Mutagenesis Study on the Zebra Fish SOX9 High-Mobility Group: Comparison of Sequence and Non-Sequence Specific HMG Domains†

Nai-Wan Hsiao, Dharmaraj Samuel,‡ Yu-Nan Liu, Li-Chuan Chen, Tzu-Ying Yang, Gurunathan Jayaraman,§ and Ping-Chiang Lyu*

Department of Life Sciences, National Tsing Hua University, Hsinchu, Taiwan, ROC

Received April 29, 2003; Revised Manuscript Received June 25, 2003

ABSTRACT: A unique class of proteins, containing high-mobility group (HMG) domain(s), recognizes unusual DNA structures and/or bends specific to AT-rich linear double-stranded DNA. The DNA binding feature of these proteins is exhibited in the HMG domain(s). Although the sequence specific and non-sequence specific HMG domains exhibit very high degrees of sequence similarity, the reasons for the difference between their DNA recognition mechanisms are unclear. A series of zebra fish SOX9 HMG domain mutants was prepared in an effort to elucidate the importance of various residues on protein stability and DNA binding. This study is the first of a comprehensive mutagenesis study on a sequence specific HMG domain. Comparing how various residues influence sequence specific and non-sequence specific HMG domains helps us to rationalize their mode of action. Positively charged amino acids concentrated at the surface of sequence specific HMG domains recognize specific, linear AT-rich DNA segments. After the negative charges at the surface of the DNA are neutralized, the hydrophobic residues of the protein may intercalate DNA. Phenylalanine at position 12 plays a crucial role in the sequence specific HMG domain. The differences in pI values, the instability index, and DNA contact regions between sequence and non-sequence specific HMG domains are associated with their functional modes.

Interactions between proteins and DNA, as in DNA packaging, repair, recombination, replication, and transcription, are vital biological processes (*1*). High-resolution protein−DNA complex structures, determined using X-ray or NMR techniques, provide several important insights into the biological roles of DNA-binding proteins and their DNA recognizing mechanisms. DNA-binding proteins are classified into groups according to the structural motifs that characterize their DNA-binding domain, including helix−turn−helix, zinc finger, leucine-zipper, and helix−loop−helix groups (*2*, *3*). A high-mobility group (HMG)[1] domain is a DNA binding motif that is abundant in nonhistone components of chromatin and in specific regulators of transcription or cell differentiation. Proteins that contain HMG domain(s) are found commonly in eukaryotes (*4*). Common properties of HMG domain proteins include interactions with the minor groove of the DNA helix, binding to irregular DNA structures, and the capacity to modulate DNA structure by bending (*5*). HMG proteins have diverse functions, including the determination of the structure and/or stability of nucleosomes as well as in transcription and replication (*4*). The discovery of the mammalian sex-determining gene, *Sry*, rapidly led to the identification of *Sox* (SRY-type HMG box) genes (*6*, *7*). SOX proteins have critical functions in numerous developmental processes, including sex determination, skeleton formation, pre-B and T cell development, and neural induction (*8*). More than 20 members of the SOX family have been cloned, and all of them contain the HMG domain. The sequence of the SOX family HMG domains is more than 50% identical to that of human testis-determining factor (SRY). Zebra fish SOX9 (zfSOX9) HMG-box protein, studied in this work, binds and bends specific AT-rich double-stranded linear DNA. Human SOX9 is directly involved in chondrogenesis by binding and activating the chondrocyte specific enhancer of the *Col2a1* gene (*9*, *10*). The loss of Sox9 gene function leads to a genetic condition known as campomelic dysplasia (CD), a form of dwarfism characterized by extreme skeletal malformation, and one in which three-fourths of XY individuals either are intersexed or exhibit male to female sex reversal (*8*, *11*−*14*). Amino acid sequences of HMG domains in vertebrates are highly conserved. For example, the HMG domain of zebra fish SOX9 differs only at a single nonconserved position from that of humans (G49S). SRY and SOX9 proteins determine the sex of an embryo (*15*). To date, at least 26 different natural mutations in the SRY gene have been discovered in individuals with a 46XY karyotype (*16*). With only one exception, all these mutations occurred within the HMG domain (*17*). The functions associated with this

† This work is supported in part by a research grant (NSC92-3112-B-007-006) from the National Science Council, Taiwan, and a Program for Promoting Academic Excellence of Universities grant (89-B-FA04-1-4) from the Ministry of Education, Taiwan.

* To whom correspondence should be addressed. Telephone: +886 35742762. Fax: +886 35715934. E-mail: lslpc@life.nthu.edu.tw.

‡ Present address: Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, PA 19111, USA.

§ Present address: Centre for Protein Engineering and Biomedical Research, The Voluntary Health Services Adyar, Chennai, 600 113, India.

[1] Abbreviations: EMSA, electrophoretic mobility shift assay; CD, circular dichroism; IPTG, isopropyl thiogalactopyranoside; SRY, sex-determining region of the Y chromosome; HMG, high-mobility group; Sox, Sry-like HMG box.

10.1021/bi034678d CCC: $25.00
Published on Web 09/05/2003

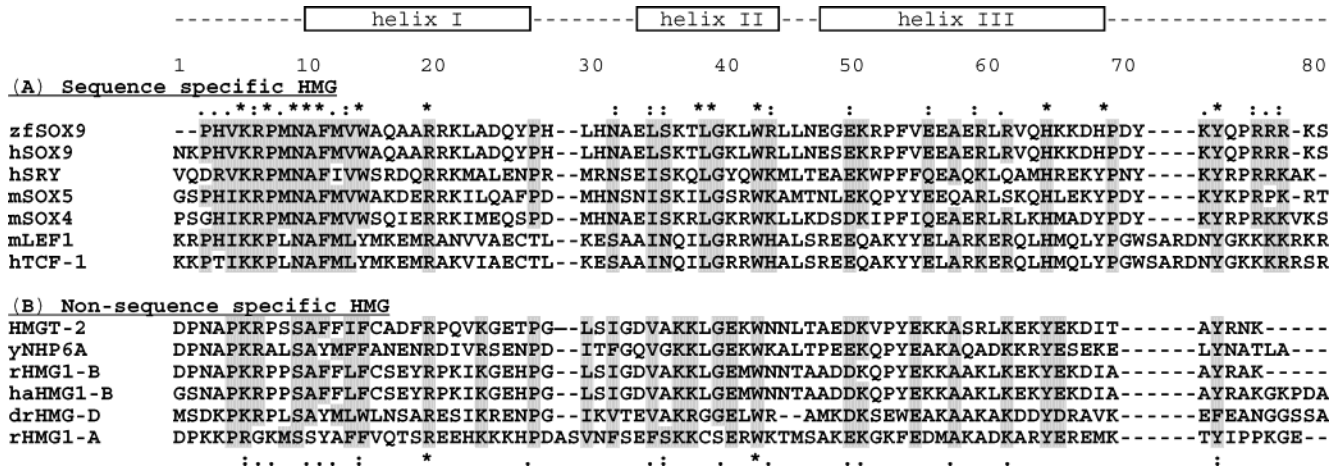

FIGURE 1: Amino acid sequence alignment of selected HMG domains: (A) DNA sequence specific domains and (B) non-sequence specific domains. Three helices are indicated as boxes. Alignment features of sequence specific HMG domains are represented at the top, and the residues conserved among both the sequence specific and non-sequence specific groups are shown at the bottom of the alignment. An asterisk denotes identical or conserved residues in all sequences. The conserved substitutions and semiconserved substitutions are denoted with colons and periods, respectively.

class of proteins exist mainly in the DNA-binding HMG domain. The HMG domain includes approximately 80 amino acids with highly conserved proline, aromatic, and basic residues (*18*).

HMG proteins are subdivided into three different categories, depending on their means of DNA recognition (*19*, *20*). HMG proteins that recognize unusual geometries of DNA molecules, such as the four-way junction or cisplatin-modified DNA, are the first category (non-sequence specific HMG) (*21*−*24*). Examples of proteins that recognize unusual DNA geometries are HMG1 and HMG2 (*25*). In the next category of HMG proteins in addition to binding with unusual DNA structures, they also bind and bend a double-stranded linear AT-rich DNA sequence (AGAACAATGG) (sequence specific HMG) (*26*). Various proteins that belong to this category are lymphoid enhancer-binding factors (LEF-1 and TCF-1) and mammalian sex-determining factors (SRY and SOX proteins). Members of the third category of HMG proteins bind unusual DNA structures with a requirement of a cofactor for sequence recognition; examples include UBF and ABF (*27*). The HMG proteins that exhibit sequence specific and non-sequence specific forms are indicated in Figure 1. Although the hydrophobic interactions between protein and DNA play an essential role in all three forms, the detailed mechanism of sequence recognition is unclear (*25*). To understand the DNA recognition mechanism of the sequence specific HMG domain, we prepared a series of mutants for zfSOX9 HMG-box protein in this work. The influence of various residues on the structure and the DNA binding ability was analyzed. The differences between sequence and non-sequence specific HMG domains were discussed as well.

EXPERIMENTAL PROCEDURES

Homology Modeling. The three-dimensional structure of the zebra fish SOX9 HMG domain was modeled using the mouse SOX5 HMG domain as a template structure (PDB entry 1I11) (*28*). The protein model was derived using the Homology module in the InsightII package (Accelrys Inc.) with an indigo^2 SGI workstation. The program replaced various residues of the mouse SOX5 HMG domain according to the primary sequence of the zfSOX9 HMG domain, to build a model that satisfied all spatial restraints. The model was further energy minimized using the Discover 3 module in InsightII, with 500 energy minimization steps. The final three-dimensional model of the zfSOX9 HMG domain was evaluated using Procheck version 3.5.4 (*29*).

Mutagenesis and Protein Expression. All mutants of the zfSOX9 HMG domain were prepared by site-directed mutagenesis using the Quick Change kit (Stratagene). Mutant and wild-type zfSOX9 HMG proteins were expressed in the pET-29a/*Escherichia coli* BL21(DE3) system after the DNA constructs of the plasmid had been verified. The bacterial cells, grown at 37 °C in LB medium to an absorbance (A_{600}) of 0.5−0.6, were induced with 2 mM IPTG to express protein. The cells were harvested by centrifugation (1000*g* for 30 min at 4 °C) after induction for 4 h. The cell lysate, after sonication, was centrifuged (17000*g* for 30 min at 4 °C) and loaded into a 8 mL native Ni column (Qiagen) pre-equilibrated with a lysis buffer [300 mM NaCl in a 50 mM phosphate buffer (pH 8.0)]. The column was repeatedly washed with a lysis buffer before the protein of interest was eluted with 150 mM imidazole (Qiagen). The protein was desalted and purified using RP-HPLC with a COSMOCIL C_{18} semipreparative column (20 mm × 25 mm, 5 μm pore size) using a linear gradient of 32 to 37% acetonitrile containing 0.1% trifluoroacetic acid. The authenticity of the sample was further confirmed with electron spray mass spectrometry.

Circular Dichroism Spectroscopy. The circular dichroism spectra of the zfSOX9 HMG domain and its mutants were obtained using an AVIV 202 spectropolarimeter. The protein sample (10 μM) in 10 mM phosphate buffer (pH 6.0), containing 100 mM KCl and 1 mM EDTA, was scanned from 260 to 200 nm at 5 °C. Protein concentrations were determined by measuring A_{280}, after serial dilutions of the pure protein (*30*). Thermal unfolding was monitored at 222 nm over a range of temperatures (2−94 °C). Guanidine hydrochloride (GdnHCl) was used for chemical-induced unfolding of the protein at 5 °C. Protein samples (10 μM) in 10 mM phosphate buffer (pH 6.0) containing 100 mM

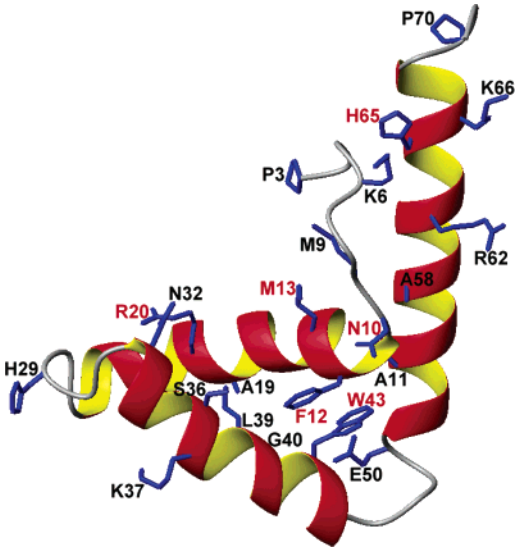

FIGURE 2: Theoretical structure of the HMG domain of zebra fish SOX9, displaying the positions and side chains of mutational sites considered in this study. All the HMG domains exhibit a common structural fold, with minor differences. Three α-helical segments form an L-shaped structure stabilized by the hydrophobic core. The long arm of the 'L is comprised of N-terminal residues coordinated against helix 3 and the C-terminus, while the short arm of the L is comprised of helices 1 and 2.

KCl, 1 mM EDTA, and GdnHCl were incubated at 5 °C overnight.

Electrophoretic Mobility Shift Assay (EMSA). The biological activities of various mutants were monitored using two different oligonucleotide probes. The SOX9-preferred *in vitro* S9WT probe 5′-TA**AGAACAATGG**GA-3′ (K_d = 12.4 ± 2.5 nm) and the *in vivo* SOX9 binding the COL2C1 probe (Col2a1 enhancer, 5′-CC**CACAATG**CC-3′) were used to test the DNA binding capacities of proteins (*26*). The affinity of the SOX9 HMG domain for S9WT is 9-fold higher than for COL2C1. The high-affinity SOX9 binding sites are shown in bold. Oligonucleotides were purchased from Biobasic (Taipei, Taiwan). DNA and aliquots of wild-type or mutant proteins were incubated in 20 μL of a reaction solution [25 mM phosphate buffer (pH 7.2)] at 4 °C for 3 h. The mixtures were loaded into a nondenaturing 20% polyacrylamide gel and electrophoresed in 1× TBE at 150 V for 3 h after addition of 4 μL of 6× DNA loading dye. The PAGE gel was separately stained with ethidium bromide and Coomassie blue (*31*, *32*). The DNA binding capacities were quantified by comparing the shift and the amount of DNA in different wells with respect to that of the wild-type protein. The IS-1000 Imaging System (Alpha innotech) (*33*−*35*) was used to analyze the PAGE gel. The band intensities were measured using TotalLab version 1.11 software from Phoretix. We have used the "manual function" of the "1D image function" in TotalLab to chose, create, and analyze the lanes. The image rectangle was selected for background subtraction. The measurements window updates the band intensities automatically (*36*, *37*). The percentage binding was estimated by comparing the complex and the free form DNA band intensities.

RESULTS

Homology Modeling of the zfSOX9 HMG Domain. High-resolution three-dimensional structures of sequence specific HMG proteins determined to date include those of SOX5 and SOX4 in free forms, and both male sex-determining factor (SRY) and lymphoid enhancer-binding factor 1 (LEF-1) in complex (*28*, *38*−*40*). The sequence of the SOX5 HMG domain is 52% identical to that of zfSOX9, and was chosen as a template for modeling the three-dimensional structure of the SOX9 HMG domain. The model has a typical structural fold with other HMG domains at which three α-helical segments form an L-shaped structure stabilized by hydrophobic interactions. The long arm of the L is comprised of N-terminal residues coordinated against helix 3 and the C-terminus, while the short arm includes helices 1 and 2 (Figure 2).

Circular Dichroism Spectroscopic Analysis of Protein Stability. The circular dichroism (CD) spectra of the wild-type zfSOX9 HMG domain and mutant proteins exhibit double minima at 208 and 222 nm, which are characteristic of an α-helical structure. Although no mutant leads to major reorganization of the fold, several mutated proteins exhibited quantitative differences with respect to the wild type. The comparison of CD spectra between mutants and the wild type is shown in Figure 3. The ellipticity at 222 nm was used to estimate the helical content of the proteins (Table 1). The

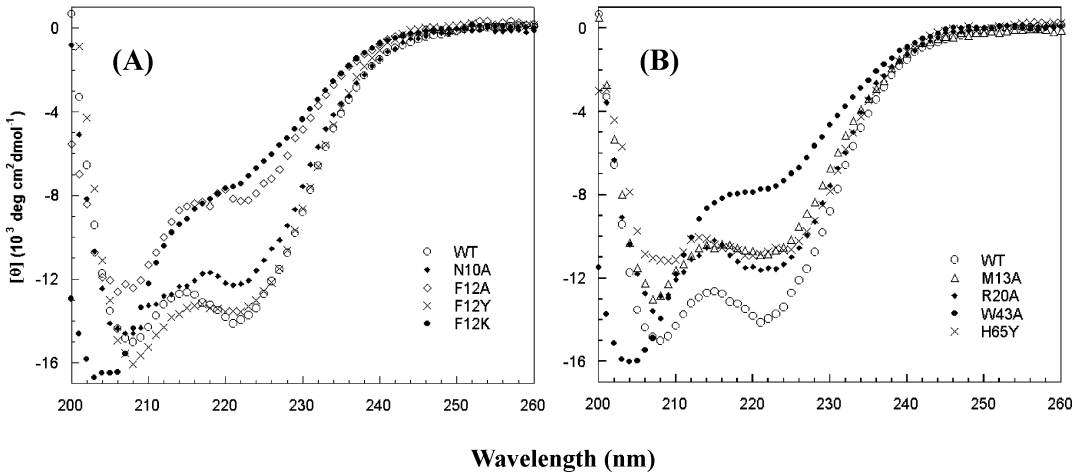

FIGURE 3: CD spectra of wild-type zfSOX9 HMG and selected mutant proteins. CD spectra illustrate the influence of various mutants on the α-helical content of the zfSOX9 HMG domain. Substitutions that involve hydrophobic residues are very important in determining the total helical content of the protein.

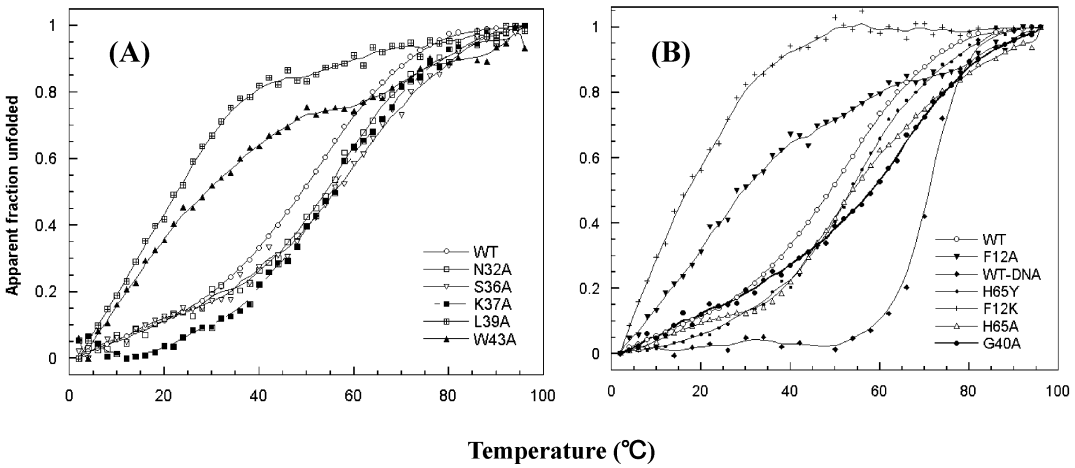

FIGURE 4: Stability of mutants of the zfSOX9 HMG domain, monitored by CD thermal denaturation. The apparent fraction of the protein unfolded is plotted vs temperature for various mutants with different stabilities. The stability of the HMG domain enormously increased upon binding with its consensus DNA sequence.

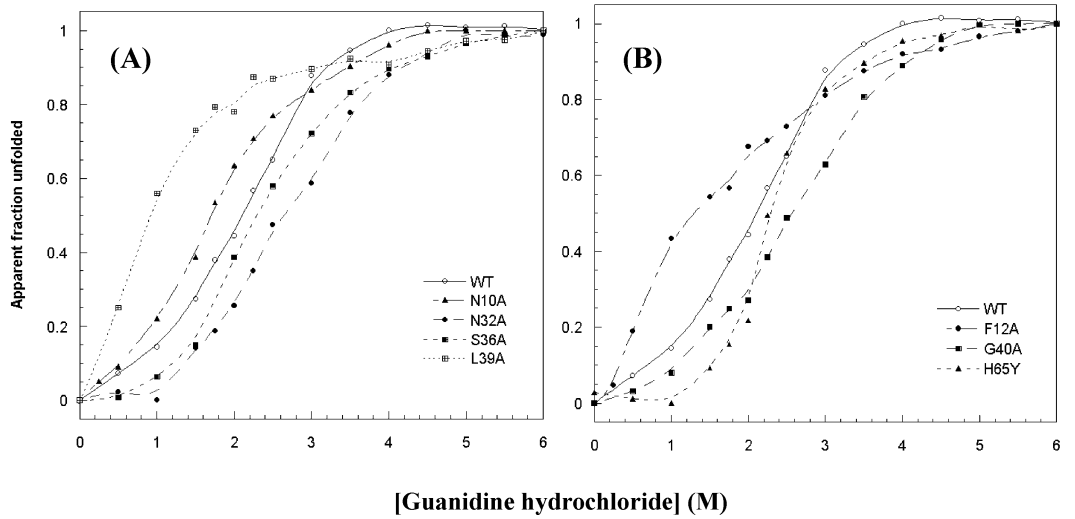

FIGURE 5: Chemical denaturation of the wild-type zfSOX9 HMG domain and its mutants by GdnHCl, monitored using circular dichroism spectroscopy techniques. The system does not behave as a perfect two-state model because N-terminal residues in the long arm of the L-shaped protein are very flexible.

thermal stability of the proteins was determined from the change in the circular dichroism signal at 222 nm over a range of temperatures. Thermal and chemical denaturations are monitored with T_m and C_m, respectively. Figures 4 and 5 present the unfolding profiles of the mutant proteins, which exhibit moderate differences in stability, obtained in temperature and GdnHCl denaturation experiments, respectively. Most of the mutants show a negligible influence on the helical content (±5%). However, seven mutations (F12A, F12K, M13A, W43A, K66A, A11L/A58L, and H65Y) exhibit a significant decrease in helicity. On the other hand, the helical content of S36A and H65A are significantly increases. Table 1 summarizes the biophysical analysis of zfSOX9 HMG mutants and comparison to other classes of HMG domains.

DNA Binding Assay. The capacities of the mutants to bind with DNA, relative to the wild type, were measured using the electrophoretic mobility shift assay (EMSA) (Figure 6). Table 1 presents the binding data for the S9WT probe and the Col2C1 probe. Wild-type HMG protein binds weakly with the hSOX9 *in vivo* Col2a1 enhancer DNA motif (C/G)(T/A)CAA(T/A)(C/G) (Col2C1 probe) (*26*). The probe, S9WT, bears the high-affinity binding site defined for other SOX proteins [(A/T)(A/T)CAA(A/T), termed SOXCON (*26*)]. Although the binding of DNA with protein occurs in 1:1 ratio, a band corresponding to traces of free DNA was observed. To quantify the binding efficiencies of various mutants, a higher protein concentration was used for all the experiments (a 1:2 DNA:protein ratio). Wild-type zfSox9 HMG−S9WT probe interaction, considered to be 100% (++++), was compared with interactions of various mutants. Mutants with less than 10% binding were considered to be nonbinding mutants and represented with a dash. Other binding interactions (10−40, 40−70, 70−90, and >90%) were specified with +, ++, +++, and ++++, respectively. In the case of the COl2C1 probe, binding was marked as + or −. DNA binding was completely affected by mutation at position 12 or 43. Replacement of phenylalanine with alanine or lysine affected the protein stability and the DNA binding, whereas tyrosine substitution did not affect the secondary structure or the stability of the protein; however, the DNA binding was affected (Figure 3). Other biologically significant mutants (M9A, N10A, M13A, R20A, W43A, H65Y, and K66A) showed direct structure−function relationships (Fig-

Table 1: Summary of the Biophysical Properties of HMG-Box Mutants[a]

zfSOX9 HMG-box mutant	[θ] (deg cm^2 dmol^{-1}) at 222 nm (5 °C)	T_m (°C)	[GdnHCl]$_{1/2}$ (M) (5 °C)	EMSA (S9WT) (5 °C)[b]	EMSA (COL2C1) (5 °C)[c]	sequence specific HMG conserved[d]	*SRY* disease (*57*)	*SOX9* disease (*57*)	*SRY-DNA* binding interface (*39*)[e]	sequence specific HMG mutation[f]	non-sequence specific HMG mutation[f]
wild type	−13935	48	2.0	100	+						
P3D	−12413	48	2.0	63.7	ND						
K6A	**−17880**	48	ND	94.6	+	#					K6A−(*41*−*43*) K6E−(*43*) K6R−(*27*) K6G±(*27*) K12A−(*41*)
M9A	−12600	48	**1.6**	75.0	+		M9I M9R		√		P15A±(*41*)
N10A	**−12228**	**43**	**1.6**	46.6	−	#			√		S16A−(*41*)
F12A	**−8257**	**<30**	**<0.5**	8.2	−	#	F12V	F12S	√	F12L−(*48*)	Y28A−(*42*)
F12Y	−13552	46	ND	**1.8**	−					F12A−(*47*)	Y28D−(*77*)
F12K	**−7415**	**<30**	ND	**1.4**	−						F12S−(*27*) F12Y−(*27*) F18A+(*41*) Y119A±(*58*) A16F−(Pt)(*46*) F16A−(Pt)(*46*)
M13A	**−10836**	47.5	2.0	61.2	−		I13T		√	I68T−(*78*) I13A−(*47*) I13T−(*47*)	M29A−(*42*) M29A−(*77*) M29D±(*77*) F13V−(*27*) F13S−(*43*) F19A−(*41*)
A19Q	−13343	48	2.0	99.6	ND						
R20A	**−11510**	48	ND	18.7	−	#	R20N		√		R26A+(*41*) R20G−(*43*) R36A−(*42*) R20E−(*27*) R20Q−(*27*) R127A−(*58*)
H29A	**−12148**	48	2.0	65.4	+						
N32A	−12555	**55**	**2.5**	70.8	+	#					V32A−(*45*) V32F−(*45*) I38A+(*41*) F48A−(*42*) F37W−(Pt)(*46*) F37A−(Pt)(*46*) I37F+(Pt)(*46*) I37A−(Pt)(*46*)
S36A	**−15301**	**55**	**2.3**	74.7	+	#	S36G		√		
K37A	−14057	**55**	2.0	86.5	+		K37		√		K43A+(*41*) K37E±(*43*) K53A−(*42*) K114A−(*58*)
L39A	−12440	**<30**	**<0.5**	82.6	+	#					
G40A	−14225	**55**	**2.6**	86.2	ND	#	G40R			G40R−(*60*)	G46A−(*41*)
W43A	**−7695**	**<30**	ND	**0.3**	−	#		W43R	√		Y150A−(*58*)
E50A	−14090	50	2.0	91.4	ND	#					
R62A	−13175	48	ND	87.2	+				√		K78A−(*42*) K68A−(*41*) K62E±(*43*)
H65A	**−16554**	**52**	ND	74.6	+	#				H65Y−(*48*)	Y65S±(*43*)
H65Y	**−10737**	**50**	**2.3**	56.1	−						Y81A±(*42*)
H65W	**−14826**	**52**	ND	63.2	+						F172A±(*58*)
K66A	**−11204**	**52**	ND	72.1	ND				√		
P70A	−13037	48	2.0	77.7	+	#	P70L	P70R		P70R−(*48*)	
K73A	−12320	48	2.1	71.8	+				√		
Y74A	−12514	48	2.0	44.5	ND	#			√		Y74S±(*43*) Y88A±(*42*) F179A±(*58*)
P3D/A19Q	−12880	48	2.1	95.5	ND						
A11L/A58L	**−9212**	**35**	**1.8**	**3.5**	ND						

[a] Experimental errors for the percent helix, T_m, and C_m are ±3%, ±3 °C, and ± 0.1 M, respectively. ND denotes an undetermined biophysical parameter. Significant mutants are shown in bold letters. [b] Wild-type zfSox9 HMG−S9WT probe interaction, considered as 100% (++++). The mutants exhibiting less than 10% of standard binding were considered nonbinding mutants and given the symbol −. Other binding interactions (10−40, 40−70, 70−90, and >90%) were specified with +, ++, +++, and ++++, respectively. The experimental error is ±5%. [c] Protein−COL2C1 probe interactions are denoted with a + or −. [d] # indicates the sequence specific HMG conserved residue. [e] Residues at the SRY−DNA interface are denoted with the √ symbol. [f] The last two columns include various HMG mutants reported in the literature. The literature references are given in parentheses. The interactions of mutant proteins with DNA are indicated by +, −, or ±, corresponding to an increased, decreased, or insignificant effect, respectively.

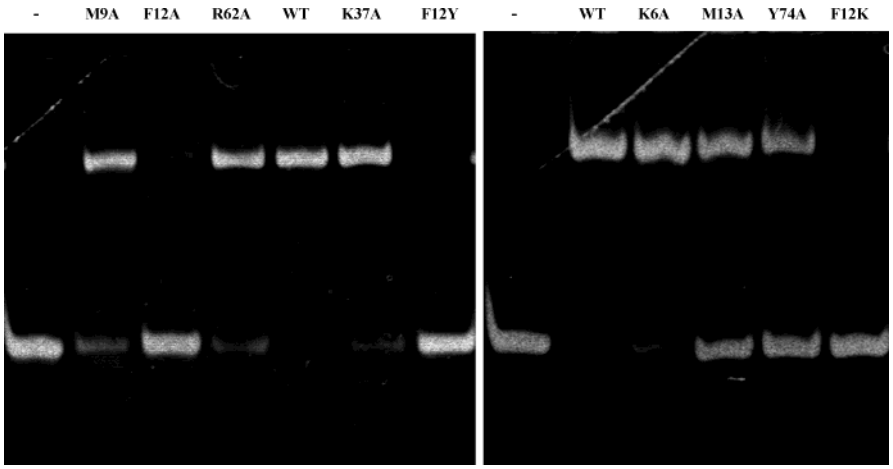

FIGURE 6: Electrophoretic mobility shift assay (EMSA) of wild-type and mutant zfSOX9 HMG domains with the SOX9 cognate *in vitro* DNA sequence (A/T)(A/T)CAA(A/T) (S9WT). A fixed concentration of S9WT was incubated with the wild type and various mutants. Representative EMSAs of wild type and selected mutants exhibit contrasting interactions with DNA. An Alpha Innotech Alfaimager IS-1000 Imaging System (*33*−*35*) was used to measure the band intensities. Table 1 represents the intensity of the wild type and other mutants.

ure 3). Alanine mutation of Lys6, Lys37, Arg62, and Tyr74 did not influence structure or function; interestingly, mutation of these positions imparts major changes in the case of non-sequence specific HMG domains (Table 1).

DISCUSSION

Mutagenesis Study of the zfSOX9 HMG Domain. Alanine scanning analysis of *zfSOX9 HMG-box* revealed that Asn10, Phe12, Met13, Arg20, and Trp43 are very important in maintaining the three-dimensional fold as well as in preserving the DNA binding ability of the protein. Replacement of either one of these residues leads to a less stable protein incapable of binding to DNA. These residues are also important in non-sequence specific HMG-box proteins (*27*, *41*−*43*). Residue Asn10 is replaced with serine in non-sequence specific HMG-box proteins. The stability of sequence specific *SOX9 HMG-box* was increased by replacing Asn32, Ser36, and Lys37 with alanine, but their DNA binding abilities were slightly decreased. In the model structure, these residues are very close to the hydrophobic core in the short arm of the L-shaped protein, so alanine substitutions strengthen the hydrophobic core to increase the protein stability. Analysis of SRY−DNA complexes revealed that Lys37 and Ser36 are located close to thymidine bases of the DNA to interact electrostatically (*39*). In the case of alanine mutants, alanine may interact with DNA but cannot neutralize charges at the DNA surface (*44*). In the case of non-sequence specific proteins, the unconserved positions 32 and 36 are mostly occupied with hydrophobic residues (Figure 1) (*41*, *42*, *45*, *46*). Substitution of the conserved Lys37 with alanine leads to a more stable protein with weakened DNA binding ability (*27*, *41*−*43*, *47*−*49*). The inference of these observations is that the sequence specific HMG-box domain requires more electroneutralization of DNA phosphate groups than non-sequence specific counterparts.

The stability of the HMG-box domain is directly related to the hydrophobic core of the molecule (*50*). Though the alanine mutations are directly related to the stability of the protein, DNA binding is less affected, except for mutation at Phe12, Met13, and Trp43. In the theoretical model at major wing, Leu39 and Trp15 form a portion of the hydrophobic core. The overall stability of the molecule is decreased as leucine is replaced with a less hydrophobic alanine (Figure 4). Previously published complex structures (PDB entries 1HRY and 2LEF) revealed that the DNA binding was unaffected because leucine or alanine at position 39 along with Trp15 opens their hydrophobic contacts to interact with the DNA molecule (*40*, *51*). Residues Met9 and Leu61 constitute another portion of the hydrophobic core in the long arm of the theoretical structure. Mutation of the methonie to alanine at position 9 may have the same effect as that of the L39A mutation. Attempts to increase the hydrophobic core size lead to adverse effects. An A11L/A58L double mutant decreased the secondary structure, stability, and DNA binding ability.

The mutants P3D, A19Q, H29A, R62A, K66A, K73A, and Y74A did not affect the secondary structure or stability of the sequence specific HMG-box protein. DNA binding was also unperturbed for these mutants except for a mild effect in the P3D and Y74A mutants (Table 1). All these residues found to effectively interact with DNA base pairs that conjoin the HMG-box recognizing consensus sequence (*48*). Though the effect of R62A was marginal in the case of the sequence specific domain, it affected the DNA binding capacity of non-sequence specific HMG-box proteins (*41*−*43*). Residues Lys6 and Glu50 are highly conserved among the HMG-box proteins (Figure 1). Alanine mutageneses of these residues had no effect on the structure or DNA binding ability of *SOX9 HMG-box.* Analysis of the location of these residues in the three-dimensional spaces of various DNA−HMG box complexes revealed that they are mostly projected away from the DNA binding-interface (*40*, *51*). Therefore, these residues might be involved in secondary interaction like bringing another portion of DNA closer or bridging other DNA binding proteins to effect the biological activity (*52*−*56*). In the case of non-sequence specific HMG-box proteins, Lys6 is found to be involved in DNA binding (*43*).

Biologically Significant Mutants. One of the mutants reported in the case CD patient is H65Y (*48*). Interestingly, tyrosine at position 65 (Tyr65) was found to be highly conserved among the non-sequence specific HMG-box proteins (*48*). McDowall et al. (*48*) have proposed that H65Y might have influenced Pro8, which in turns affects the

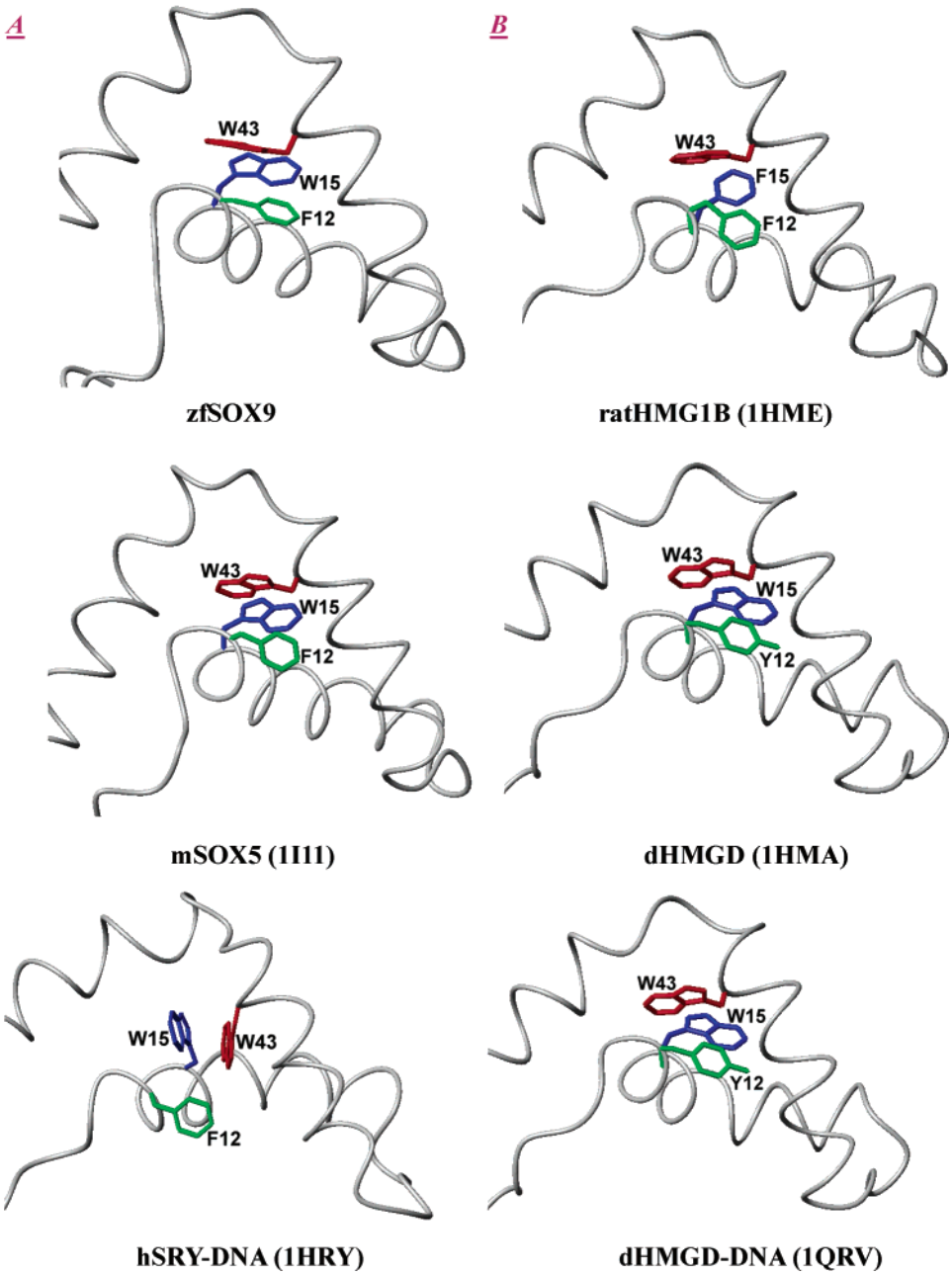

FIGURE 7: Hydrophobic interactions of (A) sequence specific and (B) non-sequence specific HMG domains. A portion of the hydrophobic core consists of phenylalanine or tryptophan at positions 12, 15, and 43 of the primary structure of the sequence specific HMG domains which are opened up to facilitate DNA−protein interaction. The last figure in each panel depicts the orientations of the side chains in the proteins involved in DNA complexes. The other figures represent the free forms.

orientation of Arg7. Mutation of Arg7 was found to affect the DNA binding in non-sequence as well as sequence specific HMG-box proteins (*42*, *57*). In other words, the presence of His at position 65 helps to keep Arg7 in the correct orientation required for electroneutralization. The mechanism was further supported by substitution of histidine with the bulkier tryptophan, leading to a stable protein capable of binding to DNA (Table 1). But the argument was contradicted by alanine and tyrosine substitutions, where tyrosine retards the DNA binding and a less bulky alanine leads to a stable protein retaining DNA binding ability (Table 1) (*48*). From the above results, it could be argued that histidine may not only influence other residues but also play a direct role. The structural studies presently underway in our laboratory may reveal the actual role of His65. A proline residue at position 70 was found to be important in orienting the C-terminal tail to interact with DNA, and its P70R mutant is found in the CD patients (*48*). Interestingly, Pro70 is conserved in sequence specific HMB-box proteins only (Figure 1). The alanine mutant of Pro70 affected the EMSA of longer length DNA, SOXCORE (21mer), and not the consensus sequence (14mer) alone (Table 1) (*48*). The residues probably influenced by the P70A mutant are Lys73 and Tyr74 (*48*). Alanine mutations of these sites scarcely influence the protein stability or DNA binding (*42*, *43*, *58*). From the structural analysis, it could be argued that these residues might have interacted with the DNA bases that flank the consensus sequences. DNA interaction at positions such as Lys73 and Tyr74 may help to match the larger energy required for complex formation by sequence specific HMG-

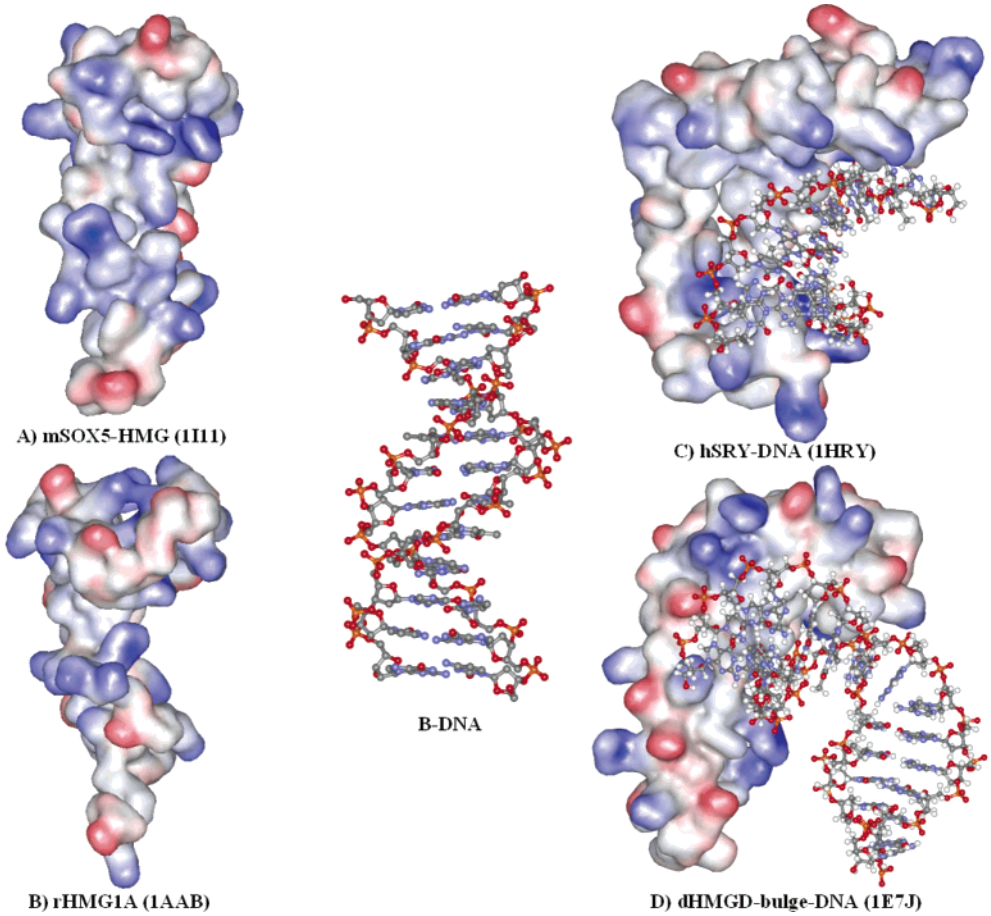

FIGURE 8: Schematic representation of the HMG domain−DNA interaction. The DNA binding interface of (A) the sequence specific HMG domain (mean pI of ∼10.38) includes several positively charged residues compared to the (B) non-sequence specific HMG domains (mean pI of ∼9.26). The DNA binding contact region of the non-sequence specific HMG domain includes only the short arm of its L-shaped structure (with 22−28 heavy atoms within 3.5 Å), whereas the sequence specific HMG domain uses both the short and long arms (with 46−145 heavy atoms within 3.5 Å) (*65*). The sequence specific HMG domain undergoes major structural reorganization, but the non-sequence specific HMG domain remains unaffected. Various proteins and protein−DNA complexes in the figure are labeled according to their Protein Data Bank name. Proteins in the free forms are shown as inverted L shapes where short arms face the viewer. The phosphate groups that cover the minor groove are neutralized by the positive charges of the sequence specific HMG domains so that they interact with hydrophobic bases closer to the minor groove of regular B-form DNA. The protein's hydrophobic side chains may act like a wedge.

box proteins. Mutants F12Y and W43A are commonly found in CD patients (*57*). Positions 12 and 43 required hydrophobic residues in sequence specific HMG-box proteins (Figure 1). Replacement of these residues with less hydrophobic residues (F12A and W43A) or with a charged residue (F12K) completely affected the DNA binding (Table 1). Structural analysis of previously determined structures of HMG-box−DNA complexes showed that these residues play important roles in DNA minor groove binding (*39*, *59*). G40R, another mutant found in the CD patient, hindered the DNA's access to the active sites of the protein chain because of its bulk and positive charge (*60*). Gly40 is a conserved residue in the case of sequence specific HMG-box proteins. Replacement of the less favored glycine from the helical region with alanine leads to a stable protein and DNA binding ability comparable to that of the wild type. In the case of non-sequence specific proteins, the G40A mutant increases the stability but completely retards DNA binding (*41*).

F12 Is the Key Residue for DNA Binding. Three residues (F12, W15, and W43) that constitute a hydrophobic core play major role in protein−DNA interactions (*27*, *41*−*43*). This hydrophobic core, located at the juncture of the minor and major wings of the L-shaped HMG domain, undergoes a major conformational reorganization when binding to DNA in sequence specific HMG-box proteins (Figure 7). Of the three residues at the hydrophobic wedge, F12 was found to be very important. Replacement of functionally important phenylalanine with tyrosine in the zfSOX9 HMG domain hindered DNA binding. The F12Y mutant did not change the structure or stability of the protein (Figures 3 and 4). Comparing the primary structure of the sequence and non-sequence specific HMG box revealed that F12 is strictly conserved in sequence specific HMG box proteins, whereas in some non-sequence specific proteins, it is replaced with tyrosine (Figure 1). Phenylalanine at position 12 enables the sequence specific HMG domain to bind linear DNA, but tyrosine fails to do so (Figure 6). However, residues in the corresponding position of the non-sequence specific domain interact with deformed or unusual DNA structures, where phenylalanine or tyrosine can fulfill the role (Figure 1) (*27*, *61*). These results demonstrate that F12 is crucial for the DNA binding of sequence specific HMG domain proteins.

Role of Basic Residues in DNA Recognition. Sequence and non-sequence specific HMG domains possess the same number of conserved hydrophobic residues, but a different number and distribution of charged residues (Figure 1). The

Table 2: Theoretical pI, Instability Index, and Number of Heavy Atoms Involved in the DNA−Protein Interface of HMG-Box Domains

HMG	theoretical pI	instability index	no. of heavy atoms within 3.5 Å
sequence specific			
zfSOX-9b	10.43	55.05	
hSOX-9	10.47	60.61	
hSRY	10.67	65.49	46 (1HRY) 92 (1J46)
mSOX-5	10.10	81.51	
mSOX-4	10.28	76.78	
mLEF-1	10.36	63.62	145 (2LEF)
hTCF-1	10.36	66.86	
non-sequence specific			
yNHP6A	8.14	40.56	
rHMG1-B	9.38	16.88	
haHMG1-B	9.49	12.45	
drHMG-D	9.47	42.03	28 (1QRV) 22 (1E7J)
rHMG1-A	9.68	41.45	24 (1CKT)
HMGT-2	9.44	28.07	

positively charged residues distributed over the surface of the HMG domain form a clamp to coordinate with the minor groove of cognate DNA (Figure 8). Theoretical isoelectric points (pI) for different HMG domains are calculated to quantify these differences. The mean theoretical pI values calculated for sequence and non-sequence specific HMG domains were 10.38 and 9.26, respectively (*43*) (Table 2). The higher basicity of the sequence specific HMG proteins implies that they can neutralize phosphate groups in linear cognate DNA to facilitate the interaction of hydrophobic residues (*44*). In the case of unusual DNA structures, fewer negative charges on the DNA must be neutralized before the nonpolar side chains of the protein effectively interact with bent DNA (*51*). The extent to which charged residues affect the recognition of the DNA sequence is further demonstrated by the following example. Although the stability of the sequence specific zfSOX9 HMG domain was slightly increased via replacement of Asn32 and Lys37 with alanine, the ability to bind DNA was varied. The alanine mutants at these positions might have failed to neutralize negative charges on the DNA surface (*44*). In non-sequence specific HMG domains, hydrophobic residues occupy the position of unconserved Asn32. Similarly, positively conserved basic residues at positions 21, 41, and 60 and the C-terminal residues (75−77) of the sequence specific HMG domain are mainly replaced with neutral or negatively charged residues in the non-sequence specific HMG domain (Figure 1).

Protein−DNA Interface of HMG Domains. The primary structures of sequence and non-sequence specific HMG domains exhibit a high degree of sequence similarity and a conserved three-dimensional fold (Figures 1 and 2) (*62*). In the case of the sequence specific HMG domain, the interactions with linear DNA should contribute to the larger energy requirement for DNA bending (*44*). The formation of recognition contacts, such as hydrogen bonds, ion pair contacts, and nonpolar contacts between protein and DNA, is energetically favorable (*63*). The sequence specific HMG domains have more residues at the protein−DNA interfaces than the non-sequence specific counterparts (Table 2). The numbers of atoms that help to coordinate these proteins with DNA are very different in each case. Although non-sequence specific HMG domains interact with DNA at a few definite numbers of positions, sequence specific HMG proteins interact at several sites to satisfy the large energy requirement (Table 2) (*44*). A simple comparison of mutants determined that the non-sequence specific HMG domains lost their activities when mutated at several key positions (Lys6, Lys37, Arg62, or Tyr74), whereas the sequence specific HMG domain could tolerate mutations at these positions (*27*, *41*−*43*, *58*, *64*). Since sequence specific HMG domains involve several residues, a single mutation did not hinder DNA binding. Sequence specific HMG domains use both the long and short arms of the L-shaped protein, whereas non-sequence specific HMG domains use only the short arm to interact with DNA (Figure 8). In other words, the number of interactions at the protein−DNA interface of the sequence-recognizing HMG domain is almost twice that in its non-sequence specific counterpart (Table 2). The numbers of heavy atoms within 3.5 Å of sequence specific and non-sequence specific HMG domains are 46−145 and 22−28, respectively (*65*). The energy released by electrostatic and van der Waals contacts helps to satisfy the large energy requirement for DNA interaction (*63*).

In their free forms, non-sequence specific HMG domains are more stable than sequence specific HMG domains. The thermal denaturation of the non-sequence specific HMGB1 protein, a plot of mean residue ellipticity *versus* temperature, followed a flat line at low temperatures, whereas that of the sequence specific zfSOX9 HMG domain followed a sloping line at lower temperatures (corresponding to a lower stability) because of the flexibility of its long arm (*50*, *66*). HMG domains attain a stable, conformational state when complexed with DNA (Figure 4B). The stabilities of the sequence specific zfSOX9 HMG domain in free and complex forms are very different; the T_m values are 50 and 70 °C, respectively. The difference between the unstable higher-energy state and the complexed low-energy state could be utilized for DNA interactions.

Theoretical Stability and Biological Role of HMG Domains. The thermodynamic stability of a protein may be quantified by its susceptibility to proteolytic cleavage (*67*−*69*). The proteolytic cleavage of proteins *in vivo* is theoretically calculated as the instability index (*70*). A larger instability index calculated from the primary sequence implies a less stable protein. Although both forms of HMG proteins include several conserved residues, the average instability indices have contrasting values, 67 and 30, for sequence and non-sequence specific HMG domains, respectively. The lower stability of sequence specific HMG domains has important biological significance. The half-life of proteins governs their functions *in vivo* (*71*). A larger instability index (shorter half-life *in vivo*) of a sequence specific HMG domain implies that its function is controlled during the developmental process. Non-sequence specific HMG proteins in cells perform "housekeeping", like protecting the B-form DNA and identifying deformed DNA, such as four-way junction, cisplatin-modified DNA or DNA bulge, to shield, repair or trigger antitumor activity (*53*, *72*) However, proteins such as SOX9 are active only during a cell's developmental process (*8*, *54*, *57*, *73*). Although the sequence specific HMG domain can also bind with a deformed portion of DNA, like the non-sequence specific HMG domains, the shorter lifetime (half-life *in vivo*) of

the sequence specific HMG domain controls its function (*74*−*76*).

The highly basic sequence specific HMG domains (pI ~10.38) interact with DNA with a molecular wedge mechanism (*25*, *59*). Positively charged amino acids neutralized negative charges at the surface of the DNA (*44*). The hydrophobic core of the protein is then opened to intercalate the DNA (*25*). Mutation analysis along with the differences between the pI values, the instability indices, and DNA contact regions of the sequence and non-sequence specific HMG domains rationalize their binding modes.

ACKNOWLEDGMENT

We are grateful to Dr. Hanna S. Yuan and Dr. Bon-chu Chung, Institute of Molecular Biology, Academia Sinica R.O.C., for their generous gifts of construct and helpful discussions.

REFERENCES

1. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. (1994) *Molecular Biology of the Cell*, 3rd ed, Garland: New York.
2. Harrison, S. C. (1991) A structural taxonomy of DNA-binding domains, *Nature 353*, 715−719.
3. Travers, A. A. (1993) *DNA-Protein Interactions*, 1st ed., Chapman & Hall, New York.
4. Wegner, M. (1999) From head to toes: the multiple facets of Sox proteins, *Nucleic Acids Res. 27*, 1409−1420.
5. Grosschedl, R., Giese, K., and Pagel, J. (1994) HMG domain proteins: architectural elements in the assembly of nucleoprotein structures, *Trends Genet. 10*, 94−100.
6. Berta, P., Hawkins, J. R., Sinclair, A. H., Taylor, A., Griffiths, B. L., Goodfellow, P. N., and Fellous, M. (1990) Genetic evidence equating SRY and the testis-determining factor, *Nature 348*, 448−450.
7. Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1991) Male development of chromosomally female mice transgenic for Sry, *Nature 351*, 117−121.
8. Pevny, L. H., and Lovell-Badge, R. (1997) Sox genes find their feet, *Curr. Opin. Genet. Dev. 7*, 338−344.
9. Kent, J., Wheatley, S. C., Andrews, J. E., Sinclair, A. H., and Koopman, P. (1996) A male-specific role for SOX9 in vertebrate sex determination, *Development 122*, 2813−2822.
10. Morais da Silva, S., Hacker, A., Harley, V., Goodfellow, P., Swain, A., and Lovell-Badge, R. (1996) Sox9 expression during gonadal development implies a conserved role for the gene in testis differentiation in mammals and birds, *Nat. Genet. 14*, 62−68.
11. Bell, D. M., Leung, K. K., Wheatley, S. C., Ng, L. J., Zhou, S., Ling, K. W., Sham, M. H., Koopman, P., Tam, P. P., and Cheah, K. S. (1997) SOX9 directly regulates the type-II collagen gene, *Nat. Genet. 16*, 174−178.
12. Lefebvre, V., Huang, W., Harley, V. R., Goodfellow, P. N., and de Crombrugghe, B. (1997) SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene, *Mol. Cell. Biol. 17*, 2336−2346.
13. Ng, L. J., Wheatley, S., Muscat, G. E., Conway-Campbell, J., Bowles, J., Wright, E., Bell, D. M., Tam, P. P., Cheah, K. S., and Koopman, P. (1997) SOX9 binds DNA, activates transcription, and coexpresses with type II collagen during chondrogenesis in the mouse, *Dev. Biol. 183*, 108−121.
14. Lefebvre, V., Li, P., and de Crombrugghe, B. (1998) A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene, *EMBO J. 17*, 5718−5733.
15. Jimenez, R., and Burgos, M. (1998) Mammalian sex determination: joining pieces of the genetic puzzle, *BioEssays 20*, 696−699.
16. Cameron, F. J., and Sinclair, A. H. (1997) Mutations in SRY and SOX9: testis-determining genes, *Hum. Mutat. 9*, 388−395.
17. Tajima, T., Nakae, J., Shinohara, N., and Fujieda, K. (1994) A novel mutation localized in the 3′ non-HMG box region of the SRY gene in 46,XY gonadal dysgenesis, *Hum. Mol. Genet. 3*, 1187−1189.
18. Laudet, V., Stehelin, D., and Clevers, H. (1993) Ancestry and diversity of the HMG box superfamily, *Nucleic Acids Res. 21*, 2493−2501.
19. Bianchi, M. E. (1995) *DNA-protein: Structural Interactions*, Oxford University Press: New York.
20. Bustin, M. (2001) Revised nomenclature for high mobility group (HMG) chromosomal proteins, *Trends Biochem. Sci. 26*, 152−153.
21. Hill, D. A., and Reeves, R. (1997) Competition between HMG-I(Y), HMG-1 and histone H1 on four-way junction DNA, *Nucleic Acids Res. 25*, 3523−3531.
22. Bianchi, M. E., Beltrame, M., and Paonessa, G. (1989) Specific recognition of cruciform DNA by nuclear protein HMG1, *Science 243*, 1056−1059.
23. Pil, P. M., and Lippard, S. J. (1992) Specific binding of chromosomal protein HMG1 to DNA damaged by the anticancer drug cisplatin, *Science 256*, 234−237.
24. Ferrari, S., Harley, V. R., Pontiggia, A., Goodfellow, P. N., Lovell-Badge, R., and Bianchi, M. E. (1992) SRY, like HMG1, recognizes sharp angles in DNA, *EMBO J. 11*, 4497−4506.
25. Thomas, J. O., and Travers, A. A. (2001) HMG1 and 2, and related "architectural" DNA-binding proteins, *Trends Biochem. Sci. 26*, 167−174.
26. Mertin, S., McDowall, S. G., and Harley, V. R. (1999) The DNA-binding specificity of SOX9 and other SOX proteins, *Nucleic Acids Res. 27*, 1359−1364.
27. Saito, K., Kikuchi, T., and Yoshida, M. (1999) The mechanism of sequence non-specific DNA binding of HMG1/2-box B in HMG1 with DNA, *Protein Eng. 12*, 235−242.
28. Cary, P. D., Read, C. M., Davis, B., Driscoll, P. C., and Crane-Robinson, C. (2001) Solution structure and backbone dynamics of the DNA-binding domain of mouse Sox-5, *Protein Sci. 10*, 83−98.
29. Morris, A. L., MacArthur, M. W., Hutchinson, E. G., and Thornton, J. M. (1992) Stereochemical quality of protein structure coordinates, *Proteins 12*, 345−364.
30. Pace, C. N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995) How to measure and predict the molar absorption coefficient of a protein, *Protein Sci. 4*, 2411−2423.
31. Iakoucheva, L. M., Walker, R. K., van Houten, B., and Ackerman, E. J. (2002) Equilibrium and stop-flow kinetic studies of fluorescently labeled DNA substrates with DNA repair proteins XPA and replication protein A, *Biochemistry 41*, 131−143.
32. Zaitseva, E. M., Zaitsev, E. N., and Kowalczykowski, S. C. (1999) The DNA binding properties of *Saccharomyces cerevisiae* Rad51 protein, *J. Biol. Chem. 274*, 2907−2915.
33. Grosshans, D. R., Clayton, D. A., Coultrap, S. J., and Browning, M. D. (2002) LTP leads to rapid surface expression of NMDA but not AMPA receptors in adult rat CA1, *Nat. Neurosci. 5*, 27−33.
34. Peretz, D., Williamson, R. A., Kaneko, K., Vergara, J., Leclerc, E., Schmitt-Ulms, G., Mehlhorn, I. R., Legname, G., Wormald, M. R., Rudd, P. M., Dwek, R. A., Burton, D. R., and Prusiner, S. B. (2001) Antibodies inhibit prion propagation and clear cell cultures of prion infectivity, *Nature 412*, 739−743.
35. Haggiag, S., Chebath, J., and Revel, M. (1999) Induction of myelin gene expression in Schwann cell cultures by an interleukin-6 receptor-interleukin-6 chimera, *FEBS Lett. 457*, 200−204.
36. Marone, M., Mozzetti, S., De Ritis, D., Pierelli, L., and Scambia, G. (2001) Semiquantitative RT-PCR analysis to assess the expression levels of multiple transcripts from the same sample, *Biol. Proced. Online 3*, 19−25.
37. Morrison, D., Woodford, N., Barrett, S. P., Sisson, P., and Cookson, B. D. (1999) DNA banding pattern polymorphism in vancomycin-resistant *Enterococcus faecium* and criteria for defining strains, *J. Clin. Microbiol. 37*, 1084−1091.
38. van Houte, L. P., Chuprina, V. P., van der Wetering, M., Boelens, R., Kaptein, R., and Clevers, H. (1995) Solution structure of the sequence-specific HMG box of the lymphocyte transcriptional activator Sox-4, *J. Biol. Chem. 270*, 30516−30524.
39. Werner, M. H., Huth, J. R., Gronenborn, A. M., and Clore, G. M. (1995) 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.
40. Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R., and Wright, P. E. (1995) Structural basis for DNA bending by the architectural transcription factor LEF-1, *Nature 376*, 791−795.

41. Taudte, S., Xin, H., and Kallenbach, N. R. (2000) Alanine mutagenesis of high-mobility-group-protein-1 box B (HMG1-B), *Biochem. J. 347* (Part 3), 807−814.
42. Allain, F. H., Yen, Y. M., Masse, J. E., Schultze, P., Dieckmann, T., Johnson, R. C., and Feigon, J. (1999) Solution structure of the HMG protein NHP6A and its interaction with DNA reveals the structural determinants for non-sequence-specific binding, *EMBO J. 18*, 2563−2579.
43. Stros, M., and Muselikova, E. (2000) A role of basic residues and the putative intercalating phenylalanine of the HMG-1 box B in DNA supercoiling and binding to four-way DNA junctions, *J. Biol. Chem. 275*, 35699−35707.
44. Lebrun, A., Shakked, Z., and Lavery, R. (1997) Local DNA stretching mimics the distortion caused by the TATA box-binding protein, *Proc. Natl. Acad. Sci. U.S.A. 94*, 2993−2998.
45. Payet, D., Hillisch, A., Lowe, N., Diekmann, S., and Travers, A. (1999) The recognition of distorted DNA structures by HMG-D: a footprinting and molecular modelling study, *J. Mol. Biol. 294*, 79−91.
46. He, Q., Ohndorf, U. M., and Lippard, S. J. (2000) Intercalating residues determine the mode of HMG1 domains A and B binding to cisplatin-modified DNA, *Biochemistry 39*, 14426−14435.
47. Weiss, M. A., Ukiyama, E., and King, C. Y. (1997) The SRY cantilever motif discriminates between sequence- and structure-specific DNA recognition: alanine mutagenesis of an HMG box, *J. Biomol. Struct. Dyn. 15*, 177−184.
48. McDowall, S., Argentaro, A., Ranganathan, S., Weller, P., Mertin, S., Mansour, S., Tolmie, J., and Harley, V. (1999) Functional and structural studies of wild-type SOX9 and mutations causing campomelic dysplasia, *J. Biol. Chem. 274*, 24023−24030.
49. Preiss, S., Argentaro, A., Clayton, A., John, A., Jans, D. A., Ogata, T., Nagai, T., Barroso, I., Schafer, A. J., and Harley, V. R. (2001) Compound effects of point mutations causing campomelic dysplasia/autosomal sex reversal upon SOX9 structure, nuclear transport, DNA binding, and transcriptional activation, *J. Biol. Chem. 276*, 27864−27872.
50. Crane-Robinson, C., Read, C. M., Cary, P. D., Driscoll, P. C., Dragan, A. I., and Privalov, P. L. (1998) The energetics of HMG box interactions with DNA. Thermodynamic description of the box from mouse Sox-5, *J. Mol. Biol. 281*, 705−717.
51. Ohndorf, U. M., Rould, M. A., He, Q., Pabo, C. O., and Lippard, S. J. (1999) Basis for recognition of cisplatin-modified DNA by high-mobility-group proteins, *Nature 399*, 708−712.
52. Bianchi, M. E., and Beltrame, M. (1998) Flexing DNA: HMG-box proteins and their partners, *Am. J. Hum. Genet. 63*, 1573−1577.
53. Imamura, T., Izumi, H., Nagatani, G., Ise, T., Nomoto, M., Iwamoto, Y., and Kohno, K. (2001) Interaction with p53 enhances binding of cisplatin-modified DNA by high mobility group 1 protein, *J. Biol. Chem. 276*, 7534−7540.
54. Marshall, O. J., and Harley, V. R. (2000) Molecular mechanisms of SOX9 action, *Mol. Genet. Metab. 71*, 455−462.
55. Dailey, L., and Basilico, C. (2001) Coevolution of HMG domains and homeodomains and the generation of transcriptional regulation by Sox/POU complexes, *J. Cell. Physiol. 186*, 315−328.
56. Courey, A. J. (2001) Cooperativity in transcriptional control, *Curr. Biol. 11*, R250−R252.
57. Koopman, P. (1999) Sry and Sox9: mammalian testis-determining genes, *Cell. Mol. Life Sci. 55*, 839−856.
58. Hu, C. H., Wang, J. M., and Tseng, H. B. (1998) The first high-mobility-group box of upstream binding factor assembles across-over DNA junction by basic residues, *Biochem. J. 333* (Part 1), 51−56.
59. Murphy, E. C., Zhurkin, V. B., Louis, J. M., Cornilescu, G., and Clore, G. M. (2001) Structural basis for SRY-dependent 46-X,Y sex reversal: modulation of DNA bending by a naturally occurring point mutation, *J. Mol. Biol. 312*, 481−499.
60. Tang, Y., and Nilsson, L. (1999) Effect of G40R mutation on the binding of human SRY protein to DNA: a molecular dynamics view, *Proteins 35*, 101−113.
61. Sassoon, J., Lilie, H., Baumann, U., and Kohli, J. (2001) Biochemical characterization of the structure-specific DNA-binding protein Cmb1 from *Schizosaccharomyces pombe*, *J. Mol. Biol. 309*, 1101−1115.
62. Murphy, F. V. t., Sweet, R. M., and Churchill, M. E. (1999) The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition, *EMBO J. 18*, 6610−6618.
63. Jen-Jacobson, L., Engler, L. E., and Jacobson, L. A. (2000) Structural and thermodynamic strategies for site-specific DNA binding proteins, *Struct. Folding Des. 8*, 1015−1023.
64. Stros, M. (2001) Two mutations of basic residues within the N-terminus of HMG-1 B domain with different effects on DNA supercoiling and binding to bent DNA, *Biochemistry 40*, 4769−4779.
65. Luscombe, N. M., Laskowski, R. A., and Thornton, J. M. (2001) Amino acid−base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level, *Nucleic Acids Res. 29*, 2860−2874.
66. Stemmer, C., Schwander, A., Bauw, G., Fojan, P., and Grasser, K. D. (2002) Protein kinase CK2 differentially phosphorylates maize chromosomal high mobility group B (HMGB) proteins modulating their stability and DNA interactions, *J. Biol. Chem. 277*, 1092−1098.
67. Ellison, D., Hinton, J., Hubbard, S. J., and Beynon, R. J. (1995) Limited proteolysis of native proteins: the interaction between avidin and proteinase K, *Protein Sci. 4*, 1337−1345.
68. Wang, L., and Kallenbach, N. R. (1998) Proteolysis as a measure of the free energy difference between cytochrome *c* and its derivatives, *Protein Sci. 7*, 2460−2464.
69. Hobart, S. A., Meinhold, D. W., Osuna, R., and Colon, W. (2002) From two-state to three-state: the effect of the P61A mutation on the dynamics and stability of the factor for inversion stimulation results in an altered equilibrium denaturation mechanism, *Biochemistry 41*, 13744−13754.
70. Guruprasad, K., Reddy, B. V., and Pandit, M. W. (1990) Correlation between stability of a protein and its dipeptide composition: a novel approach for predicting in vivo stability of a protein from its primary sequence, *Protein Eng. 4*, 155−161.
71. Vanni, C., Mancini, P., Gao, Y., Ottaviano, C., Guo, F., Salani, B., Torrisi, M. R., Zheng, Y., and Eva, A. (2002) Regulation of proto-Dbl by intracellular membrane targeting and protein stability, *J. Biol. Chem. 277*, 19745−19753.
72. Pasheva, E. A., Ugrinova, I., Spassovska, N. C., and Pashev, I. G. (2002) The binding affinity of HMG1 protein to DNA modified by cis-platin and its analogs correlates with their antitumor activity, *Int. J. Biochem. Cell Biol. 34*, 87−92.
73. Bowles, J., Schepers, G., and Koopman, P. (2000) Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators, *Dev. Biol. 227*, 239−255.
74. Soullier, S., Jay, P., Poulat, F., Vanacker, J. M., Berta, P., and Laudet, V. (1999) Diversification pattern of the HMG and SOX family members during evolution, *J. Mol. Evol. 48*, 517−527.
75. Nagai, K. (2001) Molecular evolution of Sry and Sox gene, *Gene 270*, 161−169.
76. Trimmer, E. E., Zamble, D. B., Lippard, S. J., and Essigmann, J. M. (1998) Human testis-determining factor SRY binds to the major DNA adduct of cisplatin and a putative target sequence with comparable affinities, *Biochemistry 37*, 352−362.
77. Yen, Y. M., Wong, B., and Johnson, R. C. (1998) Determinants of DNA binding and bending by the *Saccharomyces cerevisiae* high mobility group protein NHP6A that are important for its biological activities. Role of the unique N terminus and putative intercalating methionine, *J. Biol. Chem. 273*, 4424−4435.
78. Peters, R., King, C. Y., Ukiyama, E., Falsafi, S., Donahoe, P. K., and Weiss, M. A. (1995) 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.

BI034678D