG-D, the homologous residue is a methionine (Met13) which extends from the concave face of the HMG-box. In fact, oxidation of methionine 13 leads to a threefold decrease in affinity of HMG-D for DNA.⁶⁰ Consequently, the concave surface of the protein was docked to the minor groove surface of the DNA, locating the Met13 near the DNA bases to allow the residue to intercalate if the resulting structure would favor it.

The three starting models of the HMG-box docked to the different DNA fragments are shown in Figure 2. Two models of the HMG-D-DNA complex were built using the disulfide cross-linked DNA structure.²⁷ This DNA is less unwound and has a narrower and more curved minor groove than the TATA-box DNA. The HMG-box could not be aligned exactly parallel to the minor groove on this DNA model because the short wing of the protein clashed sterically with the DNA. Hence, the protein was placed at various angles to the groove. The two complexes built with this DNA placed the long axis of the HMG-box at an angle of approximately -30° and +70° from the direction of the minor groove (Figs. 2c,d). The two disulfide cross-linked DNA-HMG-box complexes will be referred to as “CL-1” and “CL-2,” respectively. A third complex, shown in Figures 2a,b, was built with a DNA model derived from the TBP-DNA complex structure and will be referred to as the “TATA” model. The DNA in this complex is unwound, giving rise to a wide minor groove that can accommodate the HMG-box nearly parallel to the groove without any steric overlap, with Met13 extending to the site of putative intercalation.

We embedded the trial structures into the solvent environment and simulated them for 160 ps. A gentle equilibration protocol was used because the trial structures of the protein-DNA complexes have vulnerable contacts, such as partial intercalation. The atoms of the protein-DNA complexes were harmonically restrained to their initial positions at the beginning of the simulation and the restraints were slowly released during the equilibration stage. Brisk motions of the atoms at the beginning of the equilibration were thus avoided so that favorable contacts at the protein-DNA interface would not be damaged. Had the system been released too quickly, it may have been trapped in a metastable state far from both the initial state and a satisfactory final state. The gentle protocol allowed the protein and the DNA to adjust slowly to the presence of each other, to saturate smoothly the energy of all degrees of free-

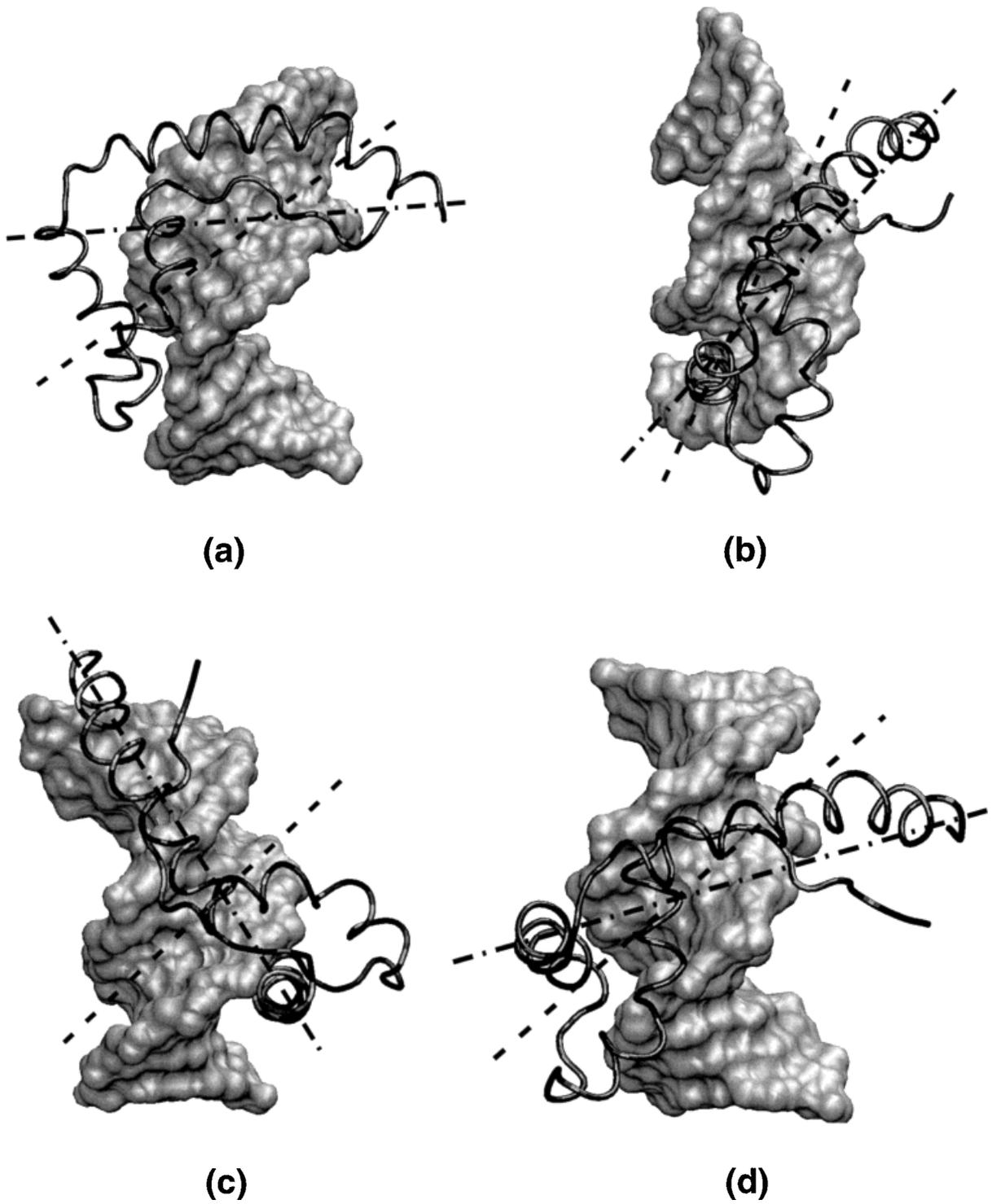


Fig. 2. Three models obtained by docking the HMG-box of HMG-D to DNA. (a) Complex of the HMG-box with DNA. (b) A view into the minor groove of the complex of the HMG-box with TATA DNA. (c,d) Complexes of the HMG-box with disulfide cross-linked

DNA ("CL-1" and "CL-2," respectively). Dashed lines are drawn along the minor groove and the "dash-and-dotted" lines represent the long axis of the protein (the axis of α -helix III).

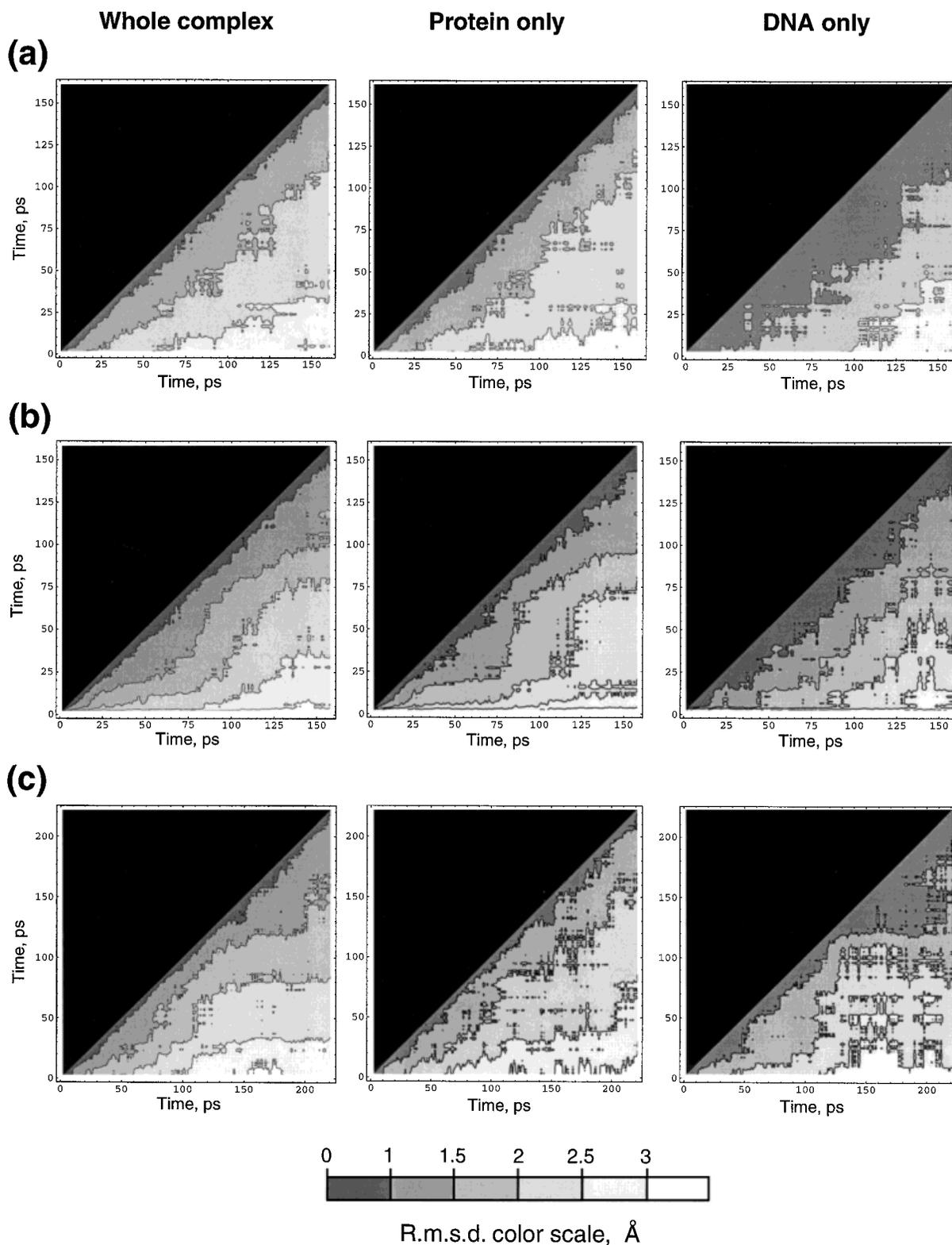


Fig. 3. 2D r.m.s.d. maps illustrating the dynamics of the trial complexes. (a) “CL-1” model, 160 ps of MD. (b) “CL-2” model, 160 ps of MD. (c) “TATA” model, 220 ps of MD. R.m.s.d. values for all atoms of the protein–DNA complex (column 1), non-hydrogen atoms of the protein (column 2), and non-hydrogen atoms of the DNA (column 3) were computed after superposition of all atoms of

the complex for each pair of snapshots during the simulation. The dark triangular patterns of low r.m.s.d. values appeared during the last stage of MD, indicating that stable conformational subfamilies were achieved by the “CL-2” and “TATA” complexes. The diagonal direction of the boundaries between the zones in the “CL-1” map indicates that the “CL-1” model lacks stability.

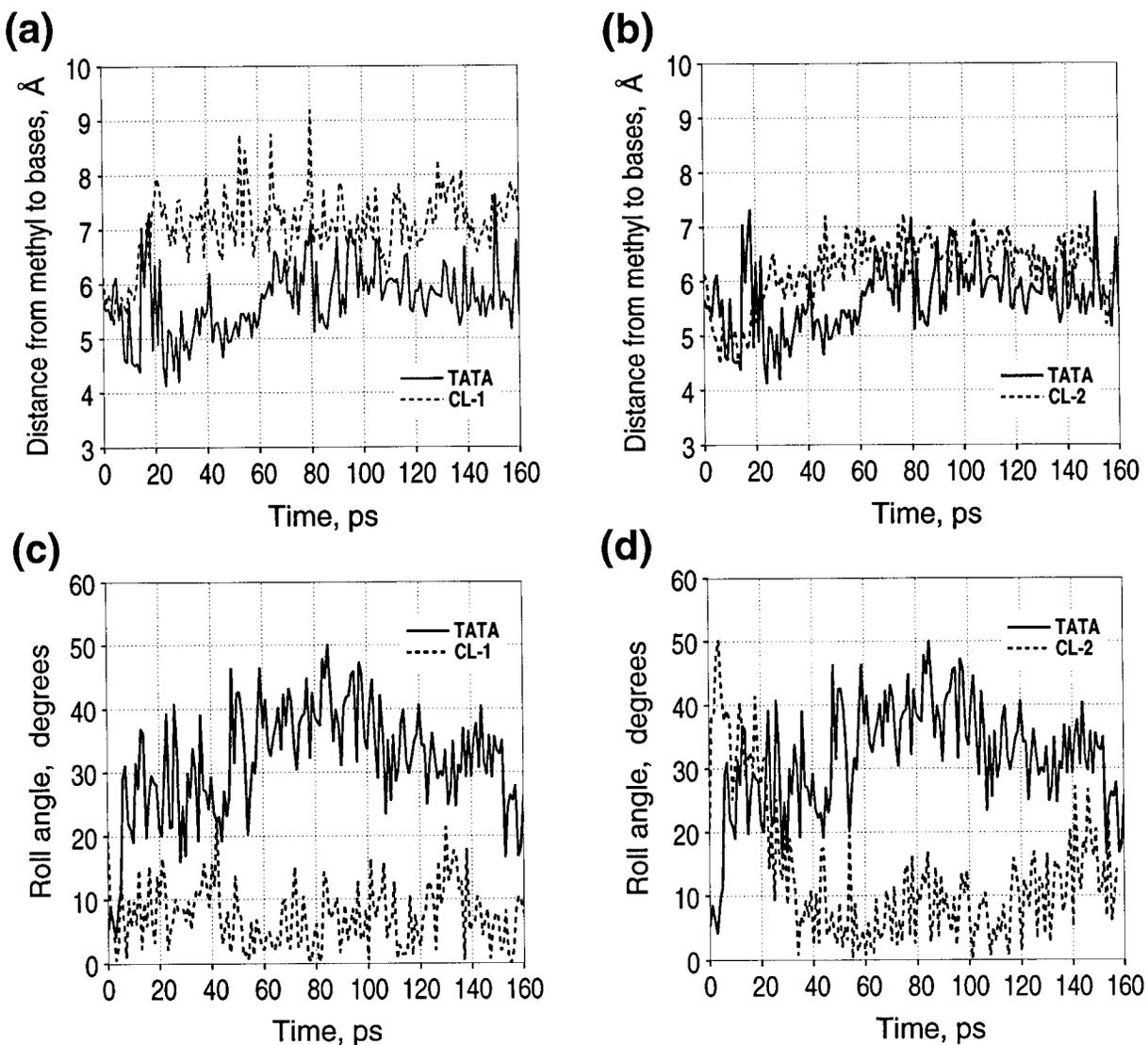


Fig. 4. Dynamics of the parameters describing Met13 partial intercalation in the three trial protein-DNA complexes. (a,b) Average distance of the carbon of the penetrating methyl group to the geometric centers of the disturbed nucleotide bases. (c,d) Roll

angle between the planes of the disturbed (6th and 7th) base pairs. The data demonstrate that the intercalation of Met13 survived only in the "TATA" model.

dom, and to start drifting to a favorable conformation of the complex.

Selection of the Best Trial Structure

The purpose of the first set of simulation runs was to distinguish the best of the three trial structures on the basis of structural changes that occurred during the simulations. The simulations of the "CL-1," "CL-2," and TATA models were compared considering three main criteria: Does the trial complex reach a stable state by the end of the simulation (160 ps)? Is the intercalation of Met13 conserved in the model? Is a good geometry of the protein and DNA maintained in the model?

To estimate the stability of the complexes, we used two-dimensional r.m.s.d. maps (Fig. 3). Each dot of the map corresponds to an r.m.s.d. between two MD snapshots. In the case where a structure stabilizes in a particular conformational substate during any stage of the dynamics, the map exhibits a triangular pattern of low r.m.s.d. protruding from the main diagonal of the map. In contrast, uniform strips parallel to the diagonal imply the absence of stability and continuous structural drift. The maps for the "CL-2" and "TATA" models display the triangular stability patterns; the map for the "CL-1" displays instability. The "CL-1" model evolved to a lower r.m.s.d. from the initial structure (approximately 2.5

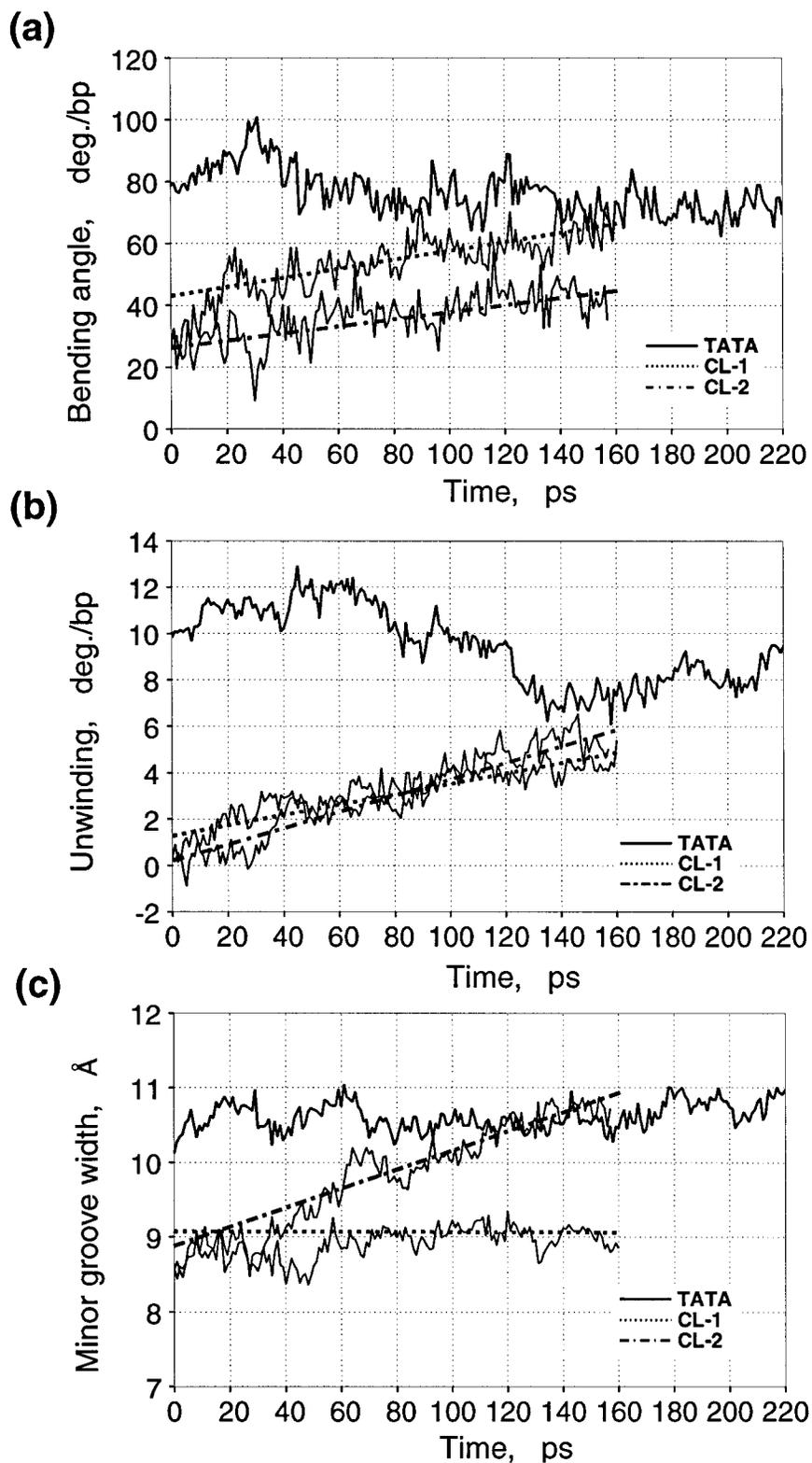


Fig. 5. Evolution of the DNA geometry parameters in the three trial protein-DNA complex structures. (a) DNA bend computed as the angle between vectors normal to the plane of the lateral base pairs. (b) The degree of DNA unwinding computed as the difference between the twist per base pair for B-form DNA (34.6°) and

the average twist per base pair in the trial structures. (c) The width of the DNA minor groove (averaged over all base pairs). The lines correspond to the best linear fit of the data and are presented solely to emphasize the trends of the parameters exhibited in the "CL-1" and "CL-2" models.

Å) than the other two complexes (approximately 3 Å), but the “CL-2” and the “TATA” models stabilized by the end of the simulation. Although the latter two trial complexes drifted further from the initial structures, they finally reached relatively stable conformations, whereas the “CL-1” complex continued to experience structural changes. This is an argument in favor of the “TATA” and “CL-2” models as better candidates for the actual HMG-D-DNA complex.

The partial intercalation of Met13 was engineered at the beginning of the simulation and the conservation of this intercalation during the 160 ps simulation was considered to be a measure of the quality of the models. The intercalation survived the first MD round in only one of the three models, the “TATA” model. During the last 80 ps of dynamics, the average distance from the putative intercalating methyl groups to DNA bases was 5.9 ± 0.5 Å for the “TATA” model, 7.2 ± 0.5 Å for the “CL-1” model, and 6.5 ± 0.4 Å for the “CL-2” model (Figs. 4a,b). A substantial value of base pair roll at the putative intercalation site was observed only in the “TATA” model (Figs. 4c,d). In the “CL-2” model, an intercalation site developed during the first 20–30 ps of the simulation, but then disappeared, and only at the end of the MD run did a significant non-zero roll angle emerge (Fig. 4d). The average value of the roll angles for the last 80 ps of dynamics was $7.6^\circ \pm 4.9^\circ$ for the “CL-1,” $10.4^\circ \pm 5.8^\circ$ for the “CL-2,” and $33.8^\circ \pm 6.9^\circ$ for the “TATA” model. These data clearly distinguish the “TATA” model from the other two.

The dynamics of the geometry of the DNA also provided a useful tool to assess the quality of the models. In all the three models, the DNA is strongly bent and unwound and has a widened minor groove. The dynamics of the corresponding parameters is shown in Figure 5. In the “TATA” model, the bend angle of the DNA stabilized at $75^\circ \pm 5^\circ$ after an initial spike (Fig. 5a). The average width of the minor groove did not significantly change, oscillating near the value of 10.4 ± 0.2 Å. The DNA unwinding angle increased at the beginning of the simulation from 10° per base pair to 12° per base pair but then decreased and finally stabilized during the last 40 ps at $7.4^\circ \pm 0.7^\circ$ per base pair. In the “CL-1” and “CL-2” models, the DNA bend and unwinding angles increased toward the values achieved by the “TATA” model (Fig. 5). The DNA minor groove width in the “CL-2” model also converged toward that of the “TATA” model, having increased by 2 Å during the simulation. In the “CL-1” complex, the groove width was stable (7.9 ± 0.3 Å).

Thus, the trial “TATA” complex of the HMG-box with DNA stabilized by the end of the simulation better than “CL-1” and as well as “CL-2.” The “TATA” model achieved better partial intercalation of Met13 than the other two models and developed a DNA geometry toward which the DNA in the “CL-2” model was converging. However, it would have required

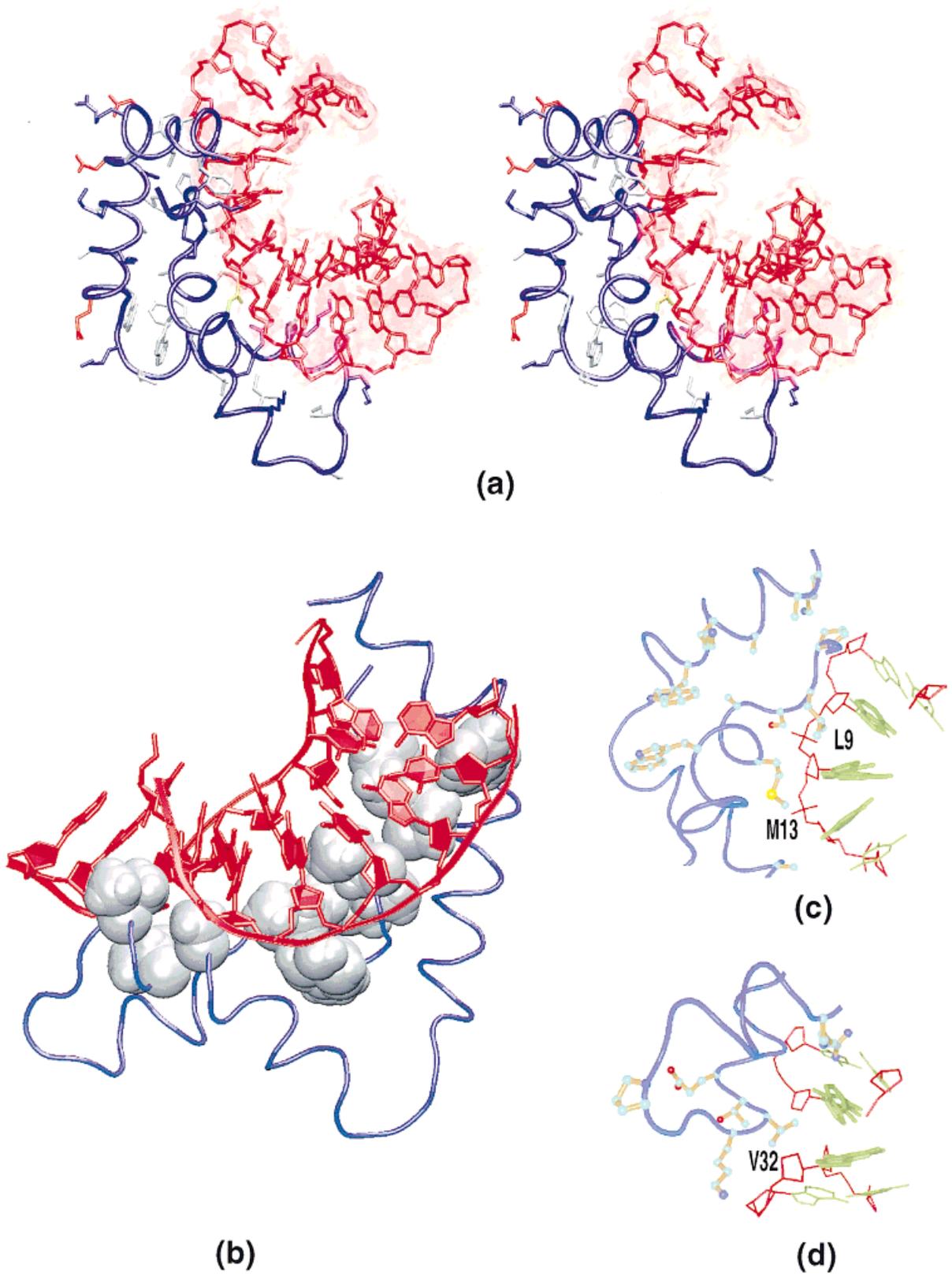
considerable additional dynamics to reach the convergence. These observations suggested that the “TATA” model is the best of the three trial complexes, and the model was selected for further structural analysis.

“TATA” Model of the HMG-box-DNA Complex

In order to collect data for the structural analysis and to assess the stability of the “TATA” model, the MD simulations with this complex were continued for an additional 60 ps. The structure remained in a stable conformational subfamily within a 1.5 Å r.m.s.d. of all heavy atoms (Fig. 3). The average r.m.s.d. of the subfamily from the initial docked structure of the “TATA” model is 3.0 ± 0.1 Å. A set of the snapshots of the structure taken from the MD trajectory every picosecond constitutes our prediction of the HMG-box-DNA complex structure. A representative snapshot is shown in Figure 6a.

The structural integrity of the protein is conserved throughout the simulation without significant deviations from normal geometry and stereochemistry or large structural distortions. The average Ramachandran map (Fig. 7b) reveals that only one residue has ϕ - ψ values in a disallowed region of the map. This residue, Ser50, is located in a short loop between helices II and III that appears to be extended in the model. Other residues in this loop, Ala45, Met46, Lys47, also occupy energetically slightly unfavorable positions in the map. To ensure that destructive unfolding was avoided, the integrity of the two hydrophobic cores (**D**, **E** in Fig. 1a) was monitored by assessing changes in their radii of gyration. The dynamics of the radii show that the cores swell at first but then stabilize. The primary hydrophobic core (**D**) swells less than the secondary (**E**), exhibiting only a 10% increment of the radius of gyration (from 6.8 to 7.5 Å) compared to a 14% increment (from 6.4 to 7.3 Å) for the secondary core. The larger change of the secondary core correlates with structural changes seen in the C-terminus of the protein.

Fig. 6. Illustration of the predicted HMG-box-DNA complex. (a) Stereo picture of a snapshot from the trajectory of the MD simulation of the “TATA” model taken after 19 ps of MD in the final round of simulations (i.e., at 179 ps after the beginning of the simulation). The C α trace of the protein backbone is shown as a blue tube and side chains of all the residues are represented as lines, connecting the heavy atoms. The positively charged, negatively charged, polar, and nonpolar residues are colored blue, red, green, and grey, respectively. The DNA is represented as a transparent surface over the lines connecting the heavy atoms. Two ribbons trace the phosphate groups of the DNA backbone. For clarity, several protein side chains and DNA atoms have been omitted. (b) Hydrophobic ridge of residues in the minor groove. The side-chain atoms of the hydrophobic ridge residues are shown as gray spheres of 1.4 van der Waals radii. For clarity, protein residues, not forming the ridge, and terminal sections of DNA have been omitted. (c) Site of Leu9 and Met13 partial intercalations. (d) Site of Val32 partial intercalation. In (c) and (d), parts of the DNA and some protein residues were removed from the diagram for clarity. The figure has been produced using the program VMD.⁵⁷



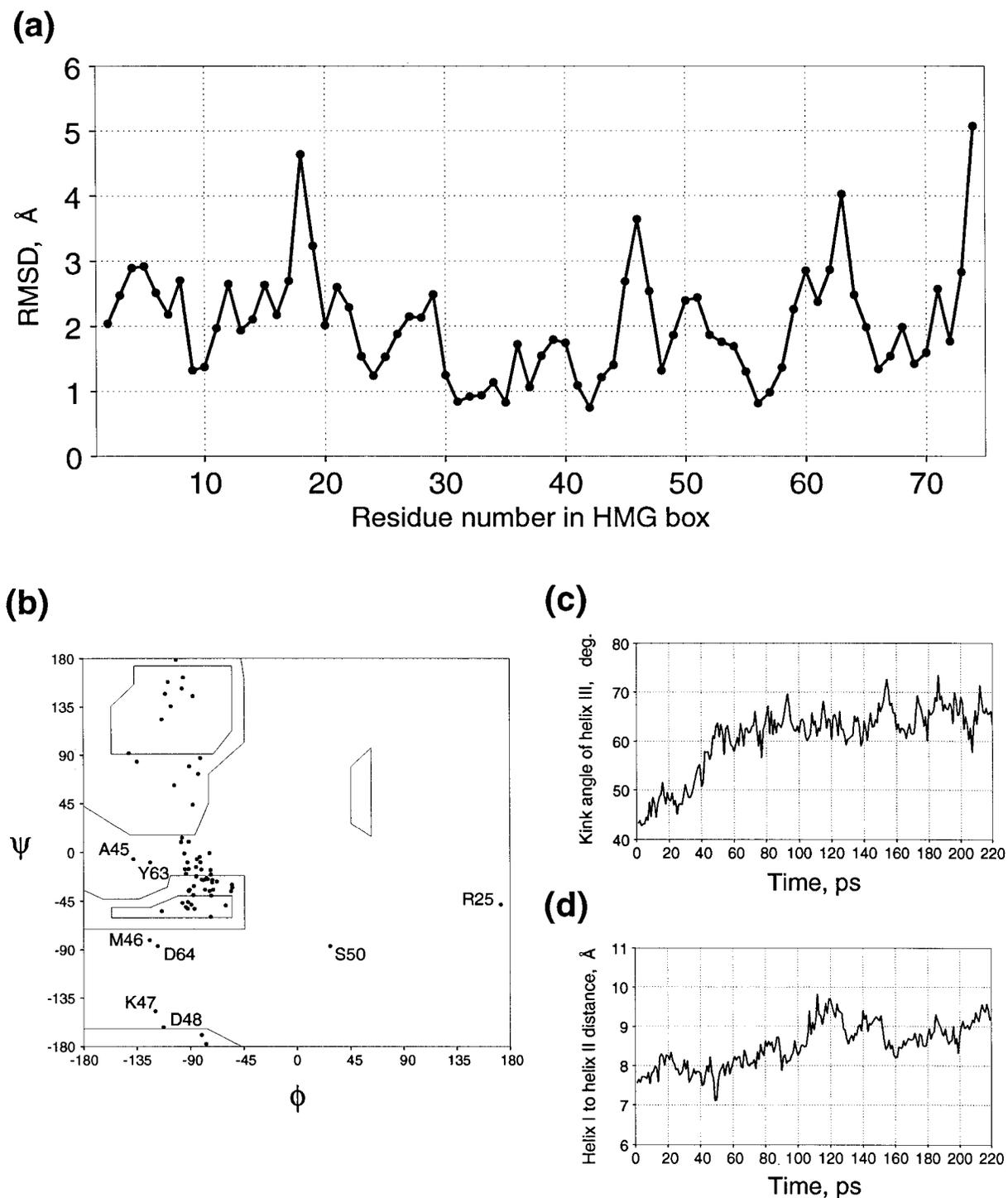


Fig. 7. Structural parameters of the HMG-box in the complex with DNA. (a) Per residue graph of the r.m.s.d. of the protein backbone from the initial fold averaged over the last 60 ps of dynamics for each residue. (b) Ramachandran plot of the average

ϕ , ψ torsion angles. (c) Dynamics of the angle between two α -helical segments of helix III. (d) Dynamics of the distance between the centers of mass of α -helices I and II of the HMG-box.

Several adjustments in the protein resulted in a structure well adapted to the DNA binding site. The r.m.s.d. profile (Fig. 7a) reveals that the largest structural changes occurred in: 1) the N-terminal

region, residues 2–21; 2) the C-terminus of helix III, residues 59–74; 3) the loop between helices II and III, residues 45–51. The overall r.m.s.d. from the initial NMR structure is 2.2 ± 0.1 Å for the backbone

atoms and $2.6 \pm 0.1 \text{ \AA}$ for all heavy atoms of the HMG-box.

In the N-terminal region, a bend of approximately 60° is observed in helix I. During the simulation, this bend evolves into a sharp kink, although the angle between two halves of the helix does not change ($63.8^\circ \pm 5.4^\circ$). The N-terminal half of helix I lies across the DNA minor groove, and the residues facing the DNA, Met13 and Asn17, protrude into the minor groove and establish direct contacts with DNA bases. The C-terminal half of the helix lies along the DNA backbone and the three residues facing the DNA, Ser18, Arg20, and Lys24, establish contacts with the phosphodiester backbone. Arg20 also contacts the DNA bases.

Helix III bends during the simulation, allowing closer interactions between the C-terminal residues and the DNA. The helix is already bent by 43° at residue 62 in the HMG-D NMR structure.³² During the first 80 ps of the dynamics, the angle between the two parts of the helix increases, stabilizing at $65^\circ \pm 3^\circ$ (Fig. 7c).

Helices I and II move apart from each other, with a corresponding increase in the distance between their centers of mass of 1.5 \AA (Fig. 7d). This movement correlates with the swelling of the main hydrophobic core mentioned above. The distal termini of α -helices I and II swivel and extend the connecting loop, permitting the protein to approach the DNA backbone more closely. The backbone appears to be attracted to the protein by a cluster of three positively charged amino acids, Arg20, Lys24, and Lys31. The N-terminus of helix III is drawn to the DNA backbone by Arg44 and Lys49, whereas the C-terminus of helix II moves outwards from the DNA. These interactions result in an extension of the loop formed by residues 45–51, connecting helices II and III, which explains the unfavorable ϕ - ψ angles observed in this loop.

The r.m.s.d. of the DNA in the “TATA” model from its initial structure is $2.2 \pm 0.1 \text{ \AA}$ for all heavy atoms. The severe DNA bend in the direction of the major groove decreases during the simulation from 80° to $71.6^\circ \pm 4.6^\circ$ (Fig. 5). The stretch of DNA that contacts the protein between the 3rd and the 10th base pairs has the largest bend of $92.0^\circ \pm 3.7^\circ$. Compared to classical B-form DNA, the DNA in the final “TATA” model is unwound by $8.2^\circ \pm 0.7^\circ$ per base pair, which is less than the unwinding angle of 9.8° per base pair measured in the initial docked model. The value of the bend is consistent with a recent estimate of 72° from DNA circularization assays (Churchill, unpublished data). The DNA minor groove is widened to accommodate the protein, with an average groove width of $10.4 \pm 0.1 \text{ \AA}$ (Fig. 5c). The profiles of the minor groove width and sugar pucker angles along the DNA are plotted in Figure 8. The sugar pucker, defined by the torsion angle δ , decreases within the protein binding site, and alternates from lower to higher values between the 9th

and 10th base pairs. The overall features of the sugar pucker and minor groove width indicate that the DNA adopts an A-form-like structure in the protein-binding site, and a B-form-like structure in the regions outside of the protein binding site.

Protein-DNA Interactions in the “TATA” Model

The contacts between the HMG-box and the DNA in the “TATA” model can be categorized as follows: first, positively charged and polar residues anchor the protein to the DNA by establishing hydrogen bonds and salt-bridges with the sugar-phosphate backbone; second, polar and charged residues establish hydrogen bonds to the atoms of the nucleotide bases; third, extensive hydrophobic interactions occur between the surface of the HMG-box and the edges of the bases and sugar-phosphate backbone in the DNA minor groove.

A summary of the hydrogen bonds and salt bridges between HMG-D and DNA is presented in Table I and Figures 9b,d. Seven positively charged residues anchor the protein to the oxygen atoms of the DNA phosphates. Four of these residues, Lys6, Arg7, Arg20, and Lys37, are highly conserved within the HMG family,¹³ whereas Lys24 is conserved only among the chromosomal HMG-box proteins, and residue 31 (Lys in HMG-D) is not conserved. Lys4 strongly interacts with the DNA by bridging the phosphodiester groups of base pairs 4 and 5. Arg7 forms a salt bridge to the sugar-phosphate group connecting Thy4 and Ade5. The plane of the guanidinium group of this arginine is parallel to the ribose ring of Thy4. This salt bridge contributes an appreciable energy to the overall HMG-box-DNA interaction (Table I).

Several interactions pin the N-terminal of helix I to the DNA. Ser10 and Tyr12 form direct hydrogen bonds with the DNA backbone, and Ser18 forms a water-mediated hydrogen bond with the DNA backbone. At the C-terminus, several residues initiate water bridges toward the DNA backbone (Lys60, Glu71) or nucleotide bases (Lys68). These bridges involve one or two water molecules.

Three amino acid residues of HMG-D establish direct hydrogen bonds to the DNA bases. The guanidinium group of Arg20 not only contacts the DNA backbone through hydrogen bonds between the Arg20 N^ϵ , $N^{\eta 1}$ atoms and Ade8 $O'_{4,}$ but the $N^{\eta 1}$ atom also forms hydrogen bonds with the N_3 atoms of Ade7 and Ade8. The hydrogen bond between Asn17 $N^{\delta 1}$ and Thy6 O_2 is buttressed by additional hydrogen bonds which Arg20 $N^{\eta 1}$ forms with Asn17 $O^{\delta 1}$ and Leu16 O, adjacent to the asparagine. Thus, Arg20 and Asn17 form a DNA-binding “fork,” stabilizing their individual contacts. At the C-terminus, Tyr63 emerges from the kinked region of helix III and extends toward the DNA; the phenolic hydroxyl alternately participates in a direct or water-mediated hydrogen bond to an acceptor atom provided by either Ade9' N_3 or Thy10' O_2 .

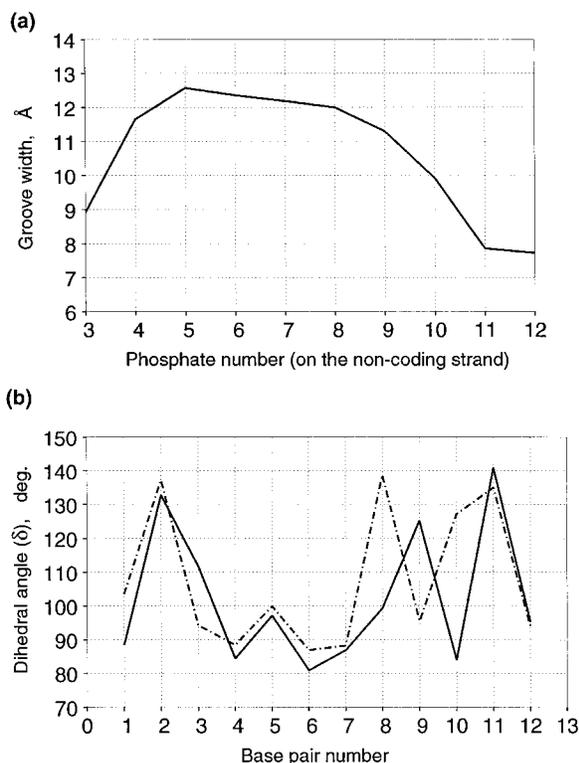


Fig. 8. Structural parameters of the DNA in the complex. (a) Minor groove width. (b) Sugar pucker (δ angle) profile. Solid line represents the noncoding, and dashed line, the coding DNA strand. The parameters indicate that the DNA at the protein binding site adopts a conformation close to that of A-form DNA.

A common feature of various HMG-box domains is an extensive ridge of hydrophobic amino acids that fits into the minor groove of the DNA upon protein binding.³⁸ In HMG-D, this ridge is formed by Pro8, Leu9, Tyr12, Met13, Val32, Val35, Ala36 (see Figs. 6b and 9e). The total surface area of HMG-D, which is buried by the DNA in the “TATA” model, is $1210 \pm 24 \text{ \AA}^2$. During the dynamics, this buried surface area decreases by approximately 4%. Figure 10a shows the contribution of individual residues to the buried surface. In addition to the hydrophobic residues just mentioned, several hydrocarbon chains of positively charged amino acids also contribute to this ridge (Lys4, Lys6, Arg7, Arg20, Lys31, Lys37, and Lys60). Most of these residues are situated in the N-terminal part of the protein, near the DNA (Fig. 10a). However, there is also a contribution from Lys60 and Tyr63 of α -helix III, which extend toward the DNA minor groove and are protected by the DNA in the final “TATA” model.

The hydrophobic interactions of HMG-boxes with DNA also include partial side-chain intercalation between the DNA bases. In the “TATA” model, three events of partial intercalation are detected (Figs. 6c,d, 9b,e). The partial insertion of side chain Met13 between the 6th and 7th base pairs of the DNA was

set up while building the trial complexes. During the simulation, another clear intercalation site developed for Leu9 between the 5th and 6th base pairs, although at the opposite DNA strand. The intercalation distance is $5.7 \pm 0.4 \text{ \AA}$ for Met13 C ϵ , $6.1 \pm 0.6 \text{ \AA}$ for Leu9 C δ^1 , and $5.2 \pm 0.4 \text{ \AA}$ for Leu9 C δ^2 . The average roll angle for the disturbed base pairs is $28.1^\circ \pm 5.3^\circ$ for the 6th helical step (Met13) and $36.5^\circ \pm 7.8^\circ$ for the 5th helical step (Leu9). The values of the roll anticorrelate (Fig. 10b): the covariance between them is equal to -0.35 for the last 60 ps of MD, and -0.55 for the last 120 ps of MD, when intercalation of Leu9 was established. It seems that the DNA does not tolerate two well-established partial intercalations at the same time and the amino acids compete for the space to intercalate. From visual inspection, Leu9 intercalation dominates, whereas Met13 contributes to the roll angle at the 6th base step by steric interaction between the methyl group and the edge of Thy6, causing a propeller twist in the 7th base pair.

A third partial intercalation site developed between the 100th to 150th picosecond of the simulation for Val32 and persisted throughout the second MD round. Val32, located at the N-terminus of helix II, disturbs the stacking of the 8th and the 9th DNA base pairs and has the largest buried surface area among all the HMG-D residues ($122 \pm 7 \text{ \AA}^2$). The distance from Val32 C γ to the base pairs is $4.5 \pm 0.3 \text{ \AA}$ and the roll angle at the 8th helical step is $24.4^\circ \pm 5.0^\circ$.

Evaluation of the Model Structure in Light of HMG-box-DNA NMR Studies

No structure of a chromosomal protein–DNA complex is available for comparison to the modeled HMG-D-DNA complex structure. However, there have been two NMR studies focusing on chromosomal protein–DNA interactions of HMG1 and HMG-D^{6,7} and two NMR structures of sequence-specific HMG-boxes bound to DNA, SRY and LEF-1.^{34,35}

The region of HMG1 box A that interacts with DNA has been defined using NMR spectroscopy.⁷ The HMG1-box A-DNA complex is in fast exchange on the NMR time scale; however, the chemical shifts of many protein amide atoms change with addition of DNA. The residues with the largest chemical shift differences were mapped and provide a picture of regions that may undergo the largest perturbations on DNA binding. These are residues† 4, 5, 8, 9, 10, 17, 18, 23, 26, 32, 33, 38, 41, 44, 45, and 68. Of these residues, the equivalent ones in the HMG-D-DNA complex model, 4, 10, 17, 18, 44, and 68, are involved in electrostatic interactions with the DNA, whereas

†The residue numbering of all HMG proteins discussed in this paper has been changed to be consistent with that of HMG-D. Sequences were aligned according to ref. 13.

TABLE I. Summary of Hydrogen Bond and Salt Bridge Interactions Between HMG-D and DNA

HMG-D residue (Donor atom)	DNA nucleotide (Acceptor atom)	Hydrogen bond length,* Å	A-H-D angle,*† degrees	Interaction energy,*‡ kcal/mol	Type of interaction§
Lys4 (N ^ε)	A ₅ (O _{P2})	2.7 ± 0.1	152 ± 14	-83.9 ± 4.0	Charge-charge
Lys4 (N ^ε)	T ₆ (O _{P1})	2.8 ± 0.3	153 ± 22	-84.3 ± 7.3	Charge-charge
Arg7	A ₅ (phosphate)	Not a hydrogen bond		-27.5 ± 4.8	Charge-charge
Lys24 (N ^ε)	A ₉ (O _{P1})	2.7 ± 0.2	163 ± 9	-86.7 ± 4.4	Charge-charge
Lys31 (N ^ε)	A ₁₀ (O _{P2})	2.7 ± 0.1	165 ± 7	-82.1 ± 2.6	Charge-charge
Lys37 (N ^ε)	T ₅ ' (O ₃ ')	3.1 ± 0.2	124 ± 18	-76.8 ± 2.8	Charge-charge
Lys37 (N ^ε)	T ₆ ' (O ₅ ')	3.2 ± 0.2	124 ± 13	-76.8 ± 2.8	Charge-charge
Lys37 (N ^ε)	T ₆ ' (O ₄ ')	3.1 ± 0.2	143 ± 17	-9.3 ± 1.8	Charge-multipole
Ser10 (O ^γ)	T ₈ ' (O _{P2})	3.6 ± 0.5	129 ± 31	-10.2 ± 5.5	Dipole-charge
Ser10 (O ^β)	T ₈ ' (O ₅ ')	3.3 ± 0.4	131 ± 30	-10.2 ± 5.5	Dipole-charge
Tyr12 (O ^η)	A ₇ ' (O ₃ ')	3.0 ± 0.2	149 ± 20	-12.5 ± 1.7	Dipole-charge
Tyr12 (O ^η)	T ₈ ' (O _{P2})	3.7 ± 0.6	145 ± 16	-12.5 ± 1.7	Dipole-charge
Asn17 (N ^{δ2})	T ₆ ' (O ₂ ')	3.0 ± 0.2	114 ± 14	-1.2 ± 0.7	Dipole-multipole
Arg20 (N ^{η1})	A ₇ ' (N ₃)	3.0 ± 0.1	155 ± 12	-13.6 ± 1.4	Charge-multipole
Arg20 (N ^{η1})	A ₈ ' (N ₃)	3.1 ± 0.2	121 ± 13	-7.6 ± 1.8	Charge-multipole
Arg20 (N ^{η1})	A ₈ ' (O ₄ ')	3.3 ± 0.2	137 ± 11	-10.7 ± 1.3	Charge-multipole
Arg20 (N ^ε)	A ₈ ' (O ₄ ')	3.0 ± 0.1	161 ± 11	-10.7 ± 1.3	Charge-multipole
DNA-binding "fork" buttressing interactions					
Arg20 (N ^{η2})	Leu16 (O)	2.8 ± 0.1	156 ± 10	-13.3 ± 1.5	Charge-dipole
Arg20 (N ^{η2})	Asn17 (O ^{δ1})	2.8 ± 0.1	148 ± 11	-15.3 ± 2.1	Charge-dipole

*All the data were averaged over the last 60 ps round of MD.

†The acceptor-hydrogen-donor angle.

‡The fifth column shows the enthalpy of interaction between the chemical groups of the donor and the acceptor atoms, as defined by the force field used. Therefore, the interaction energy is the same when different donor and/or acceptor atoms belong to the same chemical group.

§Type of interaction is indicated in each case to provide a brief explanation of the energy value observed.

8, 9, and 32 are involved in hydrophobic interactions with the DNA (Fig. 9c).

The contact residues for HMG1 nearly mimic the residues for which there is an increased buried surface area in the HMG-D-DNA model (Fig. 10a) with three exceptions: 1) residues 38, 41, and 45 of helix II do not make DNA contacts in the HMG-D-DNA model; 2) vice versa, residues 35, 36, and 37 of HMG1 do not exhibit a significant change in the chemical shift; and 3) Ala13 does not exhibit a significant chemical shift change, but would be expected to intercalate between DNA bases by analogy with SRY and LEF-1.^{34,35} In HMG-D, modification of the equivalent residue, Met13, decreases the DNA binding affinity, providing further evidence for the role of residue 13 in DNA binding, possibly by intercalation.⁶⁰ Interestingly, a large chemical shift change occurs on DNA binding for Met9 of HMG1,^{30,31} which is consistent with direct interaction with DNA, possibly through intercalation; in our model of the HMG-D-DNA complex, Leu9 and Met13 intercalate in the DNA alternately.

Structural comparison of the "TATA" model to the NMR structures of SRY-DNA³⁵ and LEF-1-DNA³⁴ complexes reveals closer similarity to the LEF-1-DNA structure (Table II). The r.m.s.d. between the TATA model and the NMR structures is 5.7 Å for the SRY-DNA complex and 4.6 Å for the LEF-1-DNA

complex for all of the protein and DNA backbone atoms, compared to the r.m.s.d. of 4.0 Å between the LEF-1 and SRY complexes. The region of greatest similarity is the short wing of the protein, for which the protein backbone r.m.s.d. are 3.0 Å for SRY and 2.2 Å for LEF-1, and the r.m.s.d. between LEF-1 and SRY is 2.9 Å. These and other r.m.s.d. values (Table II) demonstrate that the "TATA" model differs from the NMR structures to almost the same extent as the NMR structures differ from each other.

A reorientation of the protein position on the DNA occurs during the MD simulation. When the protein backbones are superimposed, the r.m.s.d. between the DNA backbones of the starting "TATA" model and the NMR structures is 7.0 Å for the LEF-1 and 8.4 Å for SRY. After the simulation, the values decrease to 5.6 and 7.3 Å, correspondingly (Table II). The orientation of the protein binding site alone approaches only that of the LEF-1 structure: the initial r.m.s.d. of 5.6 Å decreases to 4.3 Å, vs. 6.9 to 6.7 Å for SRY.

The main difference between the TATA model and SRY and LEF-1 is in helix III. The C-terminal part of helix III bends toward the DNA in all of the structures but in both SRY and LEF-1 α -helix III ends abruptly at Pro67 and interacts with the DNA minor groove with residues 68 through 74 in an extended conformation. In contrast, helix III of HMG-D bends at residue 62, but α -helical structure is conserved up

(a)

PROTEIN SEQUENCES

	1	10	20	30	40	50	60	70	80
HMG-D:	.SDKPKRPLS	AYMLWLNSAR	ESIKRENPGI	KVTEVAKRGG	ELWRAMKD--KS	EWEAKAAKAK	DDYDRAVKEF	EANG	
SRY:	.DRVKRPMN	AFIVWSRDQR	RKMALENPRM	RNSEISKQLG	YQWKMLTEAEKW	PFQEAQKLQ	AMHREKYPNY	KYR	
LEF-1:	.MHIKKPLN	AFMLYMKEMR	ANVVAECTLK	ESAAINQILG	RRWHALSREEQA	KYYELARKER	QLHMQLYPGW	SARDNYGKKK	KRKREK
HMG1:	KGDEPKKPRGKMS	SYAFFVQTSR	EEHKKKHPSV	NFSEFSKKCS	ERWKTMSAKEKG	KFEDMAKADK	ARYEREMKTY	IPP	
				DA					
	1	10	20	30	40	50	60	70	80

(b)

INTERACTIONS WITH DNA

	1	10	20	30	40	50	60	70	80
HMG-D:		P WPHHP	HH SW S	P	PH HHP	W	W W	W W	
SRY:		P P HS	HH L	P	S HSL	H	L P	LH	
LEF-1:		LPHHS	HHH HP	P	SH SL		P P	HP H	
HMG1:	Xo	XX oXXX	o oXX	X Xo	o oXX	oXoo Xo	XX o	o o o o o o o o o o	o X o
	1	10	20	30	40	50	60	70	80

(c)

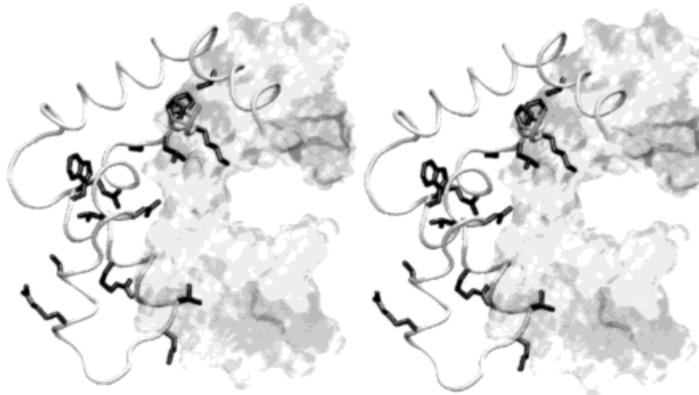


Fig. 9. Protein-DNA interactions in the predicted structure of the HMG-box-DNA complex ("TATA" model) compared to the experimentally determined structures of LEF-1-DNA³⁴ and human SRY-DNA³⁵ complexes. (a) Alignment of the HMG-box sequences for HMG-D, SRY, and LEF-1¹³ used in our modeling study (HMG-D) and NMR experiments.^{34,35} (b) The "signatures" of the protein-DNA contacts. The contacts are labeled as "P"—a polar a.a. interaction with the DNA backbone, "S"—a direct interaction of a polar a.a. with one of the DNA bases, "H"—a hydrophobic contact with the DNA, "W"—contact through a water bridge (in the "TATA" model), "L"—a "loose" contact observed in fewer than 20% of the LEF-1 or SRY NMR structures or of a distance 4–5 Å, "X"—a large, and "o"—an appreciable change in the NMR chemical shift of an HMG1 residue upon binding to DNA. (c)

Stereo diagram of our model of the HMG-D-DNA complex showing only the residues equivalent to those of HMG1 having a large NMR chemical shift change. (d) Hydrogen bond and salt bridge contacts and (e) hydrophobic contacts in the three complexes. Contacts to the DNA bases are shown as solid lines, contacts to the sugar-phosphate backbone as dashed lines. Dashed lines to circles represent contacts to the phosphates, dashed lines to cells represent contacts to the ribose rings. Residues establishing hydrophobic interactions with the DNA are colored black; the hydrogen bond or salt-bridge formers are shown in white. Ticks (">") indicate water-mediated contacts in the model structure, with the number of ticks showing the number of participating water molecules.

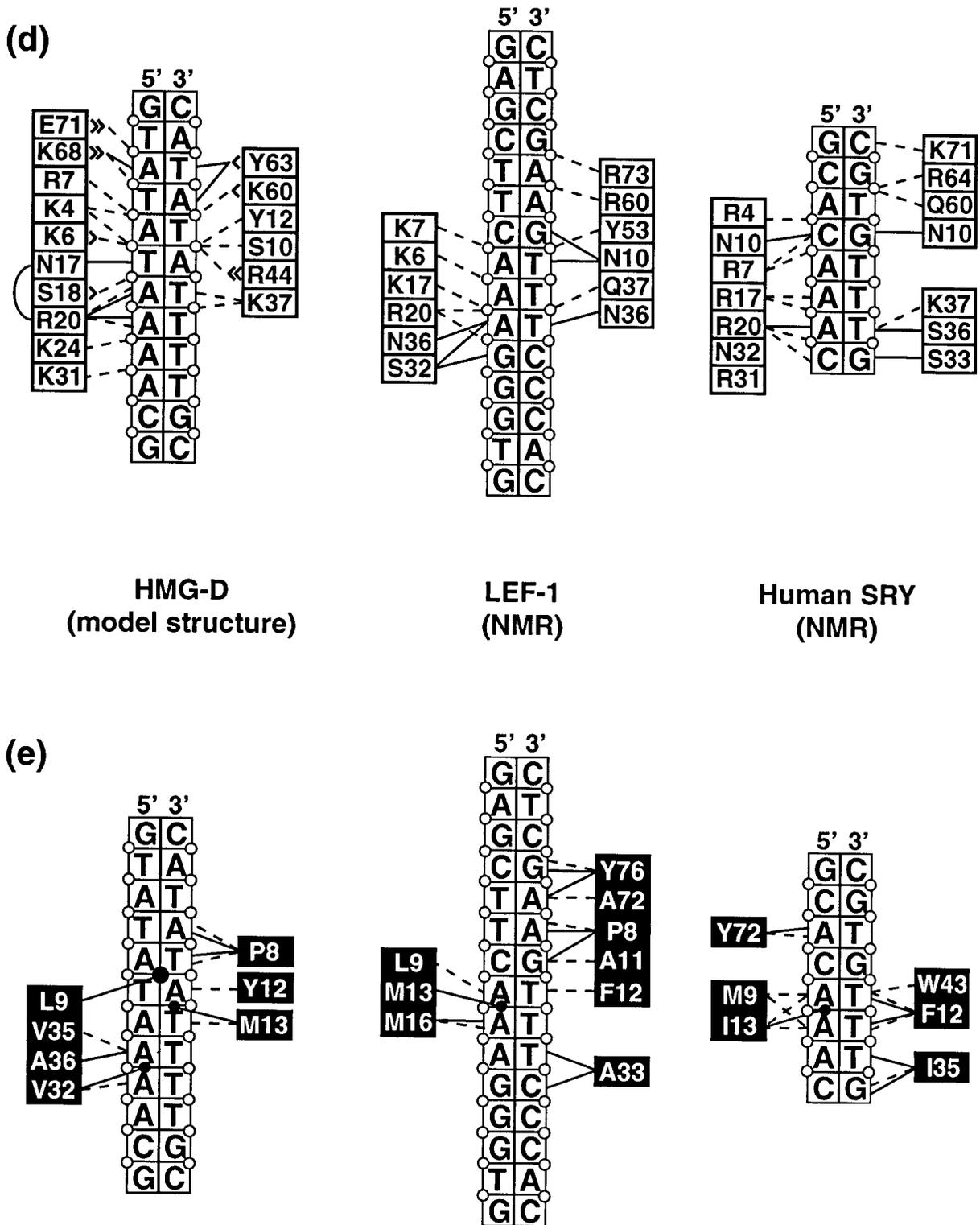


Figure 9. (Continued.)

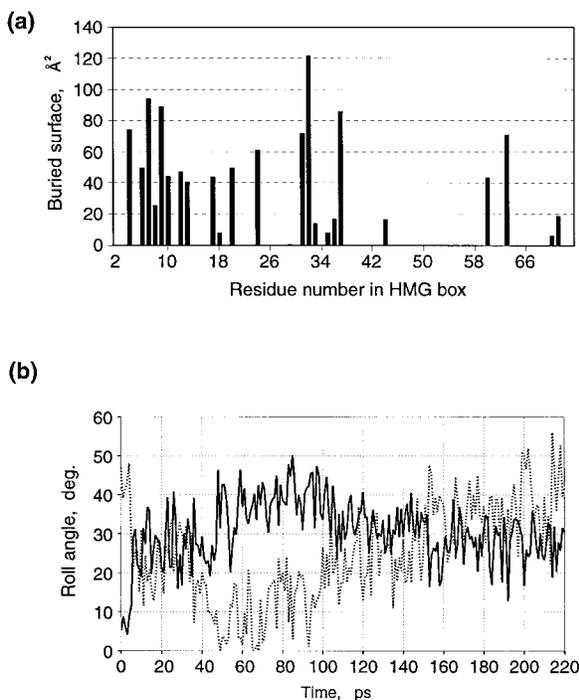


Fig. 10. Protein-DNA interactions in the predicted structure of the HMG-D-DNA complex ("TATA" model). (a) Distribution of the buried surface area of side chains among HMG-D residues (averaged over the last 60 ps of MD). (b) Comparison of DNA roll angles at the Leu9 (dotted line) and Met13 (solid line) intercalation sites.

to the C-terminus of the HMG-box, at residue 74; there is no proline at position 67 in HMG-D. This structural difference between HMG-D and the sequence-specific proteins LEF-1 and SRY is consistent with comparative sequence analysis of the sequence-specific vs. nonsequence-specific HMG-box proteins in the region of helix III.¹³ Furthermore, a slight bend at residue 62 is present in the free HMG-D structure,³² and the NH_i-NH_{i+1} NOE cross-peaks observed in the C-terminus of the HMG-box persist on formation of the HMG-D-DNA complex (Dow, Wolfe, Verdine, Jones, and Churchill; unpublished results). There is another difference among the three HMG-box structures in the C-terminal region: residues 74–86 in the LEF-1-DNA complex form a protein "tail" that crosses over the DNA backbone and interacts with the DNA major groove. This interaction is not observed for SRY and HMG-D because the proteins are truncated at residues 73 and 74, respectively. However, a similar interaction is possible for HMG-D because HMG-D has a long (≈ 30 a.a.) C-terminal "tail" with many positively charged amino acids adjacent to the HMG-box.

The DNA bending angle in the "TATA" model is 72° for the entire DNA fragment and 92° for only the protein-binding region of the DNA. The values are

closer to the DNA bend in the SRY-DNA complex ($70\text{--}80^\circ$)³⁵ than to that in the LEF-1-DNA complex ($117^\circ \pm 10^\circ$).³⁴ The interactions of the DNA major groove with the positively charged C-terminal region of LEF-1, which SRY and the current HMG-D model lack, appear to contribute to this difference.³⁴ The average DNA unwinding angles per base pair relative to canonical B-form DNA are 7.4° for the SRY-DNA complex, 2.9° for the LEF-1-DNA complex, and 8.2° for the "TATA" model. For the DNA in the protein binding site alone, the unwinding angles per base pair are 6.6° for the LEF-1-DNA complex and 12.1° for the "TATA" model. A comparison of the contacts established with DNA by LEF-1, SRY, and HMG-D reveals a similar "signature": DNA is contacted by almost the same residues in all three structures (Fig. 9b). HMG-D amino acids interacting with DNA almost always have counterparts in at least one of the other two structures. Two hydrophobic contacts are missed because the corresponding HMG-D residues are buried in the main hydrophobic core (Leu16, Trp43). When SRY and LEF-1 make base-specific contacts that contribute to sequence-specificity, the equivalent residues in the HMG-D-DNA complex interact with the DNA backbone or establish sequence-neutral hydrophobic interactions with the DNA bases. In the absence of base-specific contacts, two additional sequence-neutral interactions, partial intercalations between the DNA bases by HMG-D residues 9 and 32, could be important for stability of the nonsequence-specific complex. These additional intercalation sites are not observed in the NMR structures of the SRY- and LEF-1-DNA complexes.

DISCUSSION

Implications for the Sequence-Tolerance of Chromosomal HMG-Domain Proteins

Proteins can achieve site-specificity by forming a complementary interface and specific hydrogen bonding interactions with the DNA binding site in either the major or minor grooves (reviewed in refs. 1–3, 38). Such interactions have been visualized in detail in recent structural studies of the sequence-specific HMG-box protein-DNA complexes; the SRY and LEF-1 NMR studies show that shape complementarity and particular hydrogen bonding interactions may be important for sequence-specificity,^{34,35} as illustrated in Figures 6 and 9. Inspection of the protein-DNA contacts shows that in the LEF-1-DNA complex Asn10, Ser32, and Asn36 directly contact nucleotide bases in the binding site DNA sequence CCTTTGAA.²¹ These residues form contacts that bridge base pairs at the TG, AG (CT), and the central TT base steps, respectively (Fig. 9d). In SRY, specific base contacts are mediated by Asn10, Ser33, and Ser36. Asn10 contacts three out of the four bases of a TG base step and the serines establish hydrogen bonds with the bases at the GT step (Fig. 9d),

TABLE II. Root-mean-square Deviations[†] (Å) Between the Protein/DNA Backbone Atoms of the “TATA” Model and the NMR Structures of LEF-1-DNA and SRY-DNA Complexes

Region [‡]	Overlap [§]	HMG-D vs. LEF-1	HMG-D vs. SRY	LEF-1 vs. SRY
Overall structure	Overall structure	4.6 ± 0.1	5.7 ± 0.1	4.0 ± 0.2
Protein	Overall structure	4.7 ± 0.1	6.0 ± 0.1	4.1 ± 0.2
Protein	Protein	4.3 ± 0.1	5.6 ± 0.1	3.9 ± 0.2
Protein short wing [¶]	Protein	2.7 ± 0.1	4.5 ± 0.2	3.6 ± 0.2
Protein short wing	Protein short wing	2.2 ± 0.1	3.0 ± 0.2	2.9 ± 0.1
Protein short wing	Protein long wing	10.3 ± 1.0	10.2 ± 0.9	14.1 ± 1.2
Protein long wing [¶]	Protein	5.7 ± 0.2	6.7 ± 0.2	4.2 ± 0.2
Protein long wing	Protein long wing	6.6 ± 0.2	5.1 ± 0.1	2.3 ± 0.2
Protein long wing	Protein short wing	5.0 ± 0.2	10.1 ± 0.6	7.3 ± 0.6
N-terminal stretch [¶]	N-terminal stretch	1.2 ± 0.2	1.8 ± 0.3	1.6 ± 0.3
N-terminal stretch	Protein short wing	7.5 ± 0.7	9.4 ± 0.6	6.7 ± 0.8
N-terminal stretch	Protein long wing	5.3 ± 0.3	6.2 ± 0.3	2.4 ± 0.5
N-terminal stretch	Protein	6.1 ± 0.4	8.2 ± 0.3	4.5 ± 0.1
α-helix I [¶]	α-helix I	2.2 ± 0.1	2.3 ± 0.1	0.8 ± 0.1
α-helix I	Protein short wing	2.4 ± 0.1	2.8 ± 0.2	1.4 ± 0.1
α-helix I	Protein	2.8 ± 0.1	4.0 ± 0.2	2.1 ± 0.2
α-helix II [¶]	α-helix II	1.3 ± 0.1	1.7 ± 0.1	1.8 ± 0.1
α-helix II	Protein short wing	1.5 ± 0.1	2.7 ± 0.2	3.1 ± 0.2
α-helix II	Protein	2.1 ± 0.2	4.9 ± 0.2	4.4 ± 0.2
α-helix III [¶]	α-helix III	4.2 ± 0.2	3.8 ± 0.2	2.2 ± 0.2
α-helix III	Protein long wing	5.0 ± 0.2	5.0 ± 0.2	2.2 ± 0.2
α-helix III	Protein	5.6 ± 0.2	6.0 ± 0.2	4.1 ± 0.3
DBI [¶]	DBI	3.0 ± 0.2	4.1 ± 0.2	3.0 ± 0.2
DBI	Protein short wing	3.6 ± 0.2	4.4 ± 0.2	3.3 ± 0.2
DBI	Protein	3.5 ± 0.2	5.1 ± 0.2	3.4 ± 0.2
DNA	Overall structure	4.6 ± 0.2	5.1 ± 0.2	3.8 ± 0.2
DNA	DNA	4.0 ± 0.2	3.4 ± 0.2	3.3 ± 0.2
DNA	Protein	5.6 ± 0.3	7.3 ± 0.4	5.3 ± 0.4
DNA	DBI	5.8 ± 0.2	6.6 ± 0.4	4.5 ± 0.3
PBS [¶]	Overall structure	3.2 ± 0.1	4.0 ± 0.2	3.5 ± 0.2
PBS	DNA	3.1 ± 0.1	3.4 ± 0.3	3.5 ± 0.2
PBS	PBS	2.6 ± 0.1	1.4 ± 0.1	2.3 ± 0.1
PBS	Protein	4.3 ± 0.2	6.7 ± 0.4	5.0 ± 0.4
PBS	DBI	4.2 ± 0.1	3.2 ± 0.3	3.6 ± 0.4
DBI + PBS	Overall structure	3.5 ± 0.1	4.5 ± 0.1	3.4 ± 0.2
DBI + PBS	Protein	3.9 ± 0.2	5.7 ± 0.2	4.1 ± 0.2
DBI + PBS	DNA	4.5 ± 0.1	6.6 ± 0.2	3.8 ± 0.2
DBI + PBS	DBI + PBS	3.3 ± 0.1	3.6 ± 0.1	3.0 ± 0.2

[†]The data are averaged over the last 60 ps of MD.

[‡]Region refers to the residues/DNA bases that are compared between the “TATA” model, LEF-1-DNA, and SRY-DNA complexes.

[§]Overlap refers to the parts of the complexes that were aligned prior to r.m.s.d. calculation.

[¶]Protein regions are defined in Methods. DBI = DNA binding interface, and PBS = protein binding site.

contributing to the specificity for the GCACAAAC site recognized by SRY in the MIS promoter.^{20,37} Incidentally, SRY residues 31 and 32 establish no contact to the DNA bases, because the DNA fragment employed was only 8 bp long³⁵ and lacked a base step that might have been contacted by analogy with the LEF-1-DNA complex.

The protein regions contributing to the specificity of the HMG-box transcription factors have been examined by domain swap studies with the chromosomal protein HMG1 and human T-cell enhancer.⁶¹ Crane-Robinson and co-workers⁶¹ showed that many, if not all, of the amino acids required for the sequence-specificity of TCF1 α reside in the HMG-domain long

wing (residues 1–10 and 55–74). Of the residues that make base-specific DNA contacts in the NMR structures, only Asn10 lies in this region and residues 32 and 36 do not. However, the domain swap shows that the chimeric protein, consisting of the HMG1 short wing and TCF1 α long wing, binds to a specific DNA fragment nearly as well as the wild-type TCF1 α in the absence of specific or nonspecific competitor DNA. From these experiments alone it is difficult to distinguish between the possible contributions of the short wing residues 32 and 36, which in TCF1 α may confer slightly higher affinity and/or higher specificity for the binding site DNA than in the chimeric protein.

Comparison of the patterns of protein–DNA contacts between the transcription factor HMG proteins and the “TATA” model reveals how the nonsequence-specific HMG proteins can use a similar DNA binding mode but still retain the ability to bind to many different DNA sequences. Our analysis indicates that the amino acids occupying positions 10, 32, and 36 of the HMG-box are important for binding specificity. In HMG-D, the equivalent residues are Ser10, Val32, and Ala36. Ser10 forms a hydrogen bond with the DNA backbone. Val32 and Ala36 contribute to the hydrophobic ridge buried in the DNA minor groove. In contrast to the sequence-specific proteins, no polar contact to the DNA bases is formed at any of these positions.

The only direct polar interaction between HMG-D and the DNA bases in the “TATA” model structure may be relatively sequence-neutral. The DNA-binding “fork” Asn17–Arg20 contacts the bases of 5′-TAA-3′, whereas the equivalent residues of SRY and LEF-1, Arg/Lys17 and Arg20, interact with the DNA phosphodiester backbone,^{34,35} although Arg20 contacts the DNA bases in 20% of the SRY NMR structures. However, the first thymine base of the “TAA” segment in the HMG-D-DNA model could be replaced with a cytosine; this would also expose an oxygen to the minor groove. Both adenines could be replaced with guanines because the purine N₃ atoms are nearly equivalent and there is space in the model to accommodate the 2-amino groups. Finally, a pyrimidine could replace the last adenine. The pyrimidine O₂ oxygen would also be a good donor for the hydrogen bond and the flexible side chain of Arg20 could allow space for the oxygen to protrude into the minor groove. Thus, theoretically, any 5′-YR(Y/R)-3′ sequence could replace the TAA, lifting most restrictions on the sequence-specificity of this particular protein–DNA interaction. In fact, a slight preference for the YR base step is consistent with the preference of HMG-D for DNA containing the dinucleotide sequence “TG” known from DNA binding site selection experiments.⁶

In contrast to the sequence-specific contact made by residue 32 in the LEF-1, the equivalent residue of HMG-D, Val32, forms a sequence-neutral intercalation at the 8th base step of the DNA. This valine has the most intensive hydrophobic interactions of all residues of the HMG-D HMG-box. Remarkably, in the other nonsequence-specific HMG proteins position 32 is always (with the exception of the UBF proteins) occupied by a hydrophobic (Val, Ile) or an aromatic (Phe) residue, which are known intercalating residues.^{38,51,59,62} In the sequence-specific proteins, this amino acid is always polar (Asn, Gln, Ser, His).¹³ Thus, the contacts with DNA established by residue 32 may also be important for determining the specificity of DNA recognition; there is a “sequence-neutral” intercalation interaction with DNA in the nonsequence-specific proteins, whereas a po-

lar residue forms direct hydrogen bonds to the DNA bases in the sequence-specific proteins.^{34,35}

In the structures of LEF-1³⁴ and SRY,³⁵ residue 13 of the HMG box inserts its side chain between the DNA bases, in agreement with earlier observations.^{37,59} In our model structure, the intercalation of Met13 was pre-engineered and was conserved during the simulations. Leu9 was also found to partially intercalate at the adjacent step of the DNA helix. The two residues intercalate interchangeably, competing for the space between the 5th and 7th DNA base pairs, with Leu9 dominating. Interestingly, many chromosomal HMG proteins (especially those from the HMG1.1 subgroup)¹³ have a small (Ala) or a polar (Gln, Glu) amino acid at the 13th position; such residues are unlikely to intercalate. However, at position 9 these proteins have an aliphatic residue (Met, Leu), which are capable of intercalating DNA.^{34,35,38,62} Thus, it is possible that the characteristic partial intercalation of HMG proteins occurs at residue 9 rather than 13 in chromosomal HMG proteins. The dominance of the Leu9 intercalation found in our model is suggestive of such a possibility.

Relevance of the Methods Used and the Model Predicted

A model of the chromosomal protein HMG-D-DNA complex has been built by docking experimentally determined structures of the protein and DNA followed by MD simulations. An alternative, recently published method for structure prediction⁶³ implemented an extensive search in conformational space of the DNA bound to the protein combined with an energy minimization procedure. However, the method requires a dyad symmetry of the protein and hence is not applicable in our case because HMG-D binds DNA as a monomer.⁶ Another strategy, a variant of homology modeling (reviewed in ref. 64), consists of the reconfiguration of a previously determined structure by altering either the protein or DNA sequence, followed by minimization or molecular dynamics simulation. This method has been applied primarily to the prediction of protein structures (refs. 49, 65, and many others); there is only a single instance of its use for study of protein–DNA complexes.^{45,46} We could not use this approach because no structure of an HMG-box-DNA complex was known at the beginning of the study. Later, when the structure of the SRY-DNA complex was determined,³⁵ the coordinates were kindly provided to us by the authors prior to the release in the PDB. Preliminary results from our attempts to model the HMG-D-DNA complex on the basis of the SRY-DNA complex coordinates showed an instability of the modeled complex. Apparently, HMG-D is less similar to SRY (r.m.s.d. of 5.1 Å for the protein C^α atoms) than would be necessary for the success of the method. The coordinates of the LEF-1-DNA complex were released in the PDB only after this study had been completed.

Our approach was to prepare several reasonable docked model structures and then subject them to molecular dynamics simulations. The requirements of stability and conservation of the pre-engineered Met13 partial intercalation were used to evaluate the quality of the trial structures. Continued MD simulation proved the stability of the selected model (the "TATA" model) and provided data for structural analysis. The 1.5 Å r.m.s.d. of the resulting family of structures is within the range expected of a moderate to high resolution family of structures determined by NMR.

The methods used here are expected to work well when the structures of the protein and DNA chosen for docking are close to those in the complex, and only adjustments of the contacts at the protein-DNA interface are necessary to stabilize the structure after docking. Indeed, the relatively short MD runs employed (several hundred picoseconds) are unlikely to produce large conformational changes. On the other hand, the actual scale of the required changes is unknown a priori, and longer MD simulations did not appear to be necessary because the chosen model reached a stable structure in less than 160 picoseconds of simulation. The accuracy of the method may be improved by using more experimental constraints in the docking and MD procedures.

One region of our model that may be improved by adjustments in the modeling procedure is the interface between the C-terminal region of the protein and the DNA. This region differs between the SRY-DNA and LEF-1-DNA NMR structures and our model.^{34,35} Some of these differences may be real; for example, the α -helical structure of the C-terminus of HMG-D.^{7,32} However, the bend in α -helix III and slight differences in contacts between HMG1 and HMG-D may stem from the unbiased approach taken in positioning the water molecules and ions in this region of the model. The N-terminus and the short wing of the HMG-box were located close to the DNA, fitting well after the initial docking in the minor groove; the resulting position of the C-terminal part of helix III left a space between the protein and the DNA. Consequently, water molecules entered the space during the solvation process. In addition, one of the sodium ions, placed into a position of low electrostatic potential, was located between the C-terminal part of the protein and DNA base pairs 1 and 2, at a distance of about 5 Å from each base pair. The short simulation time did not allow the water and the ion to move enough for close protein-DNA contacts to be established and the protein structure rearrangement required to establish such contacts did not occur during the simulation. Future modeling efforts may apply the desired changes in the docked complex prior to solvation by manually adjusting the structure or by applying additional molecular dynamics or minimization procedures to only the regions of interest.

The main features of our HMG-D-DNA model are supported by considerable experimental evidence. First, the model is similar to the experimentally determined structures of other HMG-box-DNA proteins^{34,66} in the protein/DNA geometry and the "signature" of DNA-contacting residues (Fig. 9). In fact, the overall model is within the limit of observed differences between the SRY and LEF-1 structures, which are more closely related to each other in sequence and function than to HMG-D. Furthermore, the DNA geometry is consistent with a recent experimental estimate of the HMG-D-DNA bending angle. Second, the protein-DNA contacts observed in the model are consistent with the regions of HMG1 that undergo the largest changes in NMR chemical shift upon binding to DNA, with few exceptions.⁷ Finally, any differences between the number and interactions of intercalating residues in the HMG-D model are also reflected by sequence conservation of those residues in other nonsequence-specific HMG-box proteins.¹³ However, ultimate verification of the predicted features will be achieved only through forthcoming structures of the HMG-D-DNA complex determined by X-ray crystallography or NMR.

ACKNOWLEDGMENTS

We thank Drs. Scot Wolfe and Greg Verdine, who kindly provided the coordinates for the disulfide cross-linked DNA structure from their NMR study prior to publication. We thank Dr. Marius Clore for providing the coordinates of the SRY-DNA complex NMR structure prior to their release by the PDB. We thank Dr. Tom Bishop, Ilya Logunov, Dorina Kosztin, and Dr. Dong Xu for helpful discussions in the course of this work.

REFERENCES

1. Steitz, T.A. Structural studies of protein-nucleic acid interaction: The source of sequence-specific binding. *Quart. Rev. Biophys.* 23:205-280, 1990.
2. Pabo, C.O., Sauer, R.T. Transcription factors: Structural families and principles of DNA recognition. *Ann. Rev. Biochem.* 61:1053-1095, 1992.
3. Travers, A.A. DNA-protein interactions. London/New York: Chapman & Hall, 1993.
4. Von Hippel, P.H., Berg, O.G. On the specificity of DNA-protein interactions. *Proc. Natl. Acad. Sci. USA* 83:1608-1612, 1986.
5. Paull, T.T., Haykinson, M.J., Johnson, R.C. The nonspecific DNA-binding and -bending proteins HMG1 and HMG2 promote the assembly of complex nucleoprotein structures. *Genes Dev.* 7:1521-1534, 1993.
6. Churchill, M.E.A., Jones, D.N.M., Glaser, T., Hefner, H., Searles, M.A., Travers, A.A. HMG-D is an architecture-specific protein that preferentially binds to DNA containing the dinucleotide TG. *EMBO J.* 14:101-112, 1995.
7. Hardman, C.H., Broadhurst, R.W., Raine, A.R.C., Grasser, K.D., Thomas, J.O., Laue, E.D. Structure of the A-domain of HMG1 and its interactions with DNA as studied by heteronuclear three- and four-dimensional NMR spectroscopy. *Biochemistry* 34:16596-16607, 1995.
8. Bustin, M., Reeves, R. High-Mobility-Group chromosomal proteins: Architectural components that facilitate chromatin function. *Prog. Nucleic Acid Res. Mol. Biol.* 54:35-100, 1996.

9. Bianchi, M.E., Falciola, L., Ferrari, S., Lilley, M.D.J. The DNA binding site of HMG1 protein is composed of two similar segments (HMG boxes), both of which have counterparts in other eukaryotic regulatory proteins. *EMBO J.* 11:1055–1063, 1992.
10. Ner, S.S. DNA-binding proteins HMGs everywhere. *Curr. Biol.* 2:208–210, 1992.
11. Wagner, C.R., Hamana, K., Elgin, S.C.R. A high-mobility-group protein and its cDNAs from *Drosophila melanogaster*. *Mol. Cell. Biol.* 12:1915–1923, 1992.
12. Ner, S.S., Travers, A.A. HMG-D, the *Drosophila melanogaster* homologue of HMG 1 protein, is associated with early embryonic chromatin in the absence of histone H1. *EMBO J.* 13:1817–1822, 1994.
13. Baxevasanis, A.D., Landsman, D. The HMG-1 box protein family: Classification and functional relationships. *Nucl. Acids Res.* 23:1604–1613, 1995.
14. Travers, A.A., Ner, S.S., Churchill, M.E.A. DNA chaperones: A solution to a persistence problem? *Cell* 77:167–169, 1994.
15. Read, C.M., Cary, P.D., Crane-Robinson, C., Driscoll, P.C., Carillo, M.O.M., Norman, D.G. The structure of the HMG box and its interactions with DNA. In: "Nucleic Acids and Molecular Biology." Ekstein, F., Lilley, D.M.J. (eds.). Berlin: Springer-Verlag, 1995:222–250.
16. Ura, K., Nightingale, K., Wolffe, A. Differential association of HMG1 and linker histones B4 and H1 with dinucleosomal DNA—structural transitions and transcriptional repression. *EMBO J.* 15:4959–4969, 1996.
17. Ge, H., Roeder, R.G. The high mobility group protein HMG1 can reversibly inhibit class II gene transcription by interaction with the TATA-binding protein. *J. Biol. Chem.* 269:17136–17140, 1994.
18. Zappavigna, V., Falciola, L., Citterich, M.H., Mavillo, F., Bianchi, M.E. HMG1 interacts with HOX proteins and enhances their DNA binding and transcriptional activation. *EMBO J.* 15:4981–4991, 1996.
19. Grosschedl, R., Giese, K., Pagel, J. HMG domain proteins: Architectural elements in the assembly of nucleoprotein structures. *Trends Genet.* 10:94–100, 1994.
20. Harley, V.R., Jackson, D.L., Hextall, P.J., Hawkins, J.R., Berkovitz, G.D., Sockanathan, S., Lovell-Badge, R., Goodfellow, P.N. DNA-binding activity of recombinant SRY from normal males and XY females. *Science* 255:453–456, 1992.
21. Giese, K., Amsterdam, A., Grosschedl, R. DNA-binding properties of the HMG domain of the lymphoid-specific transcriptional regulator LEF-1. *Genes Dev.* 5:2567–2578, 1991.
22. Falciola, L., Murchie, A.I.H., Lilley, D.M.J., Bianchi, M.E. Mutational analysis of the DNA binding domain A of chromosomal protein HMG1. *Nucl. Acids Res.* 22:285–292, 1994.
23. Pontiggia, A., Rimini, R., Harley, V.R., Goodfellow, P.N., Lovell-Badge, R.L., Bianchi, M.E. Sex-reversing mutations affect the architecture of SRY-DNA complexes. *EMBO J.* 13:6115–6124, 1994.
24. Ferrari, S., Harley, V.R., Pontiggia, A., Goodfellow, P.N., Lovell-Badge, R., Bianchi, M.E. SRY, like HMG1, recognizes sharp angles in DNA. *EMBO J.* 11:4497–4506, 1992.
25. Treiber, D.K., Zhai, X., Jantzen, H.-M., Essigman, J.M. Cisplatin-DNA adducts are molecular decoys for the ribosomal RNA transcription factor hUBF (human upstream binding factor). *Proc. Natl. Acad. Sci. USA* 91:5672–5676, 1994.
26. Brown, S.J., Kellett, P.J., Lippard, S.J. Ixrl, a yeast protein that binds to platinated DNA and confers sensitivity to cisplatin. *Science* 261:603–605, 1993.
27. Wolfe, S.A., Ferentz, A.E., Grantcharova, V., Churchill, M.E.A., Verdine, G.L. Modifying the helical structure of DNA by design: Recruitment of an architecture-specific protein to an enforced DNA bend. *Chem. Biol.* 2:213–221, 1995.
28. Chow, C.S., Whitehead, J.P., Lippard, S.J. HMG domain proteins induce sharp bends in cisplatin-modified DNA. *Biochemistry* 33:15124–15130, 1994.
29. Giese, K., Cox, J., Grosschedl, R. The HMG box of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell* 69:185–195, 1992.
30. Weir, H.M., Kraulis, P.J., Hill, C.S., Raine, A.R.C., Laue, E.D., Thomas, J.O. Structure of the HMG box motif in the B-domain of HMG1. *EMBO J.* 12:1311–1319, 1993.
31. Read, C.M., Cary, P.D., Crane-Robinson, C., Driscoll, P.C., Norman, D.G. Solution structure of a DNA-binding domain from HMG1. *Nucl. Acids Res.* 21:3427–3436, 1993.
32. Jones, D.N.M., Searles, M.A., Shaw, G.L., Churchill, M.E.A., Ner, S.S., Keeler, J., Travers, A.A., Neuhaus, D. The solution structure and dynamics of the DNA-binding domain of HMG-D from *Drosophila melanogaster*. *Structure* 2:609–627, 1994.
33. Van Houte, L.P.A., Chuprina, V.P., van de Wetering, M., Boelens, R., Kaptein, R., Clevers, H. Solution structure of the sequence-specific HMG box of the lymphocyte transcriptional activator Sox-4. *J. Biol. Chem.* 270:30516–30524, 1995.
34. Love, J.J., Li, X., Case, D.A., Giese, K., Grosschedl, R., Wright, P.E. Structural basis for DNA bending by the architectural transcription factor LEF-1. *Nature* 376:791–795, 1995.
35. Werner, M.H., Huth, J.R., Gronenborn, A.M., Clore, G.M. Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell* 81:705–714, 1995.
36. Van de Wetering, M., Clevers, H. Sequence-specific interaction of the HMG box domain proteins TCF-1 and SRY within the minor groove of a Watson-Crick double helix. *EMBO J.* 11:3039–3044, 1992.
37. King, C.-Y., Weiss, M.A. The SRY high-mobility-group box recognizes DNA by partial intercalation in the minor groove: A topological mechanism of sequence specificity. *Proc. Natl. Acad. Sci. USA* 90:11990–11994, 1993.
38. Werner, M.H., Gronenborn, A.M., Clore, G.M. Intercalation, DNA kinking, and the control of transcription. *Science* 271:778–784, 1996.
39. Brooks C.L. III, Karplus, M., Pettitt, B.M. "Advances in Chemical Physics, Vol. LXXI." New York: John Wiley & Sons, 1988.
40. Beveridge, D.L., Ravishanker, G. Molecular dynamics studies of DNA. *Curr. Opin. Struct. Biol.* 4:246–255, 1994.
41. Brooks C.L. III, Methodological advances in molecular dynamics simulations of biological systems. *Curr. Opin. Struct. Biol.* 5:211–215, 1995.
42. De Vlieg, J., Berendsen, H.J.C., van Gunsteren, W.F. An NMR-based molecular dynamics simulation of the interaction of the *lac* repressor headpiece and its operator in aqueous solution. *Proteins* 6:104–127, 1989.
43. Eriksson, M.A.L., Hiird, T., Nilsson, L. Molecular dynamics simulations of the glucocorticoid receptor DNA-binding domain in complex with DNA and free in solution. *Biophys. J.* 68:402–426, 1995.
44. Harris, L.F., Sullivan, M.R., Popken-Harris, P.D., Hickok, D.F. Molecular dynamics simulations in solvent of the glucocorticoid receptor protein in complex with a glucocorticoid response element DNA sequence. *J. Biomol. Struct. Dyn.* 12:249–270, 1994.
45. Bishop, T.C., Schulten, K. Molecular dynamics study of glucocorticoid receptor DNA binding. *Proteins* 24:115–133, 1996.
46. Bishop, T.C., Kosztin, D., Schulten, K. How hormone receptor-DNA binding affects nucleosomal DNA: The role of symmetry. *Biophys. J.* 72:2056–2067, 1997.
47. Kosztin, D., Bishop, T.C., Schulten, K. Binding of the estrogen receptor to DNA: The role of waters. *Biophys. J.* 73:557–570, 1997.
48. Billeter, M., Giintert, P., Luginbihl, P., Wiithrich, K. Hydration and DNA recognition by homeodomains. *Cell* 85:1057–1065, 1996.
49. Li, L., Darden, T., Hiskey, R., Pedersen, L. Homology modeling and molecular dynamics simulations of the Gla domains of human coagulation factor IX and its G[12]A mutant. *J. Phys. Chem.* 100:2475–2479, 1996.
50. "Quanta." Polygen Corp., Waltham, MA, 1988.
51. Kim, Y., Geiger, J.H., Hahn, S., Sigler, P.B. Crystal struc-

- ture of a yeast TBP/TATA-box complex. *Nature* 365:512–520, 1993.
52. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690, 1984.
 53. Brünger, A.T. "X-PLOR, Version 3.1: A System for X-ray Crystallography and NMR." New Haven, CT: Yale University, Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, 1992.
 54. MacKerell A.D. Jr., Wiorkiewicz-Kuczera, J., Karplus, M. An all-atom empirical energy function for the simulation of nucleic acids. *J. Am. Chem. Soc.* 117:11946–11975, 1995.
 55. MacKerell Jr., A.D., Bashford, D., Bellott, M., Dunbrack Jr., R.L., Evanseck, J., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Reiher, I.W.E., Roux, B., Schlenkrich, M., Smith, J., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., Karplus, M. All-hydrogen Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins using the CHARMM22 Force Field. Submitted for publication, 1997.
 - MacKerell Jr., A.D., Bashford, D., Bellott, M., Dunbrack Jr., R.L., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph, D., Kuchnir, L., Kuczera, K., Lau, F.T.K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D.T., Prodhom, B., Roux, B., Schlenkrich, M., Smith, J.C., Stote, R., Straub, J., Wiorkiewicz-Kuczera, J., Karplus, M. Self-consistent parameterization of biomolecules for molecular modeling and condensed phase simulations. *FASEB Journal* 6:A143, 1992.
 56. Ravishanker, G., Swarninathan, S., Beveridge, D.L., Lavery, R., Sklenar, H. Conformational and helicoidal analysis of 30 ps of molecular dynamics on the d(CGCGAATTC-GCG) double helix. *J. Biomol. Struct. Dyn.* 6:669–699, 1989.
 57. Humphrey, W.F., Dalke, A., Schulten, K. VMD—Visual molecular dynamics. *J. Mol. Graph.* 14:33–38, 1996.
 58. Lee, B., Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* 55:379–400, 1971.
 59. Haqq, C.M., King, C.-Y., Ukiyama, E., Falsafi, S., Haqq, T.N., Donahoe, P.K., Weiss, M.A. Molecular basis of mammalian sexual determination: Activation of Müllerian inhibiting substance gene expression by SRY. *Science* 266:1494–1500, 1994.
 60. Dow, L.K., Changela, A., Hefner, H.E., Churchill, M.E.A. Oxidation of a critical methionine modulates DNA binding of the *Drosophila melanogaster* high-mobility group protein, HMG-D. *FEBS Letters* 414:514–520, 1997.
 61. Read, C.M., Cary, P.D., Preston, N.S., Lnenicek-Allen, M., Crane-Robinson, C. The DNA sequence specificity of HMG boxes lies in the minor wing of the structure. *EMBO J.* 13:5639–5646, 1994.
 62. Schumacher, M.A., Choi, K.Y., Zalkin, H., Brennan, R.G. Crystal structure of LacI member, PurR, bound to DNA: Minor groove binding by α helices. *Science* 266:763–770, 1994.
 63. Sandmann, C., Cordes, F., Saenger, W. Structure model of a complex between the Factor for Inversion Stimulation (FIS) and DNA: Modeling protein-DNA complexes with dyad symmetry and known protein structures. *Proteins* 25:486–500, 1996.
 64. Johnson, M.S., Srinivasan, N., Sowdhamini, R., Blundell, T.L. Knowledge-based protein modeling. *Crit. Rev. Biochem. Mol. Biol.* 29:168, 1994.
 65. Hu, X., Xu, D., Hamer, K., Schulten, K., Köpke, J., Michel, H. Predicting the structure of the light-harvesting complex II of *Rhodospirillum molischianum*. *Protein Sci.* 4:1670–1682, 1995.
 66. Werner, M.H., Bianchi, M.E., Gronenborn, A.M., Clore, G.M. NMR spectroscopic analysis of the DNA conformation induced by the human testis determining factor SRY. *Biochemistry* 34:11998–12004, 1995.
 67. Nicholls, A., Sharp, K.A., Honig, B. Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11:281–296, 1991.