Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Purification, crystallization and preliminary X-ray crystallographic studies of *S. cerevisiae* Hsp40 Sis1

Bingdong Sha[a]* and Douglas Cyr[b]

[a]Center for Macromolecular Crystallography, Department of Cell Biology, University of Alabama at Birmingham, Birmingham, AL 35294-0005, USA, and [b]Department of Cell Biology, University of Alabama at Birmingham, Birmingham, AL 35294-0005, USA

Correspondence e-mail: sha@orion.cmc.uab.edu

Received 5 January 1999
Accepted 30 March 1999

Heat-shock protein 70 (Hsp70), one of the major molecular chaperones, has been shown to play a central role in many cellular processes. Heat-shock protein 40 (Hsp40) works as a co-chaperone for Hsp70. Hsp40, bound by unfolded polypeptide, can interact directly with Hsp70 to stimulate the ATPase activity of Hsp70. Hsp40 can also bind to unfolded polypeptides and prevent them from aggregating *in vitro*, thus acting as an independent molecular chaperone. The *S. cerevisiae* Hsp40 Sis1 C-terminal peptide-binding domain has been crystallized. The crystals diffract to 2.7 Å and belong to space group $P4_12_12$ or $P4_32_12$ with $a = 73.63$, $c = 80.16$ Å. The structure determination by the MAD method is under way.

1. Introduction

A large group of proteins named 'molecular chaperones' play a central role in protein refolding and protein turnover (Rothman, 1989; Gething & Sambrook, 1992). Molecular chaperones can bind to and stabilize non-native proteins, help protein folding, oligomeric assembly, transport to a particular subcelluar compartment or disposal by degradation. The heat shock protein 70 (Hsp70) family is a major component of molecular chaperones. During translation or translocation, the Hsp70s and their partner proteins can stabilize nascent polypeptides until all segments of the chain necessary for folding are available. They play essential roles in protein metabolism under both stress and non-stress conditions, including protein folding and assembly, membrane translocation, uncoating of clathrin-coated vesicles and degradation of incorrectly folded proteins (Hartl, 1996; Frydman *et al.*, 1994; Clarke, 1996; Craig *et al.*, 1994; Ungermann *et al.*, 1994; Ungewickell *et al.*, 1995).

Hsp70s do not function alone. The peptide binding and release of Hsp70 are modulated by other proteins, most notably heat-shock protein 40 (Hsp40). DnaK, the Hsp70 protein of *E. coli*, is regulated by DnaJ, which is the *E. coli* Hsp40, and another cofactor GrpE. In eukaryotic cells, the mechanism for Hsp70/Hsp40 is quite different from prokaryotic system. Homologues of DnaJ exist in all compartments which contain Hsp70 (Caplan *et al.*, 1993; Ungewickell *et al.*, 1995). However, no eukaryotic GrpE homologue has been identified in eukaryotic cell cytosol. ADP dissociation is the rate-limiting step in the DnaK ATPase cycle resulting in the dependence of the prokaryotic system on an additional nucleotide exchange factor, GrpE. In the case of the eukaryotic system, since the spontaneous release of ADP from Hsp70 is more rapid than from DnaK, Hsp40 by itself can stimulate the ATPase activity of Hsp70 by up to 15-fold and, therefore, makes GrpE-like protein dispensable. In the eukaryotic system, a model has been proposed for the Hsp70 reaction cycle based on *in vitro* studies (Hohfeld *et al.*, 1995). In the absence of Hsp40, Hsp70 is predominantly in the ATP-bound form, as the intrinsic ATPase activity of Hsp70 is low. In the presence of ATP, Hsp70 alone does not form a stable complex with unfolded polypeptide (step I). Hsp40 binds to the peptide first and then interacts with Hsp70. By so doing, the ATPase activity of Hsp70 could be elevated by a factor of up to 15 (step II). Then the ADP-state Hsp70 can bind to the unfolded polypeptide presented by Hsp40 and form an Hsp70–Hsp40–unfolded polypeptide complex (step III). ADP dissociation occurs quickly and may break the complex (step IV). The released unfolded polypeptide is then subjected to the next Hsp70/Hsp40 cycle until it is renatured. The slower rate of Hsp70 cycling (15-fold compared with 50-fold in *E. coli*), activated by Hsp40 alone, may explain the optimal stabilization of translating polypeptides in the eukaryotic chains. Depletion of either Hsp40 or Hsp70 from the system greatly reduces the folding ability (Frydman *et al.*, 1994).

Eukaryotic cells use the Hsp70/Hsp40 pairs not only to protect a range of aggregation-prone unfolded polypeptides, but also for specialized cellular functions. It has been suggested that specialized Hsp40 proteins interact with Hsp70 to form functionally distinct chaperone pairs that catalyze specific cellular processes (James *et al.*, 1997; Lu & Cyr,

1999). *S. cerevisiae* has at least two DnaJ homologues which are localized in the cytosol: Sis1 and Ydj1. *SIS*1 is an essential gene for cell viability. It was first identified as a high copy number suppressor of the slow-growth phenotype of strains containing mutations in the *SIT*4 gene, which encodes a predicted serine/threonine phosphatase (Luke *et al.*, 1991). Later it was found that Sis1 could bind to ribosomes and was required for the normal initiation of translation (Zhong & Arndt, 1993). The yeast Hsp70 homologues Ssb1 and Ssb2 have been shown to be components of the ribosome-nascent chain complex (Nelson *et al.*, 1992; Pfund *et al.*, 1998), it is possible that Sis1 cooperates with Ssb1 and Ssb2 during translation process *in vivo*. Coupling with Ssb1, Sis1 is also involved in modulation of intracellular protein degradation in yeast (Ohba, 1997). Sis1 can also stimulate the ATPase activity of Ssa1 and refold denatured protein along with Ssa1 *in vitro* (Lu & Cyr, 1999). Overexpression of Sis1 can suppress the deletion of another yeast cytosolic Hsp40 Ydj1 while overexpression of Ydj1 cannot cure the lethality resulting from the deletion of Sis1 (Luke *et al.*, 1991).

Analysis of *E. coli* Hsp40 protein DnaJ suggested that DnaJ can be divided into four regions: a J domain, a G/F region, a zinc-finger domain and a peptide-binding domain (Silver & Way, 1993). The 70-amino-acid residue N-terminal J domain is conserved in all Hsp40 proteins and is the hallmark of this family. Adjacent to the J domain is a sequence rich in glycine residues and phenylalanine residues (the G/F region) which is thought to be a 'hinge' region, lacking secondary structure. The zinc-finger domain is also named the cysteine-rich repeat which is comprised of 150 amino-acid residues with four C*x*x*C*x*G*x*G motifs in it (*x* is a usually charged or polar residue). The distribution of cycteine residues is similar to that found in some zinc-binding proteins and might form two zinc-finger sub-structures. The C-terminal peptide binding domain is composed of approximately 200 residues and can bind the unfolded polypeptide as tightly as the full-length protein (Lu & Cyr, 1999). *S. cerevisiae* Hsp40 protein Ydj1 has all of the four domains that present in DnaJ. Sis1, however, lacks the zinc-finger domain.

2. Experimental

2.1. Cloning, expression and purification of Sis1

S. cerevisiae Hsp40 Sis1 was cloned into vector pet15b (Novagen). PCR was carried out using a 5′ primer GGAATTCCATATGGTCAAGGAGACAAAACTT and a 3′ primer CGCGGATCCTTAAAAATTTTCATCTATAGCA. The PCR products were digested by restriction endonucleases *Nde*I and *Bam*HI according to the instructions (New England Biolabs). The inserts were then ligated into the digested pet15b vector by T4 ligase. The nucleotide sequence was confirmed by DNA sequencing. The plasmid encoding Sis1 was finally transformed into *E. coli* strain BL21(DE3) for expression.

1 l LB medium was inoculated using the transformed *E. coli* stocks. The cells were allowed to grow at 310 K in a shaker for 12 h. The Sis1 protein could be expressed to high yield without IPTG induction. We do not fully understand the mechanism of this; it may be because the trace amount of lactose in the medium acts as a inducer for the *lac* gene. The *E. coli* cells from 1 l medium were pelleted down by centrifugation and resuspended into 100 ml 10 m*M* phosphate buffer (pH 7.2), 150 m*M* NaCl. The cell walls and membranes were broken down by sonication. The debris and insoluble materials were pelleted down at 15 000 rev min^{-1} using a Beckman JA10 rotor and the supernatant was collected. Since the Sis1 protein was histine tagged, it could be relatively easily purified by metal-chelating column. The metal-chelating beads (Pharmacia) were charged by 50 m*M* $NiSO_4$ and washed by an extensive amount of 10 m*M* phosphate buffer (pH 7.2), 150 m*M* NaCl. The collected supernatant was pumped through the Ni-charged column and the protein-bound beads were washed thoroughly using 10 m*M* phosphate buffer (pH 7.2), 0.5 *M* NaCl and 50 m*M* imidazole to remove the contaminated protein. The Sis1 protein was then eluted by 10 m*M* phosphate buffer (pH 7.2), 0.5 *M* NaCl and 200 m*M* imidazole. The eluted protein was dialyzed against 10 m*M* phosphate buffer (pH 7.2), 150 m*M* NaCl. The typical yield of Sis1 protein from 1 l medium is ∼20 mg. The Sis1 protein was further purified by a gel-filtration column of Superdex 200 (Pharmacia) mounted on an ACTA HPLC system (Pharmacia). The apparent molecular weight of Sis1 is shown to be ∼80 kDa from the elution time of the protein peak, indicating that Sis1 may form a dimer in solution.

2.2. Limited proteolysis of Sis1

We were not able to grow the crystals of full-length Sis1. The G/F region rich in glycine residues may be very flexible and interfere with protein crystallization. The full-length protein was then subjected to limited proteolysis. The reason for this was to remove the flexible G/F region and the J domain in order to obtain a stable C-terminal peptide-binding domain. Various amounts of proteinase K (Sigma) were tested for full-length Sis1 protein at a concentration of 1 mg ml^{-1} in phosphate buffer (pH 7.2), 150 m*M* NaCl. The digestion was carried out at 298 K for 1 h and the reaction was quenched by addition of 0.2 m*M* PMSF (Sigma). It was found that 0.05 µg ml^{-1} proteinase K would digest the full-length 1 mg ml^{-1} Sis1 protein to a stable fragment with molecular weight of ∼20 kDa. A mass spectrum showed that the accurate molecular weight of the fragment was 20 454 Da. N-terminal microsequencing of the fragment revealed that the first five amino-terminal residues of the fragment are S, S, S, P and T. On the basis of these results, we reasoned that the fragment is a C-terminal fragment of Sis1 from residue 171 to residue 352. The fragment was further purified by a gel-filtration column of Superdex 200 (Pharmacia). The elution time of the protein peak from the column was ∼80 min after sample injection at a flow rate of 1 ml min^{-1}. The apparent molecular weight of the Sis1 C-terminal fragment was then about 40 kDa indicating that it may still form a homodimer in solution. Typically, we can obtain 3 mg purified protein from 1 l expression culture.

2.3. Crystallization, data collection and processing of Sis1

The Sis1 C-terminal fragment was concentrated to 10 mg ml^{-1} in 10 m*M* phosphate buffer (pH 7.2), 150 m*M* NaCl and subjected to crystallization trials. Large diffraction-quality crystals (0.8 × 0.8 × 0.8 mm) were obtained within 2 d by the hanging-drop vapor-diffusion method using Linbro boxes. The well solution consisted of 1 ml 100 m*M* phosphate buffer (pH 8.15), 15% PEG 3350. The hanging drops had 3 µl protein plus 3 µl well solution. The crystals of Sis1 can be stored in 100 m*M* phosphate buffer (pH 8.15), 20% PEG 3350 for several months without any damage. We then questioned if the crystallized protein was bound with peptide. The crystals were collected, washed several times and then dissolved in water. The sample was then subjected to mass-spectrum measurement (data not shown). The results from the mass spectrum showed only one peak for the protein and no apparent peaks for the peptides. We concluded that the Sis1 C-

Table 1
Statistics of the native data set for Sis1 from SSRL beamline 7-1 synchrotron.

Resolution shells (Å)	I/σ	R_{symm}	Completeness (%)
30.0–4.90	67.5	0.021	91.8
4.90–3.89	79.4	0.030	96.6
3.89–3.40	49.2	0.053	97.3
3.40–3.09	24.9	0.107	98.6
3.09–2.87	11.8	0.220	98.9
2.87–2.70	6.17	0.377	99.5

terminal fragment which we crystallized was not bound with peptides.

The crystals are so sensitive to X-rays that freezing was necessary in order to collect a full data set from one single crystal. The crystal was transferred into the cryoprotectant containing 100 m*M* phosphate buffer (pH 8.15), 20% PEG 3350 and 20% glycerol for 5 min before it was immediately frozen at 100 K in a nitrogen gas stream using an Oxford Cryosystems Cryostream. The in-house data were collected on an R-axis IV image plate. The crystals of Sis1 diffracted to 3.2 Å using the in-house Rigaku rotating-anode X-ray generator with a mirror system. Later the crystals were taken to SSRL station beamline 7-1 (wavelength = 1.08 Å) and a 2.7 Å native data set was collected using a MAR Research image plate. The crystals were also frozen using the cryo conditions above. The crystal-to-detector distance was 200 mm and the oscillation angle was 1.5°. 50 images were collected and processed by *DENZO* and *SCALEPACK* (Minor, 1993; Otwinowski, 1993). The crystals belong to space group $P4_12_12$ or $P4_32_12$ with a = 73.63, c = 80.16 Å. The data were about 97.0% complete to 2.7 Å and R_{symm} was 0.045, the redundancy for the data is 5.5. The data statistics are shown in Table 1. The V_m value is 2.50 Å Da^{-1} for one monomer per asymmetric unit. The mosaicity for the best crystals after freezing is 0.25. The total number of unique reflections is 6231.

In order to determine the crystal structure of the Sis1 C-terminal fragment by the MAD method, the plasmid encoding Sis1 was transformed into *E. coli* strain B834(DE3) which is a methionine auxotroph. The cells were grown in the medium which consists of 2 × M9 plus glucose minimal medium supplemented with 19 amino acids (methionine not included) at 40 µg ml^{-1}, seleno-L-methionine at 40 µg ml^{-1} and vitamins (thiamine, riboflavin and pyridoxine) at 1 µg ml^{-1}. The protein purification, limited proteolysis and crystallization were carried out using the same conditions as for the native Sis1 protein. A mass spectrum of the selenomethionyl Sis1 C-terminal fragment indicated that the selenium had been incorporated into the protein. The structure determination of Sis1 C-terminal fragment by MAD is under way.

3. Discussion

Hsp40 is a chaperone family with unique features. Firstly, Hsp40 and Hsp70 differ in unfolded polypeptide-binding specificity. The unfolded peptide binds to Hsp70 in a fully extended conformation through interactions between the protein and peptide backbones (Zhu *et al.*, 1996). Synthetic peptide studies show that Hsp70 binds peptides with a minimum length of seven amino acids (Flynn *et al.*, 1991). Hsp40, on the other hand, probably binds protein folding intermediates with secondary-structural features (Langer *et al.*, 1992; Georgopoulos, 1992). The peptide-binding differences may allow Hsp40s to target the nascent proteins in an intermediate conformation state to Hsp70s. Secondly, fluorescence studies indicated that Hsp40 exposes a much greater hydrophobic surface than Hsp70 or Hsp60 and the polypeptides are in a different conformation when bound to Hsp40 compared with the Hsp40/Hsp70 system (Hartl, 1996). Thirdly, Hsp40 forms a homodimer in solution while Hsp70 functions as a monomer.

Little work has been carried out on the structural studies of Hsp40. The discovery of the crystal structure of yeast Hsp40 Sis1 will have a major impact on understanding the mechanism by which Hsp40 binds unfolded polypeptide and refolds it. At an atomic level, the crystal structure of Hsp40 may reveal a peptide-binding groove. The structure determination of the protein–peptide complex may follow. It would be of great interest to see what conformation the peptide may take in the structure and how it may interact with the protein. The stoichiometry of the Hsp40–peptide can also clearly be determined.

References

Caplan, A. J., Cyr, D. M. & Douglas, M. G. (1993). *Mol. Biol. Cell*, **4**, 555–563.
Clarke, A. R. (1996). *Curr. Opin. Struct.* **6**, 43–50.
Craig, E., Weissman, J. S. & Horwich, A. L. (1994). *Cell*, **78**, 365–372.
Flynn, G., Pohl, J., Flocco, T. & Rothman, J. (1991). *Nature (London)*, **353**, 726–730.
Frydman, J., Nimmesgern, E., Ohtsuka, K. & Hartl, F. U. (1994). *Nature (London)*, **370**, 111–117.
Georgopoulos, C. (1992). *Trends Biochem. Sci.* **17**, 295–299.
Gething, M. J. & Sambrook, J. (1992). *Nature (London)*, **355**, 33–45.
Hartl, F. U. (1996). *Nature (London)*, **381**, 571–580.
Hohfeld, J., Minami, Y. & Hartl, F. U. (1995). *Cell*, **83**, 589–598.
James, P., Pfund, C. & Craig, E. A. (1997). *Science*, **275**, 387–389.
Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M. K. & Hartl, F. U. (1992). *Nature (London)*, **356**, 683–689.
Lu, Z. & Cyr, D. M. (1999). Submitted.
Luke, M. M., Sutton, A. & Arndtt, K. T. (1991). *J. Cell Biol.* **114**, 623–638.
Minor, W. (1993). *XdisplayF* program. Purdue University, West Lafayette, Indiana, USA.
Nelson, R. J., Ziegelhoffer, T., Nicolet, C., Werner-Washburne, M. & Craig, E. A. (1992). *Cell*, **71**, 97–105.
Ohba, M. (1997). *FEBS Lett.* **409**, 307–311.
Otwinowski, Z. (1993). *Proceedings of the CCP4 study weekend: data collection and processing* edited by L. Sawyer, N. Issacs & S. Bailey, pp. 56–62. Warrington: Daresbury Laboratory.
Pfund, C., Lopez-Hoyo, N., Ziegelhofei, T., Schilke, B. A., Lopez-Beusa, P., Walter, W. A., Wiedmann, M. & Craig, E. A. (1998). *EMBO J.* **17**, 3981–3988.
Rothman, J. E. (1989). *Cell*, **59**, 591–601.
Silver, P. & Way, J. C. (1993). *Cell*, **74**, 5–6.
Ungermann, C., Neupert, W. & Cyr, D. M. (1994). *Science*, **266**, 1250–1253.
Ungewickell, E., Ungewickell, H., Holstein, S. E., Linder, R., Prasad, K., Barouch, W., Martin, B., Greene, L. E. & Eisenberg, E. (1995). *Nature (London)*, **378**, 632–635.
Zhong, T. & Arndt, K. T. (1993). *Cell*, **73**, 1175–1186.
Zhu, X., Zhao, X., Burkholder, W. F., Gragerov, A., Ogata, C. M., Gottesman, M. E. & Hendrickson, W. A. (1996). *Science*, **272**, 1606–1614.