Purification, crystallization and X-ray diffraction analysis of the DNA-binding domain of human heat-shock factor 2

Han Feng,^{a,b} Wei Liu^c and Da-Cheng Wang^{a*}

Received 27 November 2015

Accepted 1 March 2016

Edited by R. A. Pauptit, Macclesfield, England

Keywords: heat-shock factor; DNA-binding domain; transcription; HSF2.

^aNational Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China, ^bUniversity of Chinese Academy of Sciences, Beijing 100049, People's Republic of China, and ^cInstitute of Immunology, The Third Military Medical University, Chongqing 400038, People's Republic of China.
*Correspondence e-mail: dcwang@ibp.ac.cn

Cells respond to various proteotoxic stimuli and maintain protein homeostasis through a conserved mechanism called the heat-shock response, which is characterized by the enhanced synthesis of heat-shock proteins. This response is mediated by heat-shock factors (HSFs). Four genes encoding HSF1–HSF4 exist in the genome of mammals. In this protein family, HSF1 is the orthologue of the single HSF in lower eukaryotic organisms and is the major regulator of the heat-shock response, while HSF2, which shows low sequence homology to HSF1, serves as a developmental regulator. Increasing evidence has revealed biochemical properties and functional roles that are unique to HSF2, such as its DNA-binding preference and sumoylation patterns, which are distinct from those of HSF1. The structural basis for such differences, however, is poorly understood owing to the lack of available mammalian HSF structures. The N-terminal DNA-binding domain (DBD) is the most conserved functional module and is the only crystallizable domain in HSFs. To date, only HSF1 homologue structures from yeast and fruit fly have been determined. Along with extensive studies of the HSF family, more structural information, particularly from members with a remoter phylogenetic relationship to the reported structures, *e.g.* HSF2, is needed in order to better understand the detailed mechanisms of HSF biology. In this work, the recombinant DBD (residues 7–112) from human HSF2 was produced in *Escherichia coli* and crystallized. An X-ray diffraction data set was collected to 1.32 Å resolution from a crystal belonging to space group $P2_12_12_1$ with unit cell-parameters $a = 65.66$, $b = 67.26$, $c = 93.25$ Å. The data-evaluation statistics revealed good quality of the collected data, thus establishing a solid basis for the determination of the first structure at atomic resolution in this protein family.

1. Introduction

The heat-shock response (HSR) is a cellular mechanism that is highly conserved throughout prokaryotes and eukaryotes to protect cells from various proteotoxic stimuli such as heavy metals, elevated temperatures, inflammation, hyperoxia and so on (Neef *et al.*, 2011; Shamovsky & Nudler, 2008). HSR triggered upon stress is characterized by an upshifted expression level of heat-shock proteins (HSPs), which function as molecular chaperones to maintain protein homeostasis in the cell (Åkerfelt *et al.*, 2010). HSP synthesis is mediated at the transcriptional level by heat-shock factors (HSFs), which specifically recognize and bind to a conserved nucleotide sequence present in multiple copies within the promoter region of HSP-encoding genes. These *cis*-acting sequences, called heat-shock elements (HSEs), are composed of inverted

© 2016 International Union of Crystallography

repeats of the pentameric sequence nGAAn (Amin *et al.*, 1988). An intact HSE consists of multiple contiguous inverted repeats with two alternative directions, which are referred to as the head-to-head orientation (nGAAnnTTCn) or the tail-to-tail orientation (nTTCnnGAAn) (Littlefield & Nelson, 1999). In eukaryotes, HSF forms a trimer upon activation and optimally binds three nGAAn repeats, each of which is bound by one subunit (Björk & Sistonen, 2010).

A single HSF-encoding gene exists in the genome of lower eukaryotic organisms such as yeast and fruit fly, while in mammals a family of HSFs consisting of four members with diverse cellular roles, HSF1–HSF4, are expressed (Björk & Sistonen, 2010; Pirkkala *et al.*, 2001). Similar to many eukaryotic transcription factors, all mammalian HSFs contain several functional modules in their amino-acid sequences. From the N-terminus to the C-terminus, these are the DNA-binding domain (DBD), trimerization domain, regulatory domain and transactivation domain (Björk & Sistonen, 2010; Pirkkala *et al.*, 2001; Westerheide *et al.*, 2012). The DBD is the most conserved and is the only domain to have been structurally characterized. Both crystal and solution structures of HSF-DBD from *Kluyveromyces lactis* and *Drosophila melanogaster* were determined in the 1990s, revealing a winged helix–turn–helix fold of this domain (Harrison *et al.*, 1994; Vuister *et al.*, 1994). Since then, however, no structure from any other organism has been reported.

In the mammalian HSF family, HSF1 is the orthologue of the single, essential HSF in *Saccharomyces cerevisiae* and *D. melanogaster* and is the major regulator responsible for the stress-induced expression of HSPs (Björk & Sistonen, 2010; Westerheide *et al.*, 2012). HSF2, which shows lower sequence similarity, was originally discovered in tandem with HSF1 and is capable of inducing the transcription of HSPs. Further experimental data, however, supported the function of HSF2 as a developmental regulator rather than a major mediator of the heat-shock response like HSF1 (Björk & Sistonen, 2010; Westerheide *et al.*, 2012). In consistence with the functional divergence, HSF2 shows a different DNA-binding pattern from HSF1, although both factors recognize the nGAAn repeat through their DBDs. HSF1 usually binds nucleotide sequences composed of long arrays of HSE units, *e.g.* the *hsp70* promoter, but HSF2 preferentially binds those comprising shorter HSE arrays, *e.g.* the *CUP1* promoter (Kroeger & Morimoto, 1994). This variance indicates that some structural differences must exist between them in the DBD, and indeed a loop following the helix–turn–helix motif was suggested to be responsible for their different DNA-binding preferences (Ahn *et al.*, 2001). From the biochemical viewpoint, HSF2 also displays properties that are distinct from those of HSF1. HSF2 exists in the cytosol as a dimer, unlike the monomeric form of HSF1 (Björk & Sistonen, 2010; Pirkkala *et al.*, 2001). More interestingly, sumoylation, which is an important post-translational modification to mediate the transactivity of HSFs, occurs on Lys82 only in HSF2-DBD, although HSF1 contains a lysine residue at the same position (Goodson *et al.*, 2001). A modelling study suggested that two amino acids neighbouring Lys82, which forms a sharp turn in

the wing of HSF2-DBD, might serve as the structural determinant of sumoylation (Anckar *et al.*, 2006), but further experimental evidence including structural information on HSF2-DBD is needed to confirm this premise.

It should be noted that recent progress in research on HSF2 has started to change the conventional understanding of this regulator. Several lines of evidence have demonstrated that HSF2 can actively participate in transcriptional regulation of constitutively and stress-induced expression of HSPs through interplay with HSF1 (Åkerfelt *et al.*, 2010; Björk & Sistonen, 2010). These two factors can directly bind each other and form a transcriptionally competent heterotrimer upon stress stimuli or during development (Ostling *et al.*, 2007; Sandqvist *et al.*, 2009). Such cooperation may provide an efficient mechanism to fine-tune the transcription of their common target genes, but both structural and mechanistic insights into HSF1–HSF2 heterotrimerization are very limited (Björk & Sistonen, 2010; Åkerfelt *et al.*, 2010).

During the past decade functional studies of HSFs have been extensively carried out, but structural studies, which would provide more of the information that is required for better understanding of this protein family, have failed to keep pace. In this context, we report here the purification, crystallization and preliminary X-ray diffraction analyses of the DNA-binding domain (residues 7–112) of human HSF2, which is the first mammalian HSF-DBD to be crystallized.

2. Materials and methods

2.1. Macromolecule production

The gene encoding full-length human HSF2 was a generous gift from Professor Lea Sistonen at Åbo Akademi University in Finland. The nucleotide sequence corresponding to the DBD (residues 7–112) was subcloned into a pMAL-c4E plasmid (New England Biolabs) for the production of recombinant protein fused to maltose-binding protein (MBP) at the N-terminus. A *Tobacco etch virus* (TEV) protease cleavage site followed by a His₆ tag was introduced between the MBP tag and HSF2-DBD by PCR amplification. The correctness of the inserted sequence was confirmed by DNA sequencing. The *Escherichia coli* host strain used in this experiment was prepared by transforming the plasmid pRK603, which contains a TEV protease-encoding gene, into the Rosetta2 strain (Novagen, Darmstadt, Germany) before making the cells competent. The recombinant plasmid pMAL-HSF2-DBD was subsequently transformed into this strain. The bacteria were grown in LB medium containing 100 mg ml^{−1} ampicillin, 50 mg ml^{−1} kanamycin and 34 mg ml^{−1} chloramphenicol until the OD₆₀₀ approached 0.5. Expression of the fusion protein was induced with 0.3 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 4 h at 303 K. Immediately after this, intracellular production of the TEV protease was induced with anhydrotetracyclin hydrochloride (aTet) at a final concentration of 100 ng ml^{−1} for a further 2 h at 303 K, during which the MBP tag was cleaved *in vivo* from the fusion protein. The bacteria were harvested by centrifugation at

Table 1
Macromolecule-production information.

Source organism	<i>Homo sapiens</i>
DNA source	GeneBank AAA36017.1
Forward primer	GTGCCGGCTTTCCTCAGCAAGCTGT
Reverse primer	ATGATGATGATGATGATGGCCCTGGAAGTATA
Cloning vector	pMAL-c4E
Expression vector	pMAL-c4E
Expression host	<i>E. coli</i> Rosetta2
Complete amino-acid sequence of the construct produced	GHHHHHHPAFLSKLWTLVEETHTNEFITWSQNG-QSFLVLDEQRFKEILPKYFKHNNMASFVRQL-NMYGFRKVVHIDSGIVKQERDGPVEFQHPYFK-QGQDDLLENIKRKVS

5000g for 30 min at 277 K and resuspended in lysis buffer consisting of 50 mM NaH₂PO₄/Na₂HPO₄ pH 8.0, 500 mM NaCl, 25 mM imidazole before cell lysis using a high-pressure crusher at 277 K.

After centrifugation at 20 000g for 30 min at 277 K, the supernatant of the cell lysate was loaded onto an Ni²⁺-NTA chromatography column (Novagen, Darmstadt, Germany), followed by washing the column with washing buffer (50 mM NaH₂PO₄/Na₂HPO₄ pH 8.0, 500 mM NaCl, 40 mM imidazole, 5% glycerol). The recombinant HSF2-DBD protein with an N-terminal His₆ tag eluted when the imidazole concentration was increased to 200 mM. The protein was further purified by cation-exchange chromatography using a HiTrap SP HP 5 ml column (GE Healthcare, Uppsala, Sweden), eluting with a linear gradient of sodium chloride from 100 to 1000 mM, and size-exclusion chromatography using a HiLoad 16/60 Superdex 75 column (GE Healthcare, Uppsala, Sweden) equilibrated with buffer consisting of 20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM DTT, 0.2 mM EDTA. The purified protein was concentrated to 40 mg ml⁻¹ (measured from the OD₂₈₀) and stored at 193 K. Macromolecule-production information is summarized in Table 1.

Table 2
Crystallization.

Method	Hanging-drop vapour diffusion
Plate type	16-well plates
Temperature (K)	293
Protein concentration (mg ml ⁻¹)	15
Buffer composition of protein solution	20 mM Tris-HCl pH 8.0, 150 mM NaCl, 1 mM DTT, 0.2 mM EDTA
Composition of reservoir solution	0.1 M Tris-HCl pH 8.4, 1.25 M potassium sodium tartrate
Volume and ratio of drop	2 µl, 1:1
Volume of reservoir (µl)	500

2.2. Crystallization

The protein concentration was adjusted to 15 mg ml⁻¹ before crystallization. All crystallization trials were carried out using the hanging-drop vapour-diffusion method at 293 K. Initial crystallization conditions were screened using ten commercial kits from Hampton Research (California, USA) with drops set up using a Mosquito crystallization robot (TTP Labtech). The best condition found from the screening was 0.1 M Tris-HCl pH 8.5, 1.2 M potassium sodium tartrate, under which small single crystals were grown. Optimization around this condition was subsequently performed by hand, with each drop formed by mixing 1 µl protein solution and 1 µl reservoir solution and equilibrated against 500 µl reservoir solution. The optimum condition for the growth of single crystals of HSF2-DBD was 0.1 M Tris-HCl pH 8.4, 1.25 M potassium sodium tartrate. Crystallization information is summarized in Table 2.

2.3. Data collection and processing

The crystals used for data collection were directly mounted in nylon cryoloops (Hampton Research, California, USA) and flash-cooled in a stream of liquid nitrogen at 100 K. Diffrac-

Figure 1
Production of HSF2-DBD with different constructs detected by SDS-PAGE. Protein purification from constructs consisting of residues 1–112 and 7–112 is shown in (a) and (b), respectively. (a) Lanes 1–3, washed, eluted and flowthrough fractions from the Ni²⁺-NTA column; lane 4, eluate from the SP column; lane 5, eluate from the Superdex 75 column. (b) Lane 1, eluted fraction from the Ni²⁺-NTA column; lane 2, eluate from the SP column; lane 3, eluate from the Superdex 75 column. The molecular masses of the protein markers (in kDa) are given on the right-hand sides of the gels.

Table 3

Data collection and processing.

Values in parentheses are for the outer shell.

Diffraction source	BL19U1, SSRF
Wavelength (Å)	0.978455
Temperature (K)	100
Detector	Pilatus3 6M
Crystal-to-detector distance (mm)	200
Rotation range per image (°)	360
Total rotation range (°)	1.0
Exposure time per image (s)	0.1
Space group	$P2_12_12_1$
a, b, c (Å)	65.66, 67.26, 93.25
α, β, γ (°)	90, 90, 90
Mosaicity (°)	0.45
Resolution range (Å)	38.32–1.32 (1.39–1.32)
Total No. of reflections	97322
No. of unique reflections	24312
Completeness (%)	99.8 (99.8)
Multiplicity	12.0 (11.2)
$\langle I/\sigma(I) \rangle$	18.4 (3.6)
$R_{\text{r.i.m.}}$ (%)	6.6 (65.7)
Overall B factor from Wilson plot (Å ²)	15.2

tion data were collected at a wavelength of 0.97845 Å on beamline BL19U1 at Shanghai Synchrotron Radiation Facility (SSRF), People's Republic of China using a Pilatus3 6M detector. A total of 360 images were collected with 1° oscillation and 0.1 s exposure time per image. The data collected were indexed, integrated and scaled using *iMosflm* (Battye *et al.*, 2011) and *SCALA* from the *CCP4* program suite (Winn *et al.*, 2011). Statistics of data collection are summarized in Table 3.

3. Results and discussion

At the beginning of this work, we were not sure of the exact boundary of the DBD in human HSF2, and initially made a construct from residues 1 to 112 according to a secondary-structure prediction. After purification from an Ni²⁺-NTA column, a major band at 14 kDa was observed on an SDS gel (Fig. 1a) which agreed with the estimated molecular mass of

HSF2-DBD after *in vivo* removal of the MBP tag. After second and third chromatographic steps using an SP column and a Superdex 75 column, the final purity of the recombinant protein was higher than 90% as detected by SDS-PAGE (Fig. 1a). Needle-shaped crystals were observed under a single condition from approximately 500 initial conditions using commercial screening kits. Optimization of the precipitant concentration, pH and additive composition led us to obtain rod clusters at 0.2 M potassium sodium tartrate, 0.1 M sodium citrate pH 5.5, 2.0 M ammonium sulfate (Fig. 2a). Even though thorough optimization of the crystallization conditions was performed, no further improvement could be obtained. We then attempted seeding in order to grow single crystals. By using the dynamic seeding technique implemented in our laboratory (Zhu *et al.*, 2005), rod-shaped single crystals with approximate dimensions of 0.4 × 0.05 × 0.05 mm were grown using 0.2 M potassium sodium tartrate, 0.1 M sodium citrate pH 5.4, 2.2 M ammonium sulfate (Fig. 2b).

Although sufficiently large for X-ray diffraction at a synchrotron source, these crystals diffracted to merely 4–5 Å resolution (Fig. 3a). It is known that post-crystallization treatments, such as annealing, dehydration and cross-linking, can improve the diffraction quality of protein crystals (Heras & Martin, 2005), and indeed a number of cases of dramatic diffraction improvement have been reported (Haebel *et al.*, 2001; Hunsicker-Wang *et al.*, 2005; Kriminski *et al.*, 2002). Inspired by this, we attempted all of these techniques using mild operations as described in the literature. Unfortunately, no improvement in the diffraction of our crystals was achieved; in contrast, the crystals became more vulnerable to X-rays after annealing or cross-linking.

Even so, the irregular dispersion spots observed on the diffraction images (Fig. 3a) suggested that some kind of disorder might exist in the crystal. On realising this, we considered making changes to the expression construct. From another secondary-structure prediction using *Jpred4* (Drozdetskiy *et al.*, 2015), which combines *PSI-BLAST* and several predictive algorithms, we found that the N-terminal six amino

Figure 2

Crystals of HSF2-DBD obtained using constructs comprising residues 1–112 (*a, b*) and 7–112 (*c*). (*a*) Clustered rod-shaped crystals obtained using 0.2 M potassium sodium tartrate, 0.1 M sodium citrate pH 5.5, 2.0 M ammonium sulfate. (*b*) Single rod-shaped crystals grown after dynamic seeding using 0.2 M potassium sodium tartrate, 0.1 M sodium citrate pH 5.4, 2.2 M ammonium sulfate. (*c*) Rectangular crystals grown using 0.1 M Tris pH 8.4, 1.25 M potassium sodium tartrate.

Figure 3
Diffraction images of HSF2-DBD obtained on beamline BL19U1 at SSRF from constructs comprising residues 1–112 (a) and 7–112 (b).

acids in the sequence of human HSF2 might form an unstructured extension from the DBD and are likely to contribute to the poor diffraction quality of our crystals. A shorter construct from residues 7 to 112 was then produced. Using the same expression and purification protocol, recombinant HSF2-DBD from this construct could be produced with comparable purity and yield to the previous construct (Fig. 1b). Using this truncated protein, we performed another round of crystallization screening and obtained microcrystals or small single crystals under a number of conditions. The best crystals were grown using 0.1 M Tris–HCl pH 8.5, 1.2 M potassium sodium tartrate. Subsequent optimization around this condition was undertaken. Using the optimum condition consisting of 0.1 M Tris–HCl pH 8.4, 1.25 M potassium sodium tartrate, rectangular crystals with approximate dimensions of $0.2 \times 0.1 \times 0.1$ mm were obtained (Fig. 2c) which showed super diffraction quality even without the requirement for cryoprotection (Fig. 3b).

A complete diffraction data set with 360° rotation was collected to 1.32 Å resolution on beamline BL19U1 at SSRF. The crystal belonged to the orthorhombic space group $P2_12_12_1$, with unit-cell parameters $a = 65.66$, $b = 64.42$, $c = 93.25$ Å, $\alpha = \beta = \gamma = 90^\circ$. Four monomers of HSF2-DBD are likely to reside in the asymmetric unit, as estimated from the value of the Matthews coefficient (Matthews, 1968) with the highest probability ($1.84 \text{ Å}^3 \text{ Da}^{-1}$; solvent content of 33.10%). Structure determination will be performed by molecular replacement using the yeast HSF-DBD (PDB entry 3hts; Littlefield & Nelson, 1999) as the search model.

To date, only HSF-DBDs from *K. lactis* and *D. melanogaster* have been structurally characterized. Mammalian HSF1 is orthologous to those HSFs, but HSF2 showed a lower sequence similarity to them. In this context, crystallization of

the DNA-binding domain from human HSF2 established a good starting point for structure determination, which is likely to provide novel structural insights into the HSF family. In another aspect, our experience in this work highlights the importance of the expression constructs, which may greatly influence the diffraction quality of crystals even though single crystals could be grown from different constructs.

Acknowledgements

This work was supported by the National Natural Science Foundation grant 31270288 in China. We are grateful to the staff of beamline BL19U1 at SSRF for technical assistance during data collection.

References

- Ahn, S.-G., Liu, P. C. C., Klyachko, K., Morimoto, R. I. & Thiele, D. J. (2001). *Genes Dev.* **15**, 2134–2145.
- Åkerfelt, M., Morimoto, R. I. & Sistonen, L. (2010). *Nature Rev. Mol. Cell Biol.* **11**, 545–555.
- Amin, J., Ananthan, J. & Voellmy, R. (1988). *Mol. Cell. Biol.* **8**, 3761–3769.
- Anckar, J., Hietakangas, V., Denessiouk, K., Thiele, D. J., Johnson, M. S. & Sistonen, L. (2006). *Mol. Cell. Biol.* **26**, 955–964.
- Battye, T. G. G., Kontogiannis, L., Johnson, O., Powell, H. R. & Leslie, A. G. W. (2011). *Acta Cryst.* **D67**, 271–281.
- Björk, J. K. & Sistonen, L. (2010). *FEBS J.* **277**, 4126–4139.
- Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. (2015). *Nucleic Acids Res.* **43**, W389–W394.
- Goodson, M. L., Hong, Y., Rogers, R., Matunis, M. J., Park-Sarge, O.-K. & Sarge, K. D. (2001). *J. Biol. Chem.* **276**, 18513–18518.
- Haebel, P. W., Wichman, S., Goldstone, D. & Metcalf, P. (2001). *J. Struct. Biol.* **136**, 162–166.
- Harrison, C. J., Bohm, A. A. & Nelson, H. C. (1994). *Science*, **263**, 224–227.
- Heras, B. & Martin, J. L. (2005). *Acta Cryst.* **D61**, 1173–1180.

- Hunsicker-Wang, L. M., Pacoma, R. L., Chen, Y., Fee, J. A. & Stout, C. D. (2005). *Acta Cryst.* **D61**, 340–343.
- Kriminski, S., Caylor, C. L., Nonato, M. C., Finkelstein, K. D. & Thorne, R. E. (2002). *Acta Cryst.* **D58**, 459–471.
- Kroeger, P. E. & Morimoto, R. I. (1994). *Mol. Cell. Biol.* **14**, 7592–7603.
- Littlefield, O. & Nelson, H. C. (1999). *Nature Struct. Biol.* **6**, 464–470.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Neef, D. W., Jaeger, A. M. & Thiele, D. J. (2011). *Nature Rev. Drug Discov.* **10**, 930–944.
- Ostling, P., Björk, J. K., Roos-Mattjus, P., Mezger, V. & Sistonen, L. (2007). *J. Biol. Chem.* **282**, 7077–7086.
- Pirkkala, L., Nykänen, P. & Sistonen, L. (2001). *FASEB J.* **15**, 1118–1131.
- Sandqvist, A., Björk, J. K., Akerfelt, M., Chitikova, Z., Grichine, A., Vourc'h, C., Jolly, C., Salminen, T. A., Nymalm, Y. & Sistonen, L. (2009). *Mol. Biol. Cell*, **20**, 1340–1347.
- Shamovsky, I. & Nudler, E. (2008). *Cell. Mol. Life Sci.* **65**, 855–861.
- Vuister, G. W., Kim, S.-J., Orosz, A., Marquardt, J., Wu, C. & Bax, A. (1994). *Nature Struct. Biol.* **1**, 605–614.
- Westerheide, S. D., Raynes, R., Powell, C., Xue, B. & Uversky, V. N. (2012). *Curr. Protein Pept. Sci.* **13**, 86–103.
- Winn, M. D. *et al.* (2011). *Acta Cryst.* **D67**, 235–242.
- Zhu, D.-Y., Zhu, Y.-Q., Xiang, Y. & Wang, D.-C. (2005). *Acta Cryst.* **D61**, 772–775.