Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of human upstream binding factor HMG box 5 and site for binding of the cell-cycle regulatory factor TAF1

Hui Rong,[a,b] Yan Li,[a,b] Xiaowei Shi,[a,b] Xiao Zhang,[a,b] Yongxiang Gao,[a,b] Haiming Dai,[a,b] Maikun Teng,[a,b]* Liwen Niu,[a,b]* Qun Liu[c] and Quan Hao[c]

[a]Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science and Technology of China, 96 Jinzhai Road, Hefei, Anhui 230026, People's Republic of China, [b]Key Laboratory of Structural Biology, Chinese Academy of Sciences, 96 Jinzhai Road, Hefei, Anhui 230027, People's Republic of China, and [c]MacCHESS, Cornell High Energy Synchrotron Source, Cornell University, NY 14853, USA

Correspondence e-mail: mkteng@ustc.edu.cn, lwniu@ustc.edu.cn

Received 23 October 2006
Accepted 5 April 2007

PDB Reference: hUBF HMG box 5, 2hdz, r2hzdsf.

The fifth HMG-box domain in human upstream binding factor (hUBF) contributes to the synthesis of rRNA by RNA polymerase I (Pol I). The 2.0 Å resolution crystal structure of this protein has been solved using the single-wavelength anomalous dispersion method (SAD). The crystal structure and the reported NMR structure have r.m.s. deviations of 2.18–3.03 Å for the C^{α} atoms. However, there are significant differences between the two structures, with displacements of up to 9.0 Å. Compared with other HMG-box structures, the r.m.s. deviations for C^{α} atoms between hUBF HMG box 5 and HMG domains from *Drosophila melanogaster* protein D and *Rattus norvegicus* HMG1 are 1.5 and 1.6 Å, respectively. This indicates that the differences between the crystal and NMR structures of hUBF HMG box 5 are larger than those with its homologous structures. The differences between the two structures potentially reflect two states with different structures. The specific interactions between the hUBF HMG box 5 and the first bromodomain of TBP-associated factor 1 (TAF1) were studied by ultrasensitive differential scanning calorimetry and chemical shift perturbation. Based on these experimental data, possible sites in hUBF HMG box 5 that may interact with the first bromodomain of TAF1 were proposed.

1. Introduction

Human upstream binding factor (hUBF) is an HMG-box-containing factor that binds to the ribosomal RNA (rRNA) promoter and activates RNA polymerase I (Pol I) transcription through its acidic carboxy-terminal tail (Jantzen *et al.*, 1990; Tuan *et al.*, 1999). rRNA synthesis by Pol I is a key step in regulating ribosome production (Reeder, 1999; Grummt, 1999). Pol I transcription in humans requires the upstream binding factor UBF and the selectivity factor SL1 to assemble coordinately on the rRNA gene core promoter (Bell *et al.*, 1988; Comai *et al.*, 1992, 1994). The human RNA polymerase transcription factor SL1 (hSL1) is composed of TATA-binding protein (TBP) and three TBP-associated factors (TAFs) (Comai *et al.*, 1992). hSL1 cannot bind efficiently to the rRNA gene core promoter alone; its efficient binding requires hUBF (Bell *et al.*, 1990).

The hUBF is made up of hUBF1 and hUBF2 originating from alternative splicing (O'Mahony & Rothblum, 1991). The hUBF1 contains six tandemly repeated DNA-binding HMG-box domains (Jantzen *et al.*, 1990; Bachvarov & Moss, 1991). There is a deletion of 37 amino-acid residues within HMG box 2 in hUBF2 (Jantzen *et al.*, 1990; Kuhn *et al.*, 1994). Only hUBF HMG box 1 is responsible for specific binding to the rRNA gene core promoter; the other hUBF HMG-box

domains contribute to transcription activity (Jantzen *et al.*, 1992; McStay *et al.*, 1991). The HMG-box domains are usually made up of 70–80 amino-acid residues and were first recognized by sequence alignment of hUBF with a high-mobility group protein (Jantzen *et al.*, 1990).

hUBF binds to the ribosomal promoter with only a relaxed specificity (Jantzen *et al.*, 1990, 1992) and no discernible recognition sites have been defined. However, 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 DNA-binding domains of hUBF remains unclear. Recently, using a yeast two-hybrid screen (Lin *et al.*, 2002), a specific interaction between the region of hUBF spanning HMG boxes 3–6 and TAF1 (amino acids 1096–1872, including the first bromodomain) has been identified. TAF1 (previously known as $TAF_{II}250$) is a subunit of TFIID that is involved in the transcription of cell-cycle and growth-regulatory genes (Wassarman & Sauer, 2001). Further experiments have indicated that the physical interaction between hUBF and TAF1 is indispensable for the stimulation of Pol I transcription (Lin *et al.*, 2002).

Because of its biological interest, we set out to determine the crystal structure of the hUBF HMG box 5 domain, which consists of residues 479–560. The solution structure of the domain has been determined by nuclear magnetic resonance (NMR) spectroscopy (Yang *et al.*, 2003). The crystal structure and the reported NMR structure have the same topology, with an overall r.m.s.d. of 2.18–3.03 Å for the C^{α} atoms. Moreover, specific structural differences between the two structures were identified. Here, we confirmed the interaction between hUBF HMG box 5 and the first bromodomain in TAF1 (Jacobson *et al.*, 2000; PDB code 1eqf) by ultrasensitive differential scanning calorimetry.

2. Materials and methods

2.1. Protein production, crystallization and structure determination

The DNA fragment encoding the hUBF HMG box 5 domain (Swiss-Prot entry P17480) was amplified from the human brain cDNA library (Clontech) by the polymerase chain reaction using the primers 5′-GGCGCATATGGGCAAGCAAGCTGCCCGAGTCC-3′ and 5′-CGCGCTCGAGCTTCTTGGAAGAATTTG-3′. It was cloned into PET-22b using the restriction enzymes *Nde*I and *Xho*I. DNA sequencing confirmed the integrity of the cloned DNA. Expression was undertaken in the *Escherichia coli* BL21 (DE3) expression line.

A high-level expression clone was selected. For the production of wild-type protein, the transformed cells were grown in LB medium. When the OD_{600} reached approximatedly 0.6, isopropyl β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 1 m*M* and the culture was incubated at 303 K for 16 h.

The harvested cells were resuspended in lysis buffer (20 m*M* Tris–HCl pH 8.0) at 277 K, lysed by sonication and the soluble lysate was applied onto a 5 ml Ni–NTA superflow resin (Qiagen). The column was washed in two steps with loading buffer (80 m*M* Tris–HCl pH 8.0, 250 m*M* NaCl) followed by washing buffer (20 m*M* Tris–HCl pH 8.0, 250 m*M* NaCl, 50 m*M* imidazole). The protein was eluted with 500 m*M* imidazole. The eluted proteins containing hUBF HMG box 5 were pooled and then further purified by gel filtration (Superdex 75 column, Amersham Biosciences) in a buffer consisting of 20 m*M* Tris–HCl pH 8.0, 150 m*M* NaCl. The peak fractions were concentrated to 20 mg ml^{-1} and stored at 203 K. To prepare the SeMet-derivatized protein, hUBF HMG box 5 was expressed in *E. coli* strain B834 (Novagen) using M9 medium supplemented with SeMet and the amino acids leucine, isoleucine, valine, phenylalanine, lysine and threonine. The SeMet-derivatized protein was purified using a similar procedure as for the wild-type protein.

hUBF HMG box 5 was purified to near-purity as determined by SDS–PAGE analysis. The 6×His affinity tag (LEHHHHHH) was not removed prior to crystallization trials. Crystal screening was carried out by the vapour-diffusion method using tissue-culture plates and siliconized glass cover slips. Initial screens included Hampton Crystal Screens I and II (Hampton Research) with 2 µl protein solution and 2 µl precipitant solution. The crystals were grown at 293 and 277 K. The His-tagged protein crystallized in 0.1 *M* HEPES pH 7.5, 4.4 *M* NaCl and 1% 2-propanol at 277 K; the SeMet-labelled protein crystallized under the same conditions.

A native data set to 2.0 Å resolution was obtained from a wild-type protein crystal at 100 K. The data were collected on MacCHESS beamline A1 (Cornell University, USA) using an incident X-ray wavelength of 0.9363 Å and an ADSC Q210 CCD detector and were processed using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

The SeMet-labelled crystals were transiently soaked in reservoir solution supplemented with 30% glycerol as a cryoprotectant and flash-cooled. Data collection was carried out at 100 K on beamline 3W1A of the Beijing Synchrotron Radiation Facility at the Institute of High Energy Physics, Chinese Academy of Sciences. Diffraction data were collected from the SeMet-labelled protein crystal to 2.5 Å resolution, with a crystal-to-detector distance of 180 mm, and were processed using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997). There is one monomer in the crystallographic asymmetric unit, based on calculation of the Matthews coefficient (Matthews, 1968).

Because an initial attempt to solve the structure by molecular replacement using the known NMR structure of hUBF HMG box 5 (Yang *et al.*, 2003; PDB code 1l8y) as a search model did not give an appropriate solution, the structure of hUBF HMG box 5 was determined by SAD using SeMet-labelled data. The heavy-atom parameters were first determined using *SOLVE* (Terwilliger & Berendzen, 1999) and then refined; the phases were then calculated using the program *SOLVE* (Terwilliger & Berendzen, 1999). Five Se sites were found. The resulting phases were extended to 2.5 Å and improved using *RESOLVE* (Terwilliger, 2000). A model containing 53 of the 87 amino acids was automatically built by

Table 1
Crystallographic statistics.

Values in parentheses are for the highest resolution shell.

	Native wavelength	Peak wavelength
Wavelength (Å)	0.9363	0.97985
Space group	$P6_5$	$P6_5$
Molecules in ASU	1	1
Unit-cell parameters		
a (Å)	60.8	60.7
b (Å)	60.8	60.7
c (Å)	63.2	62.8
Unit-cell volume (Å)	233628	231386
No. of reflections	215209	93449
No. of independent reflections	9071	8692
Resolution limits (Å)	30–2.0	30–2.5
$R_{\rm merge}$† (%)	6.6 (18.5)	8.4 (27)
Mosaicity (°)	0.51	0.25
Completeness (%)	99.3 (93.3)	98.3 (100)
$\langle I/\sigma(I)\rangle$	35.6 (12.1)	16.3 (3.2)
Refinement		
Resolution (Å)	10.0–2.0	
Reflections used for refinement	8497	
Final *R* (%)	21.7	
Final $R_{\rm free}$ (%)	27.5	
R.m.s.d. bonds (Å)	0.01	
R.m.s.d. angles (Å)	1.3	
Mean *B* value (Å^2)	33.65	
Protein atoms	561	
Solvent atoms	91	

† $R_{\rm merge} = \sum |I_j - \langle I_j\rangle| / \sum I_j$.

RESOLVE (Terwilliger, 2003). The structure was refined at 2.0 Å resolution using *ARP*/*wARP* (Perrakis *et al.*, 1999) and *REFMAC* with manual model correction in *O* (Jones *et al.*, 1991). Figures were prepared using *PyMOL* (http://www.pymol.org) and *O* (Jones *et al.*, 1991). The crystal belongs to space group $P6_5$, with one molecule per asymmetric unit. The final refined model, with an *R* factor of 21.7% and an $R_{\rm free}$ of 27.5%, contains 65 residues. Data-collection and refinement statistics are summarized in Table 1.

2.2. Expression and purification of the first bromodomain of TAF1 (TAF1-bm1)

The DNA fragment encoding residues 1379–1487 of TAF1 (the first bromodomain of TAF1) was amplified from the human brain cDNA library (Clontech) by the polymerase chain reaction (PCR) using the primers 5′-CATATGACAGACCCTATGGTGACG-3′ and 5′-CTCGAGTTTGTCTTCTTTCTCTTTGAG-3′ (Sangon) designed on the basis of the cDNA sequence of human TAF1 (Swiss-Prot entry p21675); these oligonucleotides were introduced with *Nde*I and *Xho*I restriction sites, respectively. The reaction product was purified and then cloned into the T-vector. The positive clones were identified by restriction digestion. The DNA fragment was then ligated into the *Nde*I–*Xho*I-cleaved plasmid pET22b (Novagen). The recombinant vector was amplified in *E. coli* BL21 (DE3). Expression was induced once the optical absorption at 600 nm (A_{600}) reached 0.6 *via* addition of 0.5 m*M* IPTG for 24 h at 289 K. Cells were lysed with a sonifier and the supernatant from a 12 000*g* centrifugation was used for protein purification. The fusion protein was purified on an Ni–NTA Superflow resin column (Qiagen). The column was washed in two steps with loading buffer (50 m*M* KCl pH 7.8, 250 m*M* K_2HPO_4) followed by washing buffer (50 m*M* KCl pH 7.8, 250 m*M* K_2HPO_4, 60 m*M* imidazole). The protein was eluted with 500 m*M* imidazole. The eluted proteins containing TAF1-bm1 were pooled and dialysed in 20 m*M* Tris–HCl pH 8.0 and concentrated for ultrasensitive differential scanning calorimetry and dialysed in 50 m*M* KCl, 100 m*M* K_2HPO_4 pH 5.5 and concentrated for chemical shift perturbation.

2.3. Ultrasensitive differential scanning calorimetry (US-DSC)

US-DSC measurements (Plotnikov *et al.*, 1997) were carried on a VP DSC from Microcal (USA). The volume of the sample cell was 0.509 ml. The reference cell was filled with buffer solution (20 m*M* Tris–HCl pH 8.0) without protein. hUBF box 5 and TAF1-bm1 solutions with a concentration of 0.5 mg ml^{-1} were degassed at 298 K for half an hour. The experiment was performed in a 298 K thermostatic system.

2.4. Chemical shift perturbation

To detect the TAF1-bm1 domain-binding site on the hUBF HMG box 5 domain, 0.8 m*M* ^{15}N-labelled hUBF HMG box 5 was used. After the ^{1}H–^{15}N HSQC spectrum of free hUBF HMG box 5 had been recorded, the sample was titrated with 0.8 m*M* unlabelled TAF1-bm1. The final concentrations of both proteins at the end of the titration were approximately 0.4 m*M*. All the spectra were recorded at 298 K and assignments of the peaks for hUBF HMG box 5 were made according to Yang *et al.* (2003).

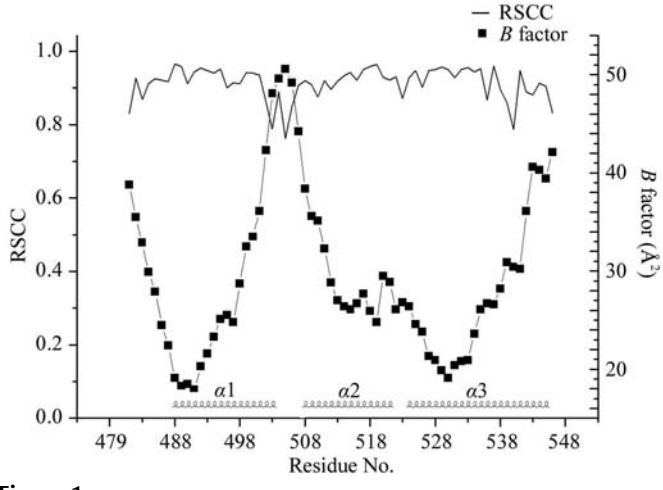

Figure 1
Real-space correlation coefficient (RSCC) calculated using the $2mF_o - DF_c$ map and main-chain *B* factors of the hUBF HMG box 5 crystal structure. The secondary structure is illustrated for reference.

3. Results and discussion

3.1. Quality of refined model

The final model of hUBF HMG box 5 consists of residues 481–546. The $2mF_o - DF_c$ map shows well defined electron density without a break for the backbone of the polypeptide. The real-space correlation coefficient (Jones *et al.*, 1991) of each residue was calculated using *O* to assess the quality of the structure model, as shown in Fig. 1. A high overall value of 0.87 indicated that the model is a good interpretation of the diffraction data. The mean *B* factor for the protein atoms is 36.5 Å^2 and agrees well with the *B* factor of 33.5 Å^2 estimated from the Wilson plot. The first two N-terminal residues, including the first plasmid-encoded residue, and the C-terminal residues 547–560 are disordered.

3.2. Overall structure

The crystal structure of hUBF HMG box 5 has the same topology as the previously reported NMR structure and other HMG-domain structures (Love *et al.*, 1995; Werner *et al.*, 1995; Xu *et al.*, 2002) and consists of loop 1 (486–487), helix 1 (residues 488–503), a type I turn (504–507), helix 2 (residues 508–521), loop 2 (522–523) and helix 3 (residues 524–545). According to the *DALI*/*FSSP* server (Holm & Sander, 1996; http://www.ebi.ac.uk/dali), this structure belongs to the HMG-domain family. Its closest relatives are HMG protein D from *Drosophila melanogaster* (r.m.s.d. 1.5 Å; Murphy *et al.*, 1999; PDB code 1qrv) and domain A of the structure-specific HMG-domain protein HMG1 from *Rattus norvegicus* (r.m.s.d. 1.6 Å; Ohndorf *et al.*, 1999; PDB code 1ckt). Other close structural neighbours include the solution structure of the sixth HMG box of mouse UBF1 (PDB code 1v63) and the solution structure of the thymus HMG-BOX protein TOX from *Mus musculus* (PDB code 2co9).

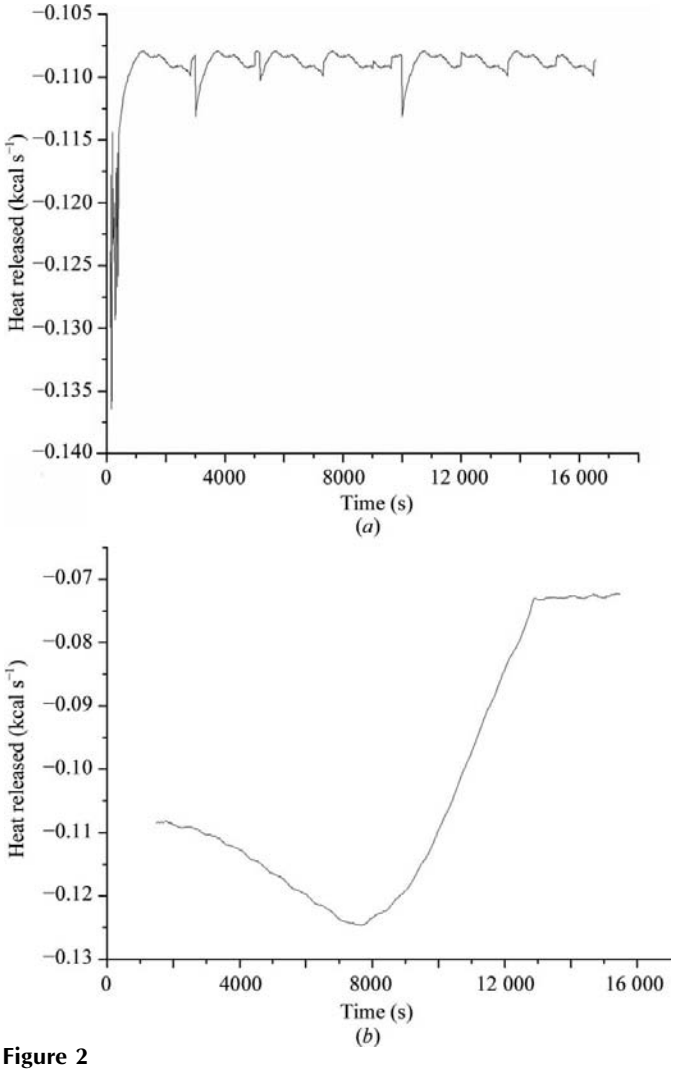

Figure 2
(*a*) Control US-DSC experiment without protein. (*b*) US-DSC profile of the interactions between TAF1-bm1 and the hUBF HMG box 5. The quantity of heat released from the reacting system is plotted against time after mixture of the two components. 1 kcal = 4.186 kJ.

3.3. Interactions between HMG box 5 and TAF1-bm1 studied by US-DSC

Fig. 2(*b*) shows that after the two protein solutions were mixed in a 1:1 molar ratio, in a first stage (from initiation to

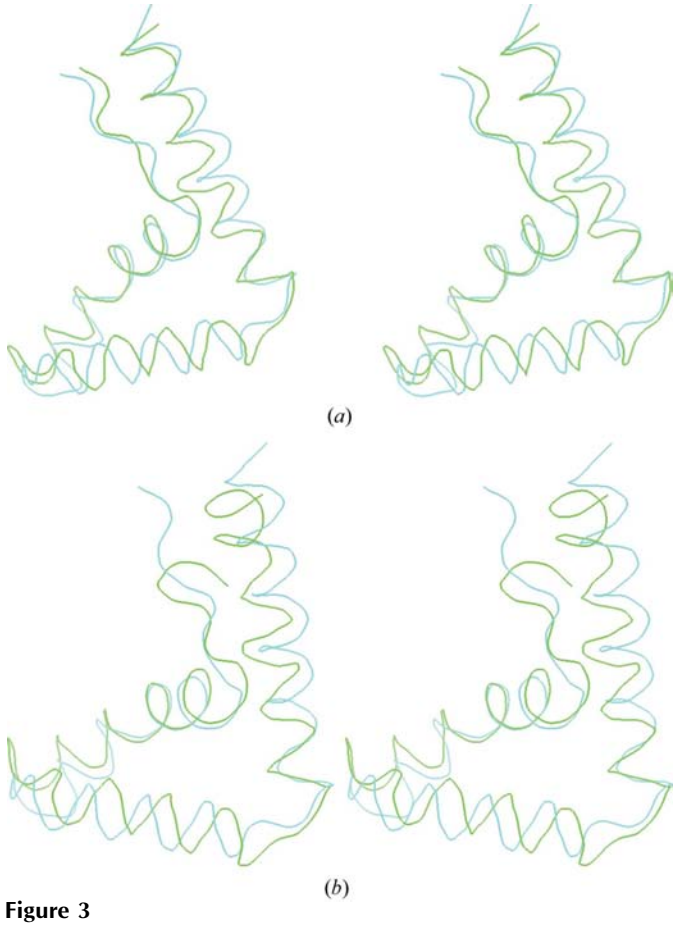

Figure 3
(*a*) Stereoview of superimposition of the C$^\alpha$ trace of the crystal structure (cyan) and model 18 from the NMR structures (green) of hUBF HMG box 5. The NMR structure contains residues 3–68. The r.m.s.d. shift for the C$^\alpha$ atoms is 2.18 Å. (*b*) Stereoview of superimposition of the C$^\alpha$ trace of the crystal structure (cyan) and model 28 from the NMR structures (green) of hUBF HMG box 5. The NMR structure contains residues 3–68. The r.m.s.d. shift for the C$^\alpha$ atoms is 3.03 Å.

approximately 7600 s) the reacting system absorbed a greater quantity of heat than the initiation of the reacting system, in a second stage (from approximately 7600 s to approximately 13 000 s) the reacting system emitted an increasing quantity of heat as time increased and in a third stage (after approximately 13 000 s) the reacting system entered equilibrium. The profile indicates that the two proteins may form a much stabler compound. It at least shows that the two proteins interact with each other. A control experiment with only buffer solution (20 m*M* Tris–HCl pH 8.0) without protein can be seen in Fig. 2(*a*).

3.4. Comparison with NMR structure

The present structure of hUBF HMG box 5 was superimposed with each of the 30 NMR structures of HMG box 5 reported previously, using *LSQMAN* (Kleywegt & Jones, 1996) to minimize the r.m.s. deviation of the two structures. The superposition resulted in r.m.s. deviations of 2.18 Å (aligned with model 18) to 3.03 Å (aligned with model 28) for C^{α} atoms (Figs. 3*a* and 3*b*). When model 18 of the NMR structures was superimposed with the crystal structure in segments (3–9, 10–26, 30–43, 46–67), the r.m.s. deviations of segments 3–9 and 46–67 were 2.75 and 2.30 Å, respectively, and were greater than r.m.s. deviations of segments 10–26 (1.60 Å) and 30–43 (1.26 Å); we therefore speculate that these structural differences between the NMR model and the crystallographic model may be the reason why the crystal structure could not be solved from the NMR model. Most of the residues in the crystal and NMR structures of hUBF HMG box 5 have similar solvent accessibility (Willard *et al.*, 2003), except for seven residues (see later). The residues that constitute the hydrophobic core are generally conserved in the crystal and NMR structures. In addition to the conserved residues identified from the NMR study, we found that residues Tyr22 and Ala36 contribute to the core. In contrast, residues that are buried in the NMR structure, Pro7 (solvent accessibility increased from 7 to 31%), Ala10 (solvent accessibility increased from 6 to 12%), Trp52 (solvent accessibility increased from 4 to 11%) and Ala56 (solvent accessibility increased from 0 to 16%), prefer to be exposed in the crystal structure. Another residue with a large increase in solvent accessibility is Glu66 (crystal, 44%; NMR, 4%). Residues 30–32 form a 3_{10}-helix in the NMR structure; while in the crystal structure these residues together with residues 33–43 form a standard α-helix. In addition, we used the *FOLD-X* server (Guerois *et al.*, 2002) to calculate the stability of the two structures. From the computation, the total energy of the crystal structure is -37.3 kJ mol^{-1}, while the total energy of each NMR structure is above 83 kJ mol^{-1}. This result is exactly consistent with the conclusion of Drenth & Haas (1992) that the crystal structure is more compact and stable than the NMR structures.

In comparison with the NMR structures, the C-terminal residues 69–82 of the crystal structure are disordered. This might be the effect of the His tag and the connecting peptides from the plasmid construction, since these peptides were not removed prior to crystallization. Although the NMR structures could provide models of residues 69–82, the quality of this segment from the NMR structures is poorer than that of other segments (Yang *et al.*, 2003).

3.5. Comparison with other HMG-domain structures

According to the *DALI*/*FSSP* server (Holm & Sander, 1996; http://www.ebi.ac.uk/dali), the overall C^{α} alignments of the crystal structure of hUBF HMG box 5 with other HMG domains indicate r.m.s.d.s of between 1.5 and 3.6 Å.

3.6. Chemical shift perturbation upon TAF1-bm1 binding

Specific chemical shift perturbations and changes in the line widths were observed in the ^{1}H–^{15}N HSQC spectrum of the HMG box 5 domain upon formation of the complex with TAF1-bm1. The chemical shift changes were observed from the beginning of the titration, when the concentration of HMG box 5 domain was approximately 0.4 m*M* and that of TAF1-bm1 was 0.04 m*M*, until the final concentrations of TAF1-bm1 and HMG box 5 domain reached approximately

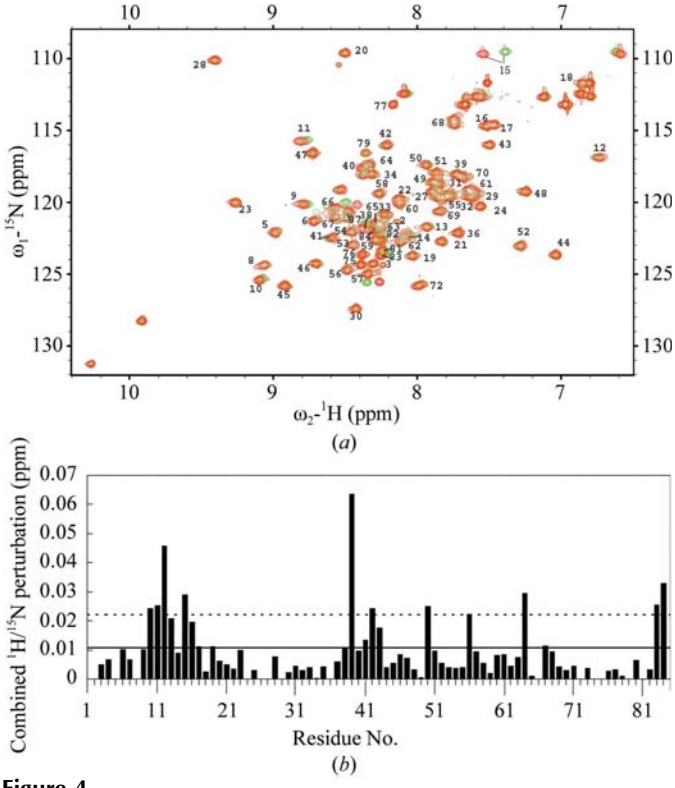

Figure 4
(*a*) Superposition of ^{1}H–^{15}N HSQC spectra of free ^{15}N-labelled hUBF HMG box 5 and of ^{15}N-labelled hUBF HMG box 5 in complex with TAF1-bm1. The hUBF HMG box 5:TAF1-bm1 ratio is approximately 1:1. The cross-peaks of free hUBF HMG box 5 are shown in red and those of complexed hUBF HMG box 5 are shown in green. All backbone NH and side-chain peaks affected by complex formation are labelled. (*b*) Histogram of the chemical shift perturbation of hUBF HMG box 5 induced by the addition of TAF1-bm1. The mean value is shown as a solid line and the mean value plus one standard deviation is shown as a dashed line. The combined ^{1}H and ^{15}N chemical shift changes were defined as $\Delta\delta$ (ppm) = $[(\Delta\delta_{HN})^2 + (\Delta\delta_N\alpha_N)^2]^{1/2}$ and the scaling factor α_N used to normalize the ^{1}H and ^{15}N chemical shift was 0.17. Residues whose cross-peaks were not observed in the presence of TAF1-bm1 were excluded.

0.4 m*M*. Superposition of the HSQC spectra of free HMG box 5 domain and of HMG box 5 domain in complex with the bromodomain of TAF1 is shown in Fig. 4(*a*).

Residues Ala10, Glu11, Glu12, Gln15, Met39, Asn42, Leu50 and Glu64 (Leu83 and Glu84 are plasmid-encoded residues) exhibited combined chemical shift changes of more than the mean value plus one standard deviation (Fig. 4*b*). Most of these residues have large changes in both proton and nitrogen chemical shifts (Fig. 4*a*). The chemical shift changes of these residues are linear with the addition of TAF1-bm1. These residues are coloured cyan in Fig. 5(*a*). In addition, residues Ile13, Gln16, Trp41, Asn43, Ala56 and Glu66 have combined chemical shift changes that are larger than the mean value and are coloured grey. Moreover, the cross-peaks of both protons of the side-chain NH_2 atom of Gln15 shifted significantly, suggesting the formation of hydrogen bonds with residues of TAF1-bm1.

3.7. HMG box 5 domain-binding surface on TAF1-bm1

The chemical shift perturbation information given above was used to map the hUBF HMG box 5 domain-binding surface on TAF1-bm1. Residues 10–13, 15, 16, 39, 41–43, 50, 56, 64 and 66 experience the most significant changes in line widths and chemical shifts upon complex formation with TAF1-bm1 (Fig. 4*b*). Most of these regions are clustered together spatially and thus constitute the binding site for TAF1-bm1. These residues are located in loop 1, helix 1, helix 2 and helix 3. Some of the perturbed residues are conserved in HMG-family members. Three hydrophobic residues, Ala10, Trp41 and Ala56, are highly conserved. In addition, two negatively charged residues, Glu64 and Glu66, are moderately conserved among HMG-family members. Glu64, Glu66 and another two negatively charged residues Glu11 and Glu12 may interact with certain residues of TAF1-bm1 *via* electrostatic interactions. In addition, there are six hydrophobic residues: Ala10, Ile13, Met39, Trp41, Leu50 and Ala56. Ala10, Met39, Leu50 and Ala56 are exposed to solvent and might

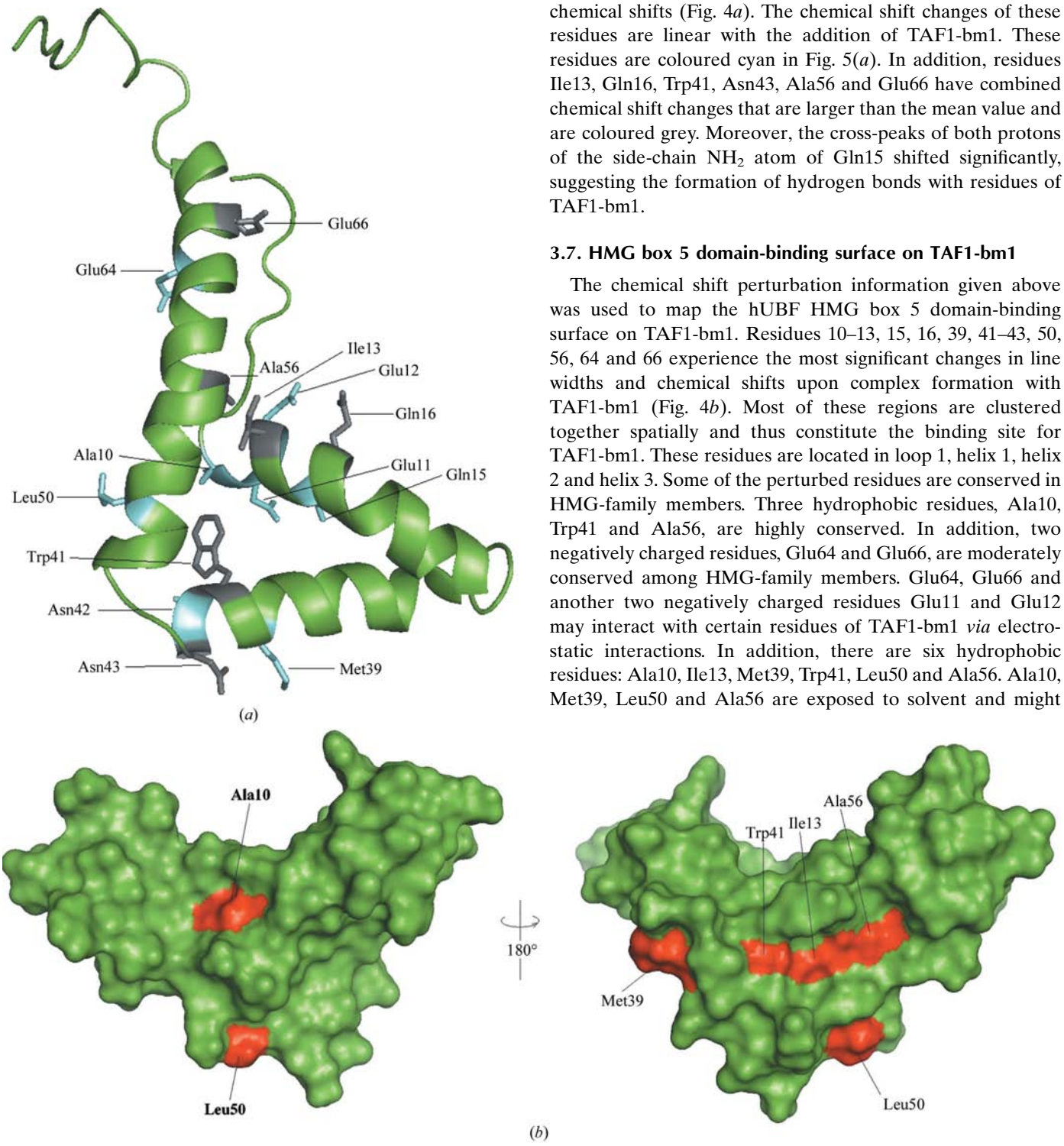

Figure 5
(*a*) Ribbon diagram showing the chemical shift perturbation of hUBF HMG box 5 (from NMR structure model 18) upon binding of TAF-bm1. The residues whose combined chemical shift changes are more than the mean value plus one standard deviation and those that are above the mean value are coloured cyan and grey, respectively. (*b*) Two views showing the surface features of hUBF HMG box 5 (from the crystal structure). The hydrophobic surface of hUBF HMG box 5 potentially involved in TAF-bm1 binding is coloured red. The right-hand view was obtained by rotating the left-hand view 180° around a vertical axis in the plane of the paper. All figures were prepared using *PyMOL* (http://www.pymol.org).

form hydrophobic interactions with TAF1-bm1. To identify functionally important amino acids on the surface of TAF-bm1, we used the *ConSurf* server (Glaser *et al.*, 2003 http://consurf.tau.ac.il). The server produced 14 conservative position spots (score = 9) including ten hydrophobic amino acids (Phe1405, Pro1408, Val1409, Tyr1417, Ile1421, Pro1424, Thr1429, Tyr1439, His1448 and Tyr1459). Most of these hydrophobic amino acids are distributed on the surface and spatially clustered together and thus may constitute a hydrophobic surface. It might be postulated that the hUBF HMG box 5 domain-binding surface on TAF1-bm1 is located on the hydrophobic concave side. Both hydrophobic and electrostatic interactions contribute to the specific recognition and binding of the hUBF HMG box 5 domain to TAF1-bm1.

Lys8 or Lys9, Glu11 or Glu12 and Arg30 in hUBF HMG box 5 correspond to the amino acids that other HMG boxes use to bind DNA. Lys8 and Lys9 are positively charged residues that cannot form hydrogen bonds to DNA like sequence-specific HMG boxes or form water-mediated hydrogen bonds to DNA like nonspecific HMG boxes. Glu11 and Glu12 are hydrophilic and cannot intercalate into DNA. The positively charged residue Arg30 is difficult to efficiently intercalate into DNA and might not form a hydrogen bond with DNA as in sequence-specific HMG boxes either (Yang *et al.*, 2003). Thus, in contrast to other hUBF HMG boxes, hUBF HMG box 5 almost shows no affinity for DNA (Yang *et al.*, 2003). This difference implies that the box 5 domain may play a different role in the activity of human upstream binding factor, functioning in the transcription of the rRNA gene. Recent reports (Bell & Tora, 1999; Bertolotti *et al.*, 1996; Saurin *et al.*, 2001) suggest that TAF1 is a regulatory protein that is involved in several aspects of Pol I and II transcription; the role of TAF1 in the transcription process is not yet well understood. The finding that hUBF HMG box 5, a domain in Pol I-specific *trans*-acting factor, interacts with TAF1-bm1 might provide more specific molecular mechanisms of UBF and TAF1 in the transcription process.

4. Conclusion

An L-shaped three-helix motif with a simple up-and-down topology is the core feature of hUBF HMG box 5. The crystal structure and the solution NMR show a similar overall folding. However, significant differences were observed between the two structures. The surface residues in the two structures showed large displacements (of up to 9.0 Å). The structures of the HMG domain of *D. melanogaster* protein D and domain A of the structure-specific HMG-domain protein HMG1 from *R. norvegicus* differ from the crystal structure of hUBF HMG box 5 by r.m.s. deviations of 1.5 and 1.6 Å, respectively, for the C^{α} atoms. This indicated that deviations between the NMR structure and crystal structure of hUBF HMG box 5 were larger from the those between hUBF HMG box 5 and other family members or homologues. The difference between the crystal and NMR structures reflects two potential states of the protein. Upon TAF1 binding, the structure of hUBF HMG box 5 might change its conformation, as revealed by a chemical shift perturbation study. Furthermore, we verified the direct interaction between the hUBF HMG box 5 domain and the TAF1-bm1 domain and identified a possible TAF1-bm1-binding site in hUBF HMG box 5. Our findings for the interaction between TAF1-bm1 and HMG box 5 might contribute to understanding the mechanism of RNA polymerase I transcription in cell-cycle regulation.

We thank Dr Yanwei Ding for his ultrasensitive differential scanning calorimetry work and Drs Yiwei Liu and Weili Yang for useful discussions. Financial support for this project to LN and MT was provided by research grants from the Chinese National Natural Science Foundation (grant Nos. 30121001, 30025012 and 30571066), the Chinese Ministry of Science and Technology (grant Nos. 2006CB806500 and 2006AA02A318) and the Chinese Academy of Sciences (grant No. KSCX1-YW-R-60).

References

Bachvarov, D. & Moss, T. (1991). *Nucleic Acids Res.* **23**, 2331–2335.
Bell, B. & Tora, L. (1999). *Exp. Cell Res.* **246**, 11–19.
Bell, S. P., Jantzen, H. M. & Tjian, R. (1990). *Genes Dev.* **4**, 943–954.
Bell, S. P., Learned, R. M., Jantzen, H. M. & Tjian, R. (1988). *Science*, **241**, 1192–1197.
Bertolotti, A., Lutz, Y., Heard, D. J., Chambon, P. & Tora, L. (1996). *EMBO J.* **15**, 5022–5031.
Comai, L., Tanese, N. & Tjian, R. (1992). *Cell*, **68**, 965–976.
Comai, L., Zomerdijk, J. C., Beckmann, H., Zhou, S., Admon, A. & Tjian, R. (1994). *Science*, **266**, 1966–1972.
Drenth, J. & Haas, C. (1992). *J. Cryst. Growth*, **122**, 107–109.
Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). *Bioinformatics*, **19**, 163–164.
Grummt, I. (1999). *Prog. Nucleic Acid Res. Mol. Biol.* **62**, 109–154.
Guerois, R., Nielsen, J. E. & Serrano, L. (2002). *J. Mol. Biol.* **320**, 369–387.
Holm, L. & Sander, C. (1996). *Nucleic Acids Res.* **24**, 206–210.
Jacobson, R. H., Ladurner, A. G., King, D. S. & Tjian, R. (2000). *Science*, **288**, 1422–1425.
Jantzen, H. M., Admon, A., Bell, S.P. & Tjian, R. (1990). *Nature (London)*, **344**, 830–836.
Jantzen, H. M., Chow, A. M., King, D. S. & Tjian, R. (1992). *Genes Dev.* **6**, 1950–1963.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Kleywegt, G. J. & Jones, T. A. (1996). *Acta Cryst.* D**52**, 826–828.
Kuhn, A., Voit, R., Stefanovsky, V., Evers, R., Bianchi, M. & Grummt, I. (1994). *EMBO J.* **13**, 416–424.
Lin, C. Y., Tuan, J., Scalia, P., Bui, T. & Comai, L. (2002). *Curr. Biol.* **12**, 2142–2146.
Love, J. J., Li, X., Case, D. A., Giese, K., Grosschedl, R. & Wright, P. E. (1995). *Nature (London)*, **376**, 791–795.
McStay, B., Frazier, M. W. & Reeder, R. H. (1991). *Genes Dev.* **5**, 1957–1968.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Murphy, F. V. IV, Sweet, R. M. & Churchill, M. E. (1999). *EMBO J.* **18**, 6610–6618.
Ohndorf, U. M., Rould, M. A., He, Q., Pabo, C. O. & Lippard, S. J. (1999). *Nature (London)*, **399**, 708–712.
O'Mahony, D. J. & Rothblum, L. I. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 3180–3184.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Perrakis, A., Morris, R. M. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.

Plotnikov, V. V., Brandts, J. M., Lin, L. N. & Brandts, J. F. (1997). *Anal. Biochem.* **250**, 237–244.
Reeder, R. (1999). *Prog. Nucleic Acid Res. Mol. Biol.* **62**, 293–327.
Saurin, A. J., Shao, Z., Erdjument-Bromage, H., Tempst, P. & Kingston, R. E. (2001). *Nature (London)*, **412**, 655–660.
Terwilliger, T. C. (2000). *Acta Cryst.* D**56**, 965–972.
Terwilliger, T. C. (2003). *Acta Cryst.* D**59**, 38–44.
Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* D**55**, 849–861.
Tuan, J. C., Zhai, W. & Comai, L. (1999). *Mol. Cell. Biol.* **19**, 2872–2879.
Wassarman, D. A. & Sauer, F. (2001). *J. Cell Sci.* **114**, 2895–2902.
Werner, M. H., Huth, J. R., Gronenborn, A. M. & Clore, G. M. (1995). *Cell*, **81**, 704–714.
Willard, L., Ranjan, A., Zhang, H., Monzavi, H., Boyko, R. F., Sykes, B. D. & Wishart, D. S. (2003). *Nucleic Acids Res.* **31**, 3316–3319.
Xu, Y., Yang, W., Wu, J. & Shi, Y. (2002). *Biochemistry*, **41**, 5415–5420.
Yang, W., Xu, Y., Wu, J., Zeng, W. & Shi, Y. (2003). *Biochemistry*, **42**, 1930–1938.