The crystal structure of the peptide-binding fragment from the yeast Hsp40 protein Sis1

Bingdong Sha[1,2]*, Soojin Lee[2] and Douglas M Cyr[2]

Background: Molecular chaperone Hsp40 can bind non-native polypeptide and facilitate Hsp70 in protein refolding. How Hsp40 and other chaperones distinguish between the folded and unfolded states of proteins to bind non-native polypeptides is a fundamental issue.

Results: To investigate this mechanism, we determined the crystal structure of the peptide-binding fragment of Sis1, an essential member of the Hsp40 family from *Saccharomyces cerevisiae*. The 2.7 Å structure reveals that Sis1 forms a homodimer in the crystal by a crystallographic twofold axis. Sis1 monomers are elongated and consist of two domains with similar folds. Sis1 dimerizes through a short C-terminal stretch. The Sis1 dimer has a U-shaped architecture and a large cleft is formed between the two elongated monomers. Domain I in each monomer contains a hydrophobic depression that might be involved in binding the sidechains of hydrophobic amino acids.

Conclusions: Sis1 (1–337), which lacks the dimerization motif, exhibited severe defects in chaperone activity, but could regulate Hsp70 ATPase activity. Thus, dimer formation is critical for Sis1 chaperone function. We propose that the Sis1 cleft functions as a docking site for the Hsp70 peptide-binding domain and that Sis1–Hsp70 interaction serves to facilitate the efficient transfer of peptides from Sis1 to Hsp70.

Addresses: [1]Center for Macromolecular Crystallography, University of Alabama at Birmingham, Birmingham, AL 35294-0005, USA and [2]Department of Cell Biology, University of Alabama at Birmingham, Birmingham, AL 35294-0005, USA.

*Corresponding author.
E-mail: Sha@cmc.uab.edu

Key words: crystal structure, Hsp40, Hsp70, molecular chaperone, Sis1

Received: **24 March 2000**
Revisions requested: **17 April 2000**
Revisions received: **24 April 2000**
Accepted: **25 April 2000**

Published: **7 July 2000**

Structure 2000, **8**:799–807

0969-2126/00/$ – see front matter

Introduction

Members of the Hsp70 and Hsp40 (DnaJ-like) families of chaperones function in specific pairs that form transient complexes with non-native regions of polypeptides to promote the folding, assembly and transport of proteins within the cell [1,2]. All members of the Hsp40 family contain a J domain, which regulates the ATP-dependent binding of peptides by Hsp70 [3–8]. The family can be divided into three sub-types: type I and type II Hsp40 proteins can also act as molecular chaperones to bind and deliver non-native proteins to Hsp70. Interestingly, type I and type II Hsp40 proteins are not functional equivalents [9,10] and exhibit major differences in chaperone activity [11]. The type I Hsp40 proteins, such as *Escherichia coli* DnaJ, human Hdj-2 and yeast Ydj1, have conserved sequence motifs, known as the zinc-finger-like region and conserved C terminus, that function via an unknown mechanism to bind non-native proteins [12–15]. Type II Hsp40 proteins, such as human Hsp40 (Hdj-1) and yeast Sis1 [9,16] lack the zinc-finger motif, but contain the conserved C terminus, and thus appear to possess a mechanism for chaperone action that is distinct from type I Hsp40 proteins. Type III Hsp40 proteins contain only the J domain and do not bind non-native polypeptides.

To investigate the structural basis for the chaperone action of type II Hsp40 proteins we crystallized [17] a fragment of Sis1, Sis1 (171–352), that retains the chaperone activity of the full-length molecule [11]. We now report the 2.7 Å X-ray crystal structure of Sis1 (171–352), which reveals that Sis1 forms a homodimer in the crystal by a crystallographic twofold axis. Sis1 monomers are elongated and consist of two domains with similar folds. Dimerization of Sis1 monomers is facilitated by a short C-terminal stretch that enables Sis1 to form a U-shaped structure that has large central cleft. Dimer formation appears critical for Sis1 chaperone function because deletion of its dimerization motif caused severe defects in its ability to assist Hsp70 in protein folding. These data are discussed in the context of a model that proposes a mechanism to describe how Hsp40 proteins bind and deliver non-native polypeptides to Hsp70.

Results

Structure of the Sis1 (171–352) monomer

A fragment of Sis1 that contains amino acid residues 171–352 and retains the chaperone function of Sis1 generated by limited proteolysis and purified [11,17]. Sis1 (171–352) was then crystallized [17] and its structure determined at 2.7 Å using the multiple anomalous dispersion (MAD) method (Table 1). The resultant electron density map from the MAD phasing was readily traceable (Figure 1a) and the mainchain was uninterrupted from residue 180–349 (Figure 1b). The first nine

Table 1

Statistics for MAD data collection and structure determination.

	Resolution (Å)	R_{symm}	Completeness (%)	<I/σ>
Data collection				
Native data	2.7	0.045 (0.377)*	97.0 (99.5)	20.6 (6.17)
Se–Met crystal data				
Remote I	3.0	0.059 (0.317)	93.5 (89.2)	17.5 (5.12)
Edge	3.0	0.063 (0.336)	92.8 (88.8)	15.9 (4.66)
Peak	3.0	0.068 (0.321)	93.1 (88.9)	16.3 (4.63)
Remote II	3.0	0.060 (0.350)	93.0 (89.0)	15.2 (4.52)
Anomalous diffraction ratios				
	λ1	λ2	λ3	λ4
λ1	0.047(0.038)	0.053	0.055	0.047
λ2		0.072(0.043)	0.032	0.055
λ3			0.074(0.046)	0.054
λ4				0.060 (0.045)

*Numbers in parenthesis are for the outer resolution shell.

residues in the N terminus and three residues in the C terminus appeared to be flexible and were not visible in the electron density map.

The structure of the Sis1 (171–352) monomer consists of eleven β strands (B1–B11) and three short α helixes (A1–A3; Figure 1c). Two distinct but structurally similar domains termed domain I, residues 180–255, and domain II, residues 260–336 were identified. In addition, Sis1 (171–352) contains a C-terminal stretch that protrudes out from the molecule. A loop region (residues 337–343) and helix A3 comprise the C-terminal stretch. Domain II is connected to domain I in a head-to-tail fashion to generate an elongated protein with the approximate dimensions of $25 \times 25 \times 90$ Å. Each domain has a core of two β sheets that are connected by a short helix. The major β sheet of domain I is formed by B1, B5 and B4. In domain II, B7, B11 and B10 comprise the major β sheet. The minor antiparallel β sheets of domain I are composed of B2 and B3 in domain I, and B8 and B9 in domain II, respectively. Domain II differs from domain I in that it has a larger major β sheet that contains an additional antiparallel strand B6 at the flank of B7, but otherwise the secondary structures of these domains is very similar. Interestingly, the primary sequence identity between these domains (not including B6) is only 24% and the sequence similarity is 32%. The secondary structures of these two domains can be superimposed remarkably well except that the junction region between B2 and B3 is eight residues longer than that between B8 and B9. If these eight amino acid residues (205–212) are ignored, the root mean square (rms) deviation of the coordinates for the mainchain atoms between the two domains is 2.19 Å once the two domains are superimposed. The structural redundancy observed in domain I and II has not been observed in the solved structures of other molecular chaperones [2]. A search for structural homologs in the Protein Data Bank (PDB) by DALI [18], using the structure of Sis1 (171–352) or domain I as a search model, did not reveal any protein structure that had a fold similar to Sis1 (171–352).

Sequence conservation for the C-terminal fragments of type II Hsp40s

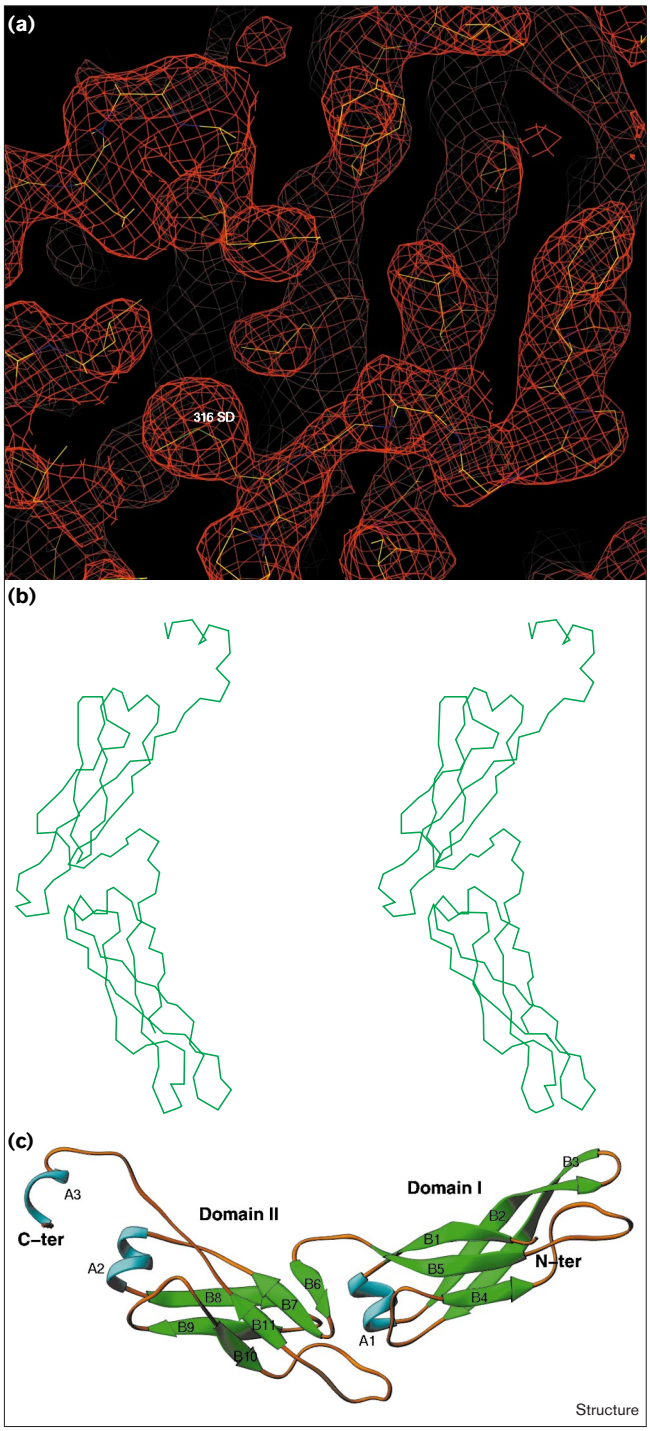
Eukaryotic organisms ranging from yeast to man encode members of the type II Hsp40 family [19]. Pairwise-BLAST analysis and multiple sequence alignment of the C-terminal regions from six different eukaryotic type II Hsp40s show that the Sis1 peptide-binding fragment shares 35–55% sequence identity with different family members (Figure 2). Thus, it is likely that the structural features exhibited by Sis1 also exist in other members of the type II Hsp40 family. The amino acid sequence of type I and type II Hsp40 proteins differ in the region that corresponds to domain I in Sis1 (171–352) and it is therefore probable that the structures of these different classes of the Hsp40 family are different.

Dimerization of Sis1 (171–352)

It has been suggested that type I and type II Hsp40 proteins function as dimers [3,9,20,21]. Consistent with these data, the Sis1 (171–352) crystal structure depicts a homodimer with the size of $30 \times 70 \times 90$ Å (Figure 3a,b). Sis1 (171–352) was crystallized with one monomer per each asymmetric unit. The respective monomers in each dimer are termed A and B and are related by a crystallographic twofold axis that is nearly perpendicular to helix A2. The dimerization of the elongated Sis1 monomers produces a U-shaped molecule that has a large cleft (~$15 \times 15 \times 60$ Å). The boundaries of this cleft are formed by the inside edges of each monomer. The major β sheets in domain II from each monomer are bent away from each other but face directly to their symmetry-related counterpart and the distance between these opposing regions is about 15 Å. Domain I of respective monomers is twisted about 130° from domain II. This twisting increases the distance between the two major

Figure 1

Sis1 monomer structure. **(a)** The electron density map around Met316 of Sis1 (171–352) after solvent flattening. The initial phases were determined by the program MADSYS [31] utilizing the anomalous scattering information from Se–Met 316. **(b)** Stereo drawing of the Cα trace of the Sis1 (171–352) monomer. Cα atoms are labeled every 20 amino acid residues. The N-terminal residues 171–179 and C-terminal residues 350–352 are not visible in the crystal structure and therefore they are not present in the figure. **(c)** The ribbon drawing of the Sis1 peptide-binding fragment structure [37]. α helices (A1–A3, blue), β strands (B1–B11, green) are numbered from the N to the C terminus. The molecule is oriented so that the major β sheets for both domain I and domain II are in front of the minor β sheets.



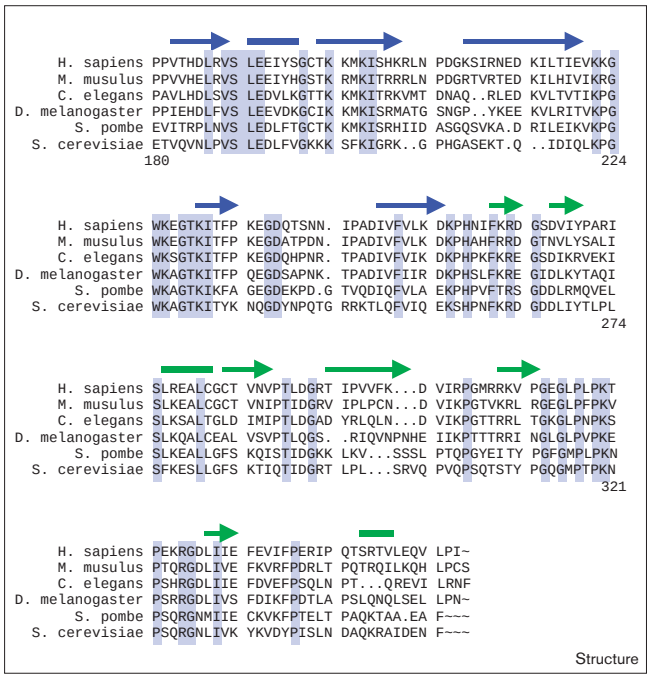
β sheets of domain I in each monomer and expands the size of the Sis1 cleft (Figure 3b).

The Sis1 (171–352) structure predicts that the C-terminal stretch, which is composed of the short helix A3 and the loop region before it, is responsible for dimerization (Figure 3c). Helix A3 from monomer B protrudes out of the monomer surface and lies on top of helix A2 from monomer A. This enables residue Asp349 on A3 from monomer B to form a salt bridge with Lys277 on A2 from monomer A (Figure 3c). In addition, a number of hydrophobic amino acid residues from A2, A3 and the loop before A3 of both monomers extend toward inside the dimer interface and form a hydrophobic core. Residues Ile348 from A3, Phe276 and Leu280 from A2, Tyr336 and Leu340 from the loop before A3 are clustered with their symmetry-related partners to stabilize the homodimer (Figure 3c). The surface area buried upon the dimer interface from each monomer is about 600 Å^2.

The functional significance of Sis1 dimer formation

To test whether or not dimer formation plays an important role in Sis1 chaperone function the C-terminal stretch proposed to be responsible for Sis1 dimerization, residues 338–352, was deleted to generate Sis1 (1–337). The oligomeric state and biochemical functions of Sis1 (1–337) were then compared with Sis1 (Figure 4). When purified Sis1 (1–337) was loaded onto a Superdex-200 gel filtration column it migrated with an apparent molecular weight of a monomer, whereas Sis1 ran with the apparent molecular weight of a dimer (Figure 4a). Sis1 can cooperate with the yeast Hsp70 protein Ssa1 to refold denatured forms of the model protein luciferase [14]. When the ability of Sis1 (1–337) to cooperate with Ssa1 to refold chemically denatured luciferase was examined, its activity was only 5% of that for Sis1 (Figure 4b). In order to for Hsp40 proteins to assist Hsp70 in protein folding they must be able to stimulate the ATPase activity of Hsp70 and interact with non-native proteins to maintain them in a folding competent conformation [1]. We observed that Sis1 (1–337), of which the J domain is intact, was fully capable of stimulating the ATPase activity of Ssa1 (Figure 4c). However, Sis1 (1–337) exhibited a severe defect in its chaperone function because it could not maintain chemically denatured luciferase in a refoldable conformation (Figure 4d). In contrast, Sis1 could maintain almost all of the denatured luciferase that was added to reactions in the folding-competent conformation for up to 3 hours after dilution from denaturant. These data demonstrate that the dimerization of Sis1 is important for its chaperone function but does not appear to play a major role in its ability to stimulate the ATPase activity of Hsp70.

Figure 2



Sequence alignment of the C-terminal regions from eukaryotic type II Hsp40 family members. Program Pileup from GCG package was utilized to align residues 180–352 of Sis1 from *S. cerevisiae* with similar regions of Hsp40 proteins from *H. sapiens* (Hdj-1), *M. musculus* (Hsp40-3), *C. elegans* (Z66513.1), *D. melanogaster* (Droj-1) and *S. pombe* (Psi protein). The amino acid residues of Sis1 are numbered below the alignment. The conserved residues in the family are marked by blue bars. The secondary structures of Sis1 are shown on top of the alignment. The structural components in domain I are denoted by blue and those in domain II are denoted by green. The α helices are represented by boxes and β strands are represented by arrows. The sequence mismatches in Sis1 (171–352) B3 and B9 regions might imply different length of the β strands or structural variations among this protein family.

Several lines of evidence suggest that defects in Sis1 (1–337) function that we report are not resultant from the misfolding of this mutant protein. Sis1 (1–337) runs as a monomer on gel filtration columns and does not form higher molecular weight aggregates (Figure 4a). When Sis1 (1–337) was subjected to limited proteolysis with proteinase K [11,17], a pattern of proteolytic fragments that was nearly identical to that of Sis1 was observed (data not shown). Furthermore, Sis1 (1–337) was fully able to stimulate the ATPase activity of Ssa1 (Figure 4c).

Structural implications for binding of non-native peptides by Sis1

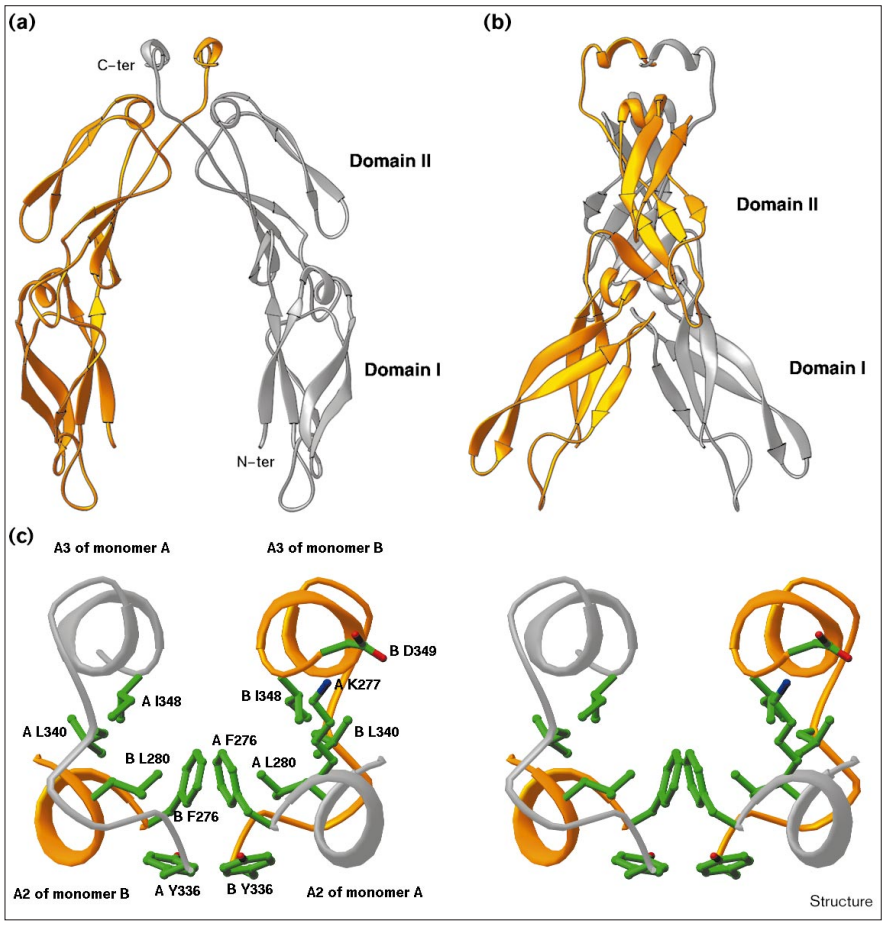
The substrate specificity and mechanism for recognition of non-native polypeptides by Hsp40 proteins is unknown. Molecular chaperone proteins are known to utilize hydrophobic patches [22] or grooves [23] to form transient complexes with hydrophobic residues exposed by non-native polypeptides. Consistent with these observations, phage display studies indicate that Sis1 prefers to bind peptides with hydrophobic sidechains (data not shown). Thus, it is likely that Sis1 binds non-native polypeptides through hydrophobic interactions. To identify sites on Sis1 that could be involved in peptide binding, the location of surface exposed hydrophobic residues were examined using GRASP presentations of the Sis1 (171–352) structure. The surface potential and hydrophobicity drawings indicate that the interior of the Sis1 cleft does not expose a large hydrophobic surface (Figures 5a,b). However, this analysis revealed a large hydrophobic depression on the surface of domain I. This region represents the largest hydrophobic region on the surface of Sis1 (171–352) and has the potential to be involved in peptide binding. Hydrophobic residues Val184 and Leu186 from B1, Phe201 and Ile203 from B2 form the rim of the hydrophobic depression and Phe251 from B5 serves as the bottom. Sequence alignment demonstrates that residues Leu186, Ile203 and Phe251 are conserved in all of the Hsp40 protein analyzed, whereas Phe201 is conservatively changed to methionine (Figure 2).

Discussion

We report the first crystal structure of an Hsp40 family member and it corresponds to the yeast Sis1 peptide-binding fragment. The structure of Sis1 (171–352) reveals that it functions as a homodimer that is constructed from elongated monomers with two distinct domains. Dimers are formed via hydrophobic interactions between the C-terminal stretches that are present on each monomer. These interactions generate an elongated molecule that has a large cleft of which the formation appears important for Sis1 chaperone function. In addition, we have identified a large hydrophobic patch that is exposed on the surface of domain I that might be involved in the binding of non-native regions of polypeptides. These data provide a testable model to explain how Sis1 functions with Ssa1 to fold proteins and protect cells from stress.

Two pieces of information suggest that the residues exposed in the hydrophobic depression on domain I might contribute to interactions between Sis1 and non-native polypeptides. First, given that domain II has sequence and structural homology to domain I, it contains a similar hydrophobic depression that is formed by Leu268 and Tyr270 from B7, Ile287 and Ile290 from B8, and Leu328 from B11. This hydrophobic depression on domain II is occupied by the aromatic sidechain of Phe261 from B6, which is unique to domain II. Second, in the crystal packing, the sidechain of Pro312 from an adjacent Sis1 dimer is inserted into the hydrophobic depression on domain I (Figures 5a,b). Thus, the hydrophobic residues exposed on the surface of domain I have the potential to interact with hydrophobic sidechains exposed by cellular proteins through van der Waals interactions.

Figure 3

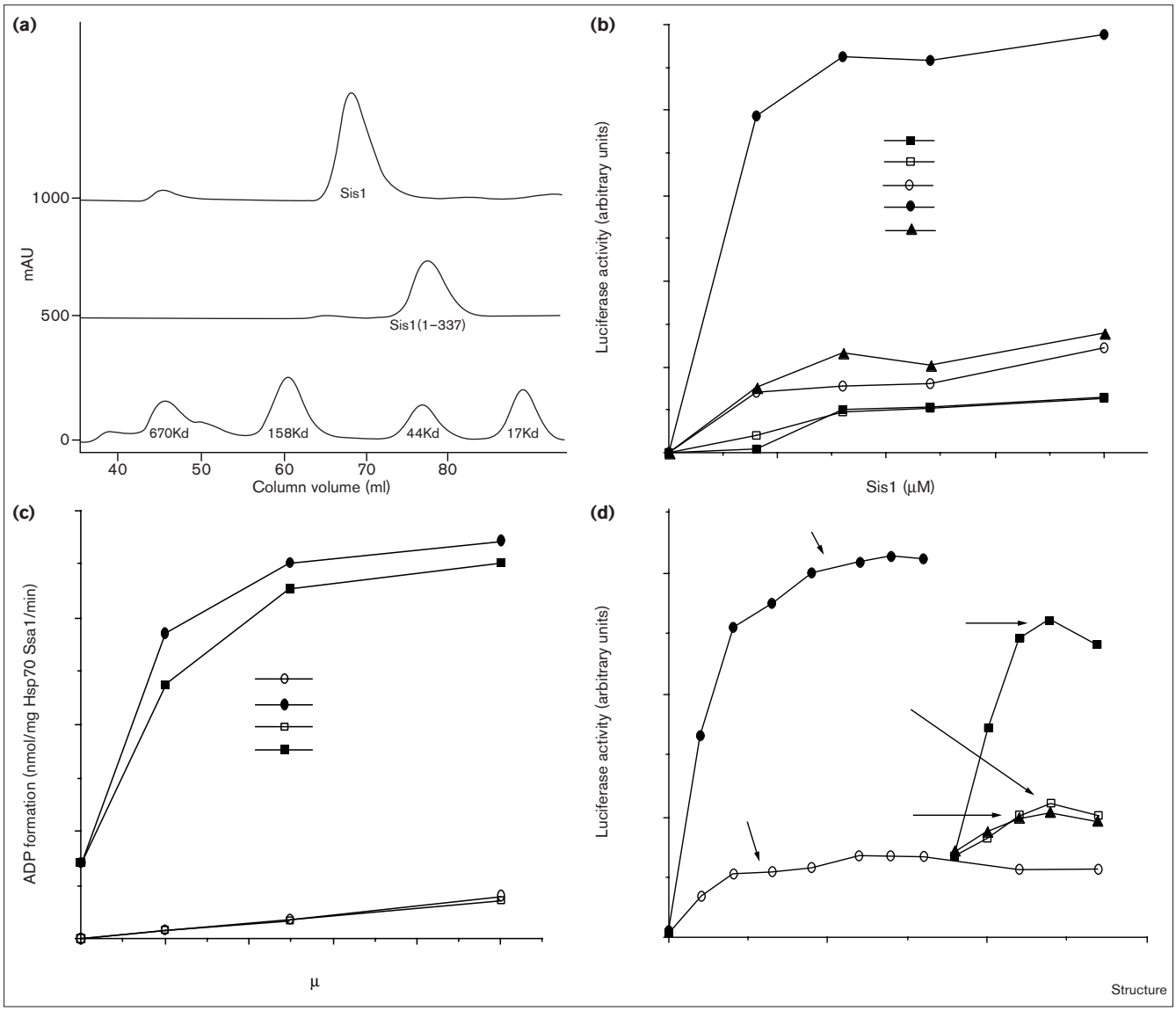


Sis1 dimer structure. **(a)** Ribbon drawing of the Sis1 dimer. Monomer A is shown in silver and monomer B is shown in gold. The two monomers are related by a vertical crystallographic twofold axis. The N terminus (residue 180) and C terminus (residue 349) of monomer A are labeled. **(b)** A view of Sis1 dimer after it is rotated 90° along the vertical axis from the orientation shown in Figure 3a. Domain I of monomer A is swung away from that of monomer B to expand the Sis1 dimer cleft. **(c)** The stereo drawing of the dimerization interface between monomer A (silver) and B (gold). The orientation of the molecule in this figure is similar to that in Figure 3a and it has a vertical twofold crystallographic axis. Carbon atoms are shown in green, nitrogen atoms are in blue and oxygen atoms are in red.

When the structure of Sis1 (171–352) is compared with that of the peptide binding domains from the molecular chaperones Hsp70 and Hsp60 a number of interesting similarities and differences are noted. First, Hsp60 and Hsp70 utilize a hydrophobic patch or groove to recognize and bind hydrophobic residues that are exposed by non-native polypeptides and the data we present suggests that Sis1 functions via a similar mechanism [22–24]. The crystal structure of the peptide-binding domain of *E. coli* DnaK complexed with the peptide NRLLLTG (in single-letter amino acid code) has been solved [23]. This structure demonstrates that Hsp70 binds extended peptides through the use of a hydrophobic channel that is formed by a β-sandwich that is covered by an α-helical lid domain [23]. Selectivity for peptide binding within this channel is conferred by a deep depression that is capable of accommodating a single sidechain from a variety of amino acid residues, but appears to prefer those of leucine, methionine or isolucine and excludes proline [23]. Sis1 monomers contain a putative peptide-binding groove that is formed by two sheets of β strands that is similar in size to the peptide-binding domain of Hsp70. However, this structure differs from that of DnaK in that it lacks a lid domain and is capable of forming complexes with proline sidechain. Structural and functional studies on *E. coli* Hsp60 reveal that GroEL utilizes a hydrophobic patch to bind non-native polypeptides [22,24]. However, the Hsp60 peptide-binding domain differs from that identified for Hsp70 and Sis1 because it is much larger and can accommodate regions of proteins that fold into hairpin loops [22,24].

Deletion of the dimerization stretch in Sis1 caused defects in the ability of this Hsp40 protein to function with Hsp70 as a molecular chaperone. We suggest a number of scenarios to explain why dimer formation is important for Sis1 chaperone function. The putative peptide-binding site on a single monomeric Sis1 molecule might be insufficient to form stable complexes with protein folding intermediates. In the Sis1 dimer crystal, the two hydrophobic depressions on domain I are spaced by ~25 Å. Therefore, the simultaneous binding of a polypeptide by each Sis1 monomer might serve to hold the substrate in an extended conformation (Figure 6a). Because Hsp70 prefers to bind

Figure 4

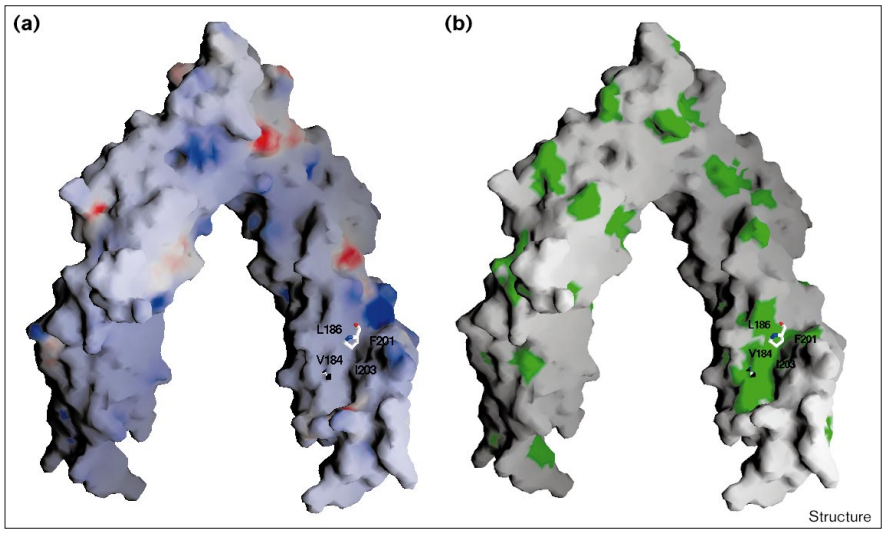


Biochemical analysis of Sis1 (1–337) function. **(a)** Analysis of Sis1 and Sis1 (1–337) mobility by gel filtration. The UV absorption traces that represent the mobility of Sis1, Sis1 (1–337) and a set of protein standards on a Superdex 200 column are shown. Minor peaks shown in the chromatographs represent protein contaminants in the respective protein preparations (data not shown). Sis1 and Sis1 (1–337) run in a single peak and therefore do not appear to exist in equilibrium between monomer and dimer or larger structures. **(b)** Folding of chemically denatured luciferase by Ssa1 facilitated by Sis1 and Sis1 (1–337). Values for spontaneous refolding of luciferase were subtracted from each value plotted. In the absence of ATP, Sis1 and Hsp70 Ssa1 are unable to cooperate in luciferase refolding [11]. To accurately measure the co-chaperone activity of Sis1 or Sis1 (1–337), the luciferase activity from Ssa1 intrinsic refolding activity has to be subtracted. When this calculation is performed, Sis1 (1–337) is shown to retain about 5% of the ability of Sis1 to assist Ssa1 in luciferase folding. **(c)** Stimulation of Hsp70 Ssa1 ATPase activity by Sis1 and Sis1 (1–337). Hydrolysis of ^{32}P-ATP to ADP by Ssa1 (0.1 mM) was monitored at 25°C. **(d)** Sis1, but not Sis1 (1–337), can maintain luciferase in a folding competent conformation. Chemically denatured luciferase (0.4 mM) was incubated at 30°C for 3 h in the presence or absence of Sis1 (1.5 mM) or Sis1 (1–337) (1.5 mM). Then, Ydj1 (1.5 mM) and Hsp70 Ssa1 (0.5 mM) were added to promote the maximal level of luciferase refolding observed under the experimental conditions utilized [11]. At the concentrations utilized in this experiment, neither Sis1 nor Sis1 (1–337) was observed to interfere with the folding of luciferase by Ydj1 and Ssa1 (data not shown).

extended peptides, the ability of Sis1 to bind regions of non-native polypeptides at these two sites simultaneously and hold them in an extended conformation might be critical for subsequent recognition of peptides by Hsp70.

The cleft formed by the Sis1 dimer might also play a crucial role in mediating interactions between Hsp40 and Hsp70 proteins. We speculate that Sis1 and Hsp70 might dock to facilitate the transfer of non-native proteins from Sis1 to

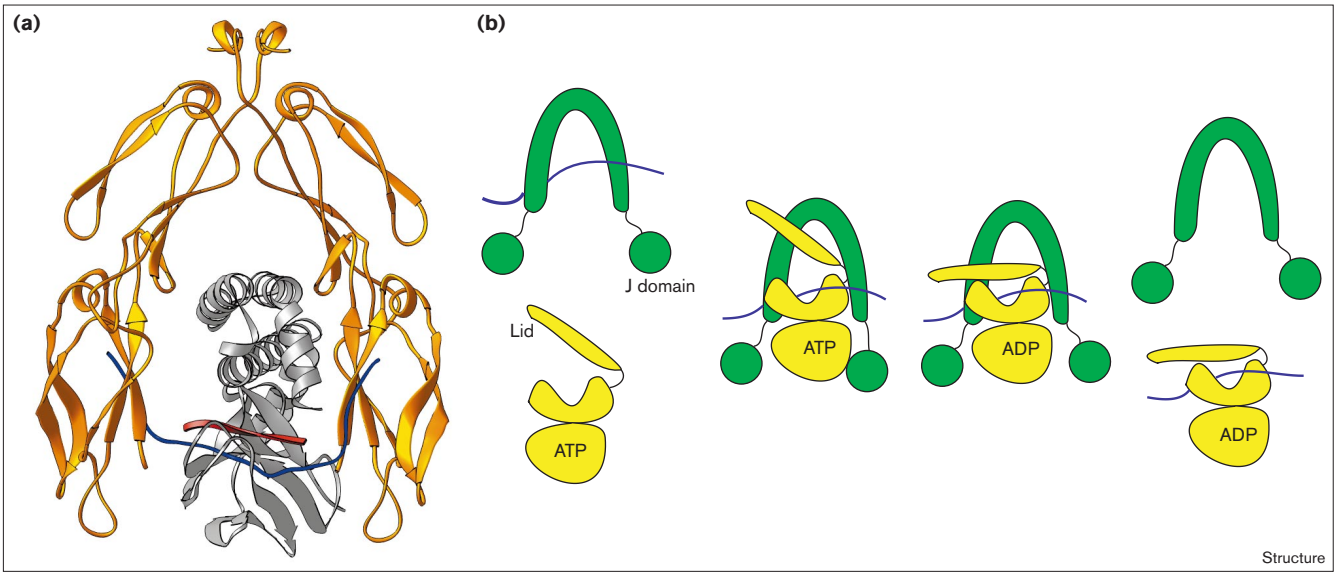
Figure 5



GRASP presentations of the Sis1 dimer. **(a)** A surface potential drawing of the Sis1 dimer determined by GRASP [38]. Blue and red denote positively and negatively charged regions, respectively. The Pro312 residue from an adjacent dimer that is involved in crystal packing is inserted into the hydrophobic depression of domain I and is shown in the ball-and-stick mode. The hydrophobic residues that are involved in forming the hydrophobic depression are labeled. Buried beneath Pro312 is residue Phe251 that forms the bottom of the hydrophobic depression.**(b)** A hydrophobicity drawing of the Sis1 dimer using GRASP [38]. The hydrophobic regions formed by the exposed carbon atoms in the sidechains of residues alanine, valine, leucine, isoleucine, methionine, phenylalanine and proline are shown in green.

Hsp70 (Figure 6a). This docking mechanism would serve to increase the efficiency of protein folding by assuring that binding of non-native proteins by Hsp70 is coordinated with the stimulation of ATP hydrolysis by the J domain of Hsp40 (Figure 6b). In support of this supposition, the peptide-binding domain of Hsp70 has been implicated in mediating the interactions with Hsp40 [25–27]. Furthermore, regions within DnaJ that lie outside of the J domain have also been shown to influence Hsp70–Hsp40 complex formation [13]. Consistent with this proposal

Figure 6



The proposed docking model to explain how Sis1 interacts with non-native polypeptide and Hsp70. **(a)** A manual model for a putative tertiary complex formed between Sis1, Hsp70 and a non-native peptide. The *E. coli* Hsp70 DnaK peptide-binding domain (silver) was manually modeled into the large cleft of Sis1 dimer (in gold) in a fashion that the helix bundle of DnaK is oriented lying through the cleft. The DnaK bound peptide is shown in red. The lid of DnaK is in closed state in the crystal structure [23]. An extended loop (25 residues, blue) was manually modeled into the Sis1 dimer to represent a bound non-native peptide. **(b)** Schematic representation of the 'docking' mechanism by which Sis1 (green) delivers a non-native peptide (blue) to Hsp70 (yellow). The cartoon drawing depicts an Hsp70 molecule that is divided into its ATPase domain [39], peptide-binding domain and the lid domain [23]. The J domain and peptide-binding fragment of Sis1 are shown schematically.

manual modeling indicates that the Sis1 dimer cleft is large enough to accommodate the peptide-binding domain of *E. coli* Hsp70 DnaK if its long helix bundle is oriented parallel to A2 and lying through the cleft (Figure 6a) [23].

Sequence alignment suggests that the structure of domain II of Sis1 is conserved between type I and II Hsp40 proteins [9]. The zinc finger motifs that are found only in type I proteins partially substitute for regions within domain I in type II family members. Thus, the structures of domain I in type I and type II Hsp40 proteins are likely to be different. The zinc finger motifs are probably involved in peptide binding by type I Hsp40s [12]. Therefore, the structures of peptide-binding sites for type I Hsp40s might be distinct from that of type II Hsp40s. The apparent structural differences between type I and type II Hsp40s might explain why they exhibit differences in their chaperone function.

Biological implications

The molecular chaperone Hsp40 is able to bind non-native polypeptide and assist Hsp70 to refold it. A fundamental question here is how Hsp40 proteins recognize and bind non-native proteins and transfer them to Hsp70. We have determined the first crystal structure of an Hsp40 member, the peptide-binding fragment of Sis1 from *Saccharomyces cerevisiae*. Sis1 functions as a homodimer composed of elongated monomers that contain two distinct domains, and we have identified a hydrophobic patch exposed on domain I that might be involved in the binding of non-native regions of polypeptides. We examined the activtity of a fragment of Sis1 (residues 1–337), which lacks the dimerization motif. This fragment lacks chaperone activity although it is able to regulate Hsp70 ATPase activity. We conclude, therefore, that dimer formation is required for Sis1 chaperone function. We also propose that the cleft in Sis1 functions as a docking site for the Hsp70 peptide-binding domain and that the Sis1–Hsp70 interaction serves to facilitate the efficient transfer of peptides from Sis1 to Hsp70.

Materials and methods

Crystallization and data collection

The purification and crystallization of native Sis1 and selenomethionyl Sis1 have been described previously [11,17]. The mass spectrum analysis of the dissolved the crystals showed that the crystals contained only Sis1 (171–352). The native data set was collected at SSRL station 7–1 ($\lambda = 1.08$ Å) using 1.5° oscillation angle and processed by DENZO and merged by Scalepack [28,29]. The native crystals belong to space group $P4_12_12$ with $a = 73.63$ Å, $c = 80.16$ Å. The redundancy of the native data set is 5.5 (4.1 for outer shell). The structure was determined using the MAD method. The MAD data sets were collected at SSRL station 1–5 using selenomethionyl Sis1 crystals. The crystals were frozen at 90K using 20% glycerol in the mother liquid as a cryoprotectant. The four energy bands that were used for MAD data collection were 11600ev (remote I), 12659ev (edge), 12654ev (peak) and 13400ev (remote II). The indexing and integration of data sets were carried out by Mosflm and the scaling and merging of the data sets were performed by Scala of the CCP4 package [30].

Phase determination

The MADSYS program package [31] was utilized to evaluate the amplitudes for the selenium-scattering component of the structure factor (Table 2). Only one methionine residue exists in the 182 amino acid residues of the Sis1 (171–352) fragment. The selenium atom was located in the Patterson map, which was calculated by F_A from the MADSYS by program Xtalview [32] and refined further by MADLSQ. Solvent flattening ($V_m = 2.5$ Å^3/Dalton, 50% solvent in the unit cell) was performed on the basis of one molecule per asymmetric unit [33]. DM [30] was utilized to carry out solvent flattening and improve the accuracy of the phases from MADSYS. The figure of merit after solvent flattening was 0.710 for all of the reflection data from 30–3.0 Å. The resultant electron density map was readily interpretable. The electron density was unbroken from amino acid residues 180–349 of the Sis1 (171–352) fragment. The average real space correlation coefficient between the all the residues in the model and electron density map is 0.785.

Model building and refinement

Model building and adjustment were carried out by program O 5.10.3 using the 3.0 Å electron density map [34]. Residues 180–349 of Sis1 (171–352) were modeled into the electron density map and this step was followed by three cycles of torsion dynamics refinement by CNS 0.3 [35] using the 2.7 Å native data collected at SSRL station 7–1. Bulk solvent correction was utilized in refinement. Six cycles of positional refinement were then carried out. Restrained individual B factor refinement was not performed until the last cycle. After each cycle of refinement, the model was manually rebuilt according to the resultant $2F_o–F_c$ and $F_o–F_c$ maps. The refinement gave reasonable rms derivation from the ideal geometry at this resolution [36] (Table 2). A Ramachandran plot of the final model by use of program Procheck revealed that 111 of the 140 nonglycine residues in the structure were in the most favorable region and one nonglycine residue is in the disallowed region. The R_{factor} and R_{free} are somewhat high and this is likely to result from the quality of our high resolution diffraction data. The R_{symm} is 0.220 for the 3.09–2.87 Å resolution shell and 0.377 for the 2.87–2.70 Å outer shell and the overall R_{symm} was 0.045. A factor that might limit the ability of Sis1 (171–352) crystals to diffract to higher resolution might be related to the fact that the Sis1 molecule has an elongated shape and dimers are formed via interactions between short C-terminal stretches of residues. Thus, rigid-body movements between domain I and domain II or individual monomers in Sis1 (171–352) dimers might occur. The Wilson plot from our various data sets showed a quite large estimated B factor ranging from 60 to 70. In addition, we have a low

Table 2

Statistics for structure refinement.

Refinement	
Resolution range (Å)	30–2.7
Number of used reflections	6347(508 used for R_{free} calculation)
R_{factor} (%)	26.5(38.2 for outer resolution shell)
R_{free} (%)	31.1(42.0 for outer resolution shell)
Number of model atoms	1347(33 water molecules)
Average value of B factors	57.39
Rms deviations from ideality [36]	
Bond lengths (Å)	0.008
Bond angles (°)	1.367
Impropers (°)	1.291
Dihedrals (°)	25.613
Structure Z-scores from program What if [40]	
1st generation packing quality	−1.66
2nd generation packing quality	−1.93
Backbone conformation	−0.06
Chi-1/Chi-2 rotamer normality	−1.39
Ramachandran plot appearance	−2.75

data/refined parameter ratio due to the high symmetry of the crystal that might also cause a higher R_{factor}. However, average real space correlation coefficients between all the residues in the refined model and the MAD map and $2F_o–F_c$ map are 0.837 and 0.891, respectively. Thus, our refined Sis1 structure fits well in the $2F_o–F_c$ map.

MAD phasing statistics: $R(F_T) = 0.031$; $R(F_A) = 0.428$; $<\Delta(\Delta\varphi)> = 34.64°$; $<\sigma(\Delta\varphi)> = 14.16°$; FOM = 0.495. $R = [\Sigma_{hkl}\Sigma_i|F_i–<F>|]/\Sigma<F>$, FT is the structure factor part of normal scattering from all the atoms. F_A is the structure factor part of normal scattering of the anomalous scatters only. $\Delta\varphi$ is the phase difference between F_T and F_A. $\Delta(\Delta\varphi)$ is the difference between two independent determinations of $\Delta\varphi$. FOM is the figure of merit for MAD phasing.

Purification and functional characterization of Sis1 (1–337)

The Sis1(1–337) deletion mutant was made using the polymerase chain reaction (PCR). Sis1 (1–337) was expressed in *E. coli* and purified as described for Sis1 [17]. To monitor the ability of Sis1 and Sis1 (1–337) to assemble into dimers, 1 ml of solutions containing 1 mg/ml of Sis1 or Sis1 (1–337) were loaded onto a Superdex 200 gel filtration column (Pharmacia). Proteins were eluted in a mobile phase composed of 20 mM Hepes, pH 7.4 and 150 mM NaCl that was pumped at 1 ml/min. The retention time of the respective proteins was monitored with a flow cell at with the wavelength set at 280 nM. The ability of Sis1 and Sis1 (1–337) to function with Hsp70 Ssa1 to refold denatured luciferase was determined as previously described [11]. Assays for the ATPase activity of Ssa1 were carried out as previously described [11].

Accession numbers

The coordinates and structure factors of Sis1 peptide-binding fragment have been deposited to Protein Data Bank with an accession number of 1C3G.

Acknowledgements

We are grateful to H. Bellamy, P. Ellis, P. Kohn and D. Mitchell for their great assistance in data collection at SSRL station 1–5 and 7–1. The work is supported by grants from NIH, HHMI and NASA to BDS and NIH and Cystic Fibrosis Foundation to DMC.

References

1. Hartl, F.U. (1996). Molecular chaperones in cellular protein folding. *Nature* **381**, 571-580.
2. Bukau, B. & Horwich, A.L. (1998). The Hsp70 and Hsp60 chaperone machines. *Cell* **92**, 351-366.
3. Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M.K. & Hartl, F.U. (1992). Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature* **356**, 683-689.
4. Wall, D., Zylicz, M. & Georgopoulos, C. (1994). The N-terminal 108 amino acid of the *E. coli* DnaJ protein stimulate the ATPase activity of DnaK and are sufficient for lamda replication. *J. Biol. Chem.* **269**, 5446-5451.
5. Schmid, D., Baici, A., Gehring, H. & Christen, P. (1994). Kinetics of molecular chaperone action. *Science* **263**, 971-973.
6. Karzai, A.W. & McMacken, R.A. (1996). A bipartite signaling mechanism involved in DnaJ-mediated activation of *E. coli* DnaK protein. *J. Biol. Chem.* **271**, 11236-11246.
7. Misselwitz, S., Staeck, O. & Rapoport, T.A. (1998). J proteins catalytically acitivate Hsp70 molecules to trap a wide range of peptide sequences. *Mol. Cell* **2**, 593-603.
8. Laufen, T., Mayer, M.P., Beisel, C., Klostermeier, D., Reinstein, J. & Bukau, B. (1999). Mechanism of regulation of Hsp70 chaperones by DnaJ cochaperones. *Proc. Natl Acad. Sci. USA* **96**, 5452-5457.
9. Luke, M., Sutton, A. & Arndt, K.T. (1991). Characterization of SIS1, a *Saccharomyces cerevisiae* homologue of bacterial DnaJ proteins. *J. Cell Biol.*, **114**, 623-638.
10. Caplan, A.J. & Douglas, M.G. (1991). Characterization of Ydj1: a yeast homologue of the bacteria dnaJ protein. *J. Cell Biol.* **114**, 609-621.
11. Lu, Z. & Cyr, D.M. (1998). Protein folding activity of Hsp70 is modified differentially by the Hsp40 co-chaperone Sis1 and Ydj1. *J. Biol. Chem.* **273**, 27824-27830.
12. Szabo, A., Korszun, R., Hartl, F.U. & Flanagan, J. (1996). A zinc finger-like domain of the molecular chaperone DnaJ is involved in binding to denatured protein substrates. *EMBO J.* **15**, 408-417.
13. Banecki, B., *et al.*, & Zylicz, M. (1996). Structure-function analysis of the zinc finger region of the DnaJ molecular chaperone. *J. Biol. Chem.* **271**, 14840-14848.
14. Lu, Z. & Cyr, D.M. (1998). The conserved carboxyl terminus and zinc finger-like domain of the co-chaperone Ydj1 assist Hsp70 in protein folding. *J. Biol. Chem.* **273**, 5970-5978.
15. Goffin, L. & Georgopoulos, C. (1998). Genetic and biochemical characterization of mutations affecting the carboxy-terminal domain of the *Escherichia coli* molecular chaperone DnaJ. *Mol. Microbiol.* **30**, 329-340.
16. Zhong, T. & Arndt, K.T. (1993). The yeast Sis1 protein, a DnaJ homologue, is required for the initiation of translation. *Cell* **73**, 1175-1186.
17. Sha, B.D. & Cyr, D.M. (1999). Purification, crystallization and preliminary X-ray crystallographic studies of *S. cerevisiae* Hsp40 Sis1. *Acta Crystallogr. D* **55**, 1234-1236.
18. Holm, L. & Sander, C. (1995). Searching protein structure databases has come of age. *Proteins* **19**, 165-173.
19. Cyr, D.M., Langer, T., & Douglas, M.G. (1994). DnaJ-like proteins: molecular chaperones and specific regulators of Hsp70. *Trends Biochem. Sci.* **19**, 176-181.
20. King, C., Eisenberg, E. & Greene, L. (1995). Polymerization of 70 kDa heat shock protein by yeast DnaJ in ATP. *J. Biol. Chem.* **270**, 22535-22540.
21. Wickner. S, Hoskins, J. & McKenney, K. (1991). Monomerization of RepA dimers by heat shock proteins activates binding to DNA replication origin. *Proc. Natl Acad. Sci.* **88**, 7903-7907.
22. Fenton, W.A., Kashi, Y., Furtak, K. & Horwich, A.L. (1994). Residues in chaperonin GroEL required for polypeptide binding and release. *Nature* **371**, 614-619.
23. Zhu, X., *et al.*, & Hendrickson, W.A. (1996). Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* **272**, 1606-1614.
24. Chen, L. & Sigler, P.B. (1999). The crystal structure of a GroEL/peptide complex: plasticity as a basis for substrate diversity. *Cell* **99**, 757-768.
25. Freeman, B.C., Myers, M.P., Schumacher, R. & Morimoto, R.I. (1995). Identification of a regulatory motif in Hsp70 that affects ATPase activity, substrate binding and interaction with HDJ-1. *EMBO J.* **14**, 2281-2292.
26. Demand, J., Luders, J. & Hohfeld, J. (1998). The carboxy-terminal domain of Hsc70 provides binding sites for a distinct set of chaperone cofactors. *Mol. Cell Biol.* **18**, 2023-2038.
27. Suh, W.C., Burkholder, W.F., Lu, C.Z., Zhao, X., Gottesman, M.E. & Gross, C.A. (1998) Interaction of the Hsp70 molecular chaperone, DnaK, with its cochaperone DnaJ. *Proc. Natl Acad. Sci.* **95**, 15223-15228.
28. Minor, W. (1993). XdisplayF program. (Purdue Univ. West Lafayette).
29. Otwinowski, Z. (1993). Proc. CCP4 study weekend: data collection and processing (compiled by Sawyer, L., Issacs, N. & Bailey, S.) 56-62.
30. CCP4, (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760-763.
31. Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**, 51-58.
32. McRee, D.E. (1992). A visual protein crystallographic software system for X11/Xview. *J. Mol. Graph.* **10**, 44-46.
33. Matthews, B.W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491-497.
34. Jones, T.A., Zhou, J.Y., Cowan, S.W. & Kjeldgard, M. (1991). Improved methods for building protein models in the electron density maps and the location of errors in these maps. *Acta Crystallogr. A* 47, 110-119.
35. Brunger, A.T., *et al.*, & Warren, G.L. (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905-921.
36. Engh, R.A. & Huber, R. (1991). Accurate bound and angel parameters for X-ray protein structure refinement. *Acta Crystallogr. A* **47**, 392-400.
37. Carson, M. (1987). Ribbons 2.0. Ribbon models for macromolecule. *J. Mol. Graph.* **5**, 103-106.
38. Nicholls, A., Sharp, K.A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* **11**, 281-296.
39. Flaherty, K.M., DeLuca-Flaherty, C. & McKay, D.B. (1991) Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. *Nature* **346**, 623-628.
40. Vriend, G. (1990). What If: a molecular modeling and drug design program. *J. Mol. Graph.* **8**, 52-56.