Solution Structure and DNA Binding Property of the Fifth HMG Box Domain in Comparison with the First HMG Box Domain in Human Upstream Binding Factor[†,‡]

Wulin Yang, Yingqi Xu, Jihui Wu,* Wangyong Zeng, and Yunyu Shi*

Laboratory of Structure Biology, School of Life Science, University of Science and Technology of China, Hefei, Anhui 230026, People's Republic of China

Received June 27, 2002; Revised Manuscript Received November 14, 2002

ABSTRACT: Human upstream binding factor (hUBF) is a nucleolar transcription factor involved in transcription by RNA polymerase I. It contains six HMG box domains. The contribution of each HMG box motif to its function is different. hUBF HMG box 1 shows a very strong binding affinity for both the four-way DNA junction and a 15 bp GC-rich rRNA gene core promoter fragment, but hUBF HMG box 5 shows a much weaker binding affinity for the four-way DNA junction and the GC-rich rRNA gene core promoter fragment. To illustrate the molecular basis of their DNA binding difference, the solution structure of box 5 was studied by NMR. The tertiary structure of box 5 shows a common flattened L-shaped fold, similar to box 1 and other HMG boxes with known structures. It is formed by intersection of three helical arms: helix 1 (residues 9−25) and helix 2 (residues 30−42) pack into each other to form the major wing, while helix 3 (residues 48−70) is aligned with the extended N-terminal segment to form the minor wing. A hydrophobic core is formed by three tryptophans (W14, W41, and W52) to maintain the fold. Although there is similarity between the two structures, negative charged electrostatic surface potential in the concave face of the molecule of box 5 exhibits great difference compared to that of box 1 and other HMG boxes with known structures. That surface is involved in DNA binding. Besides, in positions which are involved in intercalating into a DNA base pair, there are hydrophobic residues in box 1 and other HMG boxes but polar residues in box 5. These differences may contribute to the loss of the DNA binding ability of box 5.

Transcription of the tandemly repeated genes in human encoding ribosomal RNAs involves protein−protein and protein−nucleic acid interactions between the rRNA gene promoter, human RNA polymerase, and at least two transcription factors, hUBF[1] and hSL1 (*1*). The human RNA polymerase transcription factor SL1 (hSL1) is composed of TBP and three TAFs (*2*). hSL1 could not bind efficiently to the rRNA gene core promoter, and its efficient binding requires hUBF. Conversely, binding of hUBF is also enhanced by hSL1, indicating cooperative binding to the rRNA gene between the two factors (*3*).

Human upstream binding factor (hUBF) is made up of a doublet of 97 and 94 kDa, termed hUBF1 and hUBF2, respectively, which is originated by alternative splicing (*4*). Sequence analysis revealed that hUBF1 contains six tandemly repeated DNA binding domain HMG boxes (*1*, *15*). There is a deletion of 37 amino acid residues within the second of six HMG boxes in hUBF2 (*1*, *5*). The contribution of each HMG box motif to the function in DNA binding and transcription activity is different. HMG box 1 is necessary and sufficient for specific DNA binding, and other HMG domains and the C-terminal acidic tail contribute to transcription activity (*6*, *7*).

The HMG box domain is a relatively small DNA binding motif, usually 70−80 amino acid residues, which was first recognized by sequence alignment of hUBF with a high-mobility group protein. This kind of DNA binding motif has been found in many proteins (*8*−*11*), including nuclear and mitochondrial transcription factors and DNA repair proteins. HMG box proteins could be divided into two subtypes according to their DNA binding specificity. The first subtype contains only one HMG box and could bind to an AT-rich DNA sequence. Examples include the mammalian testis-determining factor SRY (*12*), the lymphoid-specific enhancer binding factor LEF1 (*13*), and the T-cell factor TCF1 (*14*). A second class of HMG box proteins that bind to DNA in no obvious sequence-specific recognition is exemplified by

[†] This work is supported by Chinese National Fundamental Research Project Grant G1999075605, Chinese National Natural Science Foundation Grants 39990600 and 30130080, the National High Technology Project of China (Grants 2001AA233021 and 2002BA711A13), and the Pilot Project of the Knowledge Innovation Program of CAS.

[‡] The coordinates of the final 30 structures obtained and the minimized average structure have been deposited in the Protein Data Bank as entries 1L8Y and 1L8Z. The ^{1}H, ^{13}C, and ^{15}N chemical shift values have been deposited in the BioMagResBank (accession number 5392).

* To whom correspondence should be addressed. Fax: +86-551-3603754. E-mail: yyshi@ustc.edu.cn or wujihui@ustc.edu.cn.

[1] Abbreviations: 2D and 3D, two- and three-dimensional, respectively; hUBF, human upstream binding factor; IPTG, isopropyl β-D-thiogalactopyranoside; HSQC, heteronuclear single-quantum spectroscopy; NOE, nuclear Overhauser effect; rmsd, root-mean-square deviation; TOCSY, total correlation spectroscopy; HMG, high-mobility group.

10.1021/bi026372x CCC: $25.00
Published on Web 01/31/2003

high-mobility group proteins 1 and 2 (HMG1 and HMG2, respectively) and the human mitochondrial transcription factor A (mtTFA) (*16*). Many proteins of both subtypes have been shown to bind with a high binding affinity to the four-way DNA junction that mimics the Holliday structure, a putative intermediate in DNA recombination (*17*).

hUBF binds to the ribosomal promoter with only a relaxed specificity (*1*, *6*); no discernible recognition site could be defined. It may represent another subclass of HMG box proteins lying between two subclasses mentioned above. This was supported by a homology study of 121 HMG boxes, which classified three (namely, HMG boxes 1, 3, and 6) of the six HMG boxes in hUBF into a middle subgroup (*10*). As the first step in RNA pol I transcription, hUBF must bind to the ribosomal promoter specifically; how this specificity is encoded in either the promoter or the multiple DNA binding domains of hUBF remains unclear. Another interesting feature of hUBF is that it binds to GC-rich sequences (*1*), whereas other HMG box proteins usually prefer AT-rich sequences, which may have greater bendability to adopt bending induced by HMG box binding.

We have previously reported the tertiary structure of hUBF HMG box 1 (*18*), which just shows a conserved structure like other HMG box proteins, including both the non-sequence-specific subtype and the sequence-specific subtype (*19*−*29*). All these HMG box domains show a common flattened L-shaped fold formed by the intersection of three helical arms. We also confirmed that HMG box 1 has a very strong binding affinity for the four-way DNA junction (dissociation constant K_D of 3.0×10^{-8} M) and a 15 bp GC-rich DNA fragment within the rRNA gene core promoter (dissociation constant K_D of 2.6×10^{-7} M) (*30*). It is believed that its structure and DNA binding properties are relevant for the role of HMG box 1 in the interaction of hUBF with the rRNA gene promoter. Each HMG box has been thought to have a discreet function in hUBF, whether the structure and DNA binding property of other HMG box domains in hUBF are similar to or different from those of the first HMG box. The relationships between their structures and function are still intriguing.

In this paper, another HMG box in hUBF, HMG box 5, was studied. It exhibits a DNA binding property very different from that of HMG box 1. The question of what cause this difference between them arises. Furthermore, we resolved its tertiary structure by NMR. hUBF box 5 shows a conserved tertiary structure like other HMG box domains, though containing a shorter loop between helix 1 and helix 2. The sequences of some other HMG boxes were compared with that of hUBF box 5. The residues at positions *X*, *Y*, and *Z* (Figure 5a), which are responsible for DNA bending and binding, are different from the residues in hUBF box 5. Also, hUBF box 5 exhibits an electrostatic surface potential different from those of all HMG box proteins with known structures. This work provides further insight into the DNA binding and bending properties of this DNA binding domain, which helps us understand the role of HMG box 5 in hUBF and the evolutionary adaptation of this sort of DNA binding domain.

MATERIALS AND METHODS

Cloning of the hUBF HMG Box 5 Gene. The DNA fragment encoding the HMG box 5 domain (amino acid residues 479−560) was amplified by PCR from the cDNA template of hUBF using the two primers 5′-GGCG CATATG GGC AAG CTG CCC GAG TCC-3′ and 5′-CGCG CTCGAG CTT CTT GGA AGA ATT TG-3′ and inserted between the *Nde*I and *Xho*I sites of PET-22b. The identity of the subcloned HMG box 5 was confirmed by DNA sequencing. The recombinant vector was transformed into *Escherichia coli* strain BL21(DE3) for expression.

Expression and Purification. A high-level expression clone was selected. For the production of unlabeled protein, the transformed cells were grown in LB medium, while for preparation of ^{13}C- and ^{15}N-labeled proteins, cultures were grown on SV40 medium containing 2.5 g/L [$^{13}C_6$]glucose and 0.5 g/L $^{15}NH_4Cl$ as the sole carbon and nitrogen sources, respectively. After the OD_{600} reached 0.6−0.8, IPTG was added to a final concentration of 1 mM. The culture was incubated for an additional 4 h. Then the cells were collected by centrifugation. After removal of the supernatant, the pellet was resuspended in 40 mL of binding buffer [20 mM Tris-HCl, 0.5 M NaCl, and 1% Triton X-100 (pH 7.6)]. The cells were broken on ice by ultrasonication and centrifuged to remove the debris. The supernatant was applied to a 5 mL Hitrap chelating column (Pharmacia), to which had been previously applied 5 mL of 0.1 M $NiSO_4$, and equilibrated with 5 bed columns of binding buffer. After application of the sample, the column was first washed with binding buffer for 5 bed columns. Washing buffer [containing 20 mM Tris-HCl (pH 7.6), 0.5 M NaCl, and 100 mM imidazole] was used to wash impurities for an additional 5 bed columns. At last, the bound recombinant protein was eluted with elution buffer [containing 20 mM Tris-HCl (pH 7.6), 0.5 M NaCl, and 1 M imidazole]. The purity of the fractions was analyzed by Tricine-SDS−PAGE (10%, w/v) (*31*). They proved to be more than 95% pure. The concentration of the pure recombinant protein was analyzed with the BCA protein assay kit (Pierce). The yield was ∼20 mg/L.

Gel Mobility Shift Assay. The four-way DNA junction binding experiment was performed according to the method of Tetsuji Ohno et al. (*32*). The four strands, each 35 nucleotides in length, were designed in this experiment: 5′-GCCTTCAACCACCGCTCAACTCAACTGCAGTCTGG-3′, 5′-GCCAGACTGCAGTTGAGTCCTTGCTAGGACG-GAGG-3′, 5′-GCCTCCGTCCTAGCAAGGGGCTGCTAC-CGGAAGGG-3′, 5′-GCCCTTCCGGTAGCAGCCTGAGCG-GTGGTTGAAGG-3′. By annealing the appropriate nucleotides and purifying by electrophoresis, we obtained the JCT3 four-way DNA junction molecules. It was labeled with [γ-^{32}P]ATP by using T4 polynucleotide kinase. A sample of purified HMG box 5 protein was incubated with a radioactively 5′-^{32}P-labeled four-way DNA junction (25 nM) in 10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 0.5 mM DTT, 1 mM $MgCl_2$, 4% glycerol, 0.5 mM EDTA, and 50 ng/μL poly(dI-dC) (binding buffer) at various concentrations (0−300 nM) in a total volume of 10 μL for 30 min at room temperature. After incubation, 1 μL of 10× gel loading buffer [250 mM Tris-HCl (pH 7.5), 0.2% bromophenol blue, and 40% glycerol] was added only to the mixture for the negative control reaction. The DNA was separated on a 4% nondenaturing polyacrylamide gel in 0.5× TBE buffer (pH 8.3) consisting of 89 mM Tris base, 89 mM boric acid, and 2 mM EDTA. After electrophoresis, the gel was scanned with a typhoon system (Amersham Pharmacia Biotech).

Surface Plasmon Resonance Assay. The interaction of the recombinant protein with the junction was further studied by using a BIACORE-1000 instrument. The binding and dissociation reactions were performed in the same buffer for the gel mobility shift assay. To investigate the interaction of the recombinant hUBF HMG box 5 with the rRNA gene core promoter, a short fragment CP was selected (*30*). The reaction was performed in the same buffer for the gel mobility shift assay. The flow rate was 5 μL/min.

NMR Sample Preparation and NMR Spectroscopy. The purified doubly labeled protein was dissolved to a final concentration of 3 mM in 45 mM potassium phosphate (pH 5.5, containing 10% D_2O) at 300 K. All NMR experiments were performed on a Bruker DMX500 spectrometer with a self-shielded *z*-axis gradient.

The following spectra were recorded at 300 K to obtain backbone and side chain resonance assignments: 2D ^1H,^{15}N-HSQC, 2D ^1H,^{13}C-HSQC, 2D TOCSY, 3D triple-resonance spectra HNCO, CBCA(CO)NH, CBCANH, C(CO)NH-TOCSY, H(CCO)NH-TOCSY, HCCH-TOCSY, HBHA-(CO)NH, 2D CB(CC)HD-COSY (aromatics), and 3D ^{13}C-separated and ^{15}N-separated NOESY (mixing time of 130 ms). The doubly labeled sample was finally lyophilized and dissolved in 99.96% D_2O, which was followed immediately with HSQC experiments in an effort to monitor the disappearance of NH signals at 290 K. After all the peaks had vanished, 2D homonuclear TOCSY and NOESY spectra were recorded at 300 K, which exhibited proton resonance from aromatic rings exclusively in the region beyond 6.0 ppm. The $\{^1H\}-^{15}N$ heteronuclear NOE was measured using ^{15}N-labeled protein in an H_2O buffer at 300 K.

NMR data processing was carried out using NMRPipe and NMRDraw software, and the data were analyzed with PIPP. All software was run on a Linux system. Linear prediction (*33*) was used to improve spectral resolution in the indirect dimensions where constant-time acquisition was used, for example, the ^{15}N evolution dimension in all the triple-resonance experiments mentioned above.

Experimental Restraints and NMR Structure Calculation. Three-dimensional ^{15}N-separated NOESY-HSQC and ^{13}C-separated NOESY-HSQC and 2D NOESY spectra in D_2O were acquired to delineate the interproton distance restraints. The intensities of the cross-peaks in spectrum were classified as strong, medium, weak, and very weak, corresponding to distance restraints of 1.8−3.0, 1.8−4.0, 1.8−5.0, and 1.8−7.0 Å, respectively. More conservative distance estimation was used for the latter two NOESY spectra, considering that the spin diffusion effect could be serious for aliphatic and aromatic protons. The chemical shift index (CSI) (*34*) was calculated for four types of nuclei: Cα, Cβ, C′, and Hα. The derived secondary structures based on the consensus CSI were converted into restraints on ϕ and ψ angles. For α-helix residues, ϕ angles were limited to $-60 \pm 40°$, while ψ was in the range of $-50 \pm 50°$. From the identified slow exchange amide protons located in regular secondary structures, hydrogen bond restraints were added. Structures were calculated using the program CNS version 1.0, employing a simulated annealing protocol for torsion angle dynamics (*35*). Simple impulsion nonbonded interactions were used during structure calculation. Structural figures were produced with MOLMOL and Rasmol 2.6.4.

RESULTS AND DISCUSSION

Binding of hUBF HMG Box 5 to DNA

A gel mobility shift assay and SPR have been carried out in an effort to analyze binding of hUBF HMG box 5 to a four-way DNA junction (Figure 1a,b). From the gel mobility shift assay, we found that box 5 has a behavior in its binding to the four-way DNA junction different from that of box 1. Under the same conditions for both experiments, a set of four shifted bands could be seen in the interaction between box 1 and the four-way DNA junction (*30*). However, there are no multiple-shift bands appearing in this experiment for box 5. Only a tiny retarded band could be seen with increasing protein concentrations. The binding affinity was measured by SPR. The equilibrium dissociation constant (K_D) of hUBF box 1 for the four-way junction is 3.0×10^{-8} M (*30*), and the binding of hUBF HMG box 5 to the four-way junction is too weak to be measured.

Another SPR experiment was carried out to study the binding of the hUBF HMG box 5 to a 15 bp GC-rich DNA fragment within the rRNA gene core promoter. This fragment has been studied by SPR in our work (*30*), which shows a rather strong binding affinity for box 1. The equilibrium dissociation constant (K_D) is 2.6×10^{-7} M. In contrast to box 1, box 5 almost shows no affinity for this DNA fragment (Figure 1c). This difference implies that these two domains play different roles in the activity of human upstream binding factor functioning in the transcription of the rRNA gene.

NMR Measurement

Chemical Shift Assignments. The 2D ^1H,^{15}N-HSQC spectrum illustrates the good dispersion of the proton and nitrogen resonances in the amide groups. Complete backbone assignments of residues G1−E84 were made for $^1H^N$, ^{15}N, $^1H_\alpha$, $^1H_\beta$, $^{13}C_\alpha$, and $^{13}C_\beta$ (in fact, the recombinant protein contains residues 478−560 of hUBF, and an N-terminal Met and a C-terminal hexahistidine-LEHHHHHH motif, which were not well assigned). Most ^{13}CO assignments were determined, except for those of residues preceding prolines. Assignments of side chain resonances were mostly completed, and were checked with C(CO)NH-TOCSY, H(CCO)NH-TOCSY, and HCCH-TOCSY experiments. Some aromatic ring proton chemical shifts were assigned by 2D TOCSY in D_2O, ^1H,^{13}C-HSQC, and CB(CC)HD-COSY experiments, and the others were assigned through NOEs between ring protons and the intraresidue H_β and H_N atoms. In all, more than 90% of the side chain resonances were assigned. All the resonance assignments are listed in the Supporting Information.

Secondary Structure Determination. The secondary structure prediction is based on the chemical shift index (CSI) and the characteristic sequential and medium-range NOE links (Figure 2). According to the CSI, there are three helical regions located at residues 10−25, 30−42, and 45−70. The first helical region (residues 10−25) is separated at residues V18 and I19 from the CSI prediction. A similar secondary structure was made from characteristic sequential and medium-range NOE links. But there are some subtle differences. The beginning of first helix seemingly extends to R9, while the end of the third helix maybe extends to A72, until it is interrupted by P73. The second helix seems to extend to M44. In the sequence of residues F26−D29, just after

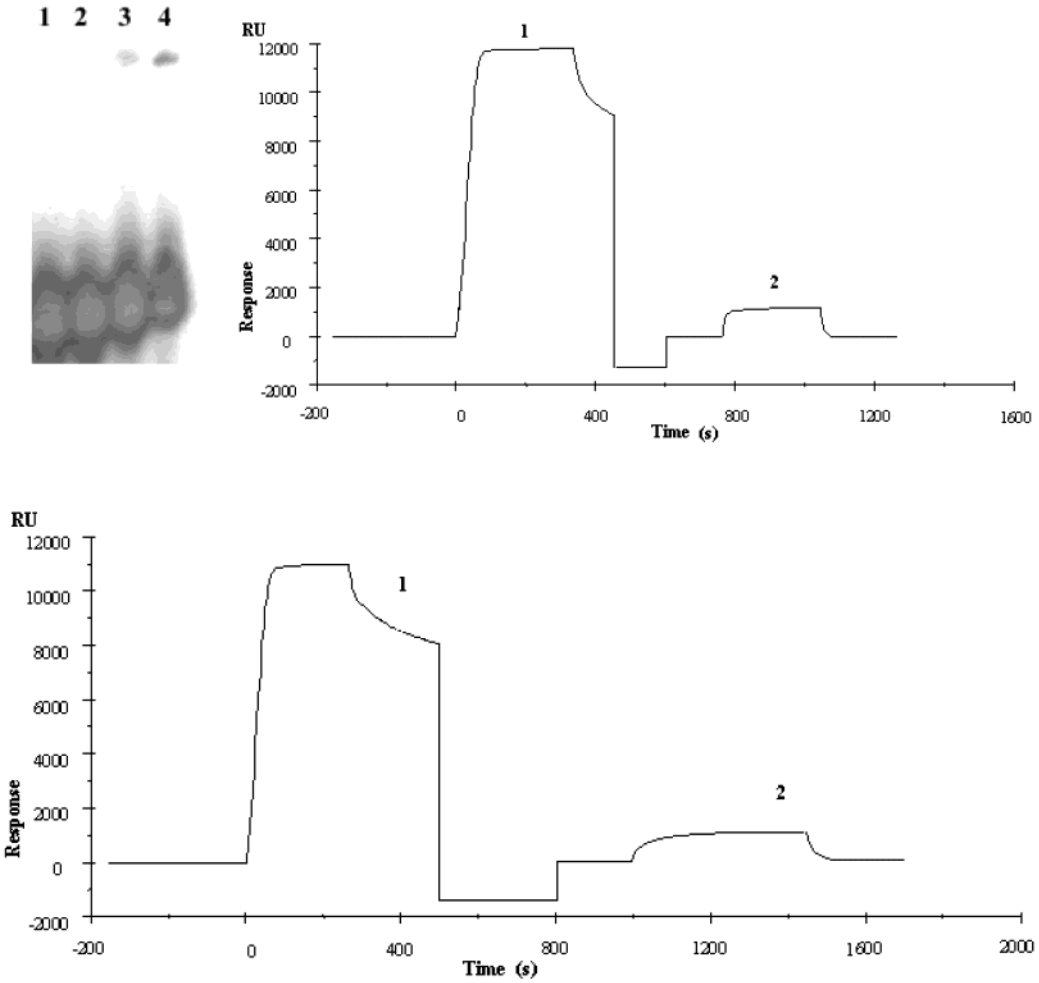

FIGURE 1: Interactions of the recombinant hUBF HMG box 5 with DNA. RU represents resonance units. (a) Gel mobility shift assay of the binding of hUBF box 5 to the four-way DNA junction (JCT3) at 25 nM. Lanes 1−4 contained 0, 100, 200, and 300 nM protein, respectively. (b) Comparison of the binding affinity of hUBF box 5 (line 2) and box 1 (line 1) for the four-way DNA junction (JCT3) by a surface plasmon resonance assay. The biotin-labeled JCT3 was fixed onto a streptavidin-coated sensor chip. Both concentrations of hUBF box 5 and box 1 are 4000 nM (*30*). (c) Comparison of the binding affinity of hUBF box 5 (line 2) and box 1 (line 1) for 15 bp linear DNA fragment CP. Both concentrations of hUBF box 5 and box 1 are 4000 nM (*30*).

FIGURE 2: Summary of sequential and medium-range NOE contacts observed in HMG box 5 of hUBF. The data are derived from ^{15}N-separated NOESY and ^{13}C-separated NOESY spectra; the mixing time is 130 ms for both. The three rows of solid bars below the sequence represent the observed sequential d_{NN}, $d_{\alpha N}$, and $d_{\beta N}$ NOE contacts; the thickness of the bars indicates the intensities of the corresponding cross-peaks in NOESY. Lines below the sequential contacts represent the $d_{\alpha N}(i,i+3)$, $d_{\alpha N}(i,i+4)$, $d_{\alpha N}(i,i+2)$, $d_{NN}(i,i+2)$, and $d_{\alpha\beta}(i,i+3)$ NOE contacts. Filled circles denote the locations of slowly exchanging amide protons. The short bars at the bottom represent consensus CSI predictions from Cα, Cβ, C′, and Hα chemical shifts; bars below the line mean an index of −1, while those above the line mean an index +1.

helix 1, a strong $d_{NN}(2,3)$ contact was found, which to some extent means a turn. Unfortunately, because there is heavy overlap in the NOESY spectrum in the region of D29, we could not observe a $d_{\alpha N}(2,4)$ contact in this sequence. The rate of exchange of amide protons with solvents was measured. Amide protons still present after 20 h of exchange were considered slowly exchanging amide protons, which may be engaged in hydrogen bonds. Nineteen slow exchange

Table 1: NMR Structure Statistics for HMG Box 5 of hUBF

(a) NMR constraints	
no. of distance constraints	1098
intraresidue	156
sequential ($\|i - j\| = 1$)	327
medium-range ($\|i - j\| < 5$)	402
long-range ($\|i - j\| \geq 5$)	213
no. of dihedral angle restraints	104
no. of hydrogen bonds	18
(b) statistics for 30 SA structures	
NOE violations	
average number >0.2 Å	3.87 ± 1.33
maximum violations (Å)	0.381
rms deviation from idealized covalent geometry	
bonds (Å)	0.0035 ± 0.0001
angles (deg)	0.53 ± 0.01
impropers (deg)	0.43 ± 0.02
Lennard-Jones potential energy (kcal/mol)	−204.29 ± 23.19
(c) coordinate precision	
pairwise rmsd for backbone atoms in residues 9−70 (Å)	0.96 ± 0.23
pairwise rmsd for heavy atoms in residues 9−70 (Å)	1.86 ± 0.27
pairwise rmsd for backbone atoms in residues 8−64 (Å)	0.64 ± 0.14
pairwise rmsd for heavy atoms in residues 8−64 (Å)	1.55 ± 0.22
rms deviation from the mean (backbone/heavy)	
residues 9−70 (Å)	0.67/1.30
residues 8−64 (Å)	0.46/1.09
residues 1−73 (Å)	1.26/1.80
(d) pairwise rmsds for backbone atoms (Å), comparison between the minimized average structure of box 5 with other HMG box proteins	
hUBF_box 5 (residues 7−57) vs hUBF box 1 (residues 14−35 and 37−65)	2.24 (*18*)
hUBF_box 5 (residues 7−57) vs rat HMG1B (residues 10−31 and 33−61)	2.72 (*19*)
hUBF_box 5 (residues 7−57) vs rat HMG1A (residues 11−32 and 35−64)	3.56 (*22*)
hUBF_box 5 (residues 7−57) vs mouse LEF-1 (residues 6−27 and 29−57)	2.74 (*25*)
hUBF_box 5 (residues 7−45 and 48−57) vs HMG-D (residues 8−29 and 31−57)	2.81 (*21*)

amide protons were identified, which were all in the predicted helical region (Figure 2).

Structure Calculations. The tertiary structure was determined from a total of 1098 restraints, including 156 intraresidue, 327 sequential, 402 medium-range, and 213 long-range, together with 104 backbone dihedral angle constraints and 18 hydrogen bond constraints (Table 1). Being unassigned, the first residue Met and the C-terminal hexahistidine tag had no NMR restraints, and they were excluded from structure calculation. From 100 calculated structures, the 30 structures with the lowest total energy were selected. All these conformations are in good agreement with the experimental data, with no distance violations larger than 0.5 Å and no angle violations of more than 5°. The Procheck program was used to check the stereochemical quality of the structures. It showed that 97.1% of the residues are in the allowed region of the Ramachandran plot.

The structure was well-defined between residues 8 and 64, with an atomic root-mean-square deviation (rmsd) from the average structure of 0.46 Å for the backbone atoms and 1.09 Å for all heavy atoms. Outside of this region, from residue P7 to the N-terminus and from residue R65 to the C-terminus, there is a gradually increasing level of disorder (Figure 3a). According to the $\{^1H\}-^{15}N$ heteronuclear NOE analysis (Figure 3b), the $\{^1H\}-^{15}N$ heteronuclear NOE values have a tendency toward negative values in both termini. In the region from residue P4 to P7 and from residue R65 to E69, a number of NOEs could still be found (Figure 3c). We suppose that the level of disorder in these residues, to some extent, is correlated with the flexibility of the peptide rather than due to a lack of NOEs.

Structure Description. Figure 4b shows a schematic ribbon model of the minimized average structure of hUBF HMG box 5. As expected from the secondary structure analysis, the three helices are located at residues 9−26, 30−42, and 48−70. It forms an L shape with helices 1 and 2 in one arm (the major wing) and helix 3 and the N-terminal residues in the other arm (the minor wing). The global structure is similar to that found for other HMG boxes. There are two loop domains connecting helix 1 and helix 2 and connecting helix 2 and helix 3. In loop 2, residues 43 and 44 are in a partially helical conformation and the following three residues (45−47) are in a type I turn conformation. The angle between helix 2 and helix 3 is ∼80°, while helix 1 intersects with helix 2 and helix 3 with angles of 35° and 75°, respectively. There is a bend of ∼38° in the middle of helix 1 (residues 18 and 19). The conserved aromatic rings of W14, W41, and W52 stack into each other, forming a hydrophobic core to stabilize the structure of the protein (Figure 4b). This core is extended into both arms and angle apexes. The packing of helix 1 into helix 2 also involves V18, I19, Y22, L34, and M37. Residue L23 make close contact with residues N28/R30 to stabilize the loop between helix 1 and helix 2. The loop between helix 2 and helix 3 is stabilized by W41/M44 and K49. The orientation of helix 1 and helix 3 depends on the contact between R9/A10 and I53, in addition to the contact between I13 and K55/A56. In the minor wing, the N-terminal segment alongside helix 1 is fixed by packing of Y63, E66, and L67 to P4 and P7. In addition, a salt bridge possibly forms between K8 and D59. This contact stabilized the minor wing.

Comparison with hUBF HMG Box 1 and Other HMG Box Proteins. The backbone rms deviations between the minimized average structure of hUBF box 5 and some other HMG boxes are listed in Table 1d. In contrast with hUBF box 1, hUBF box 5 has a greater deviation from other HMG boxes (*18*). The structure most similar to hUBF box 5 is hUBF box 1, which reflects some relations between these two domains in hUBF. Like box 1, there is an angle (only 35°) between helices 1 and 2 in hUBF box 5 somewhat smaller than that in other HMG box folds (*18*, *28*). The angle between helices 2 and 3 and bend angle of helix 1 in hUBF box 5 are almost the same as those in hUBF box 1, while the angle between helices 1 and 3 in hUBF box 5 is larger than that in box 1. They are 75° and 60°, respectively. However, this interhelical angle difference may be due to the precise nature of the NOE constraints used in the structural determination, rather than a real difference. To ensure that the structure determined for box 5 better fits the experimental NOE constraints used in structure determination than that for box 1, we have used a box 1 backbone structure as the starting structure (processed with MODELLER 4), and refined it with the restraints used for box 5. This still results in a structure that is the same as the structure that would be generated using random structure as the starting structure. The backbone rms deviation (from residue 9 to residue 70) between the minimized average structures resulting from these two approaches is 0.34, and the interhelical angles do not change.

The most distinctive change in structural feature shown by hUBF box 5 lies on loop 1 (Figure 5b). Loop 1 in hUBF

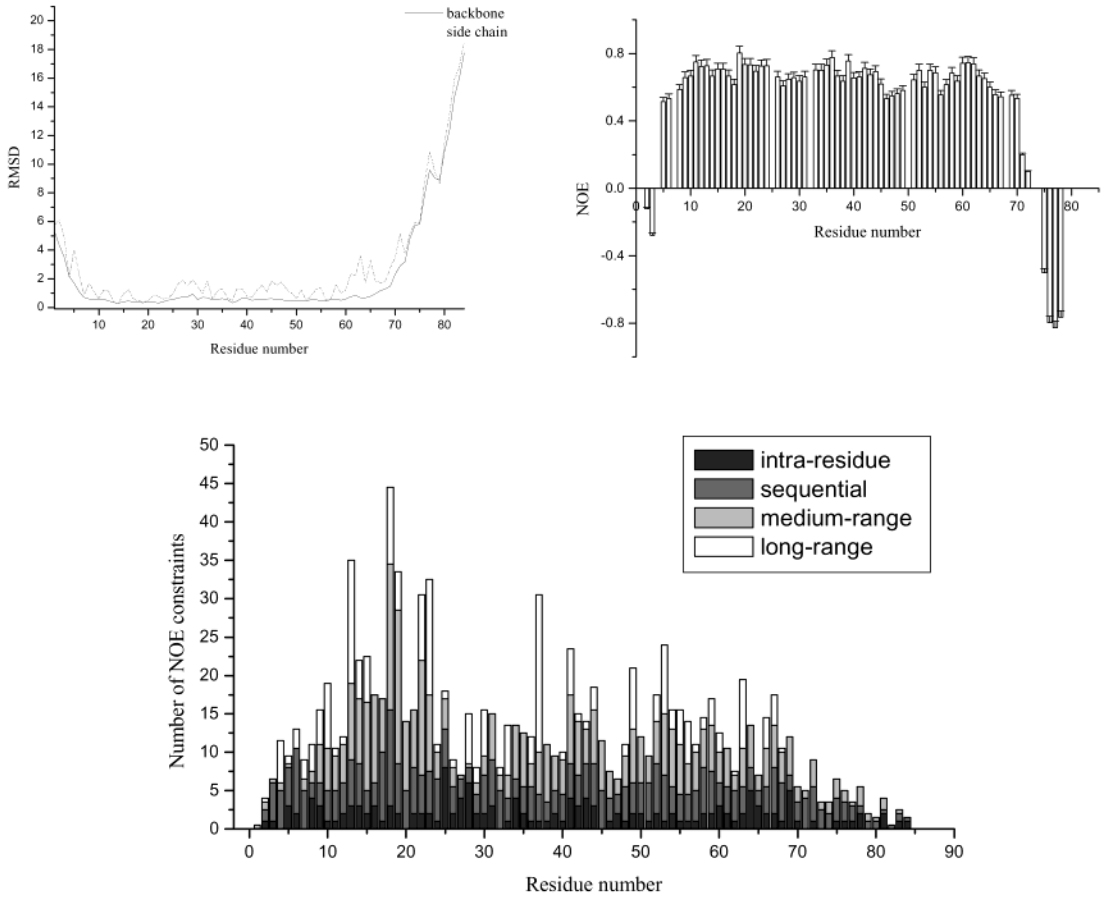

FIGURE 3: (a) Plot of the average rms deviation from mean structure vs the sequence of hUBF HMG box 5. (b) $\{^1H\}-^{15}N$ heteronuclear NOE values for individual residues of hUBF HMG box 5. Residues for which no values are shown correspond either to prolines or to residues for which data could not be extracted. (c) Plot of the number of NOE constraints per residue used in the calculation of the hUBF box 5 structure.

box 5 is shorter than that in hUBF box 1 and other HMG boxes, which makes a close contact between the C-terminal segment of helix 1 and the N-terminal segment of helix 2. This interaction has an effect on the whole structure, making it more elongated than other DNA binding HMG boxes.

Molecular Basis for Its DNA Binding Properties. All HMG boxes with known structures have a similar L-shaped structure. What causes hUBF HMG box 5 to bind to DNA in a manner so different from that of hUBF HMG box 1? According to the previously determined HMG box−DNA complexes, as exemplified by both the sequence-specific and non-sequence-specific HMG boxes (*24*−*27*), the DNA binding concave face conforms to a wide, shallow minor groove. Numerous electrostatic interaction and van der Waals contact make up the interface. In the center of the concave face, a hydrophobic wedge inserts deeply into the minor groove of the DNA duplex. This interaction introduces a kink into the bound DNA and widens the minor groove. In summary, the most important three determinants at positions *X*, *Y*, and *Z* (Figure 5) are involved in it. All known nonspecific HMG box proteins have serine at position *X*, whereas all sequence-specific HMG boxes have an asparagine at this position (*14*, *27*). The serine could form a water-mediated interaction with DNA, as in HMG-D in which serine forms water-mediated bonds to the thymine 7 ribose O4′ and adenine 6 N3 position of the DNA base. And asparagine, as in LEF1, could direct a hydrogen bond network to three of the four bases of the C·A base pair. At position *Y*, the residue is hydrophobic in both types of HMG box proteins (Met in mLEF-1, HMG-D, and NHP6A and Ile in hSRY) and could intercalate between two base pairs to cause DNA bending. At position *Z*, another intercalating residue was found in non-sequence-specific HMG boxes, such as valine in HMG-D and phenylalanine in NHP6A. But in sequence-specific HMG boxes, the residues at position *Z* (serine in mLEF1 and TCF1 and asparagine in hSRY and mSOX5) are hydrophilic and are involved in a hydrogen bond to DNA. There are no structural data for the complex of the HMG box bound to a four-way DNA junction. Fortunately, the structure of the architecture-specific complex of HMG-1 box A, a non-sequence-specific HMG box, bound to cisplatin-modified DNA could provide some insight into it (*36*). On the basis of this structure, an intercalating residue phenylalanine at position *Z* was found to partially intercalate within the already kinked DNA site. The hydrophobic intercalating residue at position *Y* is usually considered an initial and determinant step for HMG box protein binding to linear DNA. And DNA recognition at positions *X* and *Z* determines the sequence specificity in DNA binding, which was called a "sequence-neutral" mechanism for non-sequence-specific DNA recognition (*27*, *37*).

hUBF HMG box 5 shows some difference in electrostatic surface potential with respect to hUBF box 1 and all HMG boxes with known structures, including sequence-specific and

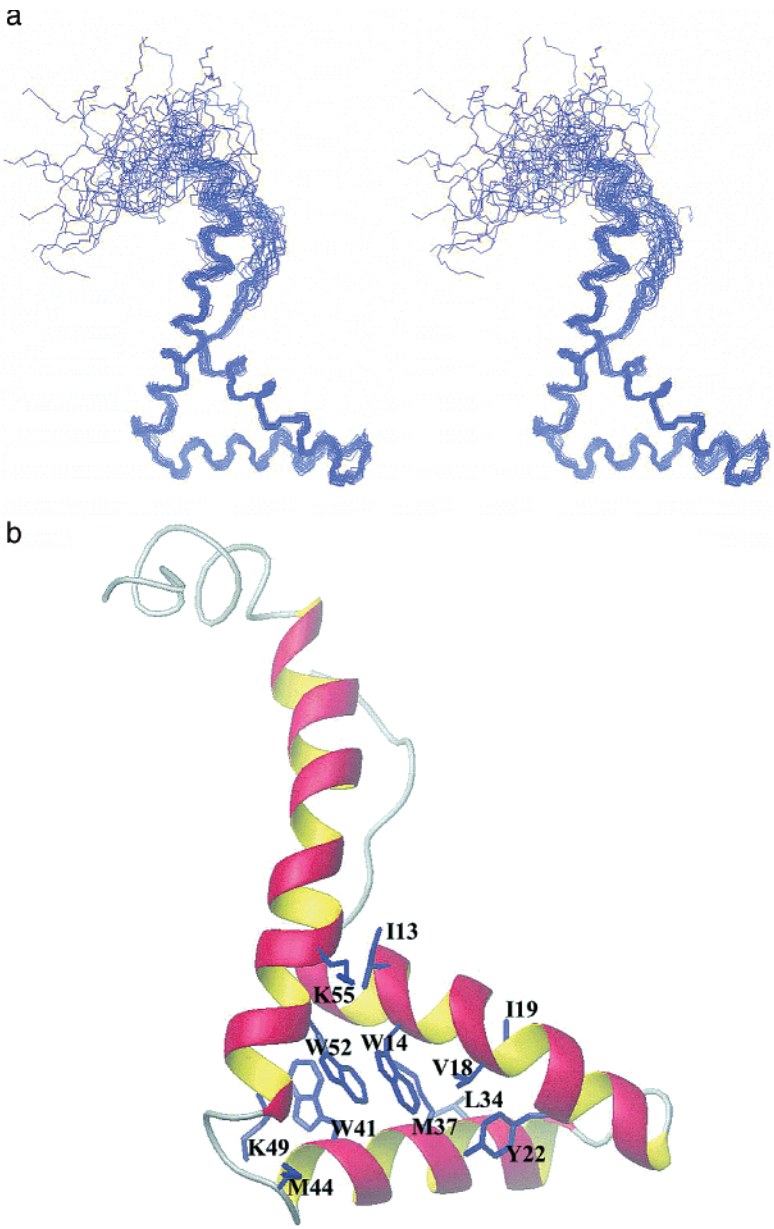

FIGURE 4: (a) Stereoview of the backbone superimposition of the 30 lowest-energy structures for hUBF HMG box 5. (b) Ribbon representation of the minimized average structure. Packing of side chains in the hydrophobic core located at the turn of the L shape, which is important for maintaining the fold.

non-sequence-specific HMG boxes (Figure 5c only showing a non-sequence-specific HMG box, HMG-D, and a sequence-specific HMG box, LEF-1, with hUBF box 1 and hUBF box 5). In the concave DNA binding face, on the basis of the complexes of HMG boxes with known structures and DNA, a hydrophobic surface is predominantly flanked by a series of basic residues that is important in interacting with the DNA phosphodiester backbone. In hUBF box 5, the existence of acidic residues on this surface must have some effect on this interaction.

Figure 5a shows the result of sequence alignment of hUBF HMG box 5 with hUBF HMG box 1 and some other HMG box proteins. In hUBF HMG box 5, three residues (R9, E12, and R30) are located at positions *X*, *Y*, and *Z*, respectively. By comparison, in hUBF HMG box 1, threonine, phenylalanine, and asparagine are at these positions. When a comparison of tertiary structure was drawn between hUBF box 5 and some other HMG boxes, it was found that the residues at these positions are almost spatially conserved. Little difference exists at positions *X* and *Y*. In hUBF HMG box 5, K8 and E11 are in more similar spatial positions than R9 and E12 with respect to residues at positions *X* and *Y* in other HMG box proteins (Figure 5b).

The hydrophobic residue phenylalanine at position *Y*, in hUBF box 1, was thought to intercalate into linear DNA and then bend it. But in hUBF HMG box 5, a glutamic acid residue is found at position *Y*, which is hydrophilic and could not be expected to intercalate into the DNA base pair. It could explain why hUBF HMG box 5, in contrast to hUBF HMG box 1, has a dramatic decrease in its binding affinity for linear DNA. hUBF box 1 has also a strong binding affinity for the four-way DNA junction (*30*). Although no intercalating residue exists at position *Y*, reasonably, the protein hUBF HMG box 5 could bind to predistorted DNA through another intercalating residue at position *Z*. But here hUBF box 5 shows a very weak affinity for the four-way

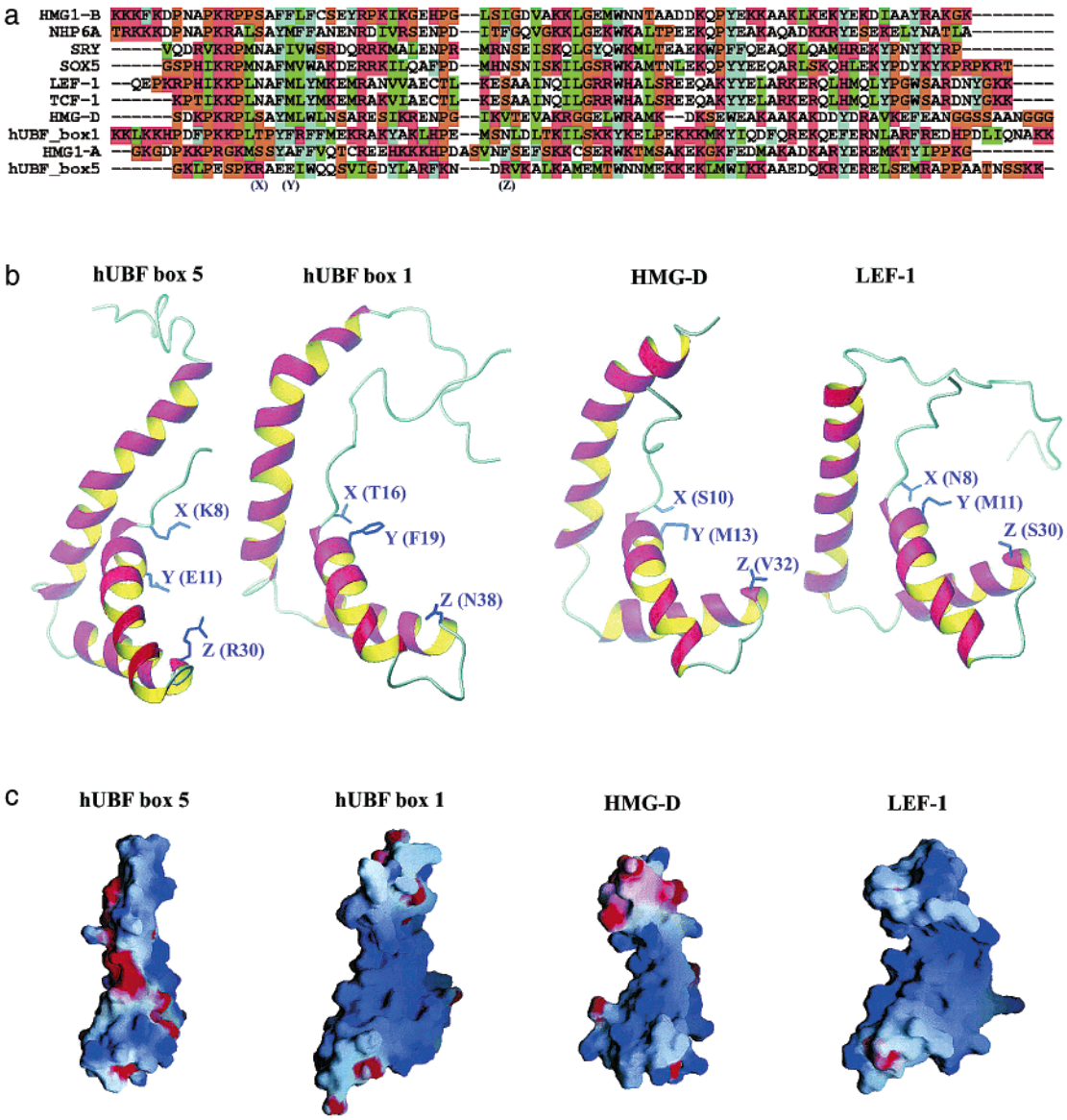

FIGURE 5: (a) Amino acid sequence and alignment of rat HMG1A, rat HMG1B, *Drosophila melanogaster* HMG-D, *Saccharomyces cerevisiae* NHP6A, hUBF HMG box 1, hUBF HMG box 5, mouse Sox-5 and mouse LEF-1, human SRY and TCF1. Three positions which are important for HMG boxes binding to DNA are denoted *X*, *Y*, and *Z*. (b) Location of residues at positions *X*, *Y*, and *Z* in the 3D structures of hUBF HMG box 5, *Drosophila melanogaster* HMG-D, mouse LEF-1, and hUBF HMG box 1. The structures are oriented to show the difference in the loop connecting helix 1 and helix 2 between hUBF HMG box 5 and hUBF HMG box 1. (c) View showing the electrostatic surface potential of hUBF box 5, hUBF box 1, *D. melanogaster* HMG-D, and mouse LEF-1 (GRASP). Blue indicates positive potential, and red indicates negative potential.

DNA junction. In hUBF HMG box 5, a positively charged residue, arginine, at position *Z*, supposedly, is difficult to efficiently intercalate into DNA. Also, it could not be supposed to form a hydrogen bond with DNA as sequence-specific HMG boxes. But in hUBF box 1, a negatively charged residue, asparagine, at this position could behave like sequence-specific HMG boxes to form a hydrogen bond with DNA. The threonine at position *X* in hUBF box 1 is similar to serine, which can form water-mediated hydrogen bonds to DNA bases like the nonspecific HMG box. But in hUBF box 5, a positively charged amino acid (K8 or R9 at this position) could not be expected to behave like serine or asparagine.

It should be noted that hUBF box 5 is more elongated (seen from Figure 5) than other HMG boxes. It seems to indicate that its complementarity to the minor groove of DNA is less appropriate than that of other HMG boxes.

The question of why HMG box 5 undergoes such a change in structure and sequence content arises. As its ability in DNA binding has vanished, what other function was carried out by it? Previous works have suggested HMG boxes 3−6 and the acidic tail in hUBF contribute to the interaction of hUBF and TBP-TAFs complex hSL1 (*6*). As one part of this segment, hUBF HMG box 5 must play a role in it. More details about it need further study.

ACKNOWLEDGMENT

We thank Dr. F. Delagio and Prof. A. Bax for providing the software NMRPipe. We thank Dr. Dan Garrett and the Lab of Chemical Physics at the National Institute of Health for making available the program PIPP. We thank Prof. A. T. Brunger for providing the program CNS. We thank Dr. R. Koradi and Prof. K. Wuthrich for providing MOLMOL.

SUPPORTING INFORMATION AVAILABLE

^{1}H, ^{13}C, and ^{15}N resonance assignments of hUBF HMG box 5 at 300 K and pH 5.5. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Jantzen, H. M., Admon, A., Bell, S. P., and Tjian, R. (1990) Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins, *Nature 344*, 830−836.
2. Comai, L., Tanese, N., and Tjian, R. (1992) The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1, *Cell 68*, 965−976.
3. Bell, S. P., Jantzen, H. M., and Tjian, R. (1990) Assembly of alternative multiprotein complexes directs rRNA promoter selectivity, *Genes Dev. 4*, 943−954.
4. O'Mahony, D. J., and Rothblum, L. I. (1991) Identification of two forms of the RNA polymerase I transcription factor UBF, *Proc. Natl. Acad. Sci. U.S.A. 88*, 3180−3184.
5. Kuhn, A., Voit, R., Stefanovsky, V., Evers, R., Bianchi, M., and Grummt, I. (1994) Functional differences between the two splice variants of the nucleolar transcription factor UBF: the second HMG box determines specificity of DNA binding and transcriptional activity, *EMBO J. 13*, 416−424.
6. Jantzen, H. M., Chow, A. M., King, D. S., and Tjian, R. (1992) Multiple domains of the RNA polymerase I activator hUBF interact with the TATA-binding protein complex hSL1 to mediate transcription, *Genes Dev. 6*, 1950−1963.
7. McStay, B., Frazier, M. W., and Reeder, R. H. (1991) xUBF contains a novel dimerization domain essential for RNA polymerase I transcription, *Genes Dev. 5*, 1957−1968.
8. Bianchi, M. E., Falciola, L., Ferrari, S., and Lilley, D. M. (1992) The DNA binding site of HMG1 protein is composed of two similar segments (HMG boxes), both of which have counterparts in other eukaryotic regulatory proteins, *EMBO J. 11*, 1055−1063.
9. Landsman, D., and Bustin, M. (1993) A signature for the HMG-1 box DNA-binding proteins, *BioEssays 15*, 539−546.
10. Baxevanis, A. D., and Landsman, D. (1995) The HMG-1 box protein family: classification and functional relationships, *Nucleic Acids Res. 23*, 1604−1613.
11. Baxevanis, A. D., Bryant, S. H., and Landsman, D. (1995) Homology model building of the HMG-1 box structural domain, *Nucleic Acids Res. 23*, 1019−1029.
12. 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., and Goodfellow, P. N. (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif, *Nature 346*, 240−244.
13. Travis, A., Amsterdam, A., Belanger, C., and Grosschedl, R. (1991) LEF-1, a gene encoding a lymphoid-specific protein with an HMG domain, regulates T-cell receptor alpha enhancer function, *Genes Dev. 5*, 880−894.
14. van de Wetering, M., Oosterwegel, M., Dooijes, D., and Clevers, H. (1991) Identification and cloning of TCF-1, a T lymphocyte-specific transcription factor containing a sequence-specific HMG box, *EMBO J. 10*, 123−132.
15. Bachvarov, D., and Moss, T. (1991) The RNA polymerase I transcription factor xUBF contains 5 tandemly repeated HMG homology boxes, *Nucleic Acids Res. 23*, 2331−2335.
16. Parisi, M. A., and Clayton, D. A. (1991) Similarity of human mitochondrial transcription factor 1 to high mobility group proteins, *Science 252*, 965−969.
17. P-ohler, J. R., Norman, D. G., Bramham, J., Bianchi, M. E., and Lilley, D. M. (1998) HMG box proteins bind to four-way DNA junctions in their open conformation, *EMBO J. 17*, 817−826.
18. Xu, Y., Yang, W., Wu, J., and Shi, Y. (2002) Solution structure of the first HMG box domain in human upstream binding factor, *Biochemistry 22*, 98−105.
19. Weir, H. M., Kraulis, P. J., Hill, C. S., Raine, A. R. C., Laue, E. D., and Thomas, J. O. (1993) Structure of the HMG box motif in the B-domain of HMG1, *EMBO J. 12*, 1311−1319.
20. Read, C. M., Cary, P. D., Crane-Robinson, C., Driscoll, P. C., and Norman, D. G. (1993) Solution structure of a DNA-binding domain from HMG1, *Nucleic Acids Res. 21*, 3427−3436.
21. Jones, D. N., Searles, M. A., Shaw, G. L., Churchill, M. E., Ner, S. S., Keeler, J., Travers, A., and Neuhaus, D. (1994) The solution structure and dynamics of the DNA-binding domain of HMG-D from *Drosophila melanogaster*, *Structure 2*, 609−627.
22. Hardman, C. H., Broadhurst, R. W., Raine, A. R. C., Grasser, K. D., Thomas, J. O., and Laue, E. D. (1995) 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.
23. van Houte, L. P. A., Chuprina, V. P., van der Watering, 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.
24. 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.
25. Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R., and Wright, P. E. (1995) Structure basis for DNA bending by the architectural transcription factor LEF-1, *Nature 376*, 791−795.
26. Allain, F. H.-T., 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.
27. Murphy, F. V., IV, Sweet, R. M., and Churchill, M. E. A. (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.
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. Murphy, E. C., Zhurkin, V. B., Louis, J. M., Comilescu, 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.
30. Yang, W., Zeng, W., Zhou, D., and Shi, Y. (2002) Cloning, expression, secondary structure characterization of HMG box 1 of hUBF from *E. coli* and its binding to DNA, *Biochim. Biophys. Acta 1598*, 147−155.
31. Schagger, H., and van Jagoe, G. (1987) Tricine-sodium dodecyl sulfate polyacrylamide gel eletrophoresis for separation of proteins in the range from 1 to 100 kDa, *Anal. Biochem. 166*, 368−373.
32. Ohno, T., Umeda, S., Hamasaki, N., and Kang, D. (2000) Binding of human mitochondrial transcription factor A, an HMG box protein, to a four-way DNA junction, *Biochem. Biophys. Res. Commun. 271*, 492−498.
33. Barkhuijsen, H., de Beer, R., Bovee, W. M., Creyghton, J. H., and van Ormondt, D. (1985) Application of linear prediction and singular value decomposition (LPSVD) to determine NMR frequencies and intensities from the FID, *Magn. Reson. Med. 2*, 86−89.
34. Wishart, D. S., and Sykes, B. D. (1994) The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data, *J. Biomol. NMR 4*, 171−180.
35. Stein, E. G., Rice, L. M., and Brunger, A. T. (1997) Torsion-angle molecular dynamics as a new efficient tool for NMR, *J. Magn. Reson. 124*, 154−164.
36. 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.
37. Thomas, J. O., and Travers, A. A. (2001) HMG1 and 2, and related 'architectural' DNA-binding proteins, *Trends Biochem. Sci. 26*, 167−174.

BI026372X