Molecular BioSystems

PAPER

Molecular chaperones DnaK and DnaJ share predicted binding sites on most proteins in the *E. coli* proteome†

Sharan R. Srinivasan,[a] **Anne T. Gillies,**[a] **Lyra Chang,**[a] **Andrea D. Thompson**[a] **and Jason E. Gestwicki***[ab]

In *Escherichia coli*, the molecular chaperones DnaK and DnaJ cooperate to assist the folding of newly synthesized or unfolded polypeptides. DnaK and DnaJ bind to hydrophobic motifs in these proteins and they also bind to each other. Together, this system is thought to be sufficiently versatile to act on the entire proteome, which creates interesting challenges in understanding the interactions between DnaK, DnaJ and their thousands of potential substrates. To address this question, we computationally predicted the number and frequency of DnaK- and DnaJ-binding motifs in the *E. coli* proteome, guided by free energy-based binding consensus motifs. This analysis revealed that nearly every protein is predicted to contain multiple DnaK- and DnaJ-binding sites, with the DnaJ sites occurring approximately twice as often. Further, we found that an overwhelming majority of the DnaK sites partially or completely overlapped with the DnaJ-binding motifs. It is well known that high concentrations of DnaJ inhibit DnaK–DnaJ-mediated refolding. The observed overlapping binding sites suggest that this phenomenon may be explained by an important balance in the relative stoichiometry of DnaK and DnaJ. To test this idea, we measured the chaperone-assisted folding of two denatured substrates and found that the distribution of predicted DnaK- and DnaJ-binding sites was indeed a good predictor of the optimal stoichiometry required for folding. These studies provide insight into how DnaK and DnaJ might cooperate to maintain global protein homeostasis.

Introduction

DnaK and DnaJ regulate protein homeostasis in *E. coli* by facilitating the folding of nascent polypeptides and the re-folding of damaged proteins.[1–3] The DnaK–DnaJ system appears to be especially important under conditions of cellular stress because mutants lacking either chaperone are particularly sensitive to elevated temperatures, antibiotics and other insults.[4,5] Survival under these conditions requires that a wide range of vulnerable polypeptides be protected from aggregation and then refolded when conditions improve. How are DnaK and DnaJ able to act on such a wide number of possible substrates? In this study, we aimed to predict how DnaK and DnaJ might interact with their substrates across the *E. coli* proteome to better understand how they coordinate global protein homeostasis.

DnaK is a member of the widely expressed and highly conserved heat shock protein 70 (Hsp70) family. Like other Hsp70s,

it is a 70 kDa chaperone that consists of two major domains: a nucleotide-binding domain (NBD) and a substrate-binding domain (SBD).[6–8] The NBD contains a nucleotide-binding cassette of the sugar kinase/Hsp70/actin superfamily, which has intrinsic ATPase activity. The SBD consists of a β-sandwich subdomain that has a hydrophobic cavity for binding to substrates[8,9] and a C-terminal, α-helical subdomain that acts as a "lid", opening and closing to accept and release these substrates. This two-domain architecture allows regulated control over substrate binding. Specifically, ATP hydrolysis in the NBD causes movements in the SBD that close the "lid" (Fig. 1). The substrates of the SBD include extended linear, hydrophobic polypeptides with lengths of at least 7 to 8 amino acids,[10,11] while recent experiments suggest that more complex, partially-folded peptides may also be bound.[12] Finally, interactions of these substrates with the SBD strongly stimulate ATPase activity, suggesting that allosteric communication is bi-directional.[13–17]

DnaJ is the founding member of the J domain-containing family of co-chaperones.[18] DnaJ is a ~40 kDa protein that is composed of a conserved, N-terminal J domain, a glycine/phenylalanine (G/F)-rich motif, a central zinc finger region and a C-terminal domain. Through its J-domain, DnaJ binds to DnaK and allosterically promotes ATP hydrolysis.[19–24]

[a] *Chemical Biology Graduate Program, University of Michigan Ann Arbor, MI 48109, USA*
[b] *Department of Pathology and Biological Chemistry and the Life Sciences Institute, University of Michigan, 210 Washtenaw Avenue, Ann Arbor, MI 48109-2216, USA. E-mail: gestwick@umich.edu; Fax: (734) 764-1247*
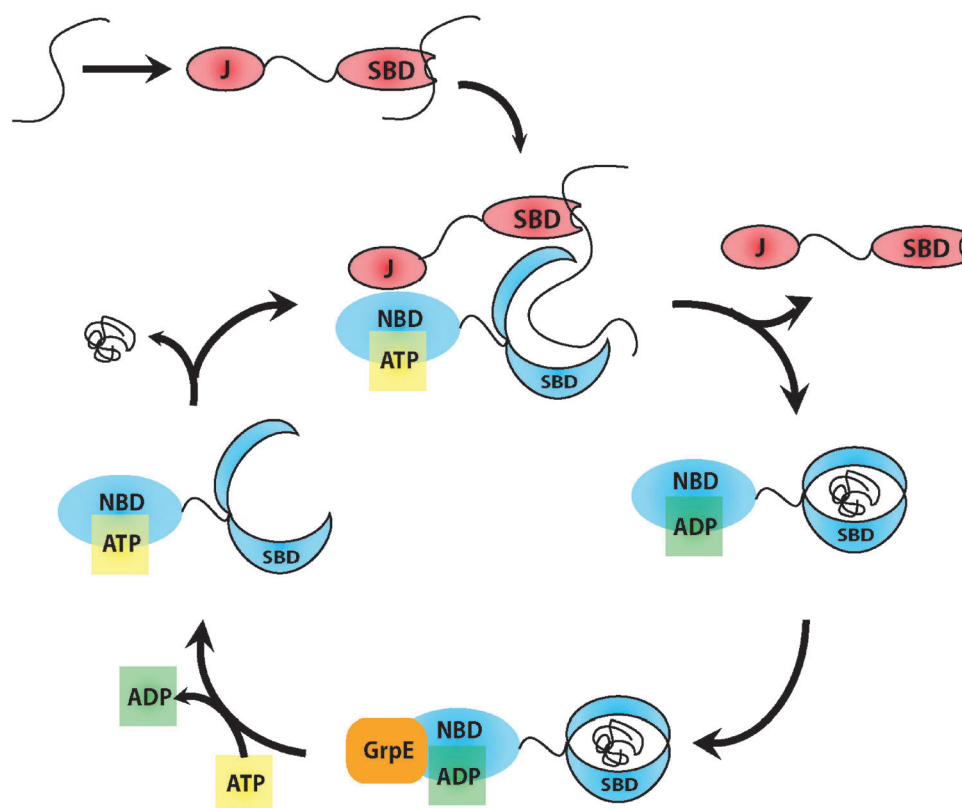
Fig. 1 **DnaK ATPase catalytic cycle**. J-domain proteins recruit client proteins, dock to NBD of DnaK, and catalyze ATP hydrolysis. This converts DnaK from the ATP bound low affinity conformation to the ADP bound conformation with high affinity for client proteins. Nucleotide exchange factors (NEFs), such as GrpE, trigger ADP release. Binding of ATP completes the catalytic cycle by causing release of client protein and co-chaperones.

A recent NMR study has clarified that the J domain interacts with DnaK in a region between the NBD and SBD.[25] The adjacent G/F motif has also been shown to help stimulate DnaK's ATPase activity, while the zinc-finger domain is implicated in contributing to the binding of unfolded peptides.[26,27] Structural studies have suggested that short, hydrophobic peptide sequences bind in a shallow groove in the zinc finger region.[28]

Earlier studies on model substrates had provided a model for how DnaK and DnaJ might cooperate to fold substrates.[29–31] In this model, DnaJ is thought to bind first and then recruit DnaK through its J domain.[23,32] Stimulation of ATP turnover by the combined action of DnaJ and interactions with substrate would then enforce tight binding (Fig. 1). This idea is supported by observations that model substrates containing DnaJ-binding motifs fused to consensus DnaK-binding regions are very effective at forming ternary complexes.[33,34] This synergy is required for the folding of denatured proteins, as evident from the strict dependence for including both DnaK and DnaJ to catalyze folding *in vitro*.[35] Further, folding is blocked by deletion of the J domain, suggesting the interactions between the chaperones are required.[24,36,37] Together, these observations have suggested that DnaK, DnaJ and substrates may need to form transient, ternary complexes to properly assist with the folding of polypeptides (Fig. 1). However, DnaJ is able to prevent protein aggregation without the assistance of DnaK,[27,38,39] suggesting that some chaperone

functions do not have the same requirements. Also, DnaJ will inhibit refolding at elevated concentrations.[27,40] Possible mechanisms to explain these observations include DnaJ oligomerization, DnaJ binding to DnaK's substrate binding cleft, and competition for shared binding sites on unfolded client substrates.[41]

Mutations in *E. coli dnaK* or *dnaJ* cause global problems in protein homeostasis, suggesting that this system acts on a relatively large percentage of the proteome.[42] How can these chaperones be versatile enough to recognize thousands of different sequences in potential substrates? Several studies have explored this question.[9,28,43,44] The major conclusion of these studies is that both molecular chaperones show a strong preference for hydrophobic residues but with only weak ability to discriminate between amino acids. Consistent with this idea, DnaK is known to make contact with both side chains and the peptide backbone, while DnaJ only binds to side chains,[44] providing a way for these chaperones to interact with hydrophobic sequences without strong preference for sequence. Such hydrophobic sequences are commonly found in the interior of folded proteins, allowing for chaperones to discriminate between folded and unfolded proteins. Together, these features appear to provide the necessary promiscuity that permits them to bind a wide range of substrates.

Although a great deal is known about how the DnaK–DnaJ system works, we were interested in the characterization of its ability to promote folding across the entire proteome.

Are DnaK and DnaJ binding sites regularly distributed across all proteins? How many sites are there and what is the spacing between these sites? How does the distribution of DnaK- and DnaJ-binding sites influence the ability of the chaperones to refold each protein? Most studies towards these questions have been necessarily limited to a handful of well-known chaperone "clients", such as luciferase, malate dehydrogenase and citrate synthase.[45–49] Do these classic enzymes accurately represent the properties of the average chaperone substrate *in vivo*? The answers to these questions are critical to understanding how the DnaK–DnaJ system maintains global protein homeostasis in the *E. coli* cytosol.

Towards that goal, we sought a computational, sequence-based approach that would allow us to predict patterns of substrate binding in the *E. coli* proteome. This process relied on the known "consensus" binding motif for DnaK defined by the Bukau group using data obtained from peptide cellulose arrays covering the sequence of several known chaperone clients.[9] Unfortunately, no such motif existed for DnaJ. Recently a binding sequence for a yeast J protein was proposed,[43] but this analysis employed several restrictions that limit the results of such a study. Thus, in this report, we first developed a position-specific, consensus-binding motif for DnaJ using analogous methods as the Bukau group to analyze data from DnaJ binding to client peptide cellulose arrays.[9,44] Then, we employed both the DnaK- and DnaJ-binding preferences to reveal *predicted* sites across the *E. coli* proteome. As previously suggested,[9,44] we found that DnaK and DnaJ are predicted to bind frequently on nearly every protein. Unexpectedly, we also found that the majority of DnaK sites seem to be partially or completely overlapped by DnaJ sites. By studying the experimental folding of two denatured substrates, we also found that the number of shared DnaK–DnaJ sites was a good predictor of the optimal chaperone stoichiometry (*e.g.* [DnaK]:[DnaJ]) needed for refolding. Based on these findings, we propose a revised model of DnaK–DnaJ action in which the chaperones bind either synergistically or competitively, depending on the distribution of sites and the relative stoichiometry of the chaperones.

Experimental procedures

Materials

Reagents were obtained from the following sources: Platinum Pfx DNA Polymerase (Invitrogen, Calsbad, CA), pMCSG7 plasmid (Midwest Center for Structural Genomics, Bethesda, MD), ATP-agarose column (Sigma, St. Louis, MO), firefly luciferase and Steady-Glo reagent (Promega, Madison, WI). All of the OD and luminescence measurements were performed using a SpectraMax M5 (Molecular Devices, Sunnyvale, CA).

DnaJ affinity matrix

Previously reported results that measured the binding of DnaJ to cellulose-bound peptide arrays[44] were used to generate amino acid frequencies for binding (Pb) and non-binding (Pn) peptides. Binding sites were identified using analysis in ImageJ. This frequency table was then used to estimate the Gibb's free energy values for binding of DnaJ to each amino acid at each position in the 13 mer peptides using the equation of Rudiger *et al.*:[9] $\Delta\Delta G_{binding} = -RT \ln(Pb/Pn)$.

Proteome-wide analyses

Proteomic libraries were obtained from UniProtKB databases. Using the DnaK and DnaJ affinity matrices, every possible 13 mer for each sequence in the proteome was parsed, shifting one amino acid at a time, evaluating each 13 mer for its $\Delta\Delta G_{binding}$. Positive binding sites were those in which the $\Delta\Delta G_{binding}$ was calculated as < -5 kJ mol^{-1} (DnaK) or < -2 kJ mol^{-1} (DnaJ). These free energy values represent a binding probability of $> 80\%$, as previously described.[30] The algorithm used is readily available for public use upon request of the authors.

Plasmids and protein purification

The *E. coli* dnaK, dnaJ, and grpE genes were amplified by PCR using Platinum Pfx DNA Polymerase and inserted into the pMCSG7 plasmid through ligation-independent cloning, as previously described.[50] His-tagged DnaK, DnaJ and GrpE were expressed and purified as described previously.[51] Firefly luciferase and malate dehydrogenase (MDH) were purchased from Promega.

Refolding assays

The ability of the DnaK–DnaJ system to refold denatured luciferase was evaluated as described.[40,52] Briefly, a concentrated stock of luciferase (8.2 μM) was denatured using 6 M guanidinium hydrochloride (GuHCl) in HEPES buffer (25 mM HEPES, 50 mM potassium acetate, 5 mM DTT, pH 7.2). This stock was incubated at room temperature for one hour, diluted to 0.2 μM in HEPES buffer lacking GuHCl, and then frozen into aliquots. Enzyme mix (10 μL) containing DnaK, DnaJ, GrpE, denatured firefly luciferase in 39 mM HEPES (170 mM potassium acetate, 1.7 mM magnesium acetate, 3 mM DTT, 12 mM creatine phosphate, 50 U/mL creatine kinase, pH 7.6) was first added into each well of a 96-well plate and then 4 μL of 3.5 mM ATP was added to start the reaction. The final concentrations were: DnaK (1 μM), GrpE (0.2 μM), denatured luciferase (8 nM), and ATP (1 mM). After one hour of incubation at 37 °C equilibrium was reached and 14 μL of 2% (v/v) SteadyGlo reagent in 50 mM glycine buffer (30 mM MgSO$_4$, 10 mM ATP and 4 mM DTT, pH 7.8) was added into each well. For each experiment, the signal from a negative control containing all of the components except DnaK was subtracted.

The MDH refolding assay was based on a previous report.[53] Briefly, MDH (0.2 μM) was denatured at 47 °C for 30 min in 20 mM MOPS buffer (2 mM magnesium acetate, 200 mM KCl, pH = 7.4). Next, 15 μL of this mixture was transferred into each well of a transparent 96-well plate and the refolding reaction was started by adding 10 μL of enzyme mix containing DnaK (1 μM), DnaJ, GrpE (0.2 μM), DTT (5 mM), and ATP (2.5 mM) dissolved in 20 mM MOPS buffer with 25 mM creatine phosphate and 87.5 U mL^{-1} creatine kinase. After 1.5 h incubation at 30 °C, 75 μL of MOPS buffer containing NADH (0.28 mM), OAA (0.5 mM), and BSA (1 mg mL^{-1}) was added to each well, and the decline of OD$_{340}$

over 5 min was recorded. The slope of this data was obtained by linear regression in GraphPad Prism 4.0.

Peptide microarrays

Glass microarrays were purchased from Jenrin Peptide Technologies (JPT, Berlin). The entire sequence of firefly luciferase was printed as 15 residue segments with overlaps of 10 residues. All peptides were arrayed in triplicate. Solutions (300 μL) containing His-tagged DnaJ (3 μM) or DnaK (5 μM) in 25 mM HEPES (100 mM NaCl, 40 mM KCl, 8 mM MgCl$_2$, 0.01% Tween 20, pH 7.4) were incubated in a sandwich apparatus (array slide and blank slide separated by 20 μm plastic spacers) for 1 h at room temperature in a humidified chamber. Arrays were then washed three times (5 min per wash) with TBS-T (50 mM Tris–HCl, 150 mM NaCl, 0.1% Tween 20, pH 7.4) at room temperature in Petri dishes. Excess buffer was removed by centrifugation prior to incubation of the slide with a TBS-T solution containing 1 : 1000 anti-His antibody conjugated with HiLyte-555 (Anaspec) for one hour at room temperature in a humidified chamber protected from light. Following three washes with TBS-T, two with water, and drying under a stream of nitrogen (all carried out in the dark), arrays were scanned with a 532 nm laser at 10 μm resolution on a GenePix 4100A Personal Scanner. Analyses were performed using GenePix Pro software and morphological background correction.

Results

DnaJ and DnaK consensus binding sequence

To probe the *E. coli* proteome for molecular chaperone binding sites, we first needed a sequence-based affinity matrix for DnaK and DnaJ. Although there has been a published matrix of affinity values for DnaK,[9,54] an equivalent matrix for DnaJ has not yet been established. This was probably due to DnaJ's propensity to bind some full-length proteins with a better affinity than smaller peptides.[55] However, for the exclusive purposes of elucidating internal binding sites for DnaJ on fully unfolded proteins (such as might be found in nascent polypeptides), we sought to build such a matrix. For DnaK, Rudiger *et al.* examined every amino acid at each position of a predicted 13 residue binding site.[9] Using a similar approach, we used previously published data on DnaJ binding to cellulose-bound peptide arrays derived from the sequences of several classic chaperone substrates[44] to estimate the affinities. Specifically, we constructed a matrix of $\Delta\Delta G_{binding}$ values for each amino acid at every position in a possible 13mer binding site (Table S1, ESI†) in a fashion similar to that done for DnaK.[9] The total $\Delta\Delta G_{binding}$ for a 13 mer peptide was calculated as the sum of each individual residue. It should be noted that although previous studies have shown a minimal preference of DnaJ for 7 mer or 8 mer substrates,[28,43] we have chosen to proceed with the longer, 13 mer substrate sites because of their use in the cellulose-bound arrays from which this matrix is derived.[44]

Using this data, we developed visual representations of the consensus binding sequences for DnaK and DnaJ (Fig. 2).[56] In concurrence with previous reports,[9] DnaK shows a preference for hydrophobic amino acids (Leu, Ile, Val, Ala, Gly), especially in the core positions 5 through 9. Additionally, DnaK has a preference for positively charged amino acids (Arg, Lys) over negatively charged amino acids (Asp, Glu). Similar to DnaK, we found that DnaJ also has a preference for hydrophobic amino acids, especially aromatic amino acids in positions 5 and 8–11. Although this DnaJ affinity matrix is derived from a relatively limited data set, it agrees well with previous suggestions about a possible consensus.[28,43,44] Further, it is noteworthy that for both chaperones, certain amino acids are overrepresented at specific positions in the 13 mer peptide, instead of displaying a site-independent preference. Together, these affinity matrices and consensus sequences positioned us to explore the binding sites in the entire *E. coli* proteome.

Proteome-wide DnaK and DnaJ binding sites

The DnaK affinity matrix has already been used to predict binding sites in individual proteins[9] and we wished to extend this search to a proteome-wide scale to understand whether classical chaperone substrates are unusual in their distribution or number of DnaK- and DnaJ-binding sites. To address this question, a tight-affinity DnaK binding site has been previously defined as $\Delta\Delta G_{binding} < -5$ kJ mol^{-1}, which corresponds to a probability of binding of $>80\%$. This value has been verified using experimental peptide-binding studies.[30] Using a similar threshold of 80% binding probability, the cutoff for a tight-affinity DnaJ-binding site was determined to be a $\Delta\Delta G_{binding} < -2$ kJ mol^{-1}. We next applied the DnaK and DnaJ affinity matrices to every 13 mer for each protein in the *E. coli* proteome, shifting by a single amino acid with each iteration. Sites that met the free energy cutoffs described above were then annotated as predicted binding sites.

We found that DnaK was predicted to bind 4215 proteins, or 98% of the total annotated *E. coli* proteome (Fig. 3A and Table S2, ESI†). Further, DnaK binds each protein an average of 23 times, although there was a large variance (Table S2, ESI†). This is in good agreement with predictions that DnaK would bind once every 36 residues, a value based on a handful of classic substrates.[9] Thus, we conclude that the model DnaK substrates (*e.g.* luciferase, citrate synthase, *etc.*) are relatively average in their number of predicted binding sites. As might be expected, one major factor in determining the absolute number of DnaK-binding sites in a protein was the length of the polypeptide. To normalize for differences in length, we calculated the contribution of DnaK binding sites to the total protein size and found that over 7% of an individual protein's sequence consists of potential DnaK binding sites (Fig. 3B and Table S2, ESI†). Additionally, DnaK's binding affinity is relatively consistent from site to site, as the average $\Delta\Delta G_{binding}$ was -7.43 ± 2.08 kJ mol^{-1} (Fig. 3C and Table S2, ESI†).

We then compared the DnaK and DnaJ binding sites. We were particularly interested in whether DnaJ might have more/fewer sites and whether the two types of sites were, on average, relatively close together or far apart in a linear peptide sequence. Our results showed that DnaJ was also predicted to interact with a large percentage of the *E. coli* proteome,
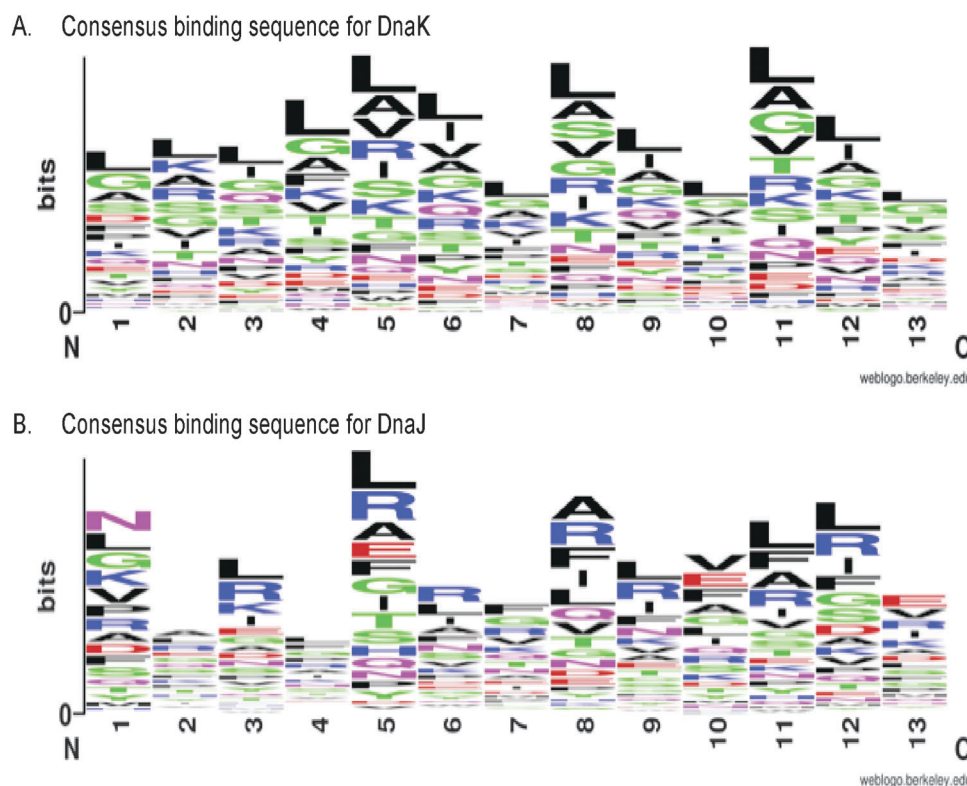
Fig. 2 **Consensus binding sequences of *E. coli* moeclular chaperones**. (A) DnaK: Hydrophobic residues, *e.g.* Leu, Ile, Gly, Val, and basic residues (Arg, Lys) are highly preferred, while acidic residues are disfavored. (B) DnaJ: Hydrophobic and aromatic residues are clearly preferred, but in a position-specific manner. Figure made using WebLogo Tool (Crooks, 2004).
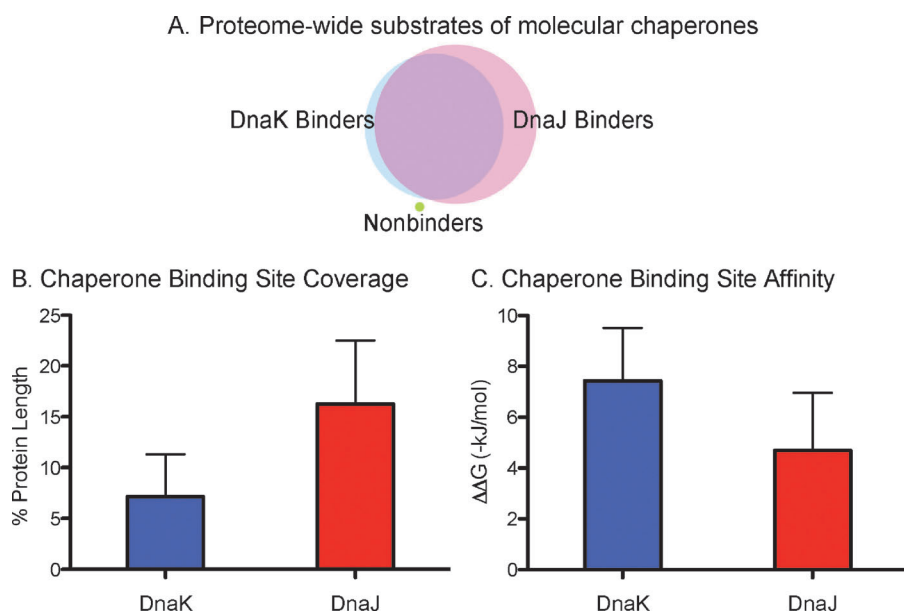

Fig. 3 **Proteome-wide mapping of DnaK and DnaJ binding sites in *E. coli*.** (A) Venn diagram of DnaK and DnaJ-binders. The majority of the proteome is bound by both chaperones, with only one peptide binding neither (UniProtID: P60995). (B) Average binding affinities for DnaK and DnaJ across the proteome. Values given as $-\Delta\Delta G_{binding}$. (C) Chaperones binding site coverage. Values are given as percent coverage of individual protein lengths, averaged over the entire proteome.

binding 4294 proteins or 99.8% of the total proteome (Fig. 3A and Table S3, ESI†). Together, proteins with either a predicted DnaK- or DnaJ-binding site account for 99.99% of the proteome, with the only outlier being a single 16 mer His operon leader peptide (UniProtID: P60995). Interestingly, of the DnaK-binding proteins, all but 9 (0.2%) also were predicted to bind DnaJ. Conversely, 87 proteins (∼2%) were predicted to bind DnaJ only (Fig. 3A and Tables S4 and S5, ESI†).

To exclude the possibility that binding sites were arising from a skewed data population, we ran our algorithm against both the natural proteome and a randomized artificial proteome using our experimentally derived affinity matrices and various scrambled matrices. The controls using the artificial proteome or the scrambled matrices all gave a significantly decreased output in binding sites (data not shown).

While one might suspect that a relatively weaker free energy cutoff for DnaJ (-2 kJ mol^{-1} vs. -5 kJ mol^{-1}) might yield sites with worse affinity, this was not the case. In fact, DnaJ was predicted to bind most proteins with good affinity (average $\Delta\Delta G_{binding}$ was -4.69 ± 2.27), only slightly weaker than DnaK (Fig. 3C and Table S3, ESI†) and indeed statistically indifferent (p-value > 0.05). Next, we were curious as to how frequently DnaJ might bind each protein. Interestingly, despite an average affinity not significantly different from DnaK, DnaJ was predicted to bind over 52 times per protein, which is more than twice the frequency of DnaK. Normalization for protein length showed that DnaJ-binding sites accounted for over 16% of a protein sequence (Fig. 3B and Table S3, ESI†). Since DnaK exhibits a symmetric affinity matrix, we also chose to examine how DnaJ's substrates would change were it to bind similarly as its larger chaperone partner. By averaging the $\Delta\Delta G_{binding}$ values for positions 1 and 13, 2 and 12, 3 and 11, and so on, we were able to render the DnaJ affinity matrix symmetric. Interestingly, although the number of substrates for DnaJ did not decrease significantly, the number of times it bound a single protein, and the total number of binding sites decreased by over 30% (data not shown).

Binding sites for DnaJ and DnaK overlap in the *E. coli* proteome

Previous studies on individual model proteins and peptides suggest that DnaK and DnaJ have similar binding sites that might partially or completely overlap in some cases.[44–49] However, the significance and generality of this finding hasn't been clear. To understand whether overlapping sites are common in a broader context, we first defined overlapping sites as those regions that contain a DnaJ-binding site within 13 residues of a DnaK site (or *vice-versa*). This distance was chosen because of structural work on substrate binding to both chaperones and predictions about steric exclusion.[34,57] We found that an astounding 95% of DnaK sites overlapped with DnaJ sites, with 26% being identical (Fig. 4A). Conversely, only 69% of DnaJ binding sites were predicted to overlap with DnaK. This result is consistent with the promiscuity of DnaJ suggested both experimentally[44,58] and computationally.[43] Curiously, we found no difference in the apparent affinity of DnaK or DnaJ for their substrates regardless of whether the binding site was unique or overlapping (Fig. 4B).

Comparison with GroEL-dependent substrates

GroEL–GroES is the other major arm of the bacterial protein folding system. GroEL forms a double-ring, cylindrical structure that is capped by GroES, forming a hollow interior chamber that promotes folding.[59] Similar to the DnaK–DnaJ system, GroEL–GroES is thought to act on a relatively large number of protein substrates. Consistent with this idea,

A. Binding Site Composition
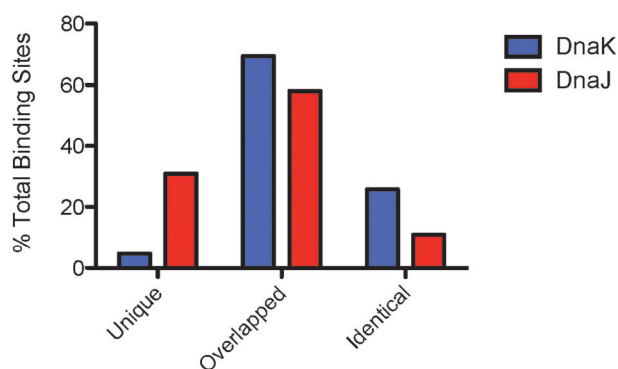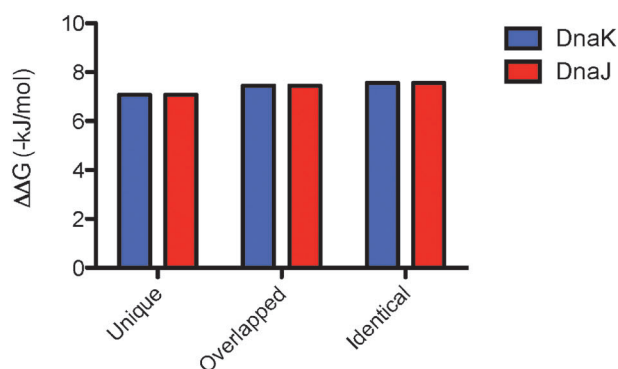

B. Binding Site Affinity Distribution

Fig. 4 **Most DnaK sites are also DnaJ sites**. (A) Chaperone binding site composition given as unique, overlapped (by corresponding chaperone), or identical, given as percent of total sites (DnaK: 95,558; DnaJ: 224,530). (B) Binding site affinity shows little to no variation across the three types of binding sites. Affinity values given as -$\Delta\Delta G_{binding}$.

proteomic studies in a temperature-sensitive GroEL mutant strain revealed 302 proteins, or about 7% of the proteome, that are dependent on this chaperone for solubility.[60] We wondered whether these specific substrates might have an unusual pattern of DnaK- or DnaJ-binding sites and applied our algorithm to this subset. However, we found that both DnaK and DnaJ bound nearly all of the GroEL substrates (DnaK: 98.3%, DnaJ: 99.7%) with no apparent differences in the predicted affinity or number of sites (data not shown). These findings suggest that the GroEL-dependent proteins also contain predicted DnaK–DnaJ sites; thus, some other feature must make these proteins particularly dependent on GroEL–GroES.

Analysis of DnaK- and DnaJ-binding sites in thermophilic archaea

The DnaK–DnaJ system is highly conserved from bacteria to mammals. However, a recent study by the Kampinga group has identified a handful of archaea that lack an obvious DnaK or DnaJ ortholog.[61] We are interested in whether the proteomes from these organisms might have distinct patterns of chaperone-binding sites. To this end, we explored the proteomes of *Thermococcus kodakaraensis* and *Methanococcus janaschii*, two archaea species lacking *dnaJ* and *dnaK* genes.[61]

Table 1 **Evolutionary conservation of chaperone binding sites**. Analysis of archaeal organisms shows binding site coverage by DnaK and DnaJ of nearly the entire proteome

Organism	DnaK homolog	DnaJ homolog	DnaK	DnaJ
E. coli	+	+	95 558 sites; 4215 proteins (98.0%)	224 530 sites; 4294 proteins (99.8%)
T. acidophilum	+	+	33 849 sites; 1480 proteins (99.9%)	89 326 sites; 1481 proteins (99.9%)
M. mazei	+	+	62 680 sites; 3297 proteins (99.8%)	169 210 sites; 3300 proteins (99.9%)
T. kodakaraensis	−	−	132 343 sites; 2305 proteins (99.9%)	191 386 sites; 2308 proteins (100%)
M. janaschii	−	−	34 317 sites; 1785 proteins (99.9%)	94 498 sites; 1786 proteins (99.9%)

As a control, we also examined the proteomes of two additional thermophilic archaea that have *dnaJ* and *dnaK* genes (*Thermoplasma acidophilum* and *Methanosarcina mazei*). Interestingly, we found that the proteomes of all four archaea contained thousands of predicted binding sites, with nearly complete coverage in every case (Table 1). This finding suggests that potential chaperone-binding sequences may have existed before the evolution of DnaK–DnaJ. In other words, the proteome within these organisms still folds and buries hydrophobic, aggregation-prone sequences, but lacks a molecular mechanism for aiding the folding of unfolded proteins or misfolded intermediates.

Revised binding model for the DnaK–DnaJ-substrate ternary complex

Previous studies have converged on two possible models for how DnaK–DnaJ might cooperate to fold substrates: one in which DnaJ and DnaK separately bind polypeptide chains[33,48] and a second in which the two chaperones compete for the same site and substrates are transferred from one chaperone to the other. The first model is often preferred based on data clearly showing enhanced folding by the DnaK–DnaJ system.[33,34,44] However, it is also commonly observed that DnaJ will block folding at higher concentrations,[27,40] suggesting that competition can also take place.

Our data on predicted DnaK and DnaJ binding sites across the *E. coli* proteome suggest that both ideas may be valid. Because DnaJ sites outnumber and overlap nearly all DnaK sites, we suggest a more generalized model, incorporating both these ideas, in which unique DnaJ sites are interspersed between sites shared by both DnaK and DnaJ sites (Fig. 5). This observation suggests that ternary complex formation will occur only if the shared DnaK/DnaJ sites are not saturated with DnaJ. Such a situation would allow DnaJ and DnaK to

Fig. 5 **Model for the arrangement of DnaK and DnaJ binding sites on unfolded substrate proteins**. Unique DnaJ sites are intersperced between common DnaK and DnaJ binding targets.

bind substrates synergistically, but only under conditions of permissive DnaK : DnaJ stoichiometry.

The optimal DnaK : DnaJ ratio for substrate refolding correlates with the composition of computationally predicted binding sites

Before testing this hybrid model, we wanted to first understand whether our computationally predicted DnaK- and DnaJ-binding sites were consistent with experimentally determined binding regions. Towards that goal, we focused on firefly luciferase, a well-studied model chaperone substrate. We designed glass arrays spotted with peptides representing the entire luciferase protein; it contained 136 15 mers with 10 residue overlaps in triplicate. Binding of His-tagged DnaK (5 μM) or DnaJ (3 μM) to these arrays was measured using fluorescent anti-His antibodies. For both DnaK and DnaJ, we found that the predicted binding sites matched remarkably well with the experimental results (Fig. 6A and B and 7B). Fig. 6C gives a quantitative comparison of DnaK and DnaJ binding across the luciferase sequence. Further, this binding data appears to be conserved even in human chaperones (data not shown).

Next, we wanted to understand how the proportion of unique and shared DnaK- and DnaJ-binding sites might contribute to the optimum stoichiometry between DnaK and DnaJ needed to achieve folding. In this task, we compared the chaperone-assisted folding of luciferase with another model substrate, malate dehydrogenase (MDH). These two substrates are particularly well suited for this question because luciferase is predicted to have 134 possible DnaJ binding sites, of which a remarkable ~81% are shared with DnaK, while MDH has 61 total DnaJ binding sites and only ~72% are predicted to be shared with DnaK. Thus, these two substrates differ in their total number of sites and their percentage of unique DnaJ binding sites (Table 2). Further, *in vitro* methods for the chaperone-mediated refolding of these substrates are well established.[40] Thus, although analyses of linear, extended sequences are clearly an over-simplification of the complexity of protein folding reactions, we thought that these substrates might serve as useful tools for probing the potential relevance of the predicted binding sites. Towards that goal, we denatured recombinant luciferase and MDH, diluted them to the final concentration previously reported to be best for refolding,[53] and then monitored the refolding reactions by monitoring the recovery of enzymatic activity (light production or malate conversion to oxaloacetate, respectively). To identify the optimal ratio between DnaK and DnaJ needed to achieve optimal refolding, we fixed the concentration of DnaK and varied the levels of DnaJ. This approach revealed that the optimal ratio between DnaK and DnaJ needed to refold luciferase was approximately 5 : 1. This result is consistent
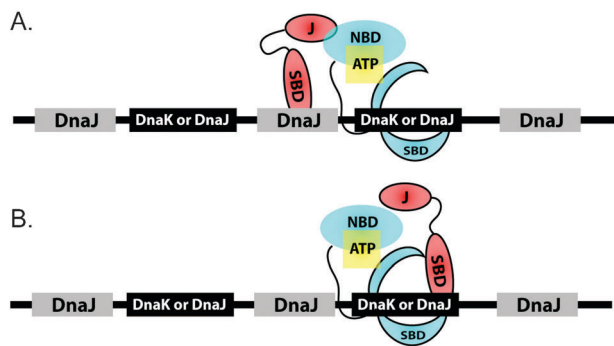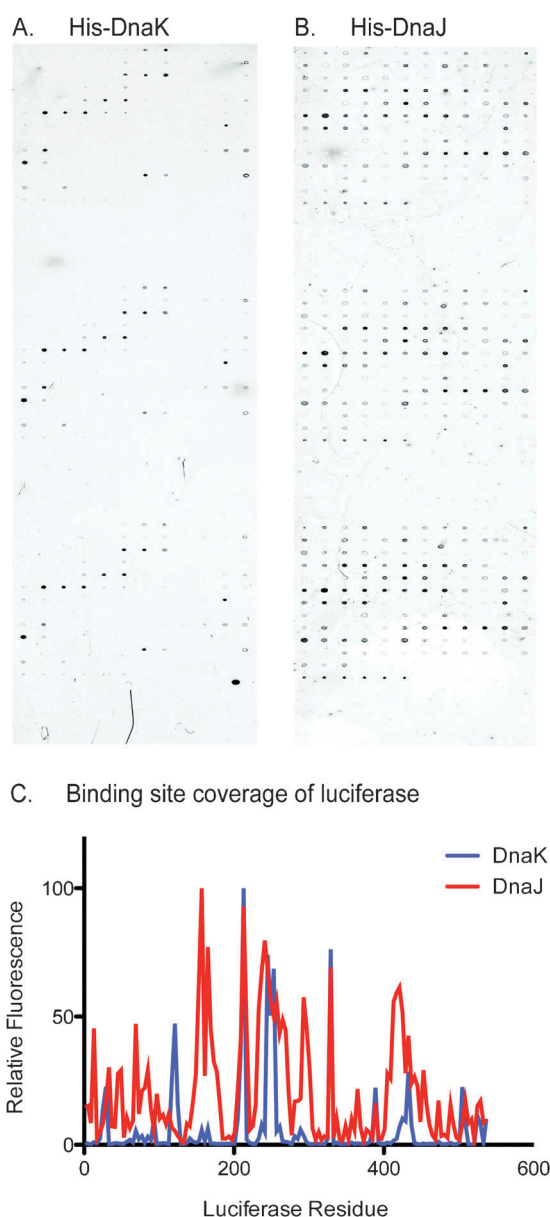
A. His-DnaK

B. His-DnaJ

C. Binding site coverage of luciferase

Fig. 6 **Glass-slide peptide microarrays confirm computational prediction of chaperone binding sites on luciferase**. (A) and (B). His-tagged DnaK and DnaJ, respectively, binding of luciferase confirm *in silico* predictions (Fig. 6). Results correlate well with previously published platforms (Rudiger, 1997; Rudiger, 2001). (C) Quantified binding site coverage for molecular chaperones on luciferase. Fluorescence is normalized to maximum values for each array.

with previous reports suggesting luciferase refolding only occurs under conditions of substoichiometric DnaJ.[21] Interestingly, the stoichiometry of these chaperones in the *E. coli* cytosol is a similar 10 : 1.[62] Next, we performed parallel experiments using denatured MDH and, interestingly, found that the optimal DnaK : DnaJ ratio was only 2 : 1 (Fig. 7A). As expected, higher concentrations of DnaJ suppressed folding of both substrates. The suppression of folding by DnaJ was more dramatic in the case of luciferase than MDH, but we hesitate to interpret this particular observation because luciferase appears to be more prone to off-pathway, irreversible aggregation.

Comparing the results of the refolding assays to the predicted number and composition of DnaK and DnaJ binding sites (Fig. 7B and C) suggest some potentially interesting connections. As discussed above, luciferase is predicted to have a relatively higher percentage of shared binding sites, when compared to MDH. Additionally, DnaJ seems to bind its sites on luciferase with a higher affinity than on MDH (Table 2). What does this mean for the refolding reactions? If DnaJ randomly "selects" binding sites, as was suggested by the relatively narrow differences between free energies, then DnaJ would be expected to compete with DnaK binding sites more often in luciferase. This might be why luciferase becomes "saturated" with DnaJ at a relatively lower concentration. Alternatively, DnaJ's higher affinity binding of luciferase compared to MDH might lead to a longer duration of interaction, preventing DnaK association, thus resulting in refolding inhibition. Although this hypothesis needs further validation using a wider range of substrates, these observations provide an intriguing initial link between computationally predicted binding sites and their possible impact on chaperone-mediated folding *in vitro*.

Discussion and conclusions

There has been great interest in understanding how the molecular chaperones DnaK and DnaJ interact with substrates to assist their folding. Pioneering work by the Bukau group and others has established consensus-binding sequences for DnaK and provided substantial insight into this bimolecular interaction on a limited number of substrates.[9,28,30,44–49] In our study, we sought to expand this analysis to understand how DnaK and its important co-chaperone, DnaJ, might interact with the entire proteome. In doing so, we developed an affinity matrix for DnaJ (Table S1, ESI†) and probed the large-scale proteome. Our results indicate that the two chaperones are predicted to interact with 99.99% of the *E. coli* proteome (Fig. 3A). While this number exceeds that of previous predictions,[63] it is likely that the potential to bind DnaK or DnaJ is advantageous to a client, particularly during stress conditions when these chaperones are most active.[4,5] The frequency of DnaK sites is in good agreement with the frequency observed using classic chaperone substrates,[9] suggesting at a proteome wide level that these classic chaperone substrates serve as a good representation of how DnaK and DnaJ interact with proteins. Furthermore, we found that both chaperones bound each substrate multiple times, with DnaJ binding sites taking up nearly twice as much space on the linear peptide sequences as sites for DnaK (16% *vs.* 7%). Unexpectedly, we found that more than 95% of predicted DnaK sites overlap with a DnaJ site (see Fig. 4A). Moreover, given the lack of substantial differences in the affinities for the different binding sites, it is unlikely that DnaK and DnaJ strongly prefer one site to another in the absence of secondary structure or other factors.

Next, we applied our algorithm to the proteomes of several archea, including two recently identified as lacking genes for homologs of DnaK or DnaJ.[61] The results indicated that all sets of proteomes contained numerous chaperone target sites, even in those without actual chaperones (Table 1). This data
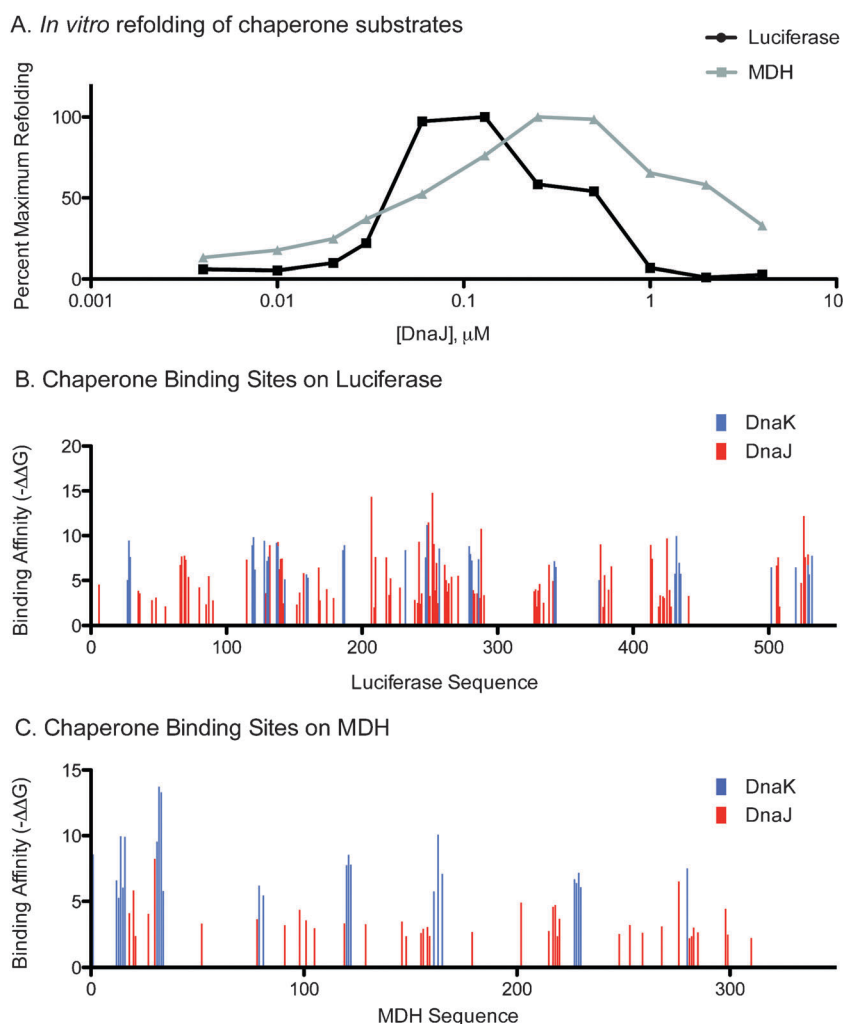
A. *In vitro* refolding of chaperone substrates

B. Chaperone Binding Sites on Luciferase

C. Chaperone Binding Sites on MDH

Fig. 7 **DnaJ differentially inhibits luciferase and MDH refolding**. (A) *In vitro* refolding assay shows DnaJ inhibiting refolding of luciferase at lower concentrations than for MDH. (B) and (C) Computational prediction of chaperone binding sites. Note luciferase contains many more DnaJ sites that bind at higher affinity than MDH.

Table 2 **Predicted DnaK and DnaJ binding sites on luciferase and MDH**. Luciferase contains over 10% more shared binding sites relative to MDH and is predicted to bind DnaJ significantly tighter than MDH

	Luciferase	MDH
Unique DnaK sites	0	1 (1.6%)
Unique DnaJ sites	25 (18.7%)	17 (27.8%)
Shared binding sites	109 (81.3%)	43 (70.5%)
Total Sites	134	61
Average $\Delta\Delta G_{binding}$ for DnaK (kJ mol^{-1}); p-value > 0.05	-7.51 ± 1.58	-7.87 ± 1.32
Average $\Delta\Delta G_{binding}$ for DnaJ (kJ mol^{-1}); p-value $= 7.86e{-}5$	-5.27 ± 2.78	-3.43 ± 1.27

supports the idea that chaperones may have emerged downstream of their clients in the evolutionary pathway. As the binding sites for yeast and human chaperones are elucidated, it will be most interesting to apply our algorithm to the corresponding proteomes for a more detailed analysis.

Finally, these studies have led us to propose a "hybrid" model for how DnaK and DnaJ bind to their peptide substrates. In this model, unique DnaJ sites are interspersed between sites that are shared by both DnaK and DnaJ (Fig. 5). This model includes aspects consistent with previous ideas, which feature either simultaneous binding of substrate

by both chaperones (*e.g.* formation of a ternary complex) or sequential transfer of client proteins from DnaJ to DnaK.[23,44] The hybrid model accommodates both these theories, and argues that the individual substrate may determine which mechanism is dominant. For example, the outcome of interactions between the DnaK–DnaJ system and its various substrates are predicted to depend on: (a) the relative stoichiometry of DnaJ and DnaK, (b) the proportion of shared sites and (c) the number of unique DnaJ sites. In contrast, the individual affinities of DnaJ and DnaK for their substrates are predicted to be relatively minor contributing factor,

though can factor in on occasion (such as with luciferase compared to MDH, Table 2). The issue of stoichiometry, though, appears to be a particularly critical component in this model because of potential competition between DnaJ and DnaK on shared sites. In this way, our model agrees with published data regarding the optimal stoichiometry of DnaK and DnaJ for proper substrate folding.[27,40] Specifically, at substoichiometric levels, DnaJ is known to accelerate substrate refolding, likely due to its help in recruiting client to DnaK and stimulating ATP turnover. However, (as shown in Fig. 7 and Table 1), our data suggests that the DnaJ-inhibition of refolding is due to DnaJ competing with DnaK for shared binding sites. While this is a thermodynamic computation analysis of binding sites, interestingly we have reached a similar conclusion to that postulated from a kinetic computational model of refolding. In 2006, Hu et al. used kinetic data regarding the on and off rates of DnaK, DnaJ, and substrate binding interactions to model refolding outcomes.[41] They concluded that competition for substrate was a plausible model. Taken together, the work of Hu and the work presented herein provide strong support. Indeed, high levels of DnaJ might compete with DnaK for shared binding sites, or alternatively block refolding by staying on the high-affinity unique sites for "too long", as substrates are unlikely to fold if chaperones remain bound to their hydrophobic regions. Though individual substrates may indeed follow either mechanism, the latter idea – at least for luciferase and MDH – appears to be very important given the difference in their binding affinities for DnaJ. Thus, the reversibility of chaperone interactions is important, which could be an advantage of having a high percentage of shared DnaK- and DnaJ-binding sites.

Acknowledgements

References

1 D. E. Feldman and J. Frydman, Curr. Opin. Struct. Biol., 2000, 10, 26–33.
2 P. Genevaux, C. Georgopoulos and W. L. Kelley, Mol. Microbiol., 2007, 66, 840–857.
3 C. Georgopoulos, Genetics, 2006, 174, 1699–1707.
4 S. Hansen, K. Lewis and M. Vulic, Antimicrob. Agents Chemother., 2008, 52, 2718–2726.
5 A. Liu, L. Tran, E. Becket, K. Lee, L. Chinn, E. Park, K. Tran and J. H. Miller, Antimicrob. Agents Chemother., 2010, 54, 1393–1403.
6 E. B. Bertelsen, L. Chang, J. E. Gestwicki and E. R. Zuiderweg, Proc. Natl. Acad. Sci. U. S. A., 2009, 106, 8471–8476.
7 S. Blond-Elguindi, S. E. Cwirla, W. J. Dower, R. J. Lipshutz, S. R. Sprang, J. F. Sambrook and M. J. Gething, Cell, 1993, 75, 717–728.
8 X. Zhu, X. Zhao, W. F. Burkholder, A. Gragerov, C. M. Ogata, M. E. Gottesman and W. A. Hendrickson, Science, 1996, 272, 1606–1614.
9 S. Rudiger, L. Germeroth, J. Schneider-Mergener and B. Bukau, EMBO J., 1997, 16, 1501–1507.
10 G. C. Flynn, J. Pohl, M. T. Flocco and J. E. Rothman, Nature, 1991, 353, 726–730.
11 A. Gragerov, L. Zeng, X. Zhao, W. Burkholder and M. E. Gottesman, J. Mol. Biol., 1994, 235, 848–854.
12 R. Schlecht, A. H. Erbse, B. Bukau and M. P. Mayer, Nat. Struct. Mol. Biol., 2011, 18, 345–351.
13 J. Jiang, E. G. Maes, A. B. Taylor, L. Wang, A. P. Hinck, E. M. Lafer and R. Sousa, Mol. Cell, 2007, 28, 422–433.
14 F. Moro, V. Fernandez-Saiz and A. Muga, Protein Sci., 2006, 15, 223–233.
15 M. Pellecchia, D. L. Montgomery, S. Y. Stevens, C. W. Vander Kooi, H. P. Feng, L. M. Gierasch and E. R. Zuiderweg, Nat. Struct. Biol., 2000, 7, 298–303.
16 S. V. Slepenkov and S. N. Witt, FEBS Lett., 2003, 539, 100–104.
17 J. F. Swain, G. Dinler, R. Sivendran, D. L. Montgomery, M. Stotz and L. M. Gierasch, Mol. Cell, 2007, 26, 27–39.
18 H. H. Kampinga and E. A. Craig, Nat. Rev. Mol. Cell Biol., 2010, 11, 579–592.
19 A. W. Karzai and R. McMacken, J. Biol. Chem., 1996, 271, 11236–11246.
20 W. L. Kelley, Trends Biochem. Sci., 1998, 23, 222–227.
21 T. Laufen, M. P. Mayer, C. Beisel, D. Klostermeier, A. Mogk, J. Reinstein and B. Bukau, Proc. Natl. Acad. Sci. U. S. A., 1999, 96, 5452–5457.
22 K. Liberek, J. Marszalek, D. Ang, C. Georgopoulos and M. Zylicz, Proc. Natl. Acad. Sci. U. S. A., 1991, 88, 2874–2878.
23 B. Misselwitz, O. Staeck and T. A. Rapoport, Mol. Cell, 1998, 2, 593–603.
24 D. Wall, M. Zylicz and C. Georgopoulos, J. Biol. Chem., 1994, 269, 5446–5451.
25 A. Ahmad, A. Bhattacharya, R. A. McDonald, M. Cordes, B. Ellington, E. B. Bertelsen and E. R. Zuiderweg, Proc. Natl. Acad. Sci. U. S. A., 2011, 108, 18966–18971.
26 B. Banecki, K. Liberek, D. Wall, A. Wawrzynow, C. Georgopoulos, E. Bertoli, F. Tanfani and M. Zylicz, J. Biol. Chem., 1996, 271, 14840–14848.
27 A. Szabo, R. Korszun, F. U. Hartl and J. Flanagan, EMBO J., 1996, 15, 408–417.
28 J. Li, X. Qian and B. Sha, Structure, 2003, 11, 1475–1483.
29 F. U. Hartl, Nat. Med., 2011, 17, 1206–1210.
30 F. Rodriguez, F. Arsene-Ploetze, W. Rist, S. Rudiger, J. Schneider-Mergener, M. P. Mayer and B. Bukau, Mol. Cell, 2008, 32, 347–358.
31 D. W. Summers, P. M. Douglas, C. H. Ramos and D. M. Cyr, Trends Biochem. Sci., 2009, 34, 230–233.
32 E. V. Pierpaoli, E. Sandmeier, H. J. Schonfeld and P. Christen, J. Biol. Chem., 1998, 273, 6643–6649.
33 W. Han and P. Christen, J. Biol. Chem., 2003, 278, 19038–19043.
34 W. Han and P. Christen, FEBS Lett., 2004, 563, 146–150.
35 A. Szabo, T. Langer, H. Schroder, J. Flanagan, B. Bukau and F. U. Hartl, Proc. Natl. Acad. Sci. U. S. A., 1994, 91, 10345–10349.
36 K. Linke, T. Wolfram, J. Bussemer and U. Jakob, J. Biol. Chem., 2003, 278, 44457–44466.
37 J. Perales-Calvo, A. Muga and F. Moro, J. Biol. Chem., 2010, 285, 34231–34239.
38 C. Y. Fan, S. Lee and D. M. Cyr, Cell Stress Chaperones, 2003, 8, 309–316.
39 Z. Lu and D. M. Cyr, J. Biol. Chem., 1998, 273, 27824–27830.
40 L. Chang, A. D. Thompson, P. Ung, H. A. Carlson and J. E. Gestwicki, J. Biol. Chem., 2010, 285, 21282–21291.
41 B. Hu, M. P. Mayer and M. Tomita, Biophys. J., 2006, 91, 496–507.
42 J. S. McCarty, A. Buchberger, J. Reinstein and B. Bukau, J. Mol. Biol., 1995, 249, 126–137.
43 P. Kota, D. W. Summers, H. Y. Ren, D. M. Cyr and N. V. Dokholyan, Proc. Natl. Acad. Sci. U. S. A., 2009, 106, 11073–11078.
44 S. Rudiger, J. Schneider-Mergener and B. Bukau, EMBO J., 2001, 20, 1042–1050.
45 J. Frydman and F. U. Hartl, Science, 1996, 272, 1497–1502.
46 J. Gamer, H. Bujard and B. Bukau, Cell, 1992, 69, 833–842.
47 J. Gamer, G. Multhaup, T. Tomoyasu, J. S. McCarty, S. Rudiger, H. J. Schonfeld, C. Schirra, H. Bujard and B. Bukau, EMBO J., 1996, 15, 607–617.
48 S. Y. Kim, S. Sharma, J. R. Hoskins and S. Wickner, J. Biol. Chem., 2002, 277, 44778–44783.
49 A. Wawrzynow and M. Zylicz, J. Biol. Chem., 1995, 270, 19300–19306.

50 L. Stols, M. Gu, L. Dieckman, R. Raffen, F. R. Collart and M. I. Donnelly, *Protein Expression Purif.*, 2002, **25**, 8–15.

51 L. Chang, Y. Miyata, P. M. Ung, E. B. Bertelsen, T. J. McQuade, H. A. Carlson, E. R. Zuiderweg and J. E. Gestwicki, *Chem. Biol.*, 2011, **18**, 210–221.

52 S. Wisen and J. E. Gestwicki, *Anal. Biochem.*, 2008, **374**, 371–377.

53 S. Diamant and P. Goloubinoff, *Biochemistry*, 1998, **37**, 9688–9694.

54 A. de Crouy-Chanel, M. Kohiyama and G. Richarme, *J. Biol. Chem.*, 1996, **271**, 15486–15490.

55 B. Feifel, H. J. Schonfeld and P. Christen, *J. Biol. Chem.*, 1998, **273**, 11999–12002.

56 G. E. Crooks, G. Hon, J. M. Chandonia and S. E. Brenner, *Genome Res.*, 2004, **14**, 1188–1190.

57 M. P. Mayer, S. Rudiger and B. Bukau, *Biol. Chem.*, 2000, **381**, 877–885.

58 J. Li and B. Sha, *Biochem. J.*, 2005, **386**, 453–460.

59 A. L. Horwich, G. W. Farr and W. A. Fenton, *Chem. Rev.*, 2006, **106**, 1917–1930.

60 E. Chapman, G. W. Farr, R. Usaite, K. Furtak, W. A. Fenton, T. K. Chaudhuri, E. R. Hondorp, R. G. Matthews, S. G. Wolf, J. R. Yates, M. Pypaert and A. L. Horwich, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 15800–15805.

61 J. Hageman, M. A. van Waarde, A. Zylicz, D. Walerych and H. H. Kampinga, *Biochem. J.*, 2011, **435**, 127–142.

62 J. C. Bardwell, K. Tilly, E. Craig, J. King, M. Zylicz and C. Georgopoulos, *J. Biol. Chem.*, 1986, **261**, 1782–1785.

63 F. U. Hartl and M. Hayer-Hartl, *Nat. Struct. Mol. Biol.*, 2009, **16**, 574–581.