PROTEINS: Structure, Function, and Genetics 31:417–433 (1998)

Interaction of Human SRY Protein With DNA: A Molecular Dynamics Study

Yun Tang and Lennart Nilsson*
Center for Structural Biochemistry, Department of Biosciences at Novum, Karolinska Institute, Huddinge, Sweden

ABSTRACT **Molecular dynamics simulations have been conducted to study the interaction of human sex-determining region Y (hSRY) protein with DNA. For this purpose, simulations of the hSRY high mobility group (HMG) domain (hSRY-HMG) with and without its DNA target site, a DNA octamer, and the DNA octamer alone have been carried out, employing the NMR solution structure of hSRY-HMG–DNA complex as a starting model. Analyses of the simulation results demonstrated that the interaction between hSRY and DNA was hydrophobic, just a few hydrogen bonds and only one water molecule as hydrogen-bonding bridge were observed at the protein–DNA interface. These two hydrophobic cores in the hSRY-HMG domain were the physical basis of hSRY-HMG–DNA specific interaction. They not only maintained the stability of the complex, but also primarily caused the DNA deformation. The salt bridges formed between the positive-charged residues of hSRY and phosphate groups of DNA made the phosphate electroneutral, which was advantageous for the deformation of DNA and the formation of a stable complex. We predicted the structure of hSRY-HMG domain in the free state and found that both hSRY and DNA changed their conformations to achieve greater complementarity of geometries and properties during the binding process; that is, the protein increased the angle between its long and short arms to accommodate the DNA, and the DNA became bent severely to adapt to the protein, although the conformational change of DNA was more severe than that of the hSRY-HMG domain. The sequence specificity and the role of residue Met9 are also discussed. Proteins 31:417–433, 1998.** © 1998 Wiley-Liss, Inc.

Key words: molecular dynamics; sex-determining region Y (SRY) protein; high mobility group (HMG) box; DNA-binding proteins; DNA bending

INTRODUCTION

Interactions between DNA and proteins are fundamental in biological processes, including DNA packaging, repair, recombination, replication, and transcription. As more and more protein–DNA complex structures are elucidated at high resolution by X-ray crystallography or NMR spectroscopy, both experimental and theoretical studies on protein–DNA interactions have advanced very quickly and have provided many significant insights into the biological roles of DNA-binding proteins and their recognition mechanisms with DNA. From the available protein–DNA complex structures, the DNA-binding proteins can be classified into a number of groups according to the structural motif characterizing their DNA-binding surfaces, such as helix-turn-helix, zinc finger, leucine-zipper.[1,2]

The human SRY (hSRY, SRY stands for sex-determining region Y) protein is a DNA-binding protein encoded by the human SRY gene,[3] located on human Y chromosome. It has been confirmed that hSRY is the testis-determining factor of humans, responsible for the testicular differentiation and, hence, the male sex.[4,5] The function of SRY in sex determination is carried out by activating the gene for Müllerian inhibiting substance (MIS). Its DNA target site is in the promoter of the MIS gene.[6] Mutations in the SRY gene usually result in gonadal dysgenesis of the XY female type (Swyer syndrome). In vitro and in vivo studies of SRY demonstrated that SRY binds specifically to the sequence AACAA-(A/T)(G/C), bends DNA by more than 70°, and is capable of transcriptional transactivation.[7,8] Cloning of the SRY gene indicated that the hSRY protein consists of 223 residues, which comprises three domains: an N-terminal domain, a central about 80 residues DNA-binding domain called the high mobility group (HMG) box, and a C-terminal domain.[3] Sequence analyses of SRY proteins from a variety of species indicate that there is little sequence similarity outside the DNA-binding HMG domain, even among primate species. On the contrary, within the DNA-binding domain, sequence identity is greater than 60% (Fig. 1).[9–11] To date, no function of the human SRY protein has been ascribed to the regions

Grant sponsor: Swedish Natural Science Research Council.
*Correspondence to: Dr. Lennart Nilsson, Center for Structural Biochemistry, Department of Biosciences at Novum, Karolinska Institute, S-141 57 Huddinge, Sweden. E-mail: lennart.nilsson@csb.ki.se
Received 30 October 1997; Accepted 16 December 1997

```
          3          12          22          32          42
hSRY      DRV**K**RPM***NA***F IVWSRDQRRK MALENPRMRN SEISKQLGYQ-
mSRY      GHV**K**RPM***NA***F MVWSRGERHK LAQQNPSMQN TEISKQLGCR-
mSOX-4    GHI**K**RPM***NA***F MVWSQIERRK IMEQSPDMHN AEISKRLGKR-
mLEF-1    MHI**K**KPL***NA***F MLYMKEMRAN VVAESTLKES AAINQILGRR-
rHMG1A    PRG**K**MSSY**A**F FVQTCREEHK KKHPDASVNF SEFSKKCSER-
rHMG1B    NAP**K**RPPS**A**F FLFCSEYRPK IKGEHPGLSI GDVAKKLGEM-
dHMGD     DKP**K**RPLS**A**Y MLWLNSARES IKRENPGIKV TEVAKRGGEL-
          43          52          62          72          82
hSRY      **W**KMLTEAEKW PFFQE**A**QKLQ AM***H***REK***YP***NY KYRPRRKAKM
mSRY      **W**KSLTEAEKR PFFQE**A**QRLK IL***H***REK***YP***NY KYQPHRRAKV
mSOX-4    **W**KLLKDSDKI PFIQE**A**ERLR LK***H***MAD***YP***DY KYRPRKKVKS
mLEF-1    **W**HALSREEQA KYYEL**A**RKER QL***H***MQL***YP***GW SARDNYGKKK
rHMG1A    **W**KTMSAKEKG KFEDM**A**KADK ARYEREMKTY IPPKGETKKK
rHMG1B    **W**NNTAADDKQ PYEKK**A**AKLK EKYEKDIAAY RAKGKPDAAK
dHMGD     **W**RAM--KDKS EWEAK**A**AKAK DDYDRAVKEF EANGGSSAAN
```

Fig. 1. Sequence alignments of representative members of the two subclasses of high mobility group (HMG) box proteins (h = human, m = mouse, r = rat, d = *Drosophila melanogaster*). SRY, SOX-4, and LEF-1 belong to a sequence-specific subfamily, whereas HMG1A, HMG1B, and HMGD belong to the other one. Several residues are conserved in almost all the HMG box family (in bold), e.g., Lys6, Ala11, Trp43, and Ala58 in hSRY (for residue number, see Fig. 2B). Some residues are conserved in most of the HMG box family; however, consensus sequences among sequence-specific subclass (in bold italic) are more than that among the nonsequence-specific one, especially in C-terminus.

outside the HMG box. More than 20 natural mutations in hSRY that lead to sex-reversal occur in the HMG box, with only one exception.[12,13] All of these strongly suggest that the function associated with the SRY protein should reside primarily in the DNA-binding HMG domain itself.

The HMG box is a new kind of eukaryotic DNA-binding protein domain[14,15] involved in functions as diverse as DNA repair, recombination, transcriptional activation, and a general role in DNA packaging. Among all of these functions, the bending of DNA or the recognition of bent DNA seems to be the central theme. These proteins can be further divided into two functional subclasses by virtue of both their subtleties of sequence and their mode of DNA binding. The first subclass comprises transcription factors that contain a single HMG box, bind DNA sequence-specifically, and are expressed in few cell types, including the SRY protein. The second subclass proteins are abundant in chromatin, which generally contains two or more tandem HMG boxes, bind DNA with little or no specificity, and are found in all cell types, typified by HMG1 and HMG2. Some residues are conserved in both subtypes, but differences are also evident (see Fig. 1). So far, five HMG box structures, four from nonsequence-specific proteins HMG1[16–18] and HMG-D,[19] and one from the sequence-specific protein Sox-4,[20] have been determined by NMR spectroscopy (for sequences, see Fig. 1). All five structures exhibit the same general fold with minor differences, in which three α-helical segments form an L-shaped structure stabilized by a hydrophobic core. The long arm of the "L" consists of the extended N-terminal section and helix 3 along with the C-terminal region, while the short one consists of helices 1 and 2 (see also Fig. 2A). These results suggest that the global fold of this DNA-binding motif HMG box should be likely to be conserved in all the HMG box family, including the hSRY protein.

Recently, the solution structure of a complex between hSRY-HMG and its DNA target site, a DNA octamer d(GCACAAAC), has been determined by multidimensional hetero-nuclear NMR spectroscopy (Fig. 2).[21] This is the first complex of an HMG box protein with DNA and it could provide insights into the sex-reversal effects of mutations in SRY and the interactions of DNA with hSRY, as well as with other HMG box proteins. From the structure of the complex, it can be seen that the hSRY-HMG domain has the same twisted L shape as the free HMG box structures resolved previously, except that the hSRY-HMG in the complex appears to have a more obtuse angle between the long and short arms. It is, however, not possible to decide if this difference arises as binding to the DNA. So it is necessary to know the free hSRY-HMG structure, which has not been determined to date. One of the most remarkable features of the complex, as evidenced by in vitro and in vivo studies before, is the structure of the DNA, which is distinguished from other protein–DNA complexes. The specific binding site occurs exclusively in the minor groove and induces a large conformational change in the DNA, i.e., Ile13 in hSRY partially intercalates into the basepairs and widens the minor groove,[21–23] then causes DNA bending ~70–80°, in order to get a perfect "induced fit." Following this complex, another complex of mouse lymphoid enhancer-binding factor (LEF-1) HMG box with DNA was also been determined by NMR.[24] The protein–DNA interactions observed in this complex are generally similar to those of SRY, but differ in several details, mainly at the C-terminus.

In the complex, the hSRY-HMG domain contains 76 residues (the residues are numbered from 1 to 76 successively, according to Ref. 21; the sequences of the hSRY-HMG domain and the DNA octamer are illustrated in Fig. 2B).

The hSRY-HMG-DNA complex provides us with some principal features of HMG-DNA interaction. It is, however, still a static view. To understand how hSRY interacts with DNA and then realizes its biological function in detail, theoretical simulation methods can be applied to the complex from dynamic and thermodynamic views. A number of mutants,

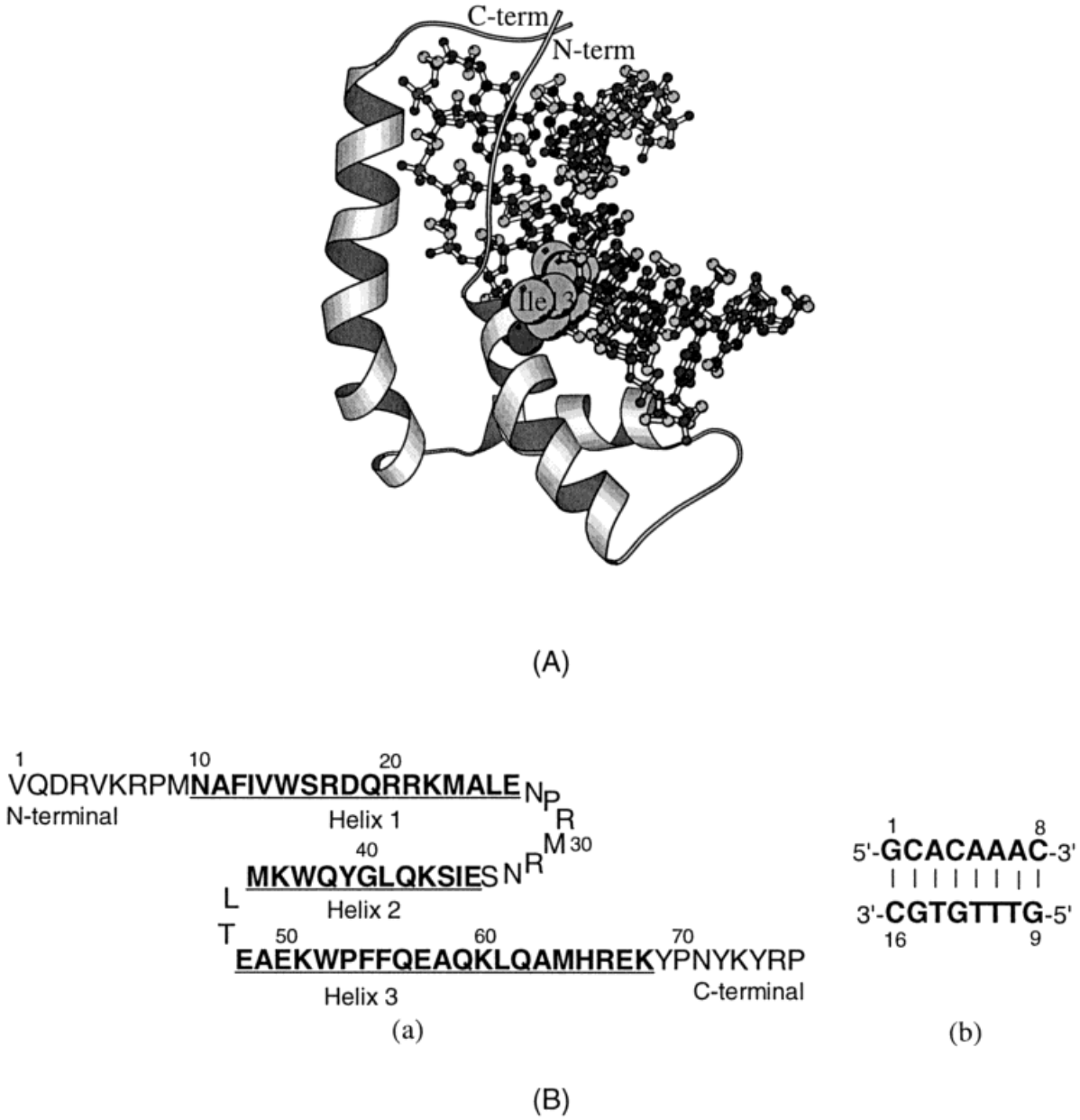

Fig. 2. (**A**) Schematic representation of the NMR solution structure of human SRY-HMG-DNA complex.[21] The DNA is represented as a ball-and-stick model and the protein is represented as a ribbon only, except for residue Ile13, represented as spacefilling model. This figure together with Figures 6, 7, 9, and 11 was generated with the program MOLSCRIPT.[56] (**B**) Schematic representation of the sequence of (a) human SRY-HMG domain and (b) DNA octamer used in the study.

influencing activity and DNA binding properties of the hSRY protein, have also been found.[25–28]

Molecular dynamics (MD) simulation is a powerful tool to analyze the structural and dynamic features of biomacromolecules.[29–31] It provides much information about atomic interactions and their time evolution at a level of detail which can both enhance and complement the experimental results. Water molecules, which have been shown to play an important role in both the specificity and affinity of protein–DNA interactions,[32] are easily incorporated in MD simulations, where they can be characterized in terms of localization and mobility, properties that may be less straightforward to assess by other means.[33] Besides proteins or nucleic acids alone, several MD simulations to date have been performed on protein–DNA complexes (for a recent review, see Ref. 34).

In the present work, we used the NMR solution structure of hSRY-HMG–DNA complex as a starting

TABLE I. Summary of the Three Molecular Dynamics Simulations

Simulation	hSRY-HMG-DNA	hSRY-HMG	DNA
No. atoms of solute	1797	1291	506
No. counterions	5 Na^+	9 Cl^-	14 Na^+
No. water molecules	6052	6934	2431
Total atoms	19958	22102	7813
Radius of watersphere	37 Å	38 Å	28 Å
Timestep	2 fs	2 fs	2 fs
Total running time	500 ps	500 ps	500 ps
Computer	DEC AXP 4100/300E	DEC AXP 4100/300E	DEC AXP 200 4/233
CPU time/10 ps	~6.50 h (4CPU)	~7.50 h (4CPU)	~10.00 h (1CPU)

model for MD simulations of the hSRY-HMG domain with and without DNA and the DNA octamer alone. All three simulations were carried out in water with electroneutralizing counterions. The results of these simulations served as a prediction for the three-dimensional structure of the free hSRY-HMG domain and to enhance our understanding of the interactions stabilizing the hSRY-HMG–DNA complex.

METHODS

All three solvated MD simulations (Table I) were run on a DEC AXP 4100/300E 4CPU parallel computer and a DEC AXP 200 4/233 workstation using the program CHARMM[35] version c25a2 and the all-atom version 22 force field (Chemistry Department, Harvard University, Cambridge, MA).[36] The TIP3P water model was used to simulate the water molecules.[37]

Simulation Details

The starting coordinates of the hSRY-HMG–DNA complex were taken from the refined, energy-minimized average NMR solution structure,[21] PDB entry code 1HRY,[38] with hydrogen atoms added by the CHARMM subroutine HBUILD.[39] Then a short adopted basis Newton-Raphson minimization (ABNR, 500 steps) was applied to the complex in vacuo to remove the unfavorable contacts, keeping harmonic constraints with a force constant of 2.0 kcal/(mol·Å^2) on heavy atoms of the complex. The starting coordinates of the hSRY-HMG domain and the DNA octamer were extracted from the complex solution structure separately, with a similar pretreatment.

After energy minimization in vacuo, the solute was inserted into the center of a preequilibrated watersphere, water molecules closer than 2.8 Å to any protein or DNA atoms were deleted, and counterions were added at random positions into the system to make the system electroneutral (Table I). Following this, another 500 steps of ABNR minimization was performed to obtain the initial structures of the subsequent molecular dynamics simulations. During these energy minimization processes, harmonic constraints with a force constant of 10.0 kcal/(mol·Å^2) were applied to the solute heavy atoms in order to allow local adjustment of the water molecules and counterions and to eliminate any residual geometrical strain.

After the initial structures were prepared, the MD simulations began with a 200 ps heating and equilibration phase, releasing the harmonic constraints implemented on the solute heavy atoms. The initial atomic velocities were assigned from a Gaussian distribution corresponding to a temperature of 300K. The nonbonded energies and forces were smoothly shifted to zero at 12.0 Å and a constant dielectric ($\epsilon = 1$) was used for electrostatic interactions. The nonbonded list including neighboring atoms within a 14.0 Å distance was updated using a heuristic testing algorithm. All hydrogens were treated explicitly employing a timestep of 2 fs for integrating the equations of motion. All bonds involving hydrogens were constrained with the SHAKE algorithm.[40] The production run of the simulations was for 300 ps, giving a total simulation time of 500 ps. Ensemble averages were determined from the production stage with coordinates and energies saved every 100 timesteps for further analysis.

The water molecules interacted with a "deformable boundary force," arising from mean field interactions of water molecules beyond the boundary.[41] Water molecules in the 2 Å shell at the edge of the system (buffer region) were treated by Langevin dynamics with a stochastic heat bath at 300K, via randomly fluctuating forces and dissipative forces using a friction coefficient of 50.0 ps^{-1} on the oxygen atoms of water molecules. For the water molecules inside the buffer region, the ordinary MD equations of motion were applied.

Because the SRY protein is flexible, the three α-helices of the protein usually become relaxed gradually during the MD process. To keep the helices normal, distance restraints were placed on the hydrogen-bonding atom pairs of the protein α-helices. There were 35 hydrogen bonds on the three helices calculated from the program HBPLUS,[42] so a total of 70 distance restraints were applied (two atom pairs per hydrogen bond were used with ranges: $r_{NH\text{-}O}$ = 1.5–2.8 Å; $r_{N\text{-}O}$ = 2.4–3.5 Å). The constraining potential between two selected atoms was a biharmonic

functional form as following:

$$E_r = \begin{cases} 0.5\, k_{min}(r_{min} - r)^2; & r < r_{min} \\ 0; & r_{min} \le r < r_{max} \\ 0.5\, k_{max}(r - r_{max})^2; & r_{max} \le r < r_{lim} \\ f_{max}\left(r - \dfrac{r_{lim} + r_{max}}{2}\right); & r_{lim} \le r; \quad r_{lim} = r_{max} + \dfrac{f_{max}}{k_{max}} \end{cases}$$

where k_{min} and k_{max} are the minimum and maximum force constants set to 5.0 kcal/(mol·Å^2), r is the calculated separation distance, r_{min} and r_{max} are the minimum and maximum distance bounds taken to be ±0.10Å around the NMR geometry, f_{max} is the maximum force set to 10.0 kcal/(mol·Å).

The normal B-DNA and normal A-DNA duplex, used for the DNA conformational comparison, were built up from X-ray diffraction data[43] by the software package QUANTA[44] with the same sequence as in the hSRY-HMG–DNA complex d(GCACAAAC). The hydrogen atoms were added by the CHARMM subroutine HBUILD,[39] followed by 500 steps of ABNR energy minimization using CHARMM, keeping harmonic constraints with a force constant of 2.0 kcal/(mol·Å^2) on the heavy atoms.

Analysis of the Simulations

After the simulation finished, an average structure was evaluated from the final 300 ps run. Deleting the water molecules and counterions, a short 500 steps ABNR energy minimization was carried out to remove the unfavorable contacts, and also with 10.0 kcal/(mol·Å^2) harmonic force constant on the heavy atoms. The minimized average structures were used in the analyses.

The criteria used for a hydrogen bond *(A···H – D)* was that the distance between the acceptor (A) atom and H atom should be less than 2.5 Å and the angle *A – H – D* should be larger than 120°.

The solvent-accessible surface areas were calculated using the definition of Lee and Richards[45] with a probe of radius 1.6 Å. The Lennard-Jones radii values were used for the complex atoms.

RESULTS

The simulations of all three systems were quite stable, as indicated by the total potential energy of the system, the temperature, and the root mean square deviations (RMSD) of the atoms from the initial structure, which remained constant during the production run (Fig. 3).

The RMSD of the averaged dynamics structures for the final 300 ps simulations compared with the NMR solution structure (shown in Fig. 4 and Table II) indicate that the sidechains of the protein are more flexible than the backbone, whereas in the DNA the backbone moves more than the bases, both in the complex and in the free state, and some conformational differences lie between the complex and the free states. The RMSD of the DNA in complex and in the free state compared with the canonical A-form and canonical B-form DNA are also listed in Table IIB, which show that the conformation of the DNA is much closer to that of the canonical A-DNA than B-DNA, especially the MD average of DNA in the free state.

The root mean square fluctuations (RMSF) of the three structures for the final 300 ps simulations (Fig. 5) show that the internal mobility of the protein and DNA is limited and uniform, both in the complex and in the free state, and that the backbone of the protein in the complex is less mobile than that in the free state, whereas the backbone of the DNA in the complex is a little more mobile than that in the free state.

The solvent-accessible surface area, which is 6,705 Å^2 for the free hSRY-HMG protein, is reduced by 1,264 Å^2 (about 19%) after binding to DNA.

Dynamics of Structural Features of the hSRY-HMG–DNA Complex

At first, 15 snapshots of the hSRY-HMG–DNA complex taken with an interval of 20 ps from the production run, together with the NMR structure, were superimposed in Figure 6. From Figure 6 we can see that, compared with the NMR structure, throughout the dynamics trajectory the hSRY-HMG–DNA complex had a similar backbone, the first strand of DNA (GCACAAAC) trended along with the N-terminus and the first helix of hSRY-HMG, and the second strand of DNA along with the second and third helices of the protein, the C-terminal region crossed above the minor groove of DNA. However, some local regions of the HMG box were more mobile than others during the simulation, which is difficult to observe by experiment. The most obviously different regions were the two terminal regions, especially in the parts around Pro8 and Met9, and around Lys73. The sidechains of Met9 and Lys73 swayed considerably (Fig. 4A). In addition, the loops between helix1 and helix 2, helix 2 and helix 3, also showed some changes. Meanwhile, the three helices were less mobile and their relative orientations and positions were stable. Contrasted with the protein, the motion of the DNA octamer was more gentle (Fig. 4C); only the two ends moved slightly and some bases twisted, e.g., bases A3 and T14.

The intramolecular fluctuations of the complex during the final 300 ps simulation (Fig. 5) were very uniform for the overall structure, with an RMSF value around 0.75 Å for most of the residues, except for the two ends of each fragment with increased flexibility. In Figure 5, three particular phenomena were observed. The first is the short loop region, residues 29–33 of hSRY-HMG domain, with higher

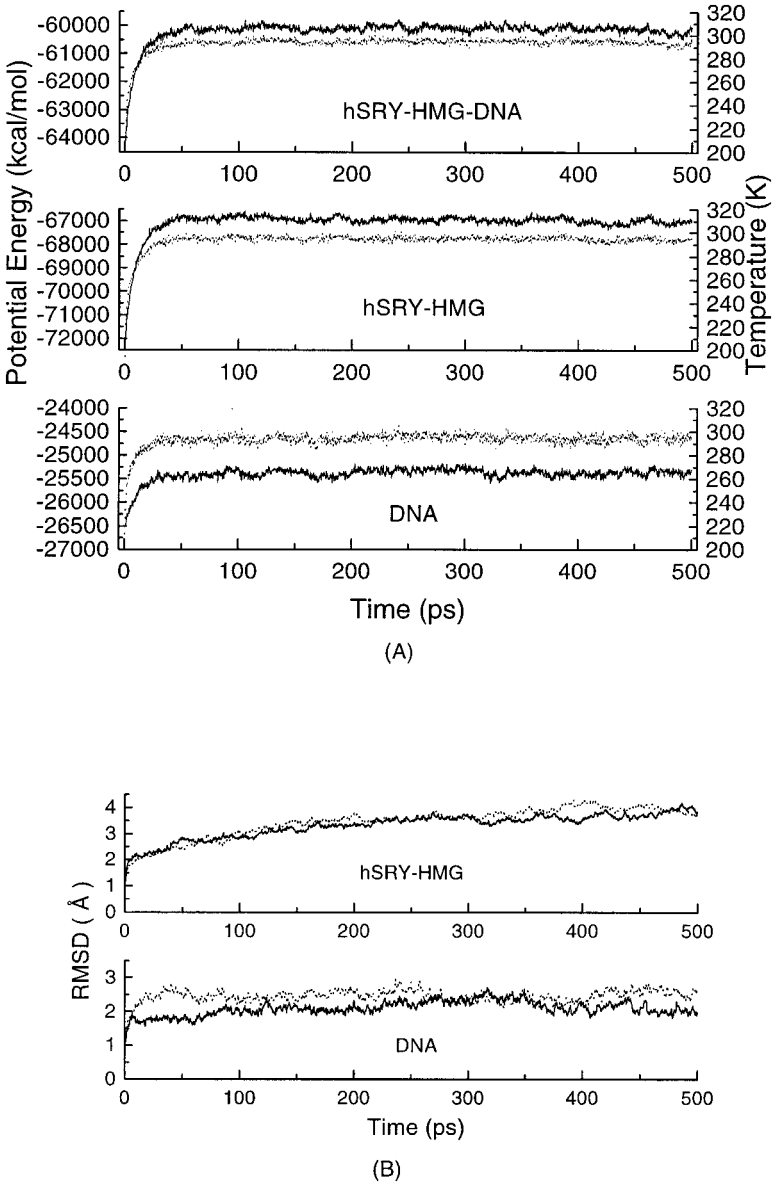

Fig. 3. (**A**) The total potential energy (kcal/mol) of the systems (solid line) and the temperature (K, dotted line) as functions of time (ps). (**B**) Time evolution of the RMSD from the initial structures of the simulations. Solid lines represent the complex, whereas dotted lines mean in free state.

RMSF values up to 1.5 Å. This region has no contacts with the DNA segment and each dynamics trajectory has a different conformation (Fig. 6), showing more flexibility whether complexed or free. The second one is the fluctuations of the loop region around Pro8 and Met9. This region has a high RMSD value compared with the NMR structure (Fig. 4A), but here only with the same low RMSF value as most of the other residues, which demonstrates that a conformation change took place in this part (Fig. 6). Finally, the fluctuations of the hSRY-HMG were lower and more uniform in the complex than in the free state, whereas those of the DNA were slightly higher in the complex than in the free state, which is beneficial to the forming of the stable and complementary complex.

The simulation confirmed the structural complementarity of DNA with hSRY-HMG and the stability of the complex observed in the NMR structure. The interaction energy between hSRY-HMG domain and the DNA was -519 kcal/mol for the NMR structure, and -567 kcal/mol for the average dynamics structure of complex. The force maintaining the stability of the complex primarily came from the hydrophobic cores of the hSRY-HMG domain, instead of hydrogen bonds and water bridges observed in other protein–DNA complexes.[32] The big hydrophobic core consists of aromatic residues Phe12, Trp15, Trp43, Phe54, and Phe55, and aliphatic residues Met9, Ala11, Ile13, Val14, Met45, Leu46, and Ala58 (Fig. 7A). All of these hydrophobic residues are located at the

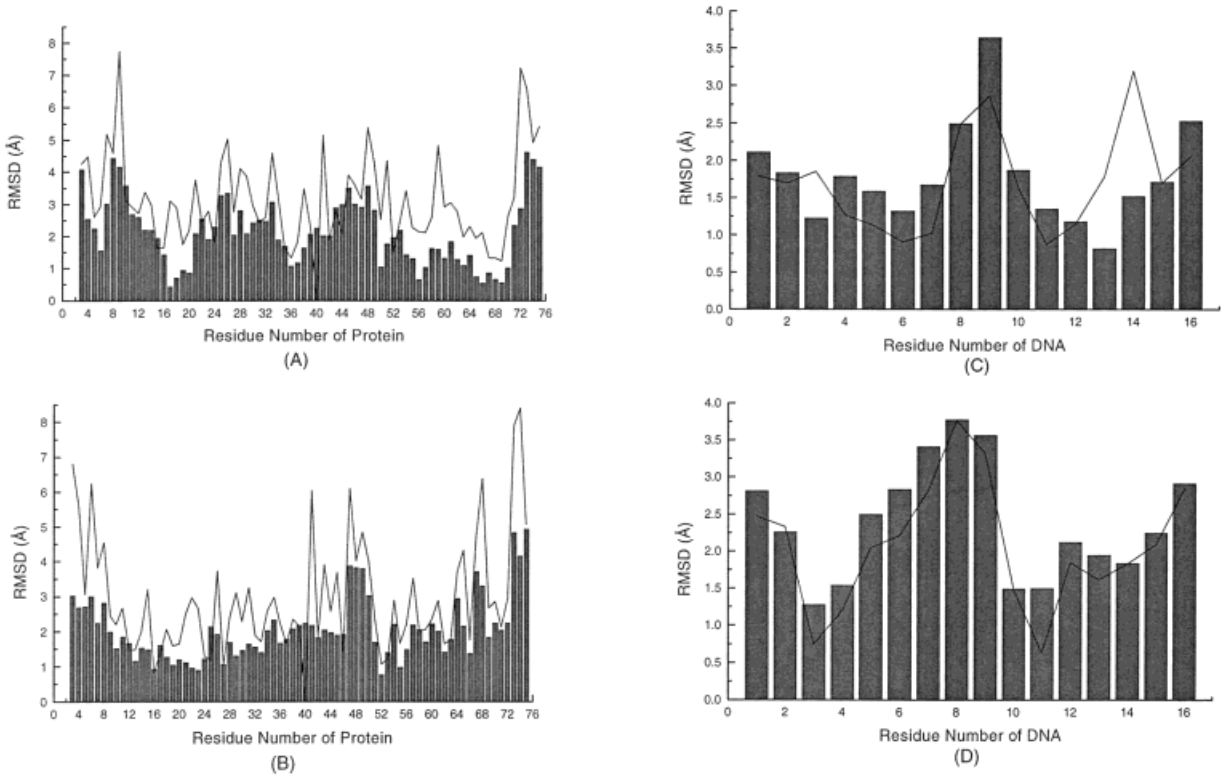

Fig. 4. The residue-averaged RMSD of (**A**) hSRY-HMG domain in complex; (**B**) hSRY-HMG domain in the free state; (**C**) DNA segment in complex; and (**D**) DNA segment in the free state throughout the dynamics production run (200–500 ps), compared with the NMR structure of the complex. The shaded bars are for the backbone, the lines for the sidechains or bases.

TABLE II. Atomic RMSD (Å) of the Averaged Dynamics Structures

(A) hSRY-HMG Domain						
	Protein in complex			Protein in free state		
	Backbone	Sidechains	All	Backbone	Sidechains	All
Compared with NMR structure	2.44	3.71	3.47	2.27	3.71	3.44
(B) DNA						
	DNA in complex			DNA in free state		
	Backbone	Base	All	Backbone	Base	All
Compared with NMR structure	2.15	1.98	2.10	2.71	2.34	2.56
Compared with canonical A-DNA	1.88	1.79	1.86	1.15	0.90	1.06
Compared with canonical B-DNA	5.29	3.06	4.47	5.11	3.13	4.37

junction of the three helices and pack them together. This hydrophobic core, which remained stable during the entire simulation, has an exposed surface on the concave surface of the HMG box. Residue Ile13 lies just at the apex of the concave surface, like a hydrophobic wedge. Meanwhile, another small hydrophobic region at the C-terminal and N-terminal regions, composed of residues Val5, Tyr69, Tyr72, and Tyr74 (Fig. 7A), links the two termini together and acts on DNA through residue Tyr74 interacting with the base of A3 and pushing the base toward the major groove. In contrast to the stable conformation of hSRY-HMG, the conformation of the DNA in the complex primarily depended on the structural shape of hSRY-HMG domain.

Contacts at the Protein–DNA Interface

The DNA was docked primarily on the hydrophobic surface of the protein. As observed in the NMR structure, there were several residues (Phe12, Ile13, Ser36, and Tyr74) in contact with the DNA bases (Fig. 7), where Ile13 partially intercalated between A5 and A6, and the aromatic sidechains of Phe12 and Tyr74 packed against the bases A5, T12 and G13, A3 and T14, respectively. However, some features different from the NMR structure and some new features

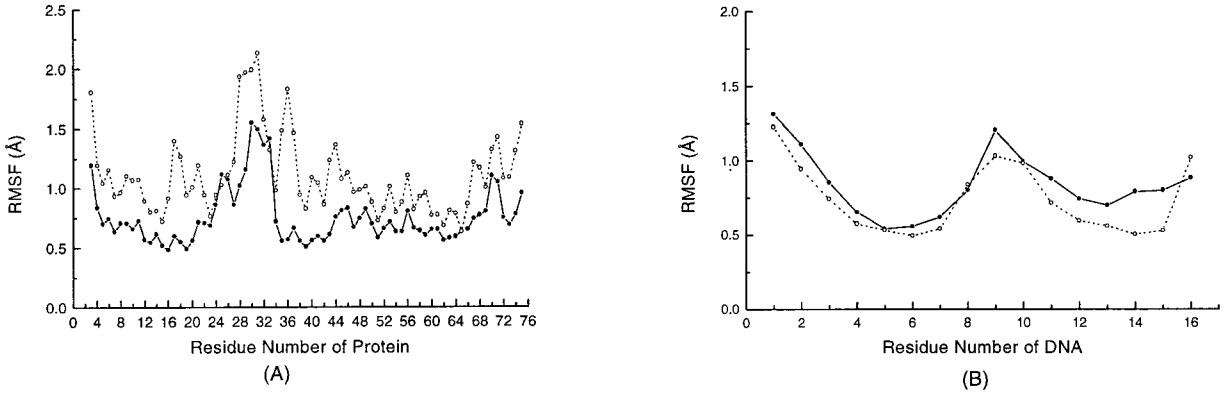

Fig. 5. The residue-averaged RMSF of the simulated structures: (**A**) the hSRY-HMG domain and (**B**) the DNA segment during the final 300 ps simulations. The solid line and circle are the structure in complex, the dotted line and open circle are the structure in the free state.

were also obtained. The most notable feature is the orientation of residue Met9. In the NMR structure, Met9 contacts with the backbone of DNA, but in our simulation a conformational change took place in Met9 and Met9 had no contact with DNA. Its sidechain instead joined in the hydrophobic core. Phe12 lay at the horizontal level with Ile13 and helped Ile13 to deform the DNA. Tyr74 directly contacted base A3 in a face-to-edge fashion. In the NMR structure, Ser33 and Ile35 are also in contact with the DNA bases, i.e., the hydroxyl group of Ser33 forms a hydrogen bond with the N2 atom of G9, and Ile35 is packed against basepairs 7 and 8. But in our simulation, Ser33 and Ile35 contacted with the DNA sugar–phosphate backbone rather than bases.

Further, we analyzed the hydrogen bonds at the protein–DNA interface. All possible hydrogen bonds were calculated from the whole production stage (300 ps). The direct and water-mediated hydrogen bonds present at the hSRY-HMG–DNA interface are listed in Table III, five direct hydrogen bonds and one water bridge. Among the five direct hydrogen bonds, only one was formed between the protein and a DNA base, that is, the hydroxyl group of Ser36 formed a hydrogen bond with the O2 atom of T10 (Fig. 7C), which is in accordance with the NMR structure, while the others were present between the protein and DNA phosphate backbone. However, the average presence times of these direct hydrogen bonds were not long. At the same time, only one water molecule was present as a hydrogen-bonding bridge between the HD21 atom of Asn10 and the O2 atom of C4, as well as the O2 atom of T14 (Fig. 7B), in which all three atoms of the water molecule were used. The average distances between the HD21 atom of Asn10 and the O2 atom of C4 and the O2 atom of T14 were 4.61 Å and 4.04 Å, respectively. No other water-mediated hydrogen bonds were detected between the protein and DNA atoms. Although there was one water bridge between the hSRY-HMG domain and the DNA bases, a large number of different water molecules were detected to participate in the formation of the water bridge as time evolved, and each water molecule had a very short average presence time, which indicated that the water-mediated hydrogen bond was very weak and the water molecule was exchanged by other waters very quickly. In the NMR structure, five direct hydrogen bonds were also found at the protein–DNA interface. However, only two hydrogen bonds were still present in the dynamics structure (Table III) and the other three had been changed during the dynamics simulation. Compared with the NMR structure, in which the carboxyamide group of Asn10 is involved in electrostatic interactions with the N2 atom of G13, the O2 atom of C4, and the N3 atom of A5, our simulation gave a different interaction. The calculation also showed that the hydroxyl group of Tyr74 did not form a fixed hydrogen bond with DNA, whereas in the NMR structure the hydroxyl group of Tyr74 may form a hydrogen bond with the O2 atom of T14.

Around the hydrophobic cores there are many positively charged residues. These residues formed electrostatic interactions with the phosphate groups of DNA as salt bridges, i.e., Arg4, Arg7, Arg17, and Arg20 form salt bridges to the phosphates of C4, A5, A7, and C8, respectively, while Lys37, Lys44, Lys51, Arg66, Lys73, and Arg75 form salt bridges to the phosphates of T11, T12, T14, G15, C16, and A3,

Fig. 6. Stereoview showing a best-fit superposition of the 15 dynamics snapshots and the NMR structure of the hSRY-HMG–DNA complex. The trace of Cα atoms (residues 3–75) of hSRY-HMG are shown in red, all nonhydrogen atoms of the DNA in blue, the NMR solution structure of the complex in green.

Fig. 7. The interactions of hSRY-HMG domain with DNA. (**A**) A whole stereoview; the hydrophobic cores are colored blue while the residues contacted with DNA bases are green. (**B**, **C**) Detailed views; residues of hSRY-HMG domain are colored blue, hydrogen bonds purple, and the red sphere in (B) represents a water bridge.

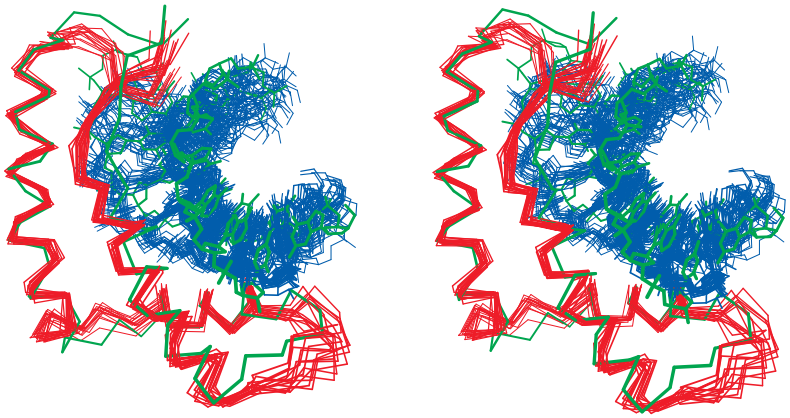

Figure 6.

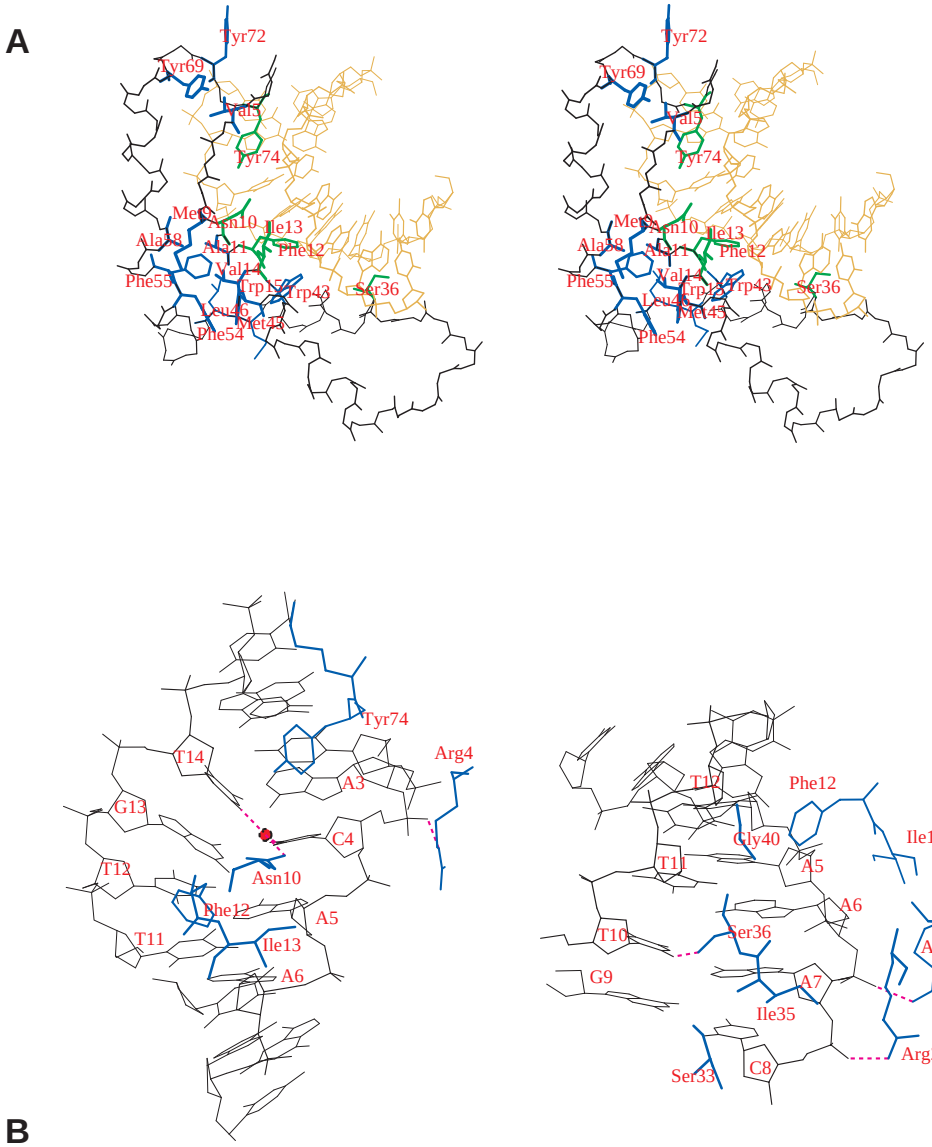

Figure 7.

TABLE III. List of Hydrogen Bonds at the Protein–DNA Interface

(A) Direct hydrogen bonds present during more than 10ps of the simulation

Hydrogen bond	Average time (ps)
Arg4-HE . . . C4-O2P	19.7[a]
Arg17-HH12 . . . A7-O2P	149.9
Arg20-HH22 . . . C8-O2P	13.3
Ser36-HG1 . . . T10-O2	13.4[a]
Lys73-HZ1 . . . C16-O2P	37.3

(B) Water-mediated hydrogen bonds

Bridge	Average time (ps)
Asn10-HD21 . . . w . . . C4-O2	0.295
Asn10-HD21 . . . w . . . T14-O2	0.263

[a]The hydrogen bond is present in the NMR structure.

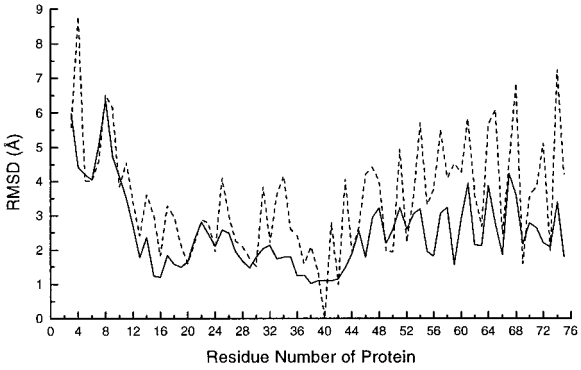

Fig. 8. The residue-averaged RMSD of hSRY-HMG domain in the free state compared with the same domain in complex for the final 300 ps simulations. The solid line is the backbone and the dashed line the sidechains.

respectively. In the NMR structure, Lys44, Lys51, and Arg75 are not involved in electrostatic interactions. Most of the phosphate groups of DNA interacted with positively charged residues, so that the counterions were far from the DNA surface during the simulation. In contrast with these, all residues with negative charge are scattered on the convex surface without contact with DNA. Furthermore, there also existed some hydrophobic interactions between the protein residues and the DNA sugar–phosphate backbone. The hydrophobic sidechains of Ile35, Trp43, Leu46, Phe55, and the hydrophobic parts of several Arg sidechains (Arg4, Arg7, Arg17, Arg20, Arg66, and Arg75) formed hydrophobic interactions with the sugar–phosphate backbone of DNA, while in the NMR structure, Met9, Phe12, and Ala24 are involved in this type of interaction, but not Ile35, Leu46, Phe55, or Arg75.

Conformation of hSRY-HMG Domain in the Free State

The RMSD of the averaged dynamics structure of the hSRY-HMG domain in the free state compared with the averaged dynamics structure of the same domain in the complex is 2.78 Å for the backbone atoms, 4.10 Å for the sidechain atoms, and 3.85 Å for the whole domain, which illustrated that some conformational differences existed between hSRY-HMG domain in complex with DNA and in the free state.

At first, a conformational comparison between the two average dynamics structures of the complexed and free hSRY-HMG was carried out. Figure 8 shows the RMSD between these hSRY-HMG structures averaged over residues, and Figure 9A shows the superimposed conformations of the hSRY-HMG domain in different states.

In the free hSRY-HMG domain, the RMSDs of most of residues are uniform (Fig. 4B), with some exceptions, especially residues 47–50 between helix 2 and helix 3. The conformation change observed in the sidechain of Met9 in the complex (Fig. 4A) did not occur in the free state and hence the loop around Pro8 and Met9 was stable. From Figures 8 and 9A it is clear that the major difference between the two conformations is in the long arm of the L-shaped conformation, i.e., the extended N-terminus and helix 3. In the complex, the hSRY-HMG box had a more obtuse angle between its long and short arms. However, the conformations of the short arms, helix 1 and helix 2, were very similar.

We also compared the conformation of the average dynamics structure of the hSRY-HMG domain in the free state with that of the other determined free HMG boxes, that is, HMG1A,[18] HMG1B,[16,17] and HMGD.[19] Figure 9B shows a superimposed comparison of these structures. The structures of HMG1A and HMGD used for comparison are the individual NMR structures closest to their corresponding minimized average structures, and the two HMG1B structures are the minimized average structures from the PDB (entry codes 1HME and 1NHM). Clearly, the structure of hSRY-HMG was in very good agreement with the others, except for the N- and C-termini, which might be related with the sequence-specific binding with DNA. The angle between the long and short arms is similar in all structures. The pairwise RMSD of the Cα atoms (for the fitted residues of each structure) was small (only about 2.0 Å, Table IV), whereas it was 2.5 Å for the same atoms of the complex.

Comparison of the conformation of the hSRY-HMG domain with the nonsequence-specific HMG domains indicated that the primary structural difference between the two HMG box subfamilies is at the N- and C-termini (Figs. 1, 9B). In the sequence-specific subfamily, there is a proline residue at site 70, which induces a turn causing the C-terminal region to bend toward the N-terminus, such that residues Tyr69, Tyr72, Tyr74, and Val5 pack together and form the second hydrophobic core. However, the C-terminus of the sequence-nonspecific HMG box is extended along with helix 3 because of

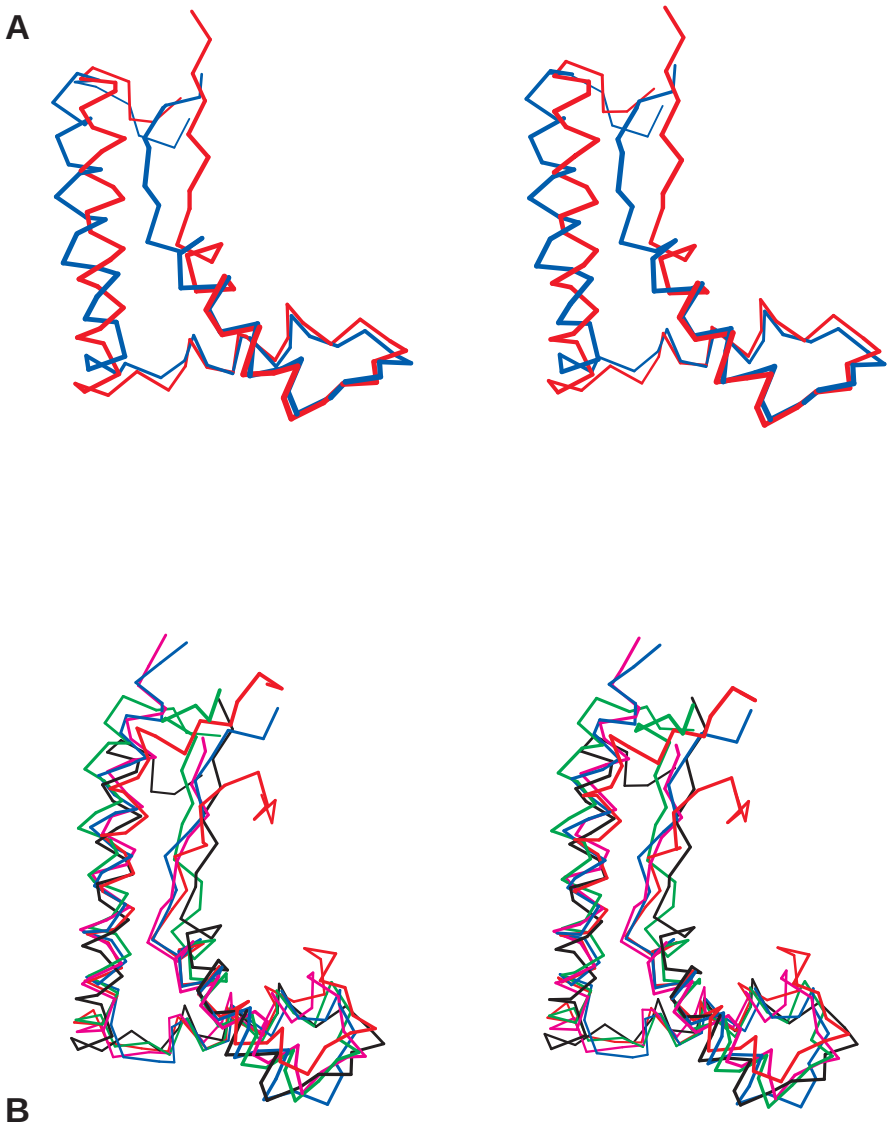

Fig. 9. (**A**) Stereoview of the superimposed Cα backbones of hSRY-HMG domain in the free state (red) and in complex with DNA (blue). (**B**) Stereoview of superimposed Cα backbones of hSRY-HMG domain in the free state (black) with other determined free HMG box, red for HMG1A,[18] both green and purple for HMG1B,[16,17] and blue for HMGD.[19]

the absence of Pro70, without forming the second hydrophobic core, even in contact with DNA. The second hydrophobic core thus is important for the sequence-specific binding to DNA.

Conformation of the DNA Octamer

The RMSD between the averaged dynamics structure of the bound and free DNA was 1.74 Å for backbone atoms, 1.69 Å for bases, and 1.74 Å for all DNA atoms, and in Table IIB we see that compared with the NMR structure, the RMSDs of DNA in the free state were a little bigger than in the complex, suggesting that a conformational change has occurred in the free state. Figure 10 shows the RMSD of each residue between the two average structures. The two conformations are very similar in bending and unwinding shape (Fig. 11) except for the recovery of base stacking and base pairing in the free conformation.

However, the conformational comparisons between canonical A-DNA, canonical B-DNA, and DNA octamer have given a dramatic change (Table IIB, Figs. 10, 11). The RMSDs of the DNA octamer either complexed or free vs. canonical B-DNA are very similar to the RMSD between canonical A-DNA and B-DNA 5.33 Å for the backbone atoms, 3.34 Å for bases, and 4.58 Å for all DNA atoms, while the values vs. canonical A-DNA are very small (Table IIB).

Figure 10 also illustrates that the DNA in the free state was similar to the canonical A-DNA, with very

TABLE IV. Atomic RMSD for the Fitted Residues of Each Structure

	Fitted residues	Cα atoms (Å)
hSRY-HMG vs. HMG1A	12–24, 36–45, 55–65	2.016
hSRY-HMG vs. HMG1B[a]	12–24, 36–45, 55–65	2.046
hSRY-HMG vs. HMG1B[b]	12–24, 36–45, 55–65	2.169
hSRY-HMG vs. HMGD	12–24, 36–45, 55–65	1.914

[a]Determined by Weir et al.[16]
[b]Determined by Read et al.[17]

low RMSD for each residue, both backbone and bases agreed very well (Fig. 11). On the other hand, there was a significant difference between the free DNA and the canonical B-DNA (Fig. 10), especially in that the two ends were evidently unwound (Fig. 11). Checking the conformation of furanose rings in the free DNA confirmed that it had C3′-*endo* pucker rather than C2′-*endo*, just like that in normal A-DNA. Distance monitoring also confirmed that our DNA structure in the free state was similar to canonical A-DNA, and that the DNA conformation in the complex was in between normal B- and A-DNA. The average distance of two adjacent phosphorus atoms in normal A-DNA is 5.74 Å, and in B-DNA it is 6.71 Å. Table V shows that in this respect some steps in the DNA in the complex were close to that of B-DNA, while others close to A-DNA. These results suggested that the average dynamics structure of the free DNA is in A-DNA form, whereas the structure of DNA in complex was somewhat different from the canonical A-DNA (Fig. 10).[46]

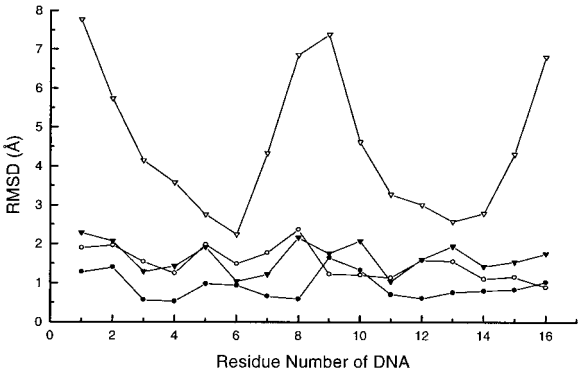

Fig. 10. The residue-averaged RMSD of the averaged dynamics structure of the DNA octamer in the free state compared with the same segment in the averaged dynamics structure of the complex (-○-); the canonical A-DNA with the same sequence (-●-); the canonical B-DNA with the same sequence (-▽-). RMSD of DNA in averaged dynamics complex structure compared with the canonical A-DNA (-▼-).

DISCUSSION

Hydrophobic Recognition Is the Physical Basis of HMG–DNA Interaction

In contrast to many other sequence-specific protein–DNA interactions, the hSRY–DNA interaction is hydrophobic. In our simulations, we can see that there are just a few hydrogen bonds at the protein–DNA interface, and only one site where a transient water molecule acts as a hydrogen-bonding bridge at the interface. The two hydrophobic cores of the hSRY-HMG domain maintained stable contacts with DNA and formed hydrophobic interactions with DNA backbone and bases.

The stereochemical characteristics of the big hydrophobic core are: aromatic residues, such as residues Phe12, Trp15, Trp43, Phe54, and Trp55, are located within the apex of the L-shaped structure and link the three helices with a hydrophobic patch exposed on the concave surface. There are also some aliphatic residues joining in the core. These aromatic sidechains could form a strong and large delocalized π interaction, stabilizing all three helices. An aliphatic residue, such as Ile13, like a hydrophobic wedge, lies at the edge of the hydrophobic region. Surrounding the hydrophobic region there are many positively charged residues, including Arg4, Arg7, Arg17, Arg20, Lys37, Lys44, Lys51, Arg66, Lys73, and Arg75. When the protein binds to DNA, the delocalized π interaction can produce a stable edge-to-face interaction with the π planes of the DNA A5 and A6 bases. Meanwhile, the aliphatic sidechain of Ile13 can partially intercalate between A5 and A6. The surrounding positively charged residues form strong salt bridges with the phosphate groups of DNA, i.e., C4, A5, A7, C8, T11, T12, T14, G15, C16, and A3, respectively. The hydrophobic sidechains of these residues also formed hydrophobic interactions with the sugar–phosphate backbone of DNA. Almost all phosphate groups had a corresponding positively charged residue. Such phosphate neutralization also favors the formation of the complex.[47] All of these hydrophobic and electrostatic interactions contribute to stabilize the bent DNA and the complex. Accompanying the binding, most of the hydrophobic concave surface is buried. There is another small hydrophobic region in the C-terminus, consisting of Tyr69, Tyr72, and Tyr74, linked by Val5, which combines the two termini together. This small core is believed to be important for the sequence-specific binding to DNA. Residue Tyr74 interacts with the base of A3 and pushing the base toward the major groove, destroying both the base stacking and base pairing of A3. Contrasted with this, although Ile13 partially intercalates into bases, it only disrupts the base stacking and keeps the base pairing of A5 with T12, A6 with T11.

At the protein–DNA interface, just five direct hydrogen bonds and one water-mediated hydrogen bond were detected. The other residues form a hydrophobic interface, with an area of 1,264 Å^2 (about 19%) in the solvent-accessible surface area of hSRY-HMG domain. This demonstrated that at the

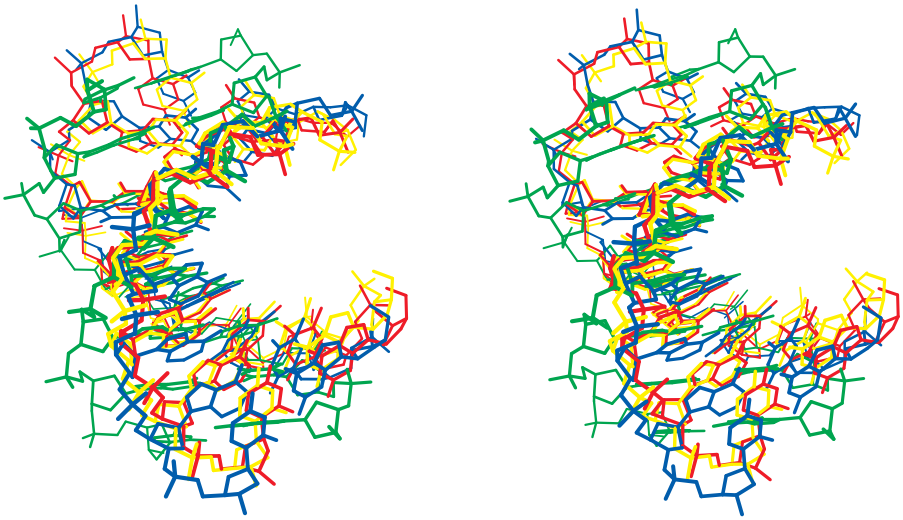

Fig. 11. Stereoview of superimposed comparison of the same DNA segment in different states. The deformed DNA in free is red, in complex, blue, the canonical A-DNA, yellow, and the canonical B-DNA, green.

TABLE V. Distance (Å) of Two Adjacent Phosphorus Atoms

Base step	DNA in free	DNA in complex	A-DNA	B-DNA
C2-A3	6.36	6.77	5.84	6.80
A3-C4	5.77	6.06	5.72	6.64
C4-A5	5.55	6.40	5.70	6.71
A5-A6	4.61	5.70	5.71	6.65
A6-A7	5.91	6.01	5.71	6.68
A7-C8	5.67	5.59	5.77	6.73
T10-T11	5.54	5.83	5.84	6.77
T11-T12	5.77	5.97	5.72	6.68
T12-G13	5.78	6.21	5.69	6.69
G13-T14	6.08	6.36	5.74	6.66
T14-G15	5.55	5.57	5.70	6.71
G15-C16	5.62	6.27	5.80	6.76

interface the main binding force was of a hydrophobic nature.

Originally, the free DNA segment was canonical B-form and highly hydrated;[21,46] however, the water molecules are released and the DNA segment approaches an A-like conformation when the hSRY-HMG-DNA complex is formed, which increases the hydrophobic contacts between protein and DNA.

Therefore, we believe that the hydrophobic cores not only keep the structural stability of the hSRY-HMG domain itself, but also make important contributions to the formation and stability of its complex with DNA. We further suggest that the hydrophobic cores are essential for hSRY binding to DNA, and hydrophobic recognition is the physical basis of hSRY-DNA-specific interaction.

The same case was found in the sequence-specific LEF-1-HMG–DNA complex.[24] Balaeff et al. predicted the nonsequence-specific HMGD–DNA complex by means of docking and MD simulation,[48] in which intensive hydrophobic interactions were also found at the interface. Furthermore, we found that these hydrophobic residues are highly conserved in almost the entire HMG box family, especially in the sequence-specific subtype; at the same time, the positively charged residues are also conserved in most of the HMG box proteins (see Fig. 1). Therefore, these features may remain the same in the entire HMG box family and we can predict that hydrophobic recognition is the physical basis of the interactions of all HMG box proteins with DNA. The second small hydrophobic region is only conserved in the sequence-specific subfamily (see also Fig. 1), which may thus be related with the DNA sequence specificity.

Based on this idea, we could understand the reason why HMG box proteins bind sequence-specifically DNA through the minor groove. In DNA, the minor groove is more hydrophobic than the major groove, especially the adenine base, which has no $-NH_2$ group in the minor groove (Fig. 12).

As to the sequence specificity of the HMG-binding DNA, we think that two key elements are required of the DNA sequence. The first is hydrophobicity, so adenine with more hydrophobicity in the minor groove is preferable; the other is flexibility, i.e., it should take less energy to be bent, which arises from the distinctive stacking properties of the ten unique base steps. Calorimetric evaluations suggest that step of CpA stacking is less stable than others.[49] Thus, the core-specific sequence for HMG box binding is CAA. If there is another hydrophobic region in the C-terminal region of the HMG box (sequence-specific HMG box), another base step A is required, i.e., ACAA. Therefore, the DNA sequence of HMG specific binding should be d(NNACAANN) (N stands

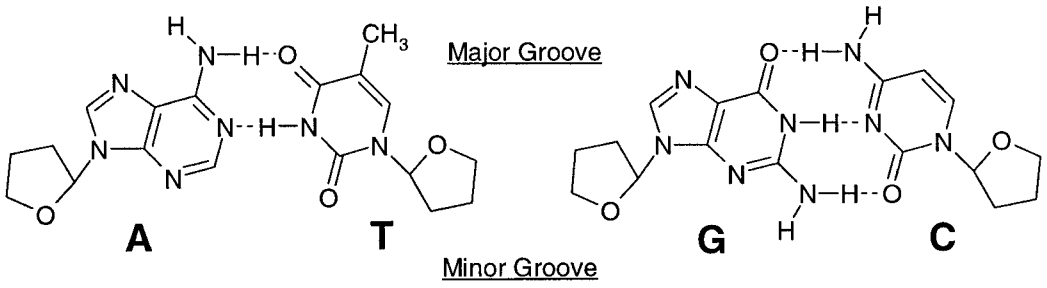

Fig. 12. Schematic illustration of Watson-Crick base pairing in DNA.

for any base, but usually there is a preference for a certain base). Pontiggia et al. showed experimentally that d(GAACAAAG) and d(TAACAATG) have similar dissociation constants and bending angles with hSRY-HMG binding, but the dissociation constant of d(GAACACAG) is reduced tenfold and the bending angle reduced by half, only 35°.[26] Harley et al. further defined the consensus DNA binding site for SRY as A/TAACAAT/A.[50] The TATA-binding proteins (TBP) have similar recognition mechanisms to the TATA box.[51]

"Induced Fit" Conformational Changes Exist in the Process of HMG–DNA Binding

Usually, there are two binding mode hypotheses concerning ligand–receptor interaction. One is the "lock and key" hypothesis, in which the ligand and receptor are static partners and the ligand has a complementary surface with that of the receptor binding site; no or few conformational changes take place upon binding. The other one is the "induced fit" hypothesis, in which ligand or receptor or both change their conformations to achieve greater complementarity of geometries and properties. Our simulations demonstrated that "induced fit" conformational changes occurred during the hSRY-HMG domain binding with DNA process.

Since the average dynamics structure of the unbound hSRY-HMG domain was found to be similar to the known structures of other free HMG boxes (Fig. 9B), we can assume that the conformational change between the two states of hSRY-HMG in the complex and in the free state was caused by the DNA binding and reflects the in vivo situations. In our simulation, the hSRY-HMG domain in the free state has an L-shaped conformation with an acute angle between its long and short arms (Fig. 9A), while the DNA segment is a linear canonical B-form in the free state (as evidenced by NMR).[21] However, in the hSRY-HMG domain the extended N-terminus, helix 3, and the extended C-terminus have flexibility, and the DNA also has conformability, which are beneficial to forming the specific complex. When the hSRY-HMG domain interacts with DNA, the protein can change the acute angle between its long and short arms to become obtuse to accommodate the DNA through the sway of the extended N- and C-termini; meanwhile, the DNA becomes severely bent to adapt to the protein, where key interactions are the partial intercalation of Ile13 and the presence of Phe12 and Tyr74 in the minor groove. As a result, hSRY and DNA form a perfect specific complex. Any change of the protein or the DNA would affect their binding by reducing the stability of the complex.

In the predicted HMGD–DNA complex, the protein-induced bend of the DNA is intermediate, 92°, between the bends observed in the DNA complexed with hSRY and LEF-1, and the angle between the long and short arms of HMGD, also appears to be increased.[48] Moreover, the same case perhaps exists in the LEF-1-HMG box binding with DNA, although its free conformation remains unknown. Because the other HMG box proteins, such as HMG1B and Sox-4, have similar conformations in the free state, we predict that conformational changes will also take place when these proteins interact with linear DNA. The "induced fit" conformational changes are likely to be the common interaction mechanism for the whole HMG box family binding DNA.

DNA Bending Analysis

DNA kinking by minor groove intercalation of one or more amino acids has only been observed for proteins involved in transcriptional regulation, which strongly indicates that such DNA kinking is a biological process involved in the control and regulation of transcription.[52] For example, after the binding of hSRY, the DNA was severely underwound and bent with expansion of the minor groove and compression of the major groove from the normal B-DNA conformation.[21,46]

In our simulation of the free DNA duplex, the normal B-DNA conformation was not recovered after a lengthy dynamics simulation; only the disrupted base pairing and base stacking were repaired, while the bending and unwinding remained. It is also found in other simulations of DNA that the CHARMM energy function tends to favor an A-like conformation.[53]

Usually, B-DNA exists in solution with high hydration. But if dehydration occurs, B-DNA will transform to A-DNA gradually; namely, the conformation

of furanose rings transforms from C2′-*endo* state in B-DNA to C3′-*endo* state in A-DNA.[54] The conformation of the bent DNA caused by hSRY-HMG binding was situated between B-DNA and A-DNA, which implied that the original B-DNA segment lost some water molecules under the influence of the hydrophobic cores of the hSRY-HMG domain. Although residue Ile13 partially intercalated between the base steps of DNA, the hydrophobic cores of HMG box may also be important to DNA bending. We could imagine that the B-DNA loses water molecules after the hydrophobic interaction with the HMG box, then under the partial intercalation of Ile13 and the affects of other residues, the bound DNA transforms to a bent and unwound conformation, gradually closer to A-DNA. A similar local stretching or unwinding mechanism has been proposed in the deformation of the TATA box induced by TBP binding.[47] Another theoretical study also showed that low dielectric interior of proteins was sufficient to cause major structural changes in DNA on association.[55] These results were in agreement with our present model. Moreover, phosphate neutralization also reduced the level of hydration of DNA and favored the transformation of DNA and the formation of the complex.[47]

In addition, given the idea that the averaged dynamics structure of hSRY-HMG domain approximates its free conformation, although both hSRY and DNA changed their conformations during the protein–DNA binding process, the protein changed its conformation to accommodate the DNA in a very limited way; there was a larger conformational change of the DNA to fit the structural shape of hSRY-HMG domain. hSRY-HMG served actually as a structural model to mould the bending conformation of DNA.

The Role of Met9

Besides the intercalating residue Ile13, some other residues also are important for the normal biological function of hSRY, such as Met9 and Gly40. Mutations of these residues would result in a decrease of the binding affinity and/or bending angles. Possible reasons for this according to our simulations are presented below for the Met9 case.

The Met9 → Ile mutation occurs naturally. Experiment confirmed that the dissociation constant was increased from 2×10^{-8} to 6.5×10^{-8}M, and the bending angle was reduced from 76° to 56°.[26]

In the complex, residue Met9 is located at the joint of the extended N-terminal and the helix 1 with the long sidechain opposing the DNA, close to the intercalating residue Ile13. Our simulations demonstrated that the deviations of the loop region around Met9 were much higher in the complex than in the free state, especially the sidechain of Met9 (Fig. 4). Meanwhile, the fluctuations of this region were low and uniform whether in complex or free (Fig. 5), which indicated that a conformational change took place in this region (Fig. 6). Moreover, this short loop region is conserved in most of the HMG box family, especially in the sequence-specific subfamily. Thus, we think that this loop region is essential to the function of hSRY and the conformation change of the region is a biological process, only taking place in the process of hSRY–DNA recognition, in which the neighbor Ile13 could be affected by the sway of Met9 and it becomes more feasible to intercalate into the basepairs of DNA.

When the Met9 → Ile mutation occurs, the sidechain of Ile is shorter and has a $-CH_3$ branch in the C_β atom, which would reduce the swing. The swing of Ile is therefore too weak to help the intercalation of Ile13. As a result, the angle of DNA bending decreases and the affinity of DNA binding reduces.

CONCLUSION

During the simulation of the complex, the major structural changes for the hSRY-HMG domain binding DNA are located in two parts: one is the short loop region around residues Pro8 and Met9, which has a large deviation compared with the NMR structure; the other is the loop of residues 29–33, which has a large fluctuation compared with the other residues. At the protein–DNA interface, just five hydrogen bonds and only one water molecule serving as a hydrogen-bonding bridge were observed. In contrast, the two hydrophobic cores of the hSRY-HMG domain not only maintain the stability of the protein itself, but also maintain the stability of the whole complex, and cause primarily the DNA conformation change. Thus, the interaction between hSRY and its DNA target site is mainly a hydrophobic interaction, and the hydrophobic cores of hSRY-HMG domain is the physical basis of hSRY-HMG–DNA-specific interaction. The salt bridges formed between the positively charged residues of hSRY and the phosphate groups of DNA make the phosphate electroneutral, and are also advantageous to the deformation of DNA and the formation of a stable complex.

The structure of hSRY-HMG domain in the free state was predicted through the simulation of the domain without DNA. Compared with the other free HMG boxes determined by NMR spectroscopy, the free hSRY-HMG domain has a very similar conformation, which is somewhat different from the domain in complex. It thus seems that both hSRY and DNA had changed their conformations to fit each other perfectly during the binding process, with the conformational change of DNA being larger than that of the hSRY-HMG domain.

ACKNOWLEDGMENTS

We thank Drs. A.A. Travers and C. Hardman (Cambridge University) for kindly sending us the

coordinates for the HMGD and HMG1A structures, respectively.

REFERENCES

1. Harrison, S.C. A structural taxonomy of DNA-binding domains. Nature 353:715–719, 1991.
2. Luisi, B. DNA–protein interaction at high resolution. In: "DNA–Protein: Structural Interactions." Lilley, D.M.J. (ed.). Oxford: IRL Press, 1995:1–48.
3. Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.-M., Lovell-Badge, R., Goodfellow, P.N. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346:240–244, 1990.
4. Werner, M.H., Huth, J.R., Gronenborn, A.M., Clore G.M. Molecular determinants of mammalian sex. Trends Biochem. Sci. 21:302–308, 1996.
5. Schafer, A.J., Goodfellow, P.N. Sex determination in humans. BioEssays 18:955–963, 1996.
6. Shah, V.C., Smart, V. Human chromosome Y and SRY. Cell Biol. Int. 20:3–6, 1996.
7. 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.
8. van de Wetering, M., Clevers, H. Sequence-specific interaction of the HMG box proteins TCF-1 and SRY occurs within the minor groove of a Watson-Crick double helix. EMBO J. 11:3039–3044, 1992.
9. Whitfield, L.S., Lovell-Badge, R., Goodfellow, P.N. Rapid sequence evolution of the mammalian sex-determining gene SRY. Nature 364:713–715, 1993.
10. Pontiggia, A., Whitfield, S., Goodfellow, P.N., Lovell-Badge, R., Bianchi, M.E. Evolutionary conservation in the DNA-binding and -bending properties of HMG-boxes from SRY proteins of primates. Gene 154:277–280, 1995.
11. Sanchez, A., Bullejos, M., Burgos, M., Hera, C., Jimenez, R., Diaz de la Guardia, R. High sequence identity between the SRY HMG box from humans and insectivores. Mamm. Genome 7:536–538, 1996.
12. Tajima, T., Nakae, J., Shinohara, N., Fujieda, K. A novel mutation localized in the 3' non-HMG box region of the SRY gene in 46X,Y gonadal dysgenesis. Hum. Mol. Genet. 3:1187–1189, 1994.
13. Hawkins, J.R. Sex determination. Hum. Mol. Genet. 3: 1463–1467, 1994.
14. Grosschedl, R., Giese, K., Pagel, J. HMG domain proteins: Architectural elements in the assembly of nucleoprotein structures. Trends Genet. 10:94–100, 1994.
15. Bianchi, M.E. The HMG-box domain. In: "DNA–Protein: Structural Interactions." Lilley, D.M.J. (ed.). Oxford: IRL Press, 1995:177–200.
16. 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 B-domain of HMG1. EMBO J. 12:1311–1319, 1993.
17. Read, C.M., Cary, P.D., Crane-Robinson, C., Driscoll, P.C., Norman, D.G. Solution structure of a DNA-binding domain from HMG1. Nucleic Acids Res. 21:3427–3436, 1993.
18. 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 interaction with DNA as studied by heteronuclear three- and four-dimensional NMR spectroscopy. Biochemistry 34:16596–16607, 1995.
19. 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.
20. van Houte, L.P.A., Chuprina, V.P., van der 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.
21. 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.
22. 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.
23. 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.
24. 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.
25. Harley, V.R., Jackson, D.I., 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.
26. Pontiggia, A., Rimini, R., Harley, V.R., Goodfellow, P.N., Lovell-Badge, R., Bianchi, M.E. Sex-reversing mutations affect the architecture of SRY-DNA complexes. EMBO J. 13:6115–6124, 1994.
27. Rimini, R., Pontiggia, A., Spada, F., Ferrari, S., Harley, V.R., Goodfellow, P.N., Bianchi, M.E. Interaction of normal and mutant SRY proteins with DNA. Phil. Trans. R. Soc. Lond. B 350:215–220, 1995.
28. Peters, R., King, C.-Y., Ukiyama, E., Falsafi, S., Donahoe, P.K., Weiss, M.A. An SRY mutation causing human sex reversal resolves a general mechanism of structure-specific DNA recognition: Application to the four-way DNA junction. Biochemistry 34:4569–4576, 1995.
29. Karplus, M., Petsko, G.A. Molecular dynamics simulations in biology. Nature 347:631–639, 1990.
30. van Gunsteren, W.F., Berendsen, H.J.C. Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry. Angew. Chem. Int. Ed. Engl. 29:992–1023, 1990.
31. Brooks C.L. III Methodological advances in molecular dynamics simulations of biological systems. Curr. Opin. Struct. Biol. 5:211–215, 1995.
32. Schwabe, J.W.R. The role of water in protein–DNA interactions. Curr. Opin. Struct. Biol. 7:126–134, 1997.
33. Levitt, M., Park, B.H. Water: Now you see it, now you don't. Structure 1:223–226, 1993.
34. Nilsson, L. Protein–DNA interactions. In: "The Encyclopedia of Computational Chemistry." Schleyer, P.v.R. (ed.). New York: John Wiley & Sons, 1998 (in press).
35. Brooks, B.R., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. J. Comput. Chem. 4:187–217, 1983.
36. 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.
37. Jorgensen, W.L., Chandraseklar, J., Madura, J.D., Impey, R.W., Klein, M.L. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79:926–935, 1983.
38. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Mayer, E.F., Brice, J.M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The protein data bank. A computer-based archival file for macromolecular structures. J. Mol. Biol. 112:535–542, 1977.
39. Brünger, A., Karplus, M. Polar hydrogen positions in proteins: Empirical energy placement and neutron diffraction comparison. Proteins 4:148–156, 1988.
40. van Gunsteren, W.F., Berendsen, H.J.C. Algorithms for macromolecular dynamics and constraint dynamics. Mol. Phys. 34:1311–1327, 1977.
41. Brooks, C.L. III, Karplus, M. Deformable stochastic boundaries in molecular dynamics. J. Chem. Phys. 79:6312–6325, 1983.
42. McDonald, I.K., Thornton, J.M. Satisfying hydrogen bonding potential in proteins. J. Mol. Biol. 238:777–793, 1994.
43. Arnott, S., Hukins, D.W.L. Refinement of the structure of

B-DNA and implications for the analysis of X-ray diffraction data from fibres of biopolymers. J. Mol. Biol. 81:93–105, 1973.
44. QUANTA 3.3, Burlington, MA: Molecular Simulation Inc., 1992.
45. Lee, B., Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. J. Mol. Biol. 55:379–400, 1971.
46. 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.
47. Lebrun, A., Shakked, Z., Lavery, R. Local DNA stretching mimics the distortion caused by the TATA box-binding protein. Proc. Natl. Acad. Sci. USA 94:2993–2998, 1997.
48. Balaeff, A., Churchill, M.E.A., Schulten, K. Structure prediction of a complex between the chromosomal protein HMG-D and DNA. Proteins 30:113–135, 1998.
49. Delcourt, S.G., Blake, R.D. Stacking energies in DNA. J. Biol. Chem. 266:15160–15169, 1991.
50. Harley, V.R., Lovell-Badge, R., Goodfellow, P.N. Definition of a consensus DNA binding site for SRY. Nucleic Acids Res. 22:1500–1501, 1994.
51. Juo, Z.S., Chiu, T.K., Leiberman, P.M., Baikalov, I., Berk, A.J., Dickerson, R.E. How proteins recognize the TATA box. J. Mol. Biol. 261:239–254, 1996.
52. Werner, M.H., Gronenborn, A.M., Clore, G.M. Intercalation, DNA kinking, and the control of transcription. Science 271:778–784, 1996.
53. Feig, M., Pettitt, M. Experiment vs force fields: DNA conformation from molecular dynamics simulations. J. Phys. Chem. B 101:7361–7363, 1997.
54. Hartmann, B., Lavery, R. DNA structural forms. Q. Rev. Biophys. 29:309–368, 1996.
55. Elcock, A.H., McCammon, J.A. The low dielectric interior of proteins was sufficient to cause major structural changes in DNA on association. J. Am. Chem. Soc. 118:3787–3788, 1996.
56. Kraulis, P.J. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. J. Appl. Crystallogr. 24:946–950, 1991.