Alanine mutagenesis of high-mobility-group-protein-1 box B (HMG1-B)

Susann TAUDTE, Hong XIN and Neville R. KALLENBACH¹

Department of Chemistry, New York University, 31 Washington Place, New York, NY 10003, U.S.A.

We have generated a set of alanine-scanning substitutions in high-mobility-group protein 1 box B (HMG1-B; the second domain of the HMG1 nuclear protein from the rat) in order to explore the influence of specific surface side chains on its function and folding. Guanidine hydrochloride and thermal unfolding studies have been carried out to investigate the effect of substituted residues on the folding pathway. Binding to four-way junction and linear-duplex DNA has been assayed to determine which residues play an important role in DNA binding. We have identified several mutants that are more stable or bind more

tightly to the junction than the wild-type, including the particular phenylalanine side chain that is thought to intercalate into the DNA. Thus the interaction between HMG1-B and branched DNA substrates should exhibit differences from present models based on the structure of the complexes that have been solved to date.

Key words: alanine scan, four-way junction, protein–DNA interaction, protein stability, structure-specific recognition.

INTRODUCTION

A variety of DNA-binding proteins in the nucleus have been identified which participate in maintaining the structure of chromatin and which mediate gene expression, replication, recombination and repair [1]. The two most abundant classes of chromosomal proteins in the nucleus are the histones and the high-mobility-group (HMG) proteins. Studies on the structure and function of chromatin have defined the role of core histones in forming nucleosomes and the role of linker histone on higher structure formation and transcription [2–6]. The relatively small (< 30 kDa) high-mobility-group proteins were initially identified by their high electrophoretic mobility in polyacrylamide gels and solubility in 2–5% trichloroacetic acid. They have been classified into families based on their size, sequence similarities and DNA-binding properties: HMG1/2, HMG14/17 and HMG1/Y. All of these interact with DNA and other proteins to regulate gene expression and structure in chromatin [1].

The HMG1/2 proteins attracted attention with the identification of an increasing number of proteins containing one or more copies of a DNA-binding-sequence motif called an ‘HMG box’ [7,8]. The HMG box motif is defined as a set of approx. 80 residues that contains conserved prolines, aromatic and basic amino acids. HMG box sequences have been identified in relatively abundant proteins that bind DNA with little base-sequence specificity (HMG1 and HMG2), as well as in sequence-specific transcription factors (lymphoid enhancer-binding factor1, LEF1; T-cell factor; and sex-determining region of the human Y chromosome; SRY) [9].

The HMG1 and HMG2 nuclear proteins share an ability to bind DNA ‘structure specifically’, recognizing DNA molecules containing bends or kinks in the double helix [10] as well as four-

way junctions (4WJs) [11,12]. Sequence-specific HMG box proteins recognize AT-rich sequences that tend to deform the DNA helix axis. Additionally, HMG1 has been demonstrated to bend DNA [13] and is thought to be active in DNA repair, recombination, as well as assembly of DNA into chromatin fibres [14]. Despite a variety of promising clues, no single function has been identified unequivocally [15].

HMG1 is composed of two tandem HMG ‘boxes’ (designated A and B) and an acidic C-terminal tail. The N-terminal A-domain (residues 1–79) and central B-domain (residues 90–163) are both basic, pI > 10, and approx. 43% similar to one another in sequence. The highly acidic C-terminal tail contains a number of consecutive aspartic or glutamic acid residues that are thought to mediate protein–protein interactions. The structure of the B-domain has been determined by two-dimensional ¹H-NMR spectroscopy [16]. It consists of three α -helices which together account for approx. 70% of the sequence, including most of the conserved basic and aromatic residues. The structure has an unusual flattened L-shape and consists of two arms forming an angle of approx. 80°. The shorter arm is formed by helices I and II, while the longer arm consists of an extended proline-rich N-terminal region packed against helix III. The relative positions of the two arms are maintained by a cluster of conserved, predominantly aromatic, residues. The structure of the A-domain of HMG1 is very similar in overall folding to that of the B-domain [17], the most obvious difference being the orientation of helix I and the angle between helices II and III, which is smaller in the A-domain than in the B-domain (by approx. 14°).

Early experiments indicated that isolated A and B domains retain the DNA-binding affinity and structure selectivity of the entire protein when expressed independently [12]. More recent

Abbreviations used: HMG1-B, high-mobility-group-protein-1 box B; DTT, dithiothreitol; 4WJ, four-way junction; EMSA, electrophoretic mobility-shift assay; LEF1, lymphoid enhancer-binding factor 1; SRY, sex-determining region of the human Y chromosome; NHP6A, non-histone protein 6A; GdnHCl, guanidine hydrochloride; [GdnHCl]_{1/2}, transition midpoint; ΔG_u , free energy of unfolding; m, change in free energy of unfolding; T_m, melting temperature; ΔH_m , enthalpy change; ΔS_m , entropy change; ΔC_p , heat capacity change.

¹ To whom correspondence should be addressed, at Department of Chemistry, New York University, 31 Washington Place, New York, NY 10003, U.S.A. (e-mail nrk1@is.nyu.edu).

evidence suggests that the two domains have slightly different binding properties; in particular, they differ in their ability to discriminate between 4WJs and supercoiled DNA [9]. Binding by the A-domain is insensitive to competition from supercoiled DNA, whereas binding by the B-domain is not.

Structures of protein–DNA complexes of two sequence-specific HMG proteins, LEF1 and SRY, bound to their cognate DNA-recognition sequences have been published [18,19]. These studies show that the proteins recognize a bend in their cognate DNA induced by intercalation of a bulky side chain of the protein between base pairs of the duplex [20,21]. This intercalation unstacks the DNA bases and unwinds the DNA duplex, leading to a widened minor groove with a severely compressed major groove. Binding occurs via interaction of the concave surface of the protein with the outside of the bent target sequence. The complexes reveal several expected interactions between side chains and the backbone of the proteins with the DNA, including basic side chains bound with phosphates of the backbone.

Recently, the structure of a non-sequence-specific HMG-binding protein (non-histone protein 6A, NHP6A) and its interaction with DNA have been solved [22]. This structure reveals interesting differences from those of DNA–LEF1 and DNA–SRY complexes. For example, while it was possible to model a fit of the DNA structure in the LEF1 complex against the NHP6A surface, that from the SRY complex does not fit. Despite obvious complementarity between the protein and DNA surfaces, there is also little possibility for hydrogen bonding to bases in the NHP6A–DNA model. Instead, the NHP6A-binding site is largely hydrophobic, with a number of flanking basic residues interacting with the DNA phosphodiester backbone. Methionine-29 (corresponding to isoleucine-68 in SRY) is thought to intercalate into DNA, but at a site shifted by 1 bp and in reverse orientation to that found in the SRY–DNA complex.

Given the abundance of HMG-box domains and our present imperfect understanding of the determinants of their interaction with DNA, it seemed useful to undertake a mutational analysis of HMG box proteins. The replacement of particular amino acids in a protein by alanine using site-directed mutagenesis allows one to define the role of these amino acids, both in the folding process and in the binding selectivity [23–25]. Important questions are how individual side chains contribute to folding and stability, and whether or not the stability of a mutant protein is coupled to its binding activity. We describe here a partial alanine scan of the second box in HMG1 (HMG1-B) from the rat [26], which identifies key residues for folding and DNA-junction binding in the protein. Surprisingly, the phenylalanine side chain thought to insert directly into the DNA by sequence similarity arguments is insensitive to alanine replacement, suggesting that HMG1-B is not identical with SRY or LEF in its DNA interaction.

MATERIALS AND METHODS

Mutagenesis and protein expression

Alanine mutants of HMG1-B were generated by site-directed mutagenesis using the method of Kunkel et al. [27], and proteins were expressed from pHB1/*Escherichia coli* BL21(DE3)pLysS [28]. Crude proteins were purified by FPLC using an Econo-Pac CM cartridge (Bio-Rad). Proteins were loaded using a 50 mM NaCl/50 mM Tris/HCl buffer (pH 7.0) and eluted in a highly purified form by increasing the NaCl concentration to 500 mM utilizing a gradient program.

Determination of protein concentrations

Protein concentrations were determined according to Pace et al. [29].

CD measurements

CD spectra were obtained using an AVIV 60DS Circular Dichroism Spectrometer (Aviv Associates, Lakewood, NJ, U.S.A.). CD spectra of 2 μ M and 3 μ M protein solutions in a 10 mM potassium phosphate buffer (pH 7.0), containing 0.5 mM dithiothreitol (DTT) and 0.5 mM EDTA, were recorded from 260 nm to 200 nm. Thermal unfolding measurements were performed at a wavelength of 222 nm over the range of 4–94 °C.

Fluorescence measurements

Guanidine hydrochloride (GdnHCl)-induced unfolding of wild-type HMG-B and its mutant proteins was followed using an Automated Temperature/Titration Filter Fluorimeter (ATFF-212; Aviv Associates), using an excitation wavelength of 280 nm and an emission wavelength of 340 nm. Samples of 5 μ M protein solutions were titrated with a 4.8 M GdnHCl stock solution at 25 °C or 5 °C. Both solutions were buffered in 10 mM potassium phosphate (pH 7.0), containing 0.5 mM DTT and 0.5 mM EDTA. The GdnHCl concentration was increased from 0 M to 4 M in 0.1 M increments, keeping the sample volume and protein concentration constant. A 30 s stirring time after each titration step ensured complete equilibration of the sample.

Emission spectra of HMG1-B were recorded in the same buffer as the GdnHCl unfolding curves, using a Perkin–Elmer Luminescence Spectrometer LS 50 over a range from 300–420 nm.

4WJ formation

Junction-binding experiments were performed on J1, a junction with four arms each of 8 bp [30]. Oligonucleotides were obtained in HPLC-purified form from CyberSyn, Lenni, PA, U.S.A. One oligonucleotide strand (101) was radiolabelled at its 5'-termini using [γ -³²P]ATP and T4 polynucleotide kinase (Promega). The strands were annealed by combining the radiolabelled strand with a 5-fold excess of the unlabelled strands in 50 mM Tris/HCl (pH 7.5)/10 mM MgCl₂. The mixture was incubated for 2 min at 90 °C, followed by cooling to 4 °C overnight.

Duplexes were generated using the 16 nt 101 strand together with its complementary 16 nt strand, 101c.

Electrophoretic mobility-shift assay (EMSA)

EMSA of the complexes of 4WJ J1 and duplex 101 respectively, with HMG1-B and its mutant proteins, was performed as a function of DNA/protein molar ratio. DNA (15.4 nM in the final reaction) and aliquots of HMG1-B or mutant protein were incubated in 10 or 20 μ l of a reaction solution [10 mM Tris/HCl (pH 8.0), 100 mM NaCl, 12 mM MgCl₂, 10 % glycerol, 3.33 mM DTT, 100 μ g/ml BSA and 1 μ g/ml sodium poly(dI-dC)] on ice for 30 min. The complexes were subjected to electrophoresis on 12 % polyacrylamide gels in 20 mM Tris/HCl (pH 8.1), containing 10 mM acetic acid and 1 mM MgCl₂.

RESULTS

We based our choice of sites for mutation on structural data from the B domain of HMG1 [16] and two complexes of HMG

Figure 1 B-domain of rat HMG1 protein and the mutants of this study

(A) Sequence of the B-domain of rat HMG1. (B) Schematic representation of HMG-B showing the position of the three α -helices. (C) Schematic representation of mutant proteins and indication of the nucleic acid change and its position. (D) Structure of the B domain of rat HMG1 [16] displaying the positions and side chains of mutational sites.

box proteins with DNA [18,19]. A series of surface residues in the protein has been replaced by alanine to explore the influence of side chains at key positions. Amino acids substituted include polar uncharged side chains (cysteine, serine and glycine), basic (lysine and arginine), acidic (glutamic acid), as well as non-polar hydrophobic side chains (phenylalanine, proline, leucine and isoleucine; Figure 1). We avoided substitutions of internal non-polar residues that comprise the primitive hydrophobic core of HMG1-B, since these would be likely to lead to a significant decrease in protein stability.

CD spectroscopy and secondary-structure analysis

CD spectra of wild-type HMG1-B and mutant proteins show minima at 208 nm and 222 nm, indicative of a structure with significant α -helical content. In no instance is there evidence of major reorganization of the overall fold. However, there are quantitative differences (Figure 2). The values at 222 nm were used to estimate the overall α -helical content of wild-type HMG1-B and mutant proteins (Table 1). While most mutant proteins possess only a slightly different α -helical content compared with

Figure 2 CD spectra of HMG1-B and selected mutant proteins

CD spectra of HMG1-B (◆), S16A (■), R26A (▲), and I38A (×) were recorded (as described in the Materials and methods section) illustrating the difference in their α -helical content.

Table 1 Secondary-structure analysis of HMG1-B and its mutants

Recorded CD spectra have minima at 208 nm and 222 nm, indicative of a significant α -helical content. The values at 222 nm were used to estimate the α -helical content.

Mutant	$[\theta]$ ($^{\circ}$ cm 2 dmol $^{-1}$) at 222 nm	% α -Helicity	Δ % α -Helicity
HMG1-B	−15 500	70.4	0
K12A	−16 000	72.7	2.3
R13A	−21 400	97.2	26.8
P14A	−15 000	68.1	−2.3
P15A	−16 000	72.7	2.3
S16A	−10 700	48.6	−21.8
F18A	−11 900	54.0	−16.4
F19A	−15 500	70.4	0
C22A	−14 200	64.5	−5.9
S23A	−14 300	64.9	−5.5
R26A	−15 700	71.3	0.9
K30A	−15 300	69.5	−0.9
L36A	−16 100	73.1	2.7
S37A	−14 700	66.8	−3.6
I38A	−15 700	71.3	0.9
G39A	−17 600	79.9	9.5
K43A	−18 200	82.7	12.3
G46A	−12 200	55.4	−15.0
E61A	−15 400	69.9	−0.5
K68A	−14 400	65.4	−5
K81A	−15 800	71.8	1.4

wild-type ($\pm 5\%$), two mutations, S16A and F18A, which are both located at the beginning of helix I, show a significant decrease in α -helicity. On the other hand, R13A located in the N-terminal tail shows a slight increase in its α -helical content.

Denaturation studies

GdnHCl-induced unfolding

GdnHCl denaturation experiments were used to determine transition midpoints, the free energy of unfolding (ΔG_u) at 25 °C

Figure 3 GdnHCl denaturation curves monitored by fluorescence

Shown here are the unfolding curves of HMG-B (◆), C22A (■), and G46A (▲) at various concentrations of GdnHCl buffered in 10 mM potassium phosphate buffer (pH 7.0), containing 0.5 mM DTT and 0.5 mM EDTA, at 25 °C. The fraction unfolded was calculated as $f_u = (y_n - y)/(y_n - y_u)$, where y is the observed fluorescence intensity of the protein, and y_n and y_u represent linear functions of denaturant concentration for the native and the unfolded states of the protein respectively.

and/or 5 °C, and the change in free energy of unfolding with GdnHCl concentration (m). Unfolding of the wild-type protein and the mutants was followed by fluorescence (Figure 3). The fluorescence signal from HMG1-B is dominated by tryptophan, the most highly conserved residue in HMG box sequences. During unfolding the tryptophan that lies in the vertex of the arrowhead-shaped fold is exposed to the solvent, leading to an increase in fluorescence signal and shift in emission maximum (Figure 4).

The unfolding of many small proteins has been found to

Figure 4 Dependence of the tryptophan emission maximum on the GdnHCl concentration

Fluorescence emission spectra of HMG1-B were recorded at 280 nm in the presence of various GdnHCl concentrations corresponding to the native, partly denatured, and denatured states of the protein. The buried tryptophan in the core of the fold is exposed to the solvent during unfolding, leading to an increase in quantum yield and shift in emission maximum to higher wavelength (redshift). Curves from top to bottom: 4 M GdnHCl, 2 M GdnHCl, 1.5 M GdnHCl, 1 M GdnHCl and 0 M GdnHCl.

Table 2 Results from the analysis of GdnHCl unfolding curves for wild-type HMG1-B and mutant proteins

GdnHCl unfolding at 25 °C shown in regular font, GdnHCl unfolding at 5 °C shown in **bold**. A, midpoint of the GdnHCl unfolding curve; B, difference between the $[\text{GdnHCl}]_{1/2}$; C, $\Delta G = \Delta G(\text{H}_2\text{O}) - m[\text{GdnHCl}]$; D, $\Delta G = \Delta G(\text{H}_2\text{O}) - m[\text{GdnHCl}]_{1/2}$; E, difference between the $\Delta G(\text{H}_2\text{O})$.

	A [GdnHCl] _{1/2} (M)	B $\Delta[\text{GdnHCl}]_{1/2}$ (M)	C m (kcal · mol ⁻¹ · M ⁻¹)	D $\Delta G(\text{H}_2\text{O})$ (kcal/mol)	E $\Delta\Delta G(\text{H}_2\text{O})$ (kcal/mol)
HMG-B	2.093	0	2.13	4.46	0
HMG-B	1.511	0	2.47	3.74	0
K12A	1.445	−0.066	2.23	3.23	0.51
R13A	1.332	−0.179	1.98	2.64	1.1
P14A	1.599	−0.494	1.65	2.64	1.82
P15A	1.608	−0.485	1.77	2.84	1.62
S16A	1.247	−0.846	1.73	2.16	2.3
F18A	0.938	−1.155	3.59	3.32	1.14
F19A	1.582	0.071	2.17	3.43	0.31
C22A	1.19	−0.321	2.92	3.47	0.27
S23A	1.312	−0.199	2.09	2.74	1
R26A	1.484	−0.027	1.72	2.56	1.18
K30A	1.485	−0.026	2.38	3.54	0.2
L36A	1.276	−0.235	2.25	2.86	0.88
S37A	1.688	−0.405	2.16	3.64	0.82
I38A	1.631	0.12	2.11	3.4	0.34
G39A	1.728	0.217	2.46	4.24	−0.5
K43A	1.452	−0.059	2.62	3.81	−0.07
G46A	1.741	0.23	1.76	3.07	0.67
E61A	1.241	−0.27	2.49	3.09	0.65
K68A	1.573	0.062	2.07	3.26	0.48
K81A	1.463	−0.048	2.5	3.66	0.08

approximate a two-state process, with small populations of intermediate states. We assume that wild-type HMG1-B and its mutant proteins unfold thermodynamically according to a two-state mechanism. Values of $\Delta G_{\text{H}_2\text{O}}$, m, and $[\text{GdnHCl}]_{1/2}$ transition midpoint [31] were calculated for all proteins (Table 2). Wild-type HMG1-B is half unfolded at a GdnHCl concentration of 1.5 M. The mutant proteins vary in their GdnHCl transition midpoints. The free energy of unfolding in water of wild-type HMG1-B is 3.74 kcal · mol⁻¹ at 25 °C and 4.46 kcal · mol⁻¹ at 5 °C (1 cal = 4.184 J). Unfolding of most of the mutant proteins was determined at 25 °C, but several unfolding curves were monitored at 5 °C due to low stability of certain proteins.

Figure 5 Binding of HMG1-B, E61A, K68A and K81A to 4WJ J1 and duplex 101

(A) A fixed concentration of 5'-³²P-labelled J1 was incubated with increasing concentrations of HMG1-B and selected mutant proteins. Complexes were analysed by EMSA as described in the Materials and methods section. (B) A fixed concentration of 5'-radiolabelled duplex 101 was incubated with increasing concentrations of HMG1-B and selected mutant proteins. EMSA of 10 μl reaction solutions were performed. Lane 1, single strand 101; lane 2a, 4WJ J1 only; lane 2b, duplex 101 only; lanes 3, 7, 11, 15, DNA incubated with HMG1-B at increasing DNA/protein ratios, 1/25, 1/50, 1/100 and 1/300. E61A (lanes 4, 8, 12, 16), K68A (lanes 5, 9, 13, 17), and K81A (lanes 6, 10, 14, 18) were incubated with DNA at the same DNA/protein ratios as HMG1-B.

Corresponding $\Delta\Delta G$ values were also calculated: $\Delta\Delta G = (\Delta G_{\text{H}_2\text{O}})_{\text{wild-type}} - (\Delta G_{\text{H}_2\text{O}})_{\text{mutant}}$. The majority of $\Delta\Delta G$ values are positive and small, in most cases below 1 kcal/mol.

Table 3 Analysis of thermal unfolding curves of wild-type HMG-B and alanine mutants

	T _m (°C)	T _m (K)	ΔS_m (kcal · mol ⁻¹ · K ⁻¹)	ΔH_m from $\Delta S_m - T_m$ (kcal · mol ⁻¹)	dlnK/d(1/T) (K)	ΔH_m from dlnK/d(1/T) · R (kcal · mol ⁻¹)	$\Delta(T_m)$ (K)	$\Delta\Delta H_m$ from $\Delta(T_m) \cdot \Delta S_m$ (kcal · mol ⁻¹)	$\Delta\Delta H_m$ from dlnK/d(1/T) · R (kcal · mol ⁻¹)	$\Delta G(25)$ with $\Delta C_p = 1.5976 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$ (kcal · mol ⁻¹)
HMG1-B	46.03	319.18	−0.16	50.69	−25521.00	50.71	0.00	0.00	0.00	2.21
K12A	44.63	317.78	−0.15	48.96	−24632.20	48.94	−1.40	1.73	1.77	2.03
R13A	43.56	316.71	−0.14	45.79	−23069.90	45.84	−2.47	4.90	4.87	1.80
P14A	38.29	311.44	−0.13	41.81	−21047.30	41.82	−7.74	8.88	8.89	1.32
P15A	37.74	310.89	−0.14	42.63	−21445.50	42.61	−8.29	8.06	8.10	1.32
S16A	32.33	305.48	−0.11	33.55	−16884.90	33.55	−13.70	17.14	17.16	0.66
F18A	26.91	300.06	−0.12	36.84	−18543.40	36.85	−19.12	13.85	13.86	0.23
I38A	48.02	321.17	−0.17	53.01	−26521.80	52.70	1.99	−2.31	−1.99	2.43
G46A	49.91	323.06	−0.16	50.95	−25668.00	51.00	3.88	−0.26	−0.29	2.36

Table 4 Comparison of mutant stability and binding affinity relative to wild-type HMG1-B

WT, wild-type protein HMG1-B; 0, no change compared with WT; +, increase of property compared with WT; —, decrease of property compared with WT.

Mutation	Stability relative to WT	Binding affinity relative to WT	Mobility of J1-protein complex relative to WT
WT	0	0	0
K12A	—	—	—
R13A	—	—	—
P14A	—	0	0
P15A	—	0	0
S16A	—	—	0
F18A	—	+	—
F19A	+	—	0
C22A	—	—	0
S23A	—	0	—
R26A	0*	+	—
K30A	0*	0	—
L36A	—	+	0
S37A	—	+	0
I38A	+	+	0
G39A	+	+	+
K43A	—	+	+
G46A	+	—	—
E61A	—	+	—
K68A	+	—	+
K81A	0*	—	+

* Difference in stability relative to WT < ± 0.05 M GdnHCl.

Thermal unfolding

Additionally, thermal denaturation of wild-type and approximately half of the mutant proteins was monitored by circular dichroism measurement at 222 nm.

The free energy change (ΔG), free energy change at 25 °C [$\Delta G(25\text{ °C})$], melting temperature (T_m), enthalpy change (ΔH_m), and entropy change (ΔS_m), were calculated according to Pace [31]. Values of heat capacity change (ΔC_p) were estimated using only information from the primary sequence of the protein [32]:

$$\Delta C_p \cong 172 + 17.6 \times N - 164 \times SS \text{ (in cal} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}\text{)}$$

N is the number of amino acid residues and SS is the number of disulphide cross-links in the protein. We used this approximation to obtain a rough value for ΔC_p (ΔC_p of HMG1-B approx. $1.6 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{K}^{-1}$). The results for wild-type HMG1-B and mutant proteins are presented in Table 3. The melting temperature of wild-type HMG1-B is 46 °C in a 10 mM potassium phosphate buffer (pH 7.0), containing 0.5 mM DTT and 0.5 mM EDTA. Under these conditions the van't Hoff ΔH_m is $50.7 \text{ kcal} \cdot \text{mol}^{-1}$. Due to the effect of ΔC_p , the corresponding $\Delta\Delta H$ values ($\Delta\Delta H = \Delta H_{\text{wild-type}} - \Delta H_{\text{mutant}}$) vary over a small range.

Binding assays

HMG box proteins have been shown to bind with high affinity to 4WJs and more weakly to linear-duplex DNA [11,12,33,34]. We tested all wild-type and mutant proteins for their ability to bind to the 4WJ J1 and linear-duplex DNA of closely related sequence using EMSA.

Wild-type HMG1-B binds to 4WJ J1 at a DNA/protein ratio of approx. 1:75 (0.0154 μM J1 and 1.155 μM HMG1-B in the final reaction), and not to linear-duplex DNA in the range of

DNA and protein concentrations tested. All mutant proteins retained their selectivity for 4WJ DNA over linear-duplex DNA (Figure 5). The binding data are summarized in Table 4. Relative binding constants were estimated from the approximate concentration of protein required for half dissociation of the complex.

DISCUSSION

Although the actual structures of the mutant proteins have not been determined, we assume from conservation of the CD spectra that all differences in secondary structure correspond to local changes in the folded structure rather than a global rearrangement. Denaturation by GdnHCl or by temperature yields consistent information about the mutant proteins: we find the same rank order of stability for the set of HMG1-B mutants (Figure 6). The binding affinity of several mutant proteins (P14A, P15A, S23A and K30A) is very close to that of wild-type protein, indicating that the substituted side chains are not immediately involved in binding HMG1-B to 4WJ J1. Other mutant proteins show a decrease in binding affinity compared with HMG1-B. These mutant sites involve residues widely separated in the protein: at the N-terminal tail (K12A and R13A), the α -helical

(1) mutants with reduced stability

F18A < S16A < P14A < P15A < S37A

(2) mutants with slightly reduced stability

(1) < C22A < E61A < L36A < S23A < R13A

(3) mutants showing nearly wild-type stability

(2) < K12A < K43A < K81A < R26A < K30A < **HMG-B**

(4) mutants presenting an increased stability

HMG-B < K68A < F19A < I38A < G39A < G46A

Figure 6 Summary of the stability of different HMG-B proteins

Stability was determined by GdnHCl-induced unfolding and thermal unfolding; mutants are grouped into four categories: (1) mutants with reduced stability, (2) mutants with slightly reduced stability, (3) mutants showing nearly wild-type stability, and (4) mutants with increased stability.

Figure 7 Binding of R26A and HMG1-B to J1 and duplex 101

(A) Binding of HMG1-B and R26A to duplex 101. Lane 1, single-strand 101; lane 2, duplex 101; lanes 3, 5, 7, 9, HMG1-B incubated with duplex 101 at DNA/protein ratios of 1:25, 1:50, 1:75, 1:300. Lanes 4, 6, 8 and 10 show duplex 101–R26A complexes at the same ratios. Note that R26A binds selectively to duplex 101, whereas HMG1-B does not. (B) Binding of HMG1-B and R26A to J1. Lane 1, single-strand 101; lane 2, 4WJ J1 only; lanes 3–10, complexes of J1–HMG1-B and J1–R26A, respectively, incubated at the same DNA/protein ratios as in (A). Lanes 11–14: competition assay. J1 was incubated with HMG-B (lanes 11 and 13) and R26A (lanes 12 and 14) at a DNA/protein molar ratio of 1:75. A 100-fold excess of unlabelled duplex 101 was added to the reaction mixtures in lanes 11 and 12, and a 1000-fold excess of unlabelled duplex 101 was added to the reaction mixtures in lanes 13 and 14. Only at a 1000-fold excess does duplex 101 compete for binding to the protein.

regions (S16A, F19A and C22A in helix I; G46A in helix II; and K68A in helix III), and at the C-terminus of the protein (K81A). The original amino acid side chains are either thought to be directly involved in binding DNA or their substitution perturbs side chains that are.

Interestingly, we identified several mutant proteins that bind more tightly to 4WJ DNA than the wild-type HMG1-B. These include amino acid substitutions in the turn connecting helices I and II (L36A, S37A and I38A), a residue at the C terminus of helix I (R26A), and two residues at the N terminus of helix II (G39A and K43A). A reasonable possibility is that all these substitutions influence the orientation or flexibility in the angle between helices I and II, which are thought to determine the capacity of the proteins to recognize DNA in the absence of sequence specificity [22]. The resulting structure(s) can then interact more effectively with the 4WJ DNA.

Substitution of arginine in position 26 by alanine gives the most significant increase in binding affinity (2–3-fold). This mutant also binds with higher affinity to linear-duplex DNA, although it retains selectivity for 4WJ DNA over linear-duplex DNA as seen in a competition experiment (Figure 7). The 4WJ DNA–protein complex is not displaced by a 100-fold excess of duplex DNA.

Figure 8 Binding of HMG1-B and R26A to J1

Lane 1, single-strand 101; lane 2, 4WJ J1 only; lanes 3–6, HMG1-B incubated with J1 at DNA/protein ratios of 1:25, 1:50, 1:75, and 1:300. Lanes 7–10 show J1–R26A complexes at the same DNA/protein ratios.

Three additional mutant proteins that show a higher binding affinity to J1 than the wild-type are more difficult to interpret; one is a mutation in helix I (F18A), the second a mutation in helix III (E61A), and the third a site in helix I (R26A). The increase in binding affinity with the mutation of F18A is unexpected, because side chains at the analogous positions have been implicated as direct participants in DNA intercalation in both SRY– [19] and LEF1–DNA complexes [18]. The side chain corresponding to F18 contacts DNA by packing against two adjacent thymines in both the SRY and LEF1 complexes. Recent NMR data [22] suggest that the counterpart of F18 of HMG1-B in NHP6A (Y28) in combination with the adjacent side chain M29 (F19 in HMG1-B) causes a kink in the DNA. If so, one would anticipate a decrease in binding affinity on replacing F18 by an alanine. The higher-affinity binding of F18A suggests that other protein–DNA contacts must be involved in recognition of the junction, which compensate for any loss of free energy of binding upon substitution of phenylalanine by alanine. It is of course possible that the phenylalanine-to-alanine change facilitates a conformational change in the protein that accompanies binding DNA. At present we cannot distinguish between these alternatives.

The mutant protein E61A displays a slightly higher binding affinity than F18A. The glutamic acid residue 61 is located in helix III facing the connecting region between the three helices. Higher binding affinity could be simply an electrostatic effect (less repulsion due to the change to alanine), or the DNA might fit better into the hydrophobic wedge due to the smaller size of alanine.

A characteristic we observe in all the DNA–protein complexes of this study is a further shift in the mobility of the DNA–protein bands in polyacrylamide gels as the protein concentration increases. The effect is evident in Figure 8, which shows mobility differences of 4WJ–HMG1-B complexes at higher concentrations. This is consistent with the presence of more than one binding process in the interaction between HMG1-B proteins with DNA. In a separate study we have found evidence for a tetrameric binding mode involving four HMG1-B molecules and a single four-arm DNA junction (results not shown). The presence of complexes involving more than one HMG1-B protein and the 4WJ could explain some puzzling differences in mobility of the complexes reported in Table 4. Several complexes migrate more slowly or more rapidly than the wild-type reference. Cases

in which the binding is tighter while the complex migrates more slowly (e.g. R26A, E61A), or vice versa (K68A, K81A), cannot be explained by affinity differences. In these cases we believe the simplest explanation is that the mutation affects a site of protein–protein interaction, rather than the binding site for DNA. It is not useful to assign a single overall dissociation constant [35] if complexes of different stoichiometry between HMG1-B and branched DNA substrates occur.

While it is highly unlikely that branched DNA molecules represent the natural target for HMG interaction, binding 4WJ DNA is a common characteristic of HMG box proteins. One hypothesis is that binding to branched or distorted duplex structures corresponds to some facet of the authentic target, such as apposition of DNA duplexes at an angle or bend. An obvious candidate might be the two duplexes emerging from a nucleosome which form an angle of 85 ° (at a monovalent cation concentration of 5 mM) [36]. Ferrari et al. [37] proposed an induced-fit hypothesis to explain the difference: sequence-specific HMG boxes can reorganize their target DNA, while the structure-specific species cannot. Our mutagenesis results point to a different recognition mechanism.

This research was supported by grant CA 24101 from the National Cancer Institute, NIH, U.S.A.

REFERENCES

- Bustin, M. and Reeves, R. (1996) *Progr. Nucleic Acid Res. Mol. Biol.* **54**, 35–100
- Pazin, M. J. and Kadonaga, J. T. (1997) *Cell* **89**, 325–328
- Widom, J. (1998) *Curr. Biol.* **8**, R788–R791
- Struhl, K. (1998) *Genes Dev.* **12**, 599–606
- Johnson, C. A., Goddard, J. P. and Adams, R. L. P. (1995) *Biochem. J.* **305**, 791–798
- Sera, T. and Wolffe, A. P. (1998) *Mol. Cell. Biol.* **18**, 3668–3680
- Jantzen, H., Admon, A., Bell, S. and Tjian, R. (1990) *Nature (London)* **344**, 830–836
- Ner, S. S. (1992) *Curr. Biol.* **2**, 208–210
- Teo, S.-H., Grasser, K. D. and Thomas, J. O. (1995) *Eur. J. Biochem.* **230**, 943–950
- Pil, P. M., Chow, C. S. and Lippard, S. J. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 9465–9469
- Bianchi, M. E., Beltrame, M. and Paonessa, G. (1989) *Science* **24**, 1056–1059
- Bianchi, M. E., Falcicola, L., Ferrari, S. and Lilley, D. M. J. (1992) *EMBO J.* **11**, 1055–1063
- Paull, T. T., Haykinson, M. J. and Johnson, R. C. (1993) *Genes Dev.* **7**, 1521–1534
- Grosschedl, R., Giese, K. and Pagel, J. (1994) *Trends Genet.* **10**, 94–100
- Wang, H., Bloom, O., Zhang, M., Vishnubhakat, J. M., Ombrellino, M., Che, J., Frazier, A., Yang, H., Ivanova, S., Borovikova, L. et al. (1999) *Science* **285**, 248–251
- Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R. C., Laue, E. D. and Thomas, J. O. (1993) *EMBO J.* **12**, 1311–1319
- Hardman, C. H., Broadhurst, R. W., Raine, A. R. C., Grasser, K. D., Thomas, J. O. and Laue, E. D. (1995) *Biochemistry* **34**, 16596–16607
- Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R. and Wright, P. E. (1995) *Nature (London)* **376**, 791–795
- Werner, M. H., Huth, J. R., Gronenborn, A. M. and Clore, G. M. (1995) *Cell* **81**, 705–714
- Werner, M. H., Gronenborn, A. M. and Clore, G. M. (1996) *Science* **271**, 778–784
- Zlatanova, J. and van Holde, K. (1998) *FASEB J.* **12**, 421–431
- Allain, F. H.-T., Yen, Y.-M., Masse, J. E., Schultze, P., Diekmann, T., Johnson, R. C. and Feigon, J. (1999) *EMBO J.* **18**, 2563–2579
- Milla, M. E., Brown, B. M. and Sauer, R. T. (1994) *Nat. Struct. Biol.* **1**, 518–523
- Brown, B. M., Milla, M. E., Smith, T. L. and Sauer, R. T. (1994) *Nature Struct. Biol.* **1**, 164–168
- Brown, B. M. and Sauer, R. T. (1999) *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1983–1988
- Bianchi, M. E. (1991) *Gene* **104**, 271–275
- Kunkel, T. A., Bebenek, K. and McClary, J. (1991) *Methods Enzymol.* **204**, 125–139
- Chow, C. S., Barnes, C. M. and Lippard, S. J. (1995) *Biochemistry* **34**, 2956–2964
- Pace, C. N., Vajdos, F., Fee, L., Grimsley, G. and Gray, T. (1995) *Protein Sci.* **4**, 2411–2423
- Kallenbach, N. R., Ma, R. I. and Seeman, N. C. (1983) *Nature (London)* **305**, 829–831
- Pace, C. N. (1988) *Methods Enzymol.* **131**, 266–279
- Milardi, D., la Rosa, C., Fasone, S. and Grasso, D. (1997) *Biophys. Chem.* **69**, 43–51
- Bianchi, M. E. (1988) *EMBO J.* **7**, 843–849
- Falcicola, L., Murchie, A. I. H., Lilley, D. M. and Bianchi, M. E. (1994) *Nucleic Acids Res.* **22**, 285–292
- Weiss, M. A., Ukiyama, E. and King, C.-Y. (1997) *J. Biomol. Struct. Dyn.* **15**, 177–184
- Bednar, J., Horowitz, R. A., Grigoryev, S. A., Carruthers, L. M., Hansen, J. C., Koster, A. J. and Woodcock, C. L. (1998) *Proc. Natl. Acad. Sci. U.S.A.* **95**, 14173–14178
- Ferrari, S., Harley, V. R., Pontiggia, A., Goodfellow, P. N., Lovell-Badge, R. and Bianchi, M. E. (1992) *EMBO J.* **11**, 4497–4506

Received 10 December 1999/31 January 2000; accepted 25 February 2000