9-1-2011

Components of a Protein Machine: Allosteric Domain Assembly and a Disordered C-terminus Enable the Chaperone Functions of Hsp70

Robert G. Smock
University of Massachusetts Amherst, rsmock@student.umass.edu

COMPONENTS OF A PROTEIN MACHINE: ALLOSTERIC DOMAIN ASSEMBLY AND A DISORDERED C-TERMINUS ENABLE THE CHAPERONE FUNCTIONS OF HSP70

A Dissertation Presented

by

ROBERT SMOCK

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2011

Molecular and Cellular Biology

COMPONENTS OF A PROTEIN MACHINE:
ALLOSTERIC DOMAIN ASSEMBLY AND A DISORDERED C-TERMINUS ENABLE THE CHAPERONE FUNCTIONS OF HSP70

A Dissertation Presented

by

ROBERT SMOCK

Approved as to style and content by:

_____
Lila Gierasch, Chair

_____
Craig Martin, Member

_____
Jeanne Hardy, Member

_____
Paul Lahti, Member

_____
Dominique Alfandari, Acting Director
Molecular and Cellular Biology Program

ABSTRACT

COMPONENTS OF A PROTEIN MACHINE:

ALLOSTERIC DOMAIN ASSEMBLY AND A DISORDERED C-TERMINUS

ENABLE THE CHAPERONE FUNCTIONS OF HSP70

SEPTEMBER 2011

ROBERT SMOCK

B.S., University of Massachusetts Amherst

Ph.D., University of Massachusetts Amherst

Directed by Professor Lila Gierasch

Hsp70 molecular chaperones protect proteins from aggregation, assist in their native structure formation, and regulate stress responses in the cell. A mechanistic understanding of Hsp70 function will be necessary to explain its physiological roles and guide the therapeutic modulation of various disease states. To this end, several fundamental features of the Hsp70 structure-function relationship are investigated. The central component of Hsp70 chaperone function is its capacity for allosteric signaling between structural domains and tunable binding of misfolded protein substrates. In order to identify a cooperative network of sites that mediates interdomain allostery within Hsp70, a mutational correlation analysis is performed using genetic data. Evolutionarily correlations that describe an allosteric network are validated by examining roles for implicated sites in cellular fitness and molecular function. In a second component of the Hsp70 molecular mechanism, a novel function is discovered for the disordered

C-terminal tail. This region of the protein enhances the refolding efficiency of substrate proteins independently of interdomain allostery and is required in the cell upon depletion of compensatory chaperones, suggesting a previously undescribed mode of chaperone action. Finally, experiments are initiated to assess the dynamic assembly of Hsp70 domains in various allosteric states and how domain orientations may be guided through interaction with partner co-chaperone proteins.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Parts of this chapter are taken from a section of a book chapter written in collaboration with Lila Gierasch (Buchner *et al.* in press, Publication 7).

1.1 Molecular chaperones and their roles in the cell

The complex diversity of life forms indicates their elaborate and malleable underpinnings, yet all organisms share the most fundamental cellular processes. Perhaps central is the use of DNA to precisely encode the synthesis of polypeptides, and the subsequent folding of these linear chains into structured proteins that perform functions in the cell required for life (Crick 1970). In dilute solution, a protein can spontaneously fold to its native, functional structure through the formation of energetically favorable interactions within the polypeptide chain (Anfinsen 1973). However, cells are not composed of dilute solution. Molecular crowding forces competition between intramolecular collapse that drives protein folding and intermolecular collapse that drives protein aggregation (Gershenson and Gierasch 2011). Moreover, the energetic favorability of folding to the native state is altered by temperature, acidity, oxidation, genetic mutation and other naturally occurring variables that cells must tolerate over a wide range of influence.

Molecular chaperones are a broadly defined functional class of proteins that help cells to cope with these challenges (Hartl and Hayer-Hartl 2002). Chaperones reversibly bind unfolded proteins, protecting them from aggregation,

and can also act as folding catalysts by accelerating the rate of native structure formation. A wide array of chaperones is utilized under normal conditions and expressed at higher levels in response to physical stress, such as elevated temperature. Many molecular chaperones have been identified in this way and labeled as ‘heat shock proteins’.

Members of an important family of molecular chaperones, the 70 kDa heat shock proteins (Hsp70s), are ubiquitously expressed in all organisms and in most subcellular compartments. In cooperation with other chaperone systems, Hsp70 proteins facilitate the folding of newly synthesized and misfolded polypeptides and the assembly/ disassembly of specific protein complexes, thereby rescuing non-native proteins from aggregation and promoting formation of native protein structure. In addition to chaperoning folding events, Hsp70s are also involved in protein translocation across membranes and protein degradation (Mayer and Bukau 2005). Furthermore, Hsp70s act as folding sensors in the cell. The interaction of Hsp70 with misfolded protein substrates under conditions of cellular stress competes with the binding of a heat shock transcription factor, positioning Hsp70 as a central regulator for the expression of a multitude of chaperone families (Morimoto 1993; Gamer *et al.* 1996). Given their diverse physiological roles, it is not surprising that Hsp70s are implicated in a number of disease states in humans, including neurodegenerative diseases (*e.g.*, Alzheimer’s, Huntington’s and Parkinson’s diseases), cystic fibrosis, and cancer, among others. These findings have elicited interest in the therapeutic modulation of Hsp70 (Patury *et al.* 2009).

Hsp70 performs its functional roles in cooperation with other molecular chaperones. Two classes of co-chaperones, nucleotide exchange factors (NEFs) and Hsp40s (also called J-domain proteins), regulate Hsp70's cycle of substrate binding and release (Fig. 1.1). In eukaryotes, Hsp70 also binds to TPR domain proteins which promote assembly with Hsp90 or an E3 ubiquitin ligase (Mayer and Bukau 2005). Functional interaction and complementarity of Hsp70 with other molecular chaperone families have also been observed, such as with Hsp110 (GroEL) (Gragerov *et al.* 1992), Hsp100 (ClpB) (Weibezahn *et al.* 2004), trigger factor (Deuerling *et al.* 1999), SecB (Altman *et al.* 1991), and the small heat shock proteins (Lee and Vierling 2000).

It is not entirely clear how Hsp70 functions as a chaperone, although several leading models represent plausible, if incomplete, explanations. A kinetic partitioning model seems to apply to all families of molecular chaperones. In this scheme (Fig. 1.2A), substrates may spontaneously fold to their native state or become kinetically trapped in a folding intermediate. Chaperones selectively bind misfolded substrates, thereby partitioning rapid folders to the native state and kinetically trapped misfolders to the chaperone-bound state, which maintains substrate solubility and lowers the rate of aggregation. This model applies to chaperones whose affinity for misfolded substrates is unregulated, such as SecB (Hardy and Randall 1991) and small heat shock proteins (Jakob *et al.* 1993), as well as to chaperones that undergo regulated cycles of substrate binding and release, such as GroEL (Fedorov and Baldwin 1997) and Hsp70 (Slepenkov and Witt 2002). In addition to reducing the rate of substrate aggregation, some

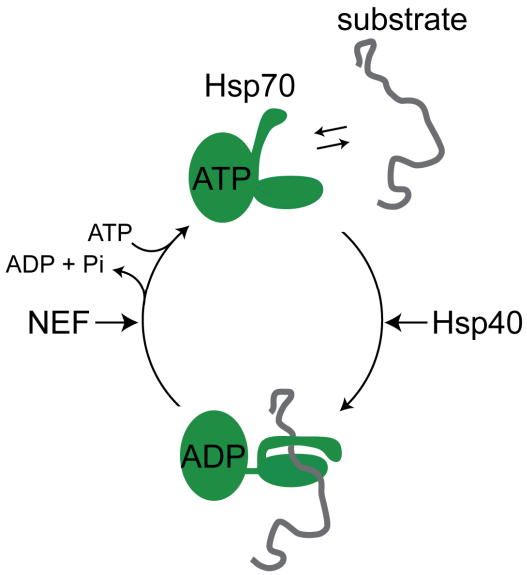

Figure 1.1. Regulated substrate binding and release in Hsp70 proteins.

Hsp70 proteins cycle between states of high and low affinity for non-native protein substrates. These transitions are mediated by the binding and hydrolysis of ATP, which is in turn regulated by nucleotide exchange factors (NEFs) and Hsp40/ J-domain proteins. NEFs are a functional class of proteins that catalyze the exchange of ADP for ATP (Mayer and Bukau 2005). Hsp40 proteins all carry a J domain capable of stimulating Hsp70 ATPase activity but are otherwise highly diversified and provide functional specificity for this chaperone system (Kampinga and Craig 2010).

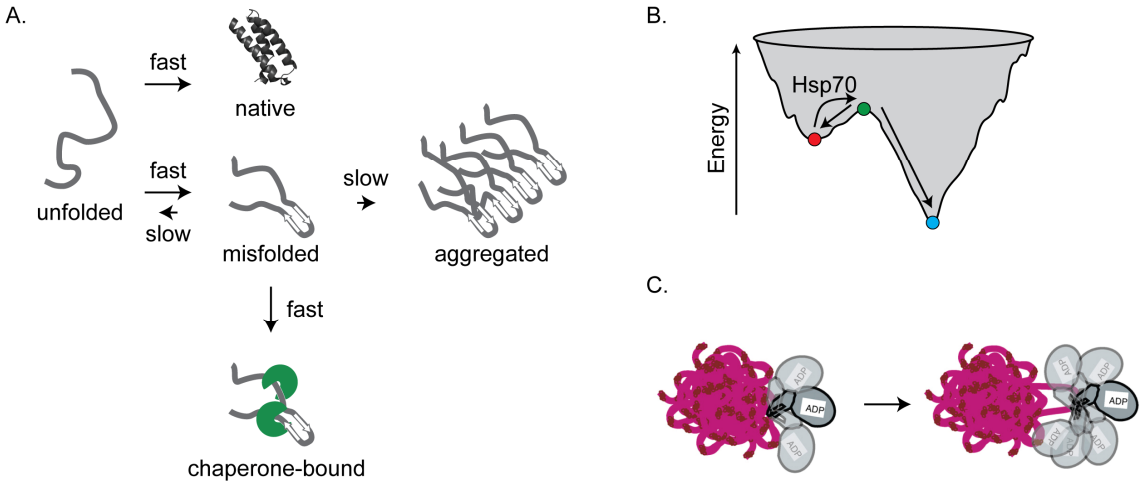

Figure 1.2. Models for Hsp70 chaperone function.

A. Kinetic partitioning. Some proteins may rapidly fold to the native state or misfold into a kinetically trapped intermediate. In the absence of chaperones, the kinetically trapped state may either slowly unfold or form aggregates. Chaperones provide an alternative fate for misfolded protein by partitioning substrates into native or sequestered states. **B.** Unfolding. A folding funnel depicts the energetic (vertical axis) and conformational (horizontal axis) landscape of protein folding. A protein may fold spontaneously to the native state (blue dot) or misfold into a kinetically trapped state (red dot). Hsp70 may unfold the substrate (green dot), giving the substrate another opportunity to fold along an energetically favorable pathway. **C.** Entropic pulling. When bound to an aggregated segment of protein, the Hsp70 molecule has increasing rotational freedom as the protein is pulled from the aggregate. This provides an entropic driving force for disaggregation. Panel C is taken from a published article for illustration (Goloubinoff and De Los Rios 2007).

chaperones can also increase the overall folding rate. Hsp70 may increase the folding rate by unfolding a kinetically trapped intermediate, allowing the substrate another opportunity to fold to its native state upon release (Fig. 1.2B) (Slepenkov and Witt 2002; Sharma *et al.* 2010). However, current models do not provide mechanistic insight into how the unfolding of soluble, kinetically trapped states occurs. This function may also allow Hsp70 to partially unfold substrates on the surface of aggregates for presentation to the Hsp100 ClpB, which is dependent on Hsp70 in an early stage of its disaggregation reaction (Weibezahn *et al.* 2004). Alternatively, the entropic pulling model of Hsp70 chaperone function posits that the occluded volume of bound Hsp70 creates an entropic cost that is greatest when the Hsp70-substrate complex is next to a surface, such as the surface of an aggregate or a membrane. This creates a driving force for Hsp70-mediated disaggregation and polypeptide translocation across membranes (Fig. 1.2C) (De Los Rios *et al.* 2006).

1.2 Hsp70 is a multidomain allosteric chaperone

To perform their many roles, Hsp70s interact with extended, hydrophobic regions of substrate proteins through their C-terminal substrate-binding domain (SBD). The affinity and kinetics of this interaction are modulated by nucleotide, which binds to the N-terminal nucleotide-binding domain (NBD). Substrate binding to the SBD catalyzes ATP hydrolysis; likewise, the ATP-bound state is characterized by a lower affinity for substrate than the ADP-bound state (Fig. 1.3A). Because of these ligand-modulated activities, the Hsp70 family represents

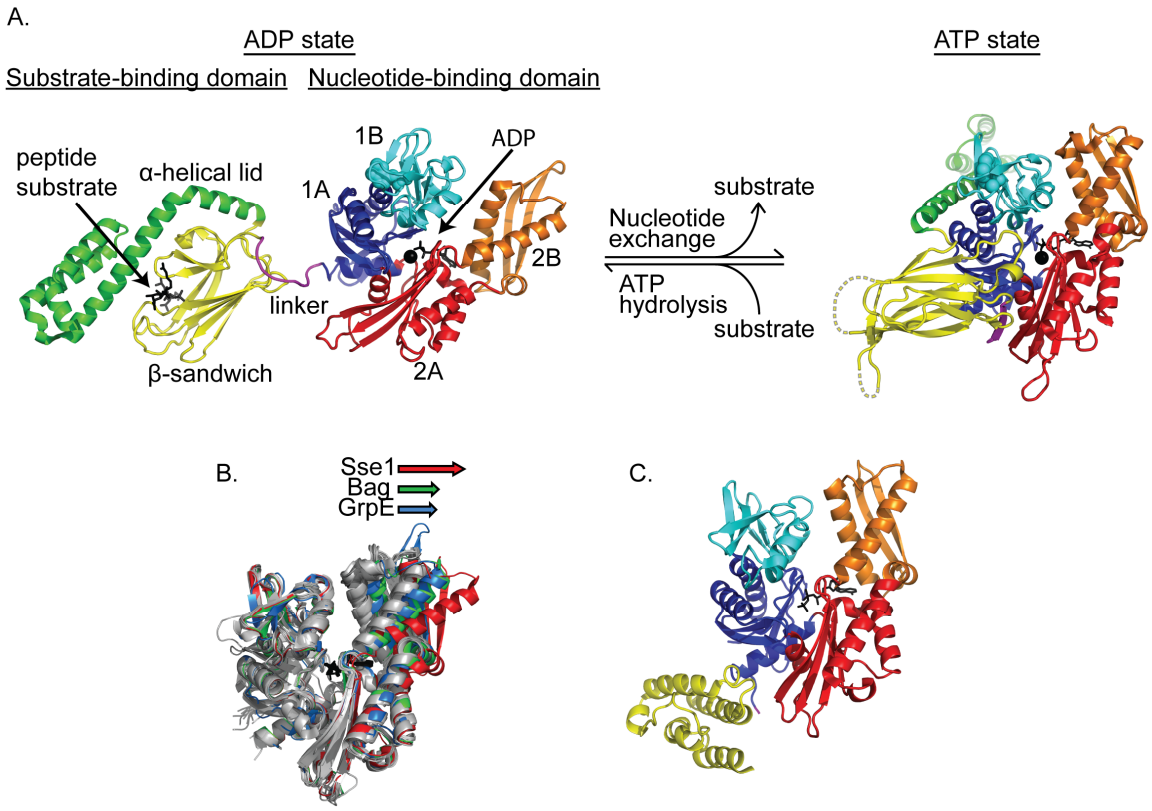

Figure 1.3. Structural basis of Hsp70 conformational transitions.

A. A simplified two-state model of the Hsp70 allosteric cycle. A structural model of the bacterial Hsp70, *E. coli* DnaK, bound to ADP and a peptide substrate is based on the crystallographic conformations of isolated domains (Zhu *et al.* 1996; Harrison *et al.* 1997); the average interdomain arrangement is based on an NMR study of a two-domain Hsp70 construct that used residual dipolar coupling analysis to orient the domains (Bertelsen *et al.* 2009). Hsp70 is colored by subdomain architecture with substrate, nucleotide and magnesium in black and Trp 102 is rendered as spheres in subdomain 1B. A crystal structure of the homologous yeast Hsp110 protein Sse1 provides insights into the ATP-bound state of Hsp70 (Liu and Hendrickson 2007). **B.** Conformations of the Hsp70 NBD in the presence or absence of NEF co-chaperones. Crystal structures of twenty-five isolated NBD fragments (gray; ADP shown as black sticks) and three NEF-bound conformations are superimposed (Harrison *et al.* 1997; Sondermann *et al.* 2001; Polier *et al.* 2008; Schuermann *et al.* 2008). The three non-homologous NEF families cause similar rotations of subdomain 2B to open the nucleotide-binding cleft and promote nucleotide exchange. **C.** A co-crystal of the bovine Hsc70 NBD and the auxilin (Hsp40) J domain depicts binding to NBD subdomains 1A and 2A (Jiang *et al.* 2007).

an excellent example of bi-directional allosterism, in which the status of ligand binding in one domain alters ligand interaction in a second domain. Mechanistically, ligand-induced conformational changes within each domain contribute to the packing and unpacking of domains in a dynamic process, resulting in timed cycles of substrate binding and release that enable physiological functions (Bukau *et al.* 2000; Hartl and Hayer-Hartl 2002; Mayer and Bukau 2005).

Nucleotides together with active site ions are instrumental in regulating the binding and release of substrate proteins. It was demonstrated in early Hsp70 biochemical characterization that binding of Mg-ATP and potassium promotes the release of a model unfolded susbstrate, reduced and carboxymethylated α-lactalbumin, based on separation of free and Hsp70-complexed species in size exclusion chromatography (Liberek *et al.* 1991; Palleros *et al.* 1993).

Considerable effort has followed to elucidate how this allosteric regulation of the SBD by the NBD is accomplished. The fluorescence signal of MABA-ATP shows unique biphasic kinetics upon binding to a two-domain Hsp70 but not to a NBD fragment. This data suggests that a second, slower step after initial ATP binding involves a conformational change that is dependent on the presence of the SBD (Ha and Mckay 1995). The rate of single turnover ATPase activity indicates that ATP-induced conformational changes occur prior to ATP hydrolysis. In other words, it is the binding of ATP but not its hydrolysis that promotes conformational change and leads to substrate release (Palleros *et al.* 1993; Ha and Mckay 1995; Theyssen *et al.* 1996). Moreover, the single turnover

rate of ATP hydrolysis is the same as the rate of the overall reaction cycle (ATP binding, hydrolysis, and release of phosphate and ADP) under steady-state conditions, hence hydrolysis is the rate-limiting step for the full cycle (Theyssen *et al.* 1996).

An early indication of the magnitude and nature of the conformational change transmitted from the ATP binding site through the molecule as a whole was offered by small angle x-ray scattering, which showed that the radius of gyration of bovine Hsc70 decreases upon addition of ATP, which was consistent with nucleotide-dependent molecular compaction (Wilbanks *et al.* 1995). This major ATP-induced conformational change is also indicated by changes in the tryptophan fluorescence of *E. coli* DnaK, which contains a single tryptophan residue at position 102 within the NBD that shows a shift in its maximum fluorescence emission wavelength, significant quenching (Palleros *et al.* 1993; Theyssen *et al.* 1996), and sequestration from solvent (Moro *et al.* 2003) only upon addition of ATP. These fluorescence changes are also evident in a DnaK mutation construct that is capable of binding ATP but defective in hydrolysis (Buchberger *et al.* 1995), again pointing to the key role of binding and not hydrolysis for allostery. Analysis of tryptophan fluorescence changes in C-terminally truncated constructs shows that they depend on both binding of Mg-ATP and the presence of the first two helices of the SBD α-helical lid, thus implicating the lid in a direct physical interaction between the domains (Buchberger *et al.* 1995; Moro *et al.* 2003).

The fact that small peptides mimic the binding of unfolded substrate proteins has been exploited in studies of the allosteric cycle of Hsp70s. Use of fluorescently labeled oligopeptides enabled study of the kinetics of substrate binding at fast time scales using stopped-flow mixing. Upon binding of ATP, Hsp70 shows reduced affinity for substrate that can be accounted for by a greatly enhanced off rate for substrate. The substrate off rate increases a thousand fold (0.0034 to 3.8 s^{-1}), shifting to a value comparable to the kinetics of the second step of ATP binding (1.5 s^{-1}) (Schmid *et al.* 1994; Theyssen *et al.* 1996).

In the reciprocal allosteric process, substrate binding stimulates ATPase activity to promote formation of the ADP state with higher substrate affinity and reduced substrate exchange rates (Flynn *et al.* 1989; Cyr *et al.* 1992; McCarty *et al.* 1995). Substrate affinity closely correlates with stimulation of ATPase activity in a series of mutant DnaK constructs that have varying affinities for substrate. This finding indicates that productive formation of a DnaK-substrate complex in the ATP state stimulates ATP hydrolysis and suggests that free energy of binding to substrate is harnessed to communicate an allosteric signal to the NBD (Mayer *et al.* 2000). Mutagenesis of the interdomain linker has revealed the critical role this region plays in transmitting an allosteric signal between domains (Vogel *et al.* 2006; Swain *et al.* 2007).

Hsp70 ATPase activity is further accelerated by functional cooperation with co-chaperones. The J domain of Hsp40 proteins enhances the rate of Hsp70 ATP hydrolysis, thus, in a manner comparable to the activation by substrate, promoting an ADP-bound state with high substrate affinity. A subset of Hsp40

proteins, such as *E. coli* DnaJ, couple ATPase stimulation with substrate delivery and therefore stabilize the interaction of Hsp70 with delivered substrates (McCarty *et al.* 1995; Laufen *et al.* 1999). Under these conditions, ADP release becomes the rate-limiting step of the catalytic cycle and is accelerated in *E. coli* by the NEF GrpE (Brehmer *et al.* 2001). The joint effects of NEF and Hsp40 proteins synergistically stimulate the overall Hsp70 ATPase reaction (McCarty *et al.* 1995).

The ability of co-chaperones to tune allostery is critical to Hsp70 physiological activities. For example, the co-chaperone-mediated acceleration of substrate binding and release cycles reduces aggregation and enhances refolding of substrate proteins (Szabo *et al.* 1994; Gassler *et al.* 1998). Experimental approaches behind these findings include monitoring substrate aggregation in the presence or absence of Hsp70 and co-chaperones by light scattering or centrifugation. The kinetics of substrate folding have most often been measured by selecting a substrate with a spectroscopic readout of function, such as denatured luciferase (Szabo *et al.* 1994; Minami *et al.* 1996).

Until quite recently, most crystallographic studies of Hsp70s were carried out on domains that had been identified by limited proteolysis and chosen to obtain well-behaved constructs. A structure of the bovine Hsc70 NBD first crystallized two decades ago depicted a bi-lobed molecule surrounding a nucleotide-binding cleft, with similarity to hexokinase (Flaherty *et al.* 1990) and actin (Flaherty *et al.* 1991). Each of the two main lobes can be further divided into two structural subdomains (Fig. 1.3A). To date, dozens of isolated NBD

structures have been crystallographically resolved, and together they illustrate roles for Hsp70 residues and active site ions in coordination of nucleotide binding and ATP hydrolysis (*e.g.*, Flaherty *et al.* 1994; Wilbanks *et al.* 1994; Johnson and McKay 1999). Overall, the different crystallographic forms are largely superimposable regardless of mutation at functionally relevant sites or the status of bound nucleotide or ions (gray structures in Fig. 1.3B).

More dramatic changes in NBD conformation were first observed in a co-crystal of *E. coli* DnaK bound to its NEF, GrpE (Harrison *et al.* 1997). The binding of GrpE to the upper lobes of DnaK (subdomains 1B and 2B) stabilizes an open nucleotide-binding cleft, which clarifies how GrpE promotes nucleotide exchange. Remarkably, a completely non-homologous eukaryotic NEF, Bag, binds to bovine Hsc70 via the same structural region; the resulting cleft opening and rotation of subdomain 2B thus appears to be a conserved mechanism of enhancing nucleotide exchange (Sondermann *et al.* 2001). *S. cerevisiae* Sse1 represents the Hsp110 family of eukaryotic NEFs that are related in both sequence and structure to Hsp70s. Sse1 and its relatives seem to act as NEFs in the same manner as the others: binding to subdomains 1B and 2B, rotating subdomain 2B, and opening the nucleotide-binding cleft (Andreasson *et al.* 2008; Polier *et al.* 2008; Schuermann *et al.* 2008). Apparently, these three evolutionarily distinct co-chaperone families have functionally converged to impart the same structural changes on Hsp70 structures (Fig. 1.3B). A fourth non-homologous NEF family, exemplified by *S. cerevisiae* HspBP1, diverges from this mechanism by causing

additional conformational changes in subdomain 1B of the Hsp70 Ssa1 (Shomura *et al.* 2005; Andreasson *et al.* 2008).

Hsp40 co-chaperones interact with partner Hsp70s through their J domains but are otherwise highly diversified (Kampinga and Craig 2010). A direct physical interaction between DnaK and the *E. coli* Hsp40 DnaJ was indicated by DnaK-DnaJ suppressor mutations using a genetics technique and measurement of their binding affinity using surface plasmon resonance (Gassler *et al.* 1998; Suh *et al.* 1998). A structure of bovine Hsc70 and the J domain of auxilin has been determined crystallographically by engineering a disulfide crosslink between the suppressor mutation sites (Jiang *et al.* 2007). The resulting J domain interface lies between Hsc70 subdomains 1A and 2A (Fig. 1.3C). It is intriguing to note that the subdomain 1A/ 2A region is a hotspot for allosteric regulation in actin, which is distantly homologous to the Hsp70 NBD (Dominguez 2004).

Several structures of isolated Hsp70 SBD fragments have been determined in complex with bound substrate peptides. The DnaK SBD is composed of two crystallographically resolved subdomains: a β-sandwich that cradles bound substrate, and an α-helical lid that forms multiple contacts along the β-sandwich (Fig. 1.3A) (Zhu *et al.* 1996). The lid encloses bound peptide in a high substrate affinity conformation of the SBD. In one structure, a specialized *E. coli* Hsp70, HscA, binds to an IscU-derived peptide substrate in the opposite N- to C-terminal orientation from other structures, which modestly repositions the helical lid (Cupp-Vickery *et al.* 2004; Tapley and Vickery 2004).

Recently, two-domain structures of Hsp70 have begun to emerge. Nonetheless, structural elucidation of the entire allosteric cycle has been elusive, as no Hsp70 has been crystallized ATP-bound and without substrate, which is the state in the allosteric mechanism that most favors interdomain docking and SBD α-helical lid opening. However, the previously mentioned NEF, Sse1, is a full-length homolog of Hsp70 that binds stably to ATP. Its crystal structure seems to shed considerable light on the Hsp70 ATP-bound state, as Sse1 reveals how an apo-SBD with an opened α-helical lid docks onto the NBD. The mode of Sse1 interdomain docking observed was consistent with the known sequestration of Trp 102 in DnaK, the important role of the interdomain linker in formation of the Hsp70 interdomain interface, and the conformational changes previously observed among the Hsp70 NBD subdomains (Fig. 1.3A) (Liu and Hendrickson 2007). Another two-domain structure captured Hsp70 in an elongated, undocked architecture but with the SBD binding to the interdomain linker of a different Hsp70 monomer in the crystal lattice (Chang *et al.* 2008). A two-domain crystal structure of bovine Hsc70 (Jiang *et al.* 2005) showed docking of its nucleotide-free NBD and substrate-bound SBD, which is not an allosteric state that is anticipated based on past biochemical and mechanistic data, albeit largely on *E. coli* DnaK. While the Hsp70 NBD and SBD of some species do not interact strongly in the absence of ATP (Swain *et al.* 2007; Bertelsen *et al.* 2009), nuclear magnetic resonance (NMR) chemical shifts of bovine Hsc70 in solution suggest that the NBD and SBD do stably dock in the nucleotide-free state of this isoform (Jiang *et al.* 2007).

Hydrogen-deuterium exchange (HDX) mass spectrometry showed that both domains of *E. coli* DnaK have a surprising degree of flexibility in solution and are mutually stabilized in the full-length protein (Rist *et al.* 2006). HDX experiments analyzed by NMR provided more detailed site-specific information. Major conformational changes expected in the ATP-bound, low substrate affinity state are revealed by residue groups that become either more ordered or disordered upon binding to ATP. Some residues of the SBD β-sandwich became ordered (showed slower exchange) in a region that likely participates directly in the interdomain interface. By contrast, residues at the other end of the β-sandwich near the substrate-binding pocket showed increased solvent accessibility, indicating that binding site destabilization leads to reduced substrate affinity in the ATP state and is mediated by interdomain docking (Swain *et al.* 2007).

NMR studies have provided a number of important insights into dynamic Hsp70 processes. Chemical shift mapping of a *T. thermophilus* DnaK NBD fragment showed that ATP binding fixes this domain in a single conformation, while ADP binding allows the domain to exchange between this conformation and a second, unique conformation (Revington *et al.* 2004). NMR residual dipolar coupling analyses of bacterial and eukaryotic NBDs revealed the ability of their subdomains to undergo dynamic reorientation in response to nucleotide binding, with lobes deviating up to 10° from their crystal structure orientations. These subdomain rotations alter accessibility to the nucleotide-binding cleft and modulate a hydrophobic surface that has been implicated in interdomain docking

(Zhang and Zuiderweg 2004; Bhattacharya *et al.* 2009). Solution studies of the DnaK SBD in the presence or absence of bound substrate (Morshauser *et al.* 1999; Pellecchia *et al.* 2000; Swain *et al.* 2006) show that substrate-bound and apo states form structures similar to the substrate-bound SBD crystal structure (Zhu *et al.* 1996) but with a slight repacking of the α-helical lid and significantly higher lid mobility in the absence of substrate. Together these data suggest that the influence of the NBD, mediated by docking, is required for significant conformational change in the SBD.

Comparative studies of NBD constructs with or without the conserved hydrophobic interdomain linker (*E. coli* DnaK 1-388 versus 1-392) provided evidence that the linker's role in allostery includes direct effects on the NBD. Even in the absence of the SBD, presence of the linker stimulates ATPase activity (Vogel *et al.* 2006; Swain *et al.* 2007). NMR chemical shifts show that the linker binds a hydrophobic surface on the NBD spanning the interface between subdomains 1A and 2A. These findings implicate the linker in mediating interdomain allosteric control and physically bringing the domains together in the ATP state (Swain *et al.* 2007). In a two-domain construct of *T. thermophilus* DnaK, the interdomain interface becomes more ordered in the presence of ATP, and the effects of ATP or peptide binding are distributed throughout the molecule (Revington *et al.* 2005). Similarly in an *E. coli* DnaK two-domain construct, widespread chemical shift changes are observed in the presence of ATP relative to isolated domains (Swain et al., 2007). By contrast, the chemical shifts of each isolated domain fragment are entirely superimposable with the two-domain

construct when bound to ADP and substrate, indicating that the domains do not stably interact in this state (Swain *et al.* 2007). More detailed characterization of the ADP state by residual dipolar coupling reveals that the NBD and SBD can rotate semi-independently in 35°cones tethered by a flexible interdomain linker (Bertelsen *et al.* 2009).

1.3 Statement of thesis

Collectively, a variety of biophysical techniques have uncovered several features of Hsp70 function and their mechanistic roles in this dynamically allosteric machine. As we learn more about Hsp70, it becomes increasingly clear that ligands and cofactors induce a multitude of conformations and show species-specific differences. What features define functional specificity among Hsp70 isoforms, and which are part of the universal Hsp70 mechanism? Looking beyond a two-state model of Hsp70 function, what is the nature of conformational sub-populations in the native state ensemble and how do they influence allostery? Do conformations mediated by co-chaperones reflect intrinsic Hsp70 motions or are they induced by co-chaperone binding? An enduring and fundamentally important question regarding the biophysics of Hsp70 function is an extension of the protein folding problem: how does Hsp70 catalyze protein folding and serve other functional roles? Multiple, kinetically regulated cycles of binding to exposed, hydrophobic segments of protein substrates and their release into aqueous solution are critical to these functions. Theoretical models of the energetic process have been proposed, but testing the precise mechanistic

effect of Hsp70 activity on substrate proteins has been experimentally challenging.

In Chapter 2 of this thesis, unifying aspects of the allosteric mechanism are investigated across the Hsp70 family as a whole. Rather than performing measurements on a vast number of Hsp70 isoforms, this issue is approached by analyzing naturally occurring cooperative mutations in the evolutionary record of the protein family. Hsp70 presents an excellent case study for using this type of analysis to reveal evolutionary linkages across protein domains. Chapter 3 describes the discovery of a sequence motif that is conserved within a large and diverse subset of Hsp70 proteins. This conserved and disordered region enhances chaperone function independently of ligand binding, allostery, co-chaperone interaction and other protein interactions, leading to a new model for the chaperone mechanism that can account for these unexpected features. Finally in Chapter 4, initial experiments are developed to characterize ensembles of domain assembly in various conformational states of Hsp70.

CHAPTER 2

AN INTERDOMAIN SECTOR MEDIATES ALLOSTERY IN HSP70 MOLECULAR CHAPERONES

This chapter is the result of collaborations with Annie Marcelino, Olivier Rivoire, Bill Russ, Joanna Swain, Stan Liebler, Rama Ranganathan and Lila Gierasch (Smock and Gierasch 2005, Publication 1; Marcelino *et al.* 2006, Publication 2; Smock *et al.* 2010, Publication 5).

2.1 Introduction

2.1.1 Correlated mutation analyses

Amino acid conservation has long been valued as a primary indicator of the importance of individual residues in the structure and function of a protein. Conservation measures mutational features at individual positions in the sequence alignment of a protein family, yet it does not describe the cooperative context of an amino acid with other residues in the protein. In several studies, the comparative analyses of correlated mutations in a protein's evolutionary history have successfully identified functionally and structurally important residues involved in cooperative interactions.

Most simply, a mutation at one site in a protein that is correlated to mutation at another site can arise from compensating side chain interactions that maintain energetic favorability. In myoglobin, mutations that switch a positively and negatively charged side chain in a given residue are highly correlated to the opposite charge reversal of neighboring residues (Neher 1994). Similarly, compensating substitutions were observed in a small set of hydrophobic core

residues in SH3 domains. When these naturally occurring mutations were introduced into the wild type Fyn SH3 domain, it was found that the destabilizing effect of mutation at a single site was ameliorated by adding in just one or two of its correlated mutations (Larson *et al.* 2000). Linkages shared between amino acids also form distributed networks that correspond to protein function. In PDZ domains, correlations are observed between the amino acid frequencies of a ligand binding site and a series of physically connected positions that extend through tertiary structure. Introducing these mutations into a wild type PSD-95 PDZ domain influences the binding properties of the domain. This observation includes mutations that do not neighbor binding residues, indicating that their perturbation is propagated through a cooperative network (Lockless and Ranganathan 1999).

One approach for measuring correlated mutations in protein families, the statistical coupling analysis (SCA), has revealed networks that underlie the conserved functional properties of single domain protein families (Lockless and Ranganathan 1999; Estabrook *et al.* 2005; Russ *et al.* 2005; Socolich *et al.* 2005). In addition to its support by a large body of experimental evidence, SCA has been extensively developed as a statistical technique adapted specifically to the unique challenges of correlated mutation analysis. For the purpose of comparison, researchers have applied the general purpose χ^2 goodness-of-fit test to the analysis of compensating substitutions in proteins (Larson *et al.* 2000; Kass and Horovitz 2002). In this approach, the χ^2 statistic is disregarded for positions in a sequence alignment that include some sparsely populated amino

acid frequencies even if the site is moderately conserved (Kass and Horovitz 2002), likely because the χ^2 statistic is too greatly influenced by less prevalent amino acid combinations. This treatment effectively removes positions containing multiple low abundance mutations and is statistically rigorous but does not fully exploit evolutionary information. Alternatively, measuring mutational substitutions with a correlation coefficient has been used to give a comparison between all pairs of sites in a sequence alignment (Neher 1994). In SCA, the dot product of mutational vectors gives a conservation-weighted correlation (Estabrook *et al.* 2005; Russ *et al.* 2005; Socolich *et al.* 2005). This provides a global measure of correlations in which all pairwise comparisons are made while less populated amino acid frequencies have a proportionally reduced influence on the analysis.

The most significant correlations are extracted using several features in SCA. First, the amino acid content of a sequence alignment is randomly shuffled in order to destroy significant correlations and to define significance thresholds for the original, unshuffled alignment. This approach provides general utility to remove noise and is used in χ^2-based analyses as well (Noivirt *et al.* 2005). Second, a correlations matrix is organized by clustering or components analysis. Rather than dealing with the correlations between isolated pairs of sites, this approach allows a researcher to interpret only coupling networks that contain a large number of residues with mutually reinforcing patterns of interaction. Third, signals that are present as a consequence of historical inheritance can be separated.

If untreated, the third condition presents a particular problem for extracting correlations that are indicative of cooperative interactions in the Hsp70 evolutionary record. High correlations are observed that result solely from sequence differences between eukaryotic and prokaryotic Hsp70s. In part, these include sites that localize to NEF binding areas. Throughout phylogeny, NEFs perform similar functions and bind to overlapping regions on Hsp70 but fall into non-homologous classes with different sequence features among eukaryotes and prokaryotes, giving rise to correspondingly differential conservation patterns in Hsp70. While of some interest, a functional interpretation of these correlations is obscured by the observation of a great number of other mutations that are also prevalent in the eukaryotic/ prokaryotic divide, resulting in the inclusion of many more apparently random, structurally dispersed sites in the same correlations set. This group of correlations as a whole is therefore difficult to interpret with regard to Hsp70 structure and function and becomes problematic if it complicates the interpretation of other desired correlation patterns, such as those that describe unifying features of the allosteric mechanism shared by all Hsp70 proteins.

Using established methods on the Hsp70 alignment, the eukaryotic/ prokaryotic division causes most or all of the significant correlations in a χ^2 analysis (Kass and Horovitz 2002), gives a strong but separable signal in SCA after matrix clustering (Estabrook *et al.* 2005; Russ *et al.* 2005; Socolich *et al.* 2005), and is identifiable as a source of uneven historical inheritance in newer applications of SCA (Appendix) (Halabi *et al.* 2009). Therefore, a variety of

correlated mutation analyses allow the characterization of cooperative interactions in proteins and SCA in particular is well suited for application to the Hsp70 family.

In SCA, the basic finding is that most residues evolve nearly independently, while a small fraction of residues are collectively coupled to form functional units called sectors (Halabi *et al.* 2009). A characteristic of sectors is structural connectivity; a contiguous system of sector residues within the protein core often connects distant surfaces in the three-dimensional structure. Thus, at least within single protein domains, sectors provide a basis for explaining structural and functional properties of proteins.

2.1.2 Position and sequence correlations in the single domain iLBP family

In an initial study performed in collaboration with Annie Marcelino and Lila Gierasch (Marcelino *et al.* 2006), we applied a version of SCA (Estabrook *et al.* 2005; Russ *et al.* 2005; Socolich *et al.* 2005) to the family of intracellular lipid binding proteins (iLBP). The single domain β-sheet iLBP proteins solubilize and transport hydrophobic molecules in eukaryotic cells. In addition to their functional properties, it was of particular interest to us that members of the iLBP family have been used for folding studies (*e.g.*, Clark *et al.* 1997; Gunasekaran *et al.* 2004), since information gleaned from sequence analysis can potentially be related to experimental knowledge of key structural interactions in the protein family. In this section, I briefly describe the correspondence of iLBP position correlations to the structure and function of the protein family, note the distribution of iLBP

sequences into ligand binding classes, and begin to characterize how these two pieces of information can be related to one another.

Two large sectors of mutually reinforcing position correlations were found in the iLBP family. One sector comprises a predominantly hydrophobic core away from the ligand binding site and likely represents important structural relationships in the iLBP fold (Fig. 2.1A). The other sector includes a portal region where ligand enters its binding site, regions of the ligand-binding cavity, and the region where the 10-stranded β-barrel characteristic of this family closes (Fig. 2.1B). Linkages between these two sectors suggest that evolutionary pressures on this family constrain structural and functional sequence information in an interdependent fashion. This may be a consequence of the necessity of the structure to wrap around hydrophobic ligands, which confounds the typical sequestration of hydrophobic side chains in a protein folding core (Fig. 2.1C). Additionally, ligand entry and exit require these structures to have a capacity for specific conformational change during binding and release.

In addition to pairwise position correlations, pairwise sequence comparisons can also be made to more fully exploit the information content of a protein family's sequence alignment. Clustering an iLBP sequence similarity matrix organizes sequence correlations into sub-families which vary by organism, tissue localization and ligand binding specificity (Marcelino *et al.* 2006). Similarly, the binding interfaces of several domain families have been identified by parsing differentially conserved sites in constructed sequence sub-families (Lichtarge *et al.* 1996). Using SCA, the analysis of amino acid variations at correlated

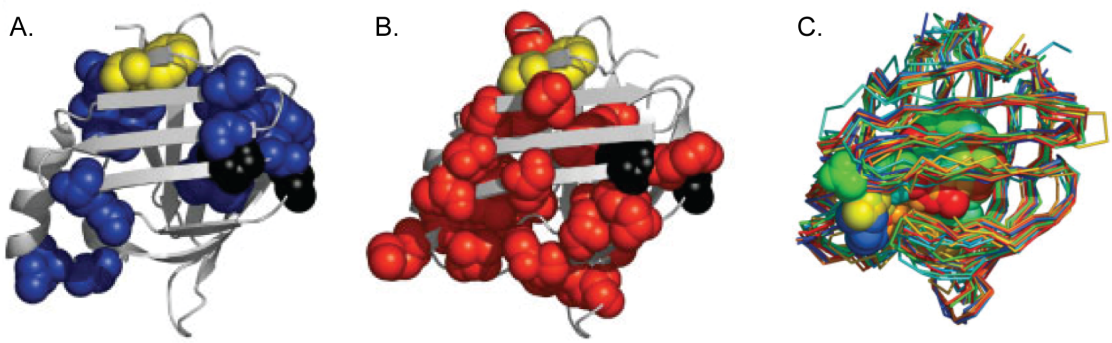

Figure 2.1. **Correlated mutations in single domain iLBP proteins.**

A. An evolutionary sector comprises a predominantly hydrophobic core away from the ligand binding site and likely represents key structural information for the iLBP fold (blue spheres; Trp 7 is shown in yellow for reference and highly conserved sites are shown in black). **B.** The other sector includes the portal region where ligand enters its binding site, regions of the ligand-binding cavity, and the region where the 10-stranded β-barrel characteristic of this family closes (red spheres). Sector residues point inwards towards the central cavity. **C.** Thirty superimposed crystal structures of iLBP-ligand complexes are each colored differently. Ligands are shown as spheres within the central cavity. The iLBP fold is structurally conserved across the protein family while ligand binding specifity is divergent. As a whole, the iLBP cavity volume is largely filled by multiple orientations of these diverse hydrophobic ligands.

positions alone separates WW domains into their correct ligand binding classes (Russ *et al.* 2005). Together, these results indicate that position and sequence correlations are related in small protein domain families that are composed of isoforms with divergent functions. Further exploration of the joint information content of position and sequence correlations in protein families can provide greater insight into their molecular evolution.

The same SCA methodology applied to the iLBP family was also used to identify structural and functional sectors in the multidomain Hsp70 family (Appendix). The larger size and functional complexity of multidomain proteins introduces additional uncertainty in the interpretation of particular roles for sectors. In the following sections, position and sequence correlations are simultaneously evaluated to extract sites underlying the divergence of homologous Hsp70 and Hsp110 proteins. This development allows for the unambiguous assignment of correlated mutations to specific functions in increasingly complex systems.

2.1.3 Cooperativity across Hsp70 domains

Allosteric coupling, the process by which spatially distant sites on proteins functionally interact, is a defining biological property of many proteins, but the underlying structural basis remains difficult to understand (Smock and Gierasch 2009). The central problem is the difficulty of detecting the pattern of cooperative functional interactions between amino acid residues in protein structures. One approach to this problem is to analyze the correlated evolution of amino acids in

a protein family. The co-evolutionary principle of protein residues that underlies sectors is not limited to the coupling of amino acids within a single domain. Indeed, allosteric coupling between two or more protein domains is a common finding in studies of cellular function. This suggests the existence of sectors – units of evolutionary selection – that are shared between different protein domains. For example, a sector spanning two domains could couple a functional site on one protein domain to a functional site on a second protein domain. Such sectors could explain conserved aspects of allosteric coupling.

A test of the concept of applying SCA to multidomain protein signaling was undertaken using a construct with a light-modulated PAS domain fused to dihydrofolate reductase (DHFR). Domain surfaces predicted to influence intradomain function based on SCA results were linked in close proximity, with the striking result that the DHFR activity in the chimera became light-dependent (Lee *et al.* 2008).

The Hsp70 molecular chaperones are a large and diverse family of allosteric two-domain proteins and present an excellent case study to test a role for sectors shared between domains. The overall family of Hsp70-like proteins comprises the Hsp70s and the Hsp110s, homologs that contain both domains and are regarded as structural models for Hsp70s (Easton *et al.* 2000; Liu and Hendrickson 2007). However, despite their sequence similarity, the Hsp110 proteins have evolved to be non-allosteric, such that the nucleotide-binding domain remains stably bound to ATP and does not appear to regulate the substrate-binding domain. Consistent with these findings, Hsp110s are incapable

of folding substrate proteins through cycles of nucleotide exchange and hydrolysis (Shaner and Morano 2007).

The functional divergence of Hsp70s and Hsp110s reveals patterns of co-evolution that can be associated with interdomain allostery. The identification of a group of co-evolving residues that show structural contiguity between the two Hsp70 domains provides testable hypotheses about allosteric function in these molecular chaperones and serves as an example that will be applicable more generally in defining interaction networks between proteins and protein domains.

2.2 Results

2.2.1 A sector associated with Hsp70 allostery

A version of SCA (Halabi *et al.* 2009) was adapted to exploit differential features of the Hsp70-like family in collaboration with Olivier Rivoire, Bill Russ, Stan Liebler and Rama Ranganathan (Smock *et al.* 2010), who developed and implemented the algorithmic details of SCA described in this section.

To identify sectors in Hsp70, SCA is used to compute a weighted correlation matrix ($\tilde{C}$) that describes the co-evolution of every pair of amino acid positions in the Hsp70/110 family. SCA was performed on a diverse alignment of 926 sequences that broadly samples all major phylogenetic groups. The essence of sector identification is to analyze the non-random correlations in the $\tilde{C}$ matrix to find collectively evolving groups of residues. One approach to do this is spectral decomposition, in which sectors are defined by the pattern of residue contribution to the top few eigenmodes of the $\tilde{C}$ matrix (Halabi *et al.* 2009).

Analysis of the Hsp70/110 protein family suggests a more targeted strategy for analysis of the $\tilde{C}$ matrix in which we take advantage of the functional divergence of allosteric mechanism between Hsp70 and Hsp110 proteins to guide sector identification. The basic idea is to simultaneously evaluate the pattern of divergence between sequences in a protein family and the pattern of co-evolution between amino acid positions (Casari *et al.*, 1995; Lichtarge *et al.* 1996). This can be done in the framework of SCA using a mathematical method known as singular value decomposition. If the pattern of sequence divergence classifies members of a protein family according to distinctions in a functional mechanism (*e.g.*, allostery), then we can identify the group of coevolving residues that correspond to this mechanism. We describe this approach here in context of the Hsp70/110 family (see Methods for details). Given a weighted binarized sequence alignment ($\tilde{X}$) comprised of M sequences (rows) and L positions (columns), we can compute the following two correlation matrices:

$$\tilde{C} = \frac{1}{M}\tilde{X}^T\tilde{X} \text{ and } \tilde{S} = \frac{1}{L}\tilde{X}\tilde{X}^T,$$

where $\tilde{C}$ is the SCA correlation matrix between positions and $\tilde{S}$ is a correlation matrix between sequences.

The grouping of position correlations into larger networks can be investigated by principal components analysis. Principal components of $\tilde{C}$ are represented by eigenvectors, which capture global patterns of pairwise position correlations, and eigenvalues, which indicate the magnitude of the pattern (the strength of the correlations). A histogram of all the eigenvalues of $\tilde{C}$ indicates the magnitude of these position correlation patterns in the Hsp70/110 alignment as a

whole (Fig. 2.2A). Their significance can be assigned by shuffling amino acids within the same alignment. The eigenvalue distribution resulting from this randomized alignment represents the magnitude of correlations that would be expected by random given the same amino acid content devoid of any meaningful correlations. Several eigenmodes can be considered highly significant.

Essentially the same result is obtained by using prior versions of SCA (Estabrook *et al.* 2005; Russ *et al.* 2005; Socolich *et al.* 2005) or by identifying position correlation groupings by hierarchical clustering rather than spectral decomposition techniques like principal components analysis (Appendix). However, a unique benefit of matrix factorization is that position and sequence correlations in $\tilde{X}$ can be related to one other through singular value decomposition. The singular value decomposition of $\tilde{X}$ is:

$$\tilde{X} = U\Sigma V^T,$$

in which columns of U are eigenvectors of $\tilde{S}$, columns of V are eigenvectors of $\tilde{C}$, and Σ is related to the eigenvalues of these matrices. Importantly, this decomposition allows a direct mapping between each principal axis of sequence variation (a column in U) and the corresponding principal axis of positional co-evolution (the same column in V). If functionally distinct sequences segregate along an axis of sequence variation, then the positions that underlie this divergence are defined in the corresponding axis of positional co-evolution.

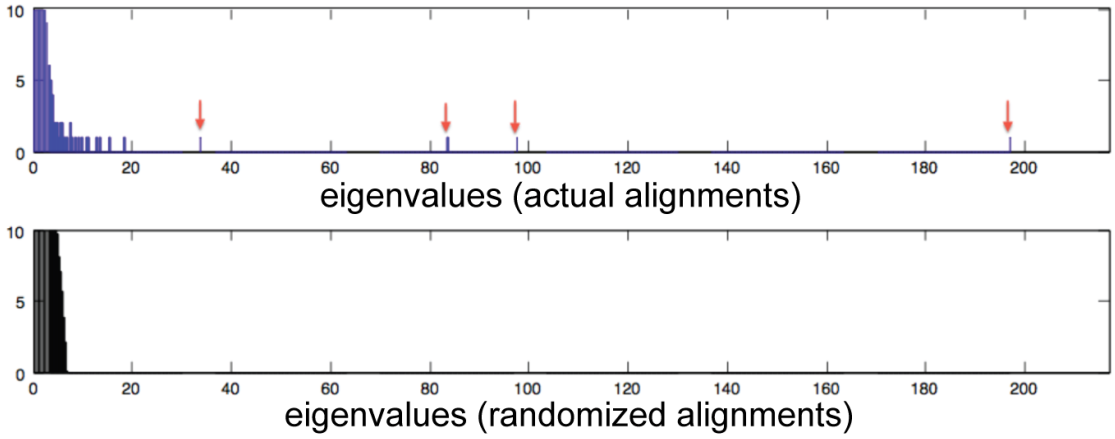

Figure 2.2. Significance of Hsp70/110 position correlation patterns.

The eigenvalues of $\tilde{C}$, the position correlations matrix, are plotted in a histogram for the actual alignment (top) and randomized alignments (bottom). From the comparison, as many as twelve eigenvalues may be considered statistically significant while four eigenvalues stand apart very clearly (red arrows).

Examination of the top principal axes of sequence variation for the Hsp70/110 family shows in fact a clear separation of the allosteric (Hsp70) and non-allosteric (Hsp110) members into two distinct clusters (Fig. 2.3A). The corresponding axis of the $\tilde{C}$ matrix reveals a protein sector comprising a small fraction (19%) of correlated Hsp70/110 positions that underlie the separation of Hsp70 and Hsp110 family members (Fig. 2.3B). The sector positions partition roughly equally between the nucleotide-binding domain (NBD in blue, Fig. 2.3B) and the substrate-binding domain (SBD in green, Fig. 2.3B) demonstrating co-evolution of residues in both domains to form a single unit of evolutionary selection. Consistent with the finding that allosteric coupling is a property of the Hsp70 sub-family, positions comprising this sector are more conserved within the Hsp70 sub-family than in the Hsp110 sub-family. Taken together, these results define an interdomain sector in the Hsp70 sub-family that is associated with the allosteric mechanism.

2.2.2 Structural interpretation of the Hsp70 sector

What is the structural interpretation of this Hsp70 sector? NMR (Swain *et al.* 2007; Bertelsen *et al.* 2009) and tryptophan fluorescence (Moro *et al.* 2003) data in DnaK, the *E.coli* Hsp70, show that in the ADP-bound state, the NBD and SBD are dissociated and largely independent. In contrast, upon ATP binding, the NBD undergoes major conformational rearrangement, participates in the interdomain interface, and promotes substrate release from the SBD (Wilbanks *et al.* 1995; Moro *et al.* 2003; Mayer and Bukau 2005; Swain *et al.* 2007).

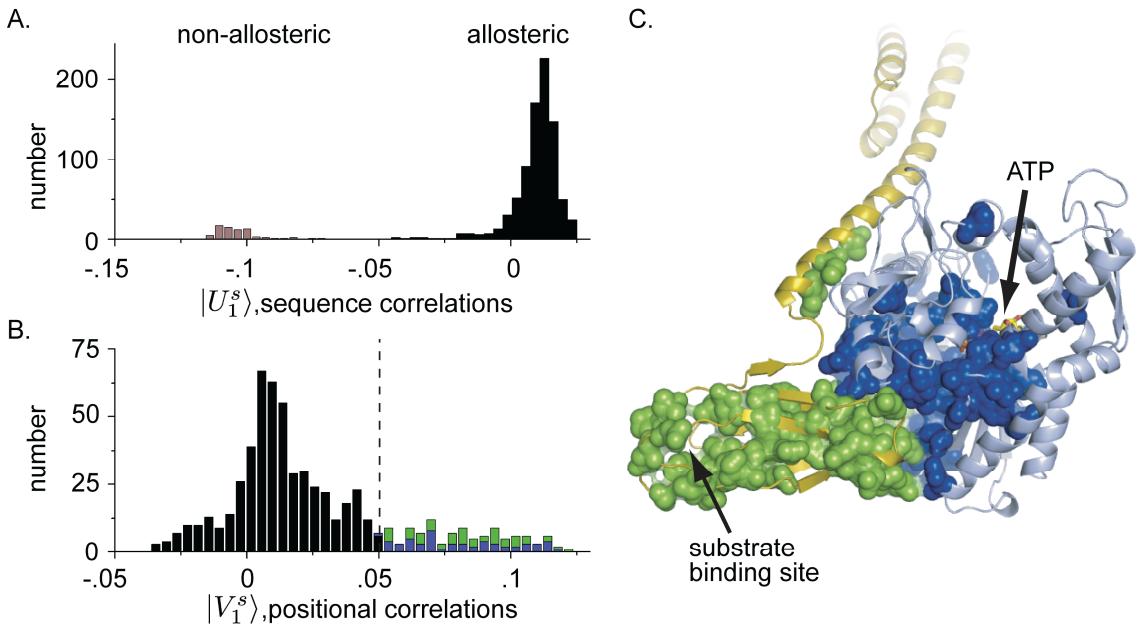

Figure 2.3. Identification of an allosteric sector in Hsp70 proteins.

A. A histogram of the Hsp70/110 sequences projected on a top axis of sequence variation derived from singular value decomposition. This axis separates the Hsp70/110 family into two sub-families that correspond to the allosteric Hsp70s and the non-allosteric Hsp110s. **B.** A histogram of Hsp70 residues projected on the corresponding axis of positional co-evolution, showing that a small fraction of residues is largely responsible for the sequence divergence shown in panel A (115 sector residues out of 605 total residues, or 19%). This defines the Hsp70 sector. The sector nearly equally comprises residues from the NBD (blue, 56 residues) and the SBD (green, 59 residues). **C.** The Hsp70 sector mapped onto a model of the ATP-bound state of E.coli DnaK. The sector forms a physically contiguous group of residues that connect the ATP binding site in the NBD (pale blue) to the substrate-binding pocket of the SBD (yellow) through the binding interface between the two domains. Sector positions are represented as spheres and colored as in panel B.

To examine the spatial arrangement of the Hsp70 sector in the ATP-bound state, we represented sector residues on an Sse1-derived model for DnaK (Liu and Hendrickson 2007). Consistent with a role in allosteric coupling, residues comprising the sector form a physically contiguous network of atoms linking the ATP-binding site to the substrate-binding site, passing through the interdomain interface (Fig. 2.3C). The physical connectivity is remarkable given that only a small fraction of overall Hsp70 residues are involved (Fig. 2.3B). Prior work showed that sparse but connected clusters of amino acids link distantly positioned functional sites within individual protein domains (Lockless and Ranganathan 1999; Socolich *et al.* 2005; Halabi *et al.* 2009). This work extends this result to show that functionally coupled but non-homologous protein domains can share a single sector that connects their respective functional sites through a protein-protein interface.

The initial homology model of the ATP-bound state of Hsp70 was created by simply threading DnaK sequence onto the structure of ATP-bound Sse1, a non-allosteric Hsp110 protein. While Sse1 shows general structural features expected for ATP-bound Hsp70 (Liu and Hendrickson 2007), direct sequence threading created poorly optimized interdomain contacts for DnaK with a large interdomain solvent cavity (Fig. 2.4A). The model was subjected to simulated annealing molecular dynamics to relax the structure into a more energetically favorable arrangement. Simulation led to the same overall conformation of the protein but the solvent cavity collapsed (Fig. 2.4B) and new regions of interdomain contacts were formed, such as in the vicinity of the NBD residue

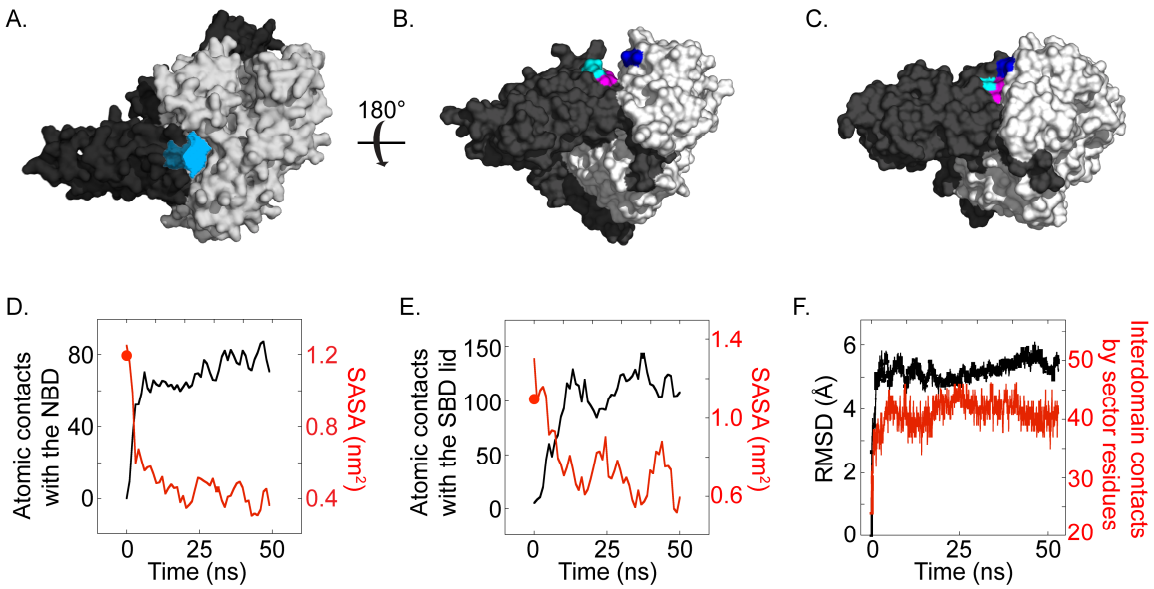

Figure 2.4. Modeling Hsp70 structure in an ATP-bound, domain-docked conformation.

A. An initial Hsp70 homology model. Threading the amino acid sequence of the *E. coli* Hsp70 DnaK directly onto an Hsp110(ATP) crystal structure (PDB 2QXL) (Liu and Hendrickson 2007) leads to poorly optimized interdomain contacts with a cavity of 286 Å^3 volume within the interface (shown in blue). **B.** Prior to simulation, the domain surfaces near sector residues D326 (blue), K414 (cyan) and N415 (magenta) are not contiguous in the DnaK(ATP) homology model. **C.** During simulation, the interdomain cavity collapses and the interdomain sector positions join into a single contiguous group across the interface (shown on the median structure of the trajectory). **D.** Molecular dynamics simulation forms specific interdomain contacts consistent with experimental data. SBD residue K414 and **E.** NBD residue W102 initially have solvent-accessible surface areas (SASA) as in isolated domain DnaK crystal structures (red circles, PDBs 1DKG and 1DKZ) (Zhu *et al.* 1996; Harrison *et al.* 1997). As the simulation progresses, both sites become buried through the formation of interdomain contacts. **F.** Overall, the number of interdomain contacts between sector residues increases during the simulation while structural changes (RMSD) reach a plateau.

D326 and SBD residues K414 and N415 (Fig. 2.4C-E), which are sector residues. Resettling the structural model in this way was also corroborated by the finding that NBD residue W102 was sequestered by the SBD lid (Fig. 2.4E), consistent with fluorescence data (Moro *et al.* 2003). In this energetically optimized model for the ATP-bound DnaK structure, contacts between sector residues increased significantly across the interdomain interface (Fig. 2.4F), further implicating their role in mediating the energetically favorable physical interaction of Hsp70 domains.

2.2.3 Functional studies of Hsp70 allostery

Does the Hsp70 sector represent the mechanism of allosteric coupling between the NBD and SBD? A number of biochemical and genetic studies on a variety of Hsp70s provide a basis for this assessment. Within the NBD, sector positions include catalytic residues E171 and D201, the mutation of which impairs ATP-induced conformational change (Johnson and McKay 1999), and T199, the mutation of which stabilizes ATP-induced conformational change (Buchberger *et al.* 1995), among other sites making direct contact with bound nucleotide (Fig. 2.5A). Studies of isolated bacterial Hsp70 NBDs have demonstrated ATP-dependent reorientation of all four subdomains (Zhang and Zuiderweg 2004; Bhattacharya *et al.* 2009), and the sector spans all of the subdomain interfaces. In particular, actin and Hsp70 retain sequence conservation at nucleotide-binding loops and adjacent crossing helices that form an interface between subdomains 1A and 2A (Bork *et al.* 1992). Actin responds

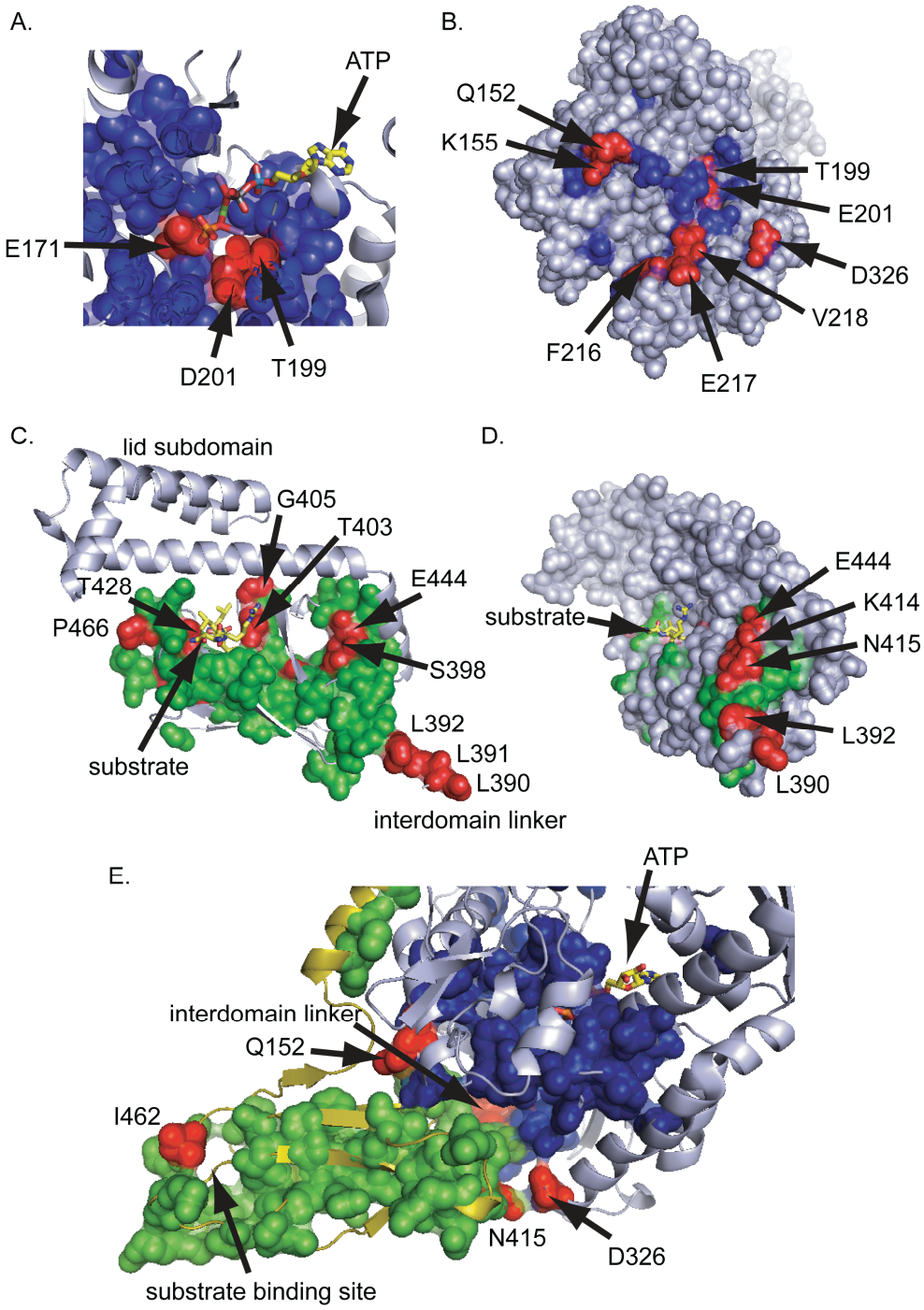
A.
ATP
E171
D201
T199
B.
Q152
K155
T199
E201
D326
V218
F216
E217
C.
lid subdomain
G405
T403
E444
T428
P466
S398
L392
L391
L390
substrate
interdomain linker
D.
E444
K414
substrate
N415
L392
L390
E.
ATP
interdomain linker
Q152
I462
N415
D326
substrate binding site

Figure 2.5. **Evidence for a role of the Hsp70 sector in allosteric coupling.**

Examination of sector positions in the NBD (blue) and in the SBD (green) reveals many experimentally characterized sites implicated in allostery (colored in red; see text for details). **A**. In the ATP binding site, sector residues include determinants of nucleotide hydrolysis and nucleotide-mediated conformational change. **B**. At a NBD surface contacting the SBD, a patch of sector residues includes several known to perturb allostery upon mutation. **C,** In the SBD, the sector contains residues both proximal and distal from the substrate-binding site, and includes the functionally critical interdomain linker (390-392). **D**, At the SBD surface that comprises the interface with the NBD, sector residues include several that display defects in allostery upon mutation. **E**. Across the interdomain interface, mutation at residues Q152 and F216 (not shown) in the NBD partially suppress the loss-of-function mutations at position 462 on the SBD. Also shown are positions 415 and 326, sector residues that contact across the interdomain interface and are previously untested with regard to allosteric function.

to bound nucleotide through an ATP-dependent shearing motion between subdomains 1A and 2A (Schuler 2001), and this structural region is a focal point of Hsp70 sector mapping. These findings are consistent with the view that at least in part, the co-evolution of sector positions may be related to the anisotropic mechanical coupling of amino acids within the protein structure. A similar empirical relationship has been noted between observed mechanical interactions and the sector mediating specificity in the S1A serine proteases (Halabi *et al.* 2009).

Moreover, the crossing helices form a solvent-accessible cleft between subdomains 1A and 2A in actin-like NBDs. In actin, the cleft mediates interaction with allosteric effector proteins (Dominguez 2004), while in Hsp70/110, the cleft is proposed to act as an intramolecular binding surface for the interdomain linker (Jiang *et al.* 2007; Liu and Hendrickson 2007; Swain *et al.* 2007). In the allosteric Hsp70 sector that our analysis describes, the cleft surface is lined with sector residues (*e.g.*, Y145, D148, K155, E217 and V218, see Fig. 2.5B), and mutation at these sites is reported to perturb Hsp70 allostery (Gassler *et al.* 1998; Vogel *et al.* 2006). The conserved interdomain linker sequence motif ^{389}VLLL^{392}, which stimulates ATPase activity when present on truncated NBD constructs (Swain *et al.* 2007) and the mutation of which impairs interdomain allostery in full-length Hsp70 (Laufen *et al.* 1999; Vogel *et al.* 2006), is in the sector. Binding of the linker to the cleft below the crossing helices is postulated to be key to formation of the domain-docked state, bringing the SBD into proximity to the NBD (Swain *et al.* 2007).

Sector positions within the SBD comprise a structurally contiguous set of atoms that extends from the substrate-binding site through the protein core to a solvent-exposed region that includes the interdomain linker (Fig. 2.5C,D). The functional significance of the sector is supported by several previous observations. Sector residue K414 is centered in the domain surface patch, making multiple interdomain contacts in the docked state. Previous work showed that this residue plays a critical role in allosteric signal transmission as mutation at this site blocked interdomain docking and allostery (Montgomery *et al.* 1999). A SBD sector position has been shown to have an epistatic relationship with sector positions in the NBD: Mutation of substrate-binding domain residue I462 is lethal in the yeast Hsp70 Ssc1 but is partially suppressed by NBD mutations Q152L and F216L (Fig. 2.5E) (Davis *et al.* 1999). There is also evidence that structural regions not essential for allosteric coupling are not involved in the interdomain sector; the SBD lid is nearly absent in sector residues, and a DnaK variant lacking the lid retains core allosteric function (Swain *et al.* 2006). Interestingly, several sector positions within the substrate-binding domain experience large NMR chemical shift changes upon binding of a peptide substrate (S398, T403, G405, T428, D431, I438, F457, L459, F457, G468) (Swain *et al.* 2006). In addition, mutation of sector residues far from the substrate-binding site (S398, G400, G443, E444, L459) reduces substrate-binding affinity (Burkholder *et al.* 1996).

The physical and functional connectivity of a single co-evolutionary sector across domains provides strong support for the proposal that the sector mediates the allosteric coupling central to the basic biological activity of Hsp70.

2.2.4 Direct experimental analysis of the interdomain sector

Knowledge of the interdomain sector provides new hypotheses for further experimental testing. For example, residues D326 and N415 are sector positions that display interdomain contact, but no previous experiment has tested their involvement in interdomain allostery (Fig. 2.4 and 2.5). Conservative mutations were introduced based on amino acid frequencies at these positions in Hsp70 sequences and their effect on interdomain allostery was measured both *in vivo* and *in vitro*. A direct test for the influence of sector mutants on organism fitness is provided by the ability of DnaK to promote *E. coli* growth at elevated temperature (Bukau and Walker 1990). For example, strains of *E. coli* in which the chromosomal copy of DnaK is deleted grow very weakly after heat shock but are rescued by expression of wild type DnaK from a plasmid (Fig. 2.6). In contrast, the D326V or N415G DnaK variants fail to complement the DnaK knockout strain upon heat shock, demonstrating that these positions are critical for Hsp70 activity.

The origin of these cellular defects was investigated by purifying the mutant DnaK proteins and employing biochemical tests of allosteric function *in vitro*. DnaK D326V and N415G are soluble, natively folded and thermally stable in the absence of ATP and substrate (Fig. 2.7). The fluorescence of the sole

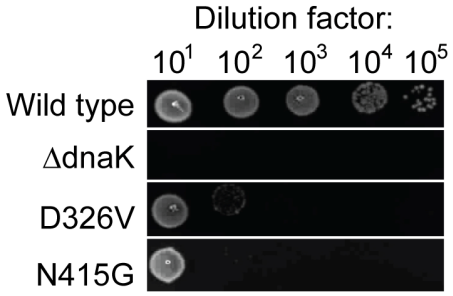

Figure 2.6. Mutation of the sector at domain surfaces leads to a cellular fitness disadvantage.

A stress response assay, showing that though wild type DnaK efficiently rescues growth in an *E.coli* Δ*dnaK* strain upon 43 ^{o}C heat shock, DnaK variants D326V and N415G do not.

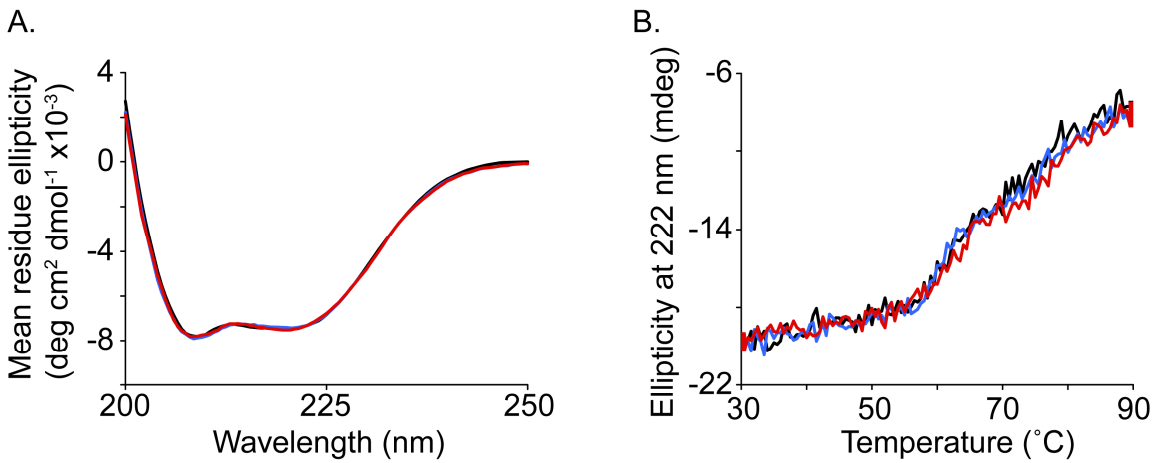

Figure 2.7. **Sector mutants DnaK D326V and N415G are well-folded.**

A. Circular dichroism (CD) spectra and **B.** thermal denaturation curves for D326V and N415G mutant DnaK proteins are indistinguishable from those of wild type, indicating that these sector mutants retain native secondary structure and are thermally stable (wild type, black; D326V, blue; N415G, red). Sector mutant CD spectra in A show less than 5% intensity difference from that of wild type and were normalized to allow comparisons of curve shapes, which are diagnostic for secondary structural content.

intrinsic tryptophan residue in DnaK is diagnostic for ATP-dependent interdomain docking because it displays a characteristic blue shift and intensity quench upon interdomain interaction (Moro *et al.* 2003). DnaK D326V and N415G show the same tryptophan fluorescence spectrum as wild type DnaK in the absence of nucleotide, indicating that W102 is in its normal chemical environment in the undocked state. However, the extents of W102 fluorescence blue shifting and intensity quenching upon addition of ATP are reduced relative to wild type (Fig. 2.8A). The same trends are observed for wild type and mutants when W102 accessibility is assessed by acrylamide quenching (Fig. 2.8B). These findings are characteristic of a specific defect in ATP-induced conformational change and domain docking in the point mutants. In addition, functional DnaK allostery entails a ~7-fold stimulation of ATPase activity upon binding of peptide to the substrate-binding domain. Relatve to wild type DnaK, D326V and N415G show significantly elevated basal ATPase rates and only ~3-fold stimulation by peptide (Fig. 2.8C). Given these data and the knowledge that ATP hydrolysis is the rate-limiting step of the reaction cycle (McCarty *et al.* 1995), the likely interpretation is that the sector mutants shift the normal DnaK conformational equilibrium from the ATP-induced, domain-docked state to the more independent domain arrangement characteristic of the ADP state (Fig. 2.8D).

These findings are consistent with the hypothesis that these two sector positions are important for stabilizing the peripheral interdomain interface and mediating allosteric communication between domains. More generally, these

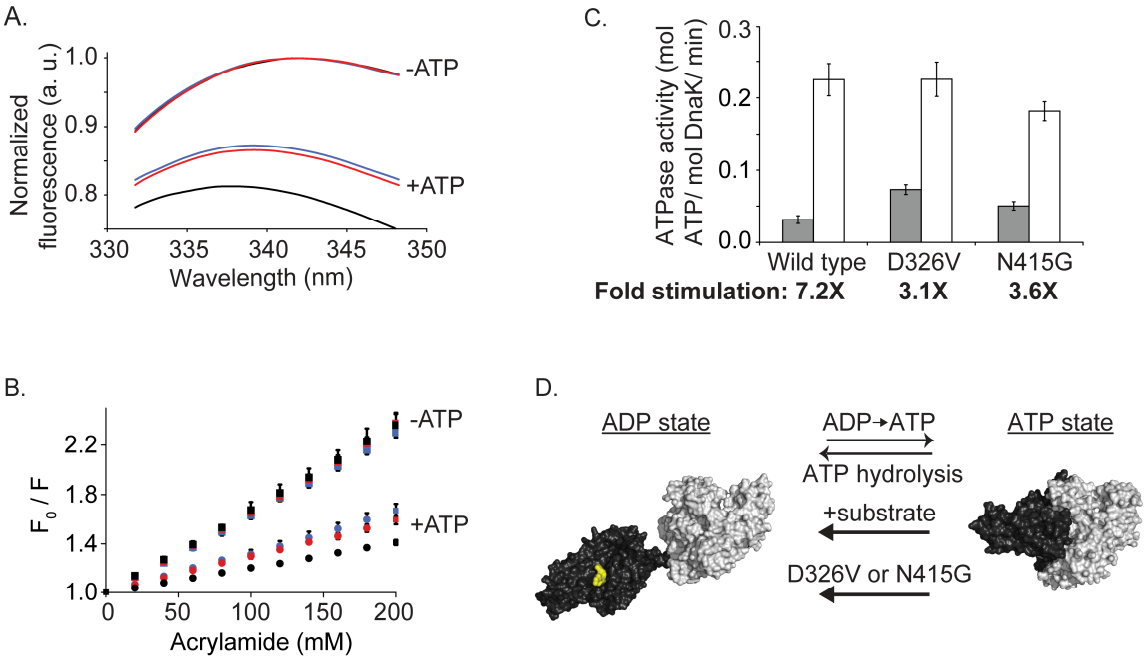

Figure 2.8. Sector mutants destabilize the docked ATP-bound state.

A. Fluorescence of the sole tryptophan residue in DnaK is diagnostic for ATP-dependent interdomain docking. DnaK D326V and N415G display normal tryptophan fluorescence in the absence of ATP but only a partial conversion of the ensemble to the ATP state based on blue shift and intensity (wild type, black; D326V, blue; N415G, red). **B.** Acrylamide quenching of W102 fluorescence intensity is similar for wild type DnaK and sector mutants in the absence of ATP. However, upon addition of ATP, sector mutants are partially defective in the extent to which they are quenched by acrylamide relative to wild type (wild type, black; D326V, blue; N415G, red). **C**. The sector mutations distal from the ATP binding site cause elevated basal ATPase activities (grey) and reduced stimulation by the substrate peptide p5 (p5-stimulated rates in white; fold stimulation below plot). **D**. These data can be explained by a two-state model in which sector mutants destabilize the docked, ATP-bound DnaK conformation.

data provide further evidence that Hsp70 sector analysis has predictive value in describing an interdomain allosteric network.

2.3 Discussion

In summary, we show that sequence analysis alone of the Hsp70 molecular chaperone family identifies a group of co-evolving residues, a sector, that is responsible for the core function of the Hsp70 proteins – allosteric coupling between distantly positioned functional sites on two protein domains. As per previous reports (Lockless and Ranganathan 1999; Socolich *et al.* 2005; Halabi *et al.* 2009), the sector is sparse, such that only a small fraction of total amino acids in the protein are involved, and physically contiguous, so that the ATP-binding site on the NBD is connected to the substrate-binding site on the SBD through a continuous network of interacting amino acids. The identification of the sector provides a clear basis for directing new experiment designs towards a more complete understanding of the mechanism and evolutionary divergence of allostery in these proteins. For example, Hsp70s utilize co-chaperones in team-assisted roles and many sector positions emerging as an allosteric surface for interdomain allostery in Hsp70s also play a role in J domain binding and J-mediated ATPase stimulation (Gassler *et al.* 1998; Suh *et al.* 1998; Vogel *et al.* 2006).

This work adds a key new finding with regard to the concept of protein sectors. Previous work showed that multiple quasi-independent sectors are possible within a single protein domain, each of which contributes to a different

aspect of function (Halabi *et al.* 2009). Here, we show that a single sector can also exist to functionally couple two different, non-homologous protein domains. This result emphasizes the point that sectors are simply defined as units of selection, without regard to hierarchies of structural organization. An interesting possibility that follows is that sectors could physically join and co-evolve across protein-protein interfaces in order to mediate the coupling of activities between proteins – the essence of signal transmission and allosteric regulation. Indeed, this idea has been recently used to design a synthetic two-domain allosteric protein (Lee *et al.* 2008). It will be interesting to further test the notion that connection of protein sectors is a process through which allostery between proteins might evolve.

2.4 Methods

2.4.1 Statistical coupling analysis

Hsp70/110 sequences were obtained by combining the non-redundant results of PSI-BLAST (Altschul *et al.* 1997) searches queried with *E. coli* DnaK, human Hsc70 and yeast Sse1. Sequences were aligned by ClustalW (Thompson *et al.* 1994) and non-Hsp70/110 sequences were removed based on their anomalous sequence length or identity. Sequences sharing greater than 95% similarity to another sequence were removed to distribute sampling. Alignment errors were corrected by comparing superimposed crystal structures and alignment gapping patterns, insertion/deletion tendencies flanking secondary structural elements, secondary structural periodicity and physicochemical

conservation (Doolittle *et al.* 1996). The final alignment was large (926 sequences) and diverse (unconserved sites approached random amino acid distributions).

As in a previous version of the statistical coupling analysis (Halabi *et al.* 2009), the alignment is binarized in an *M* by *L* matrix *X* with $X_{si} = 1$ if the most frequent amino acid at position *i* is present in sequence *s*, and $X_{si} = 0$ otherwise; *M* is the number of sequences (rows of *X*) and *L* the number of positions (columns of *X*). The definition of the SCA matrix $\tilde{C}$ involves position-specific weights ϕ_i that quantify the degree of conservation of each position *i* : $\phi_i = \ln\left[f_i(1-q^{(a_i)})/(1-f_i)q^{(a_i)}\right]$, where f_i represents the frequency of the prevalent amino acid a_i at position *i*, and $q^{(a_i)}$ a background frequency for this amino acid. A weighted alignment $\tilde{X}$ is defined with $\tilde{X}_{si} = \phi_i\left(X_{si} - f_i\right)$. The SCA matrix $\tilde{C}$ is the *L* by *L* matrix of correlations between positions given by $\tilde{C} = \tilde{X}^T\tilde{X}/M$, where $\tilde{X}^T$ denotes the transpose of $\tilde{X}$. Similarly, $\tilde{S} = \tilde{X}\tilde{X}^T/L$ gives an *M* by *M* matrix of correlations between sequences. The eigenvectors of $\tilde{C}$ and $\tilde{S}$ form the columns of two orthogonal matrices, *V* and *U*, which are related through the singular value decomposition of $\tilde{X}$: $\tilde{X} = U\Sigma V^T$, where Σ is a diagonal matrix.

An independent component analysis provides a linear transformation W^s that maps the top 3 eigenvectors of *U* into 3 maximally independent axes of sequence variations. One of these 3 directions, the *M*-dimensional vector $\left|U_1^s\right\rangle$, is found to discriminate the non-allosteric sequences from the rest of the sequences in the alignment (Fig. 2.3A). Applying the same linear transformation

W^s to the top 3 eigenvectors of $\tilde{C}$ defines a direction of positional variations, the *L*-dimensional vector $\left|V_1^s\right\rangle$, which indicates the positions underlying the discrimination. The allosteric sector is defined as the positions *i* making significant contribution to $\left|V_1^s\right\rangle$, i.e., $\left\langle i\middle|V_1^s\right\rangle \geq \varepsilon$, where ε = 0.05 corresponds to a threshold of statistical significance (Fig. 2.3B).

2.4.2 Structural modeling

The ATP-bound *S. cerevisiae* Sse1 (Hsp110) structure and a sequence alignment between Sse1 and *E. coli* DnaK (Hsp70) (Liu and Hendrickson 2007) were used to generate a homology model of DnaK(ATP) using Modeller (Sali and Blundell 1993). Molecular dynamics simulations were carried out using the Gromacs platform and Gromos96 force field (Hess *et al.* 2008). The structural model was truncated at residue 531, and ATP, magnesium and potassium ions coordinated in the active site were included. An ATP topology file provided by an earlier study (Colombo *et al.* 2008) was utilized. The system was solvated in a box with at least 12 Å spacing from protein atoms to the edge of the box. Net charge was neutralized with potassium ions and energy was minimized by steepest descents followed by short position-restrained molecular dynamics to equilibrate water molecules at 300 K. In the production molecular dynamics simulation, a simulated annealing protocol cycled through temperature gradients based on previous work in a smaller system (Lindorff-Larsen *et al.* 2005): 300 to 400 K over 150 ps, 400 to 350 K over 150 ps, 350 to 300 K over 500 ps, and 300 K held for 100 ps. Berendsen temperature coupling, Parrinello-Rahman pressure

coupling, and a periodic boundary condition were used. All trajectory analysis was performed within the Gromacs package. Atomic RMSD fluctuations were analyzed by principal component analysis, and cosine content of the first component indicated that motions within the RMSD plateau region were dominated by diffusion (Hess 2002). To avoid over-interpretation, structural clustering was performed on the trajectory such that the entire RMSD plateau region was defined as a single cluster (4.7-53 ns) to determine a median structure, and correlated motions within this region were not investigated further on the basis of cooperativity. Solvent-exposed residues are defined as sites with fractional side chain accessible surface area greater than 0.25 relative to extended Gly-Xaa-Gly peptides, as calculatated by the VADAR web server using the default parameters (Willard *et al.* 2003). Interdomain sector contacts are calculated as sector positions of separate domains that contain side chain atoms within 5.0 Å of each other.

2.4.3 Heat shock assay

Plasmid pMS119 containing a wild type *E. coli dnaK* gene insertion was used as a template for site-directed mutagenesis (Montgomery *et al.* 1999). Plasmids were transformed into temperature-sensitive *E. coli* BB1553 cells (Δ*dnaK52 sidB1*) (Bukau and Walker 1990). Single colonies were grown overnight in LB in the presence of antibiotics at 30 ºC, and each growth's optical density at 600 nm was normalized to 0.2 by dilution with LB media. Growths were serially diluted 10-fold in water pre-equilibrated at 43 ˚C, spotted onto growth

media plates pre-equilibrated at 43 ˚C, and placed in an incubator at the same temperature for 15 hours. Leaky expression off the pMS119 *tac* promoter was sufficient to achieve nearly optimal growth rescue by plasmid-encoded DnaK in LB media without IPTG induction.

2.4.4 Purification of proteins and peptides

E. coli DnaK was prepared similarly as previously described (Montgomery *et al.* 1999), except that *E. coli* BB1553 cells were used and grown at 30 ℃. In a modified two-column purification, the first anion exchange column was used with buffers at pH 7.4. In the second column, DnaK was eluted from the ATP-agarose resin with 2 mM ADP. KCl replaced NaCl in all purification buffers. Crude p5 peptide (CLLLSAPRR) was purchased from Genscript and purified by HPLC using a diphenyl column with elution at ~30% acetronitrile and 70% water; mass spectrometry confirmed the identity of the peptide.

2.4.5 *In vitro* measurements

Steady-state ATPase activity was measured in an enzyme-coupled system as previously described (Montgomery *et al.* 1999). The ATPase activity of 1 μM DnaK plus or minus 100 μM p5 at 30 ℃ was measured on a Biotek Gen5 platereader using Costar 3631 plates. Measurements were taken 3-5 times for each DnaK and auto-hydrolysis sample.

DnaK W102 fluorescence and acrylamide quenching were measured similarly as previously described (Moro *et al.* 2003). Measurements were taken at

room temperature in a Photon Technology International fluorometer at 295 nm excitation wavelength with 4 nm slit widths on both excitation and emission sides. For each sample, spectra were averaged over ten acquisitions and normalized to an intensity of 1.0 prior to the addition of ATP. Faster 15 second scans showed the same spectral trends, indicating that ATP hydrolysis was not a complicating factor during the measurement.

Circular dichroism measurements were taken as previously described (Montgomery *et al.* 1999) using a Jasco J-715 spectrophotomer. Wavelength scans were measured at 30 ℃ using 2 μM DnaK in 10 mM potassium phosphate buffer at pH 7.6. Temperature melts were measured at 222 nm using 2 μM DnaK in 10 mM potassium phosphate buffer, 1 mM $MgCl_2$ and 1 mM ADP, pH 7.6.

CHAPTER 3

A C-TERMINAL CONSERVED AND DISORDERED REGION ENHANCES THE CHAPERONE FUNCTION OF *E. COLI* DNAK

This chapter is the result of collaboration with Mandy Blackburn and Lila Gierasch (Smock *et al.* 2011, Publication 6).

3.1 Introduction

The ubiquitously distributed Hsp70 family of molecular chaperones shepherd newly synthesized polypeptide chains, protect cells from stress-induced protein aggregation, assist in protein translocation across membranes, and regulate assembly and disassembly of macromolecular complexes. These physiological functions are accomplished by a two-domain allosteric mechanism in which cycles of ATP binding and hydrolysis in the N-terminal nucleotide-binding domain (NBD) control the binding and release of hydrophobic polypeptide segments in the substrate-binding domain (SBD) (Bukau *et al.* 2000; Hartl and Hayer-Hartl 2002; Mayer and Bukau 2005).

While we have gained an increasingly detailed picture of the allosteric transition in Hsp70 chaperones and modes of substrate interaction (Rist *et al.* 2006; Liu and Hendrickson 2007; Swain *et al.* 2007; Bertelsen *et al.* 2009), it is striking that there is much less known about the function of the extreme C-terminal unstructured region (Fig. 3.1). In eukaryotic cytoplasmic Hsp70s, this C-terminal segment contains a conserved TPR-domain interaction motif that mediates binding with Chip, Hip, or Hop, co-chaperones that facilitate the

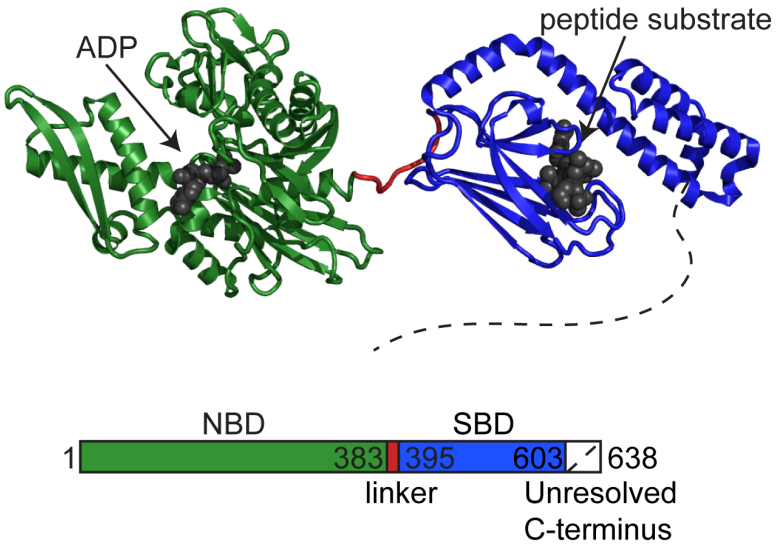

Figure 3.1. Domains of Hsp70 proteins showing the C-terminal sequence of undefined function.

In a substrate-bound conformation of *E. coli* DnaK (PDB 2KHO) (Zhu *et al.* 1996; Harrison *et al.* 1997; Bertelsen *et al.* 2009), the NBD (green) is joined to the SBD (blue) through a short interdomain linker (red). A peptide substrate is sandwiched between the SBD β-sandwich subdomain and α-helical lid, with 35 crystallographically unresolved sites in the C-terminal tail.

assembly of a variety of Hsp70 complexes (Mayer and Bukau 2005). Even though bacteria lack homologs of these Hsp70-interacting TPR domain proteins, many, including the extensively characterized *E. coli* Hsp70, DnaK, have a disordered C-terminal extension of similar length to that in eukaryotes.

To explore the functional significance of this C-terminal region, we carried out a comparative sequence analysis of Hsp70 family members, and two large and distinct groups emerged based on patterns of C-terminal conservation: those containing the eukaryotic TPR domain interaction motif and those harboring a previously undescribed C-terminal motif shared among diverse bacterial sequences, including *E. coli* DnaK, and several eukaryotic organellar Hsp70s. Surpisingly, mutation of conserved C-terminal sequence in DnaK results in significant impairment of its protein refolding activity *in vitro* without affecting interdomain allostery, interaction with co-chaperones DnaJ and GrpE or the binding of a peptide substrate, defying classical explanations for the chaperoning mechanism of Hsp70. Moreover, mutation of specific conserved sites within the DnaK C-terminus reduces the capacity of the cell to withstand stresses on protein folding caused by elevated temperature or the depletion of other chaperone proteins. These features of the C-terminal region support a model in which it acts as a disordered tether linked to a conserved, weak substrate-binding motif, and that this enhances chaperone function by transiently interacting with folding clients.

3.2 Results

3.2.1 A novel sequence motif in Hsp70 C-termini of bacterial origin

An alignment of 730 Hsp70 sequences reveals high conservation in the protein overall (53%) but comparatively low conservation in the C-terminal tail (17% conserved in alignment positions corresponding to *E. coli* DnaK residues 604-638). In addition, residue variation within C-terminal sequences is not statistically correlated to other positions in the protein, indicating mutational independence of this region from the rest of Hsp70 structure (Smock *et al.* 2010). To investigate sequence features of the C-terminal segment in greater detail, the alignment was truncated to the 604-638 region. A sequence similarity matrix was then calculated and hierarchically clustered (Fig. 3.2A). Many eukaryotic and bacterial Hsp70 sequences contain large C-terminal deletions (413 sequences including those labeled green and yellow in the dendrogram in Fig. 3.2A). However, two large and distinct sequence clusters emerge from the analysis with relatively high C-terminal conservation (labeled blue and red in the dendrogram in Fig. 3.2A). As expected, one of the clusters (157 sequences labeled blue in Fig. 3.2A) is composed of strictly eukaryotic Hsp70s that share a sequence motif for TPR domain interaction, including human paralogs Hsc70 and Hsp70-1.

Surprisingly, a second, separately conserved cluster (165 sequences labeled red in Fig. 3.2A) contains Hsp70 sequences from 148 evolutionarily diverse gram-negative and gram-positive bacteria, twelve eukaryotes, of which at least eleven localize to the mitochondria or plastid, and five archaea. This cluster includes *E. coli* DnaK, the most extensively characterized protein in this group,

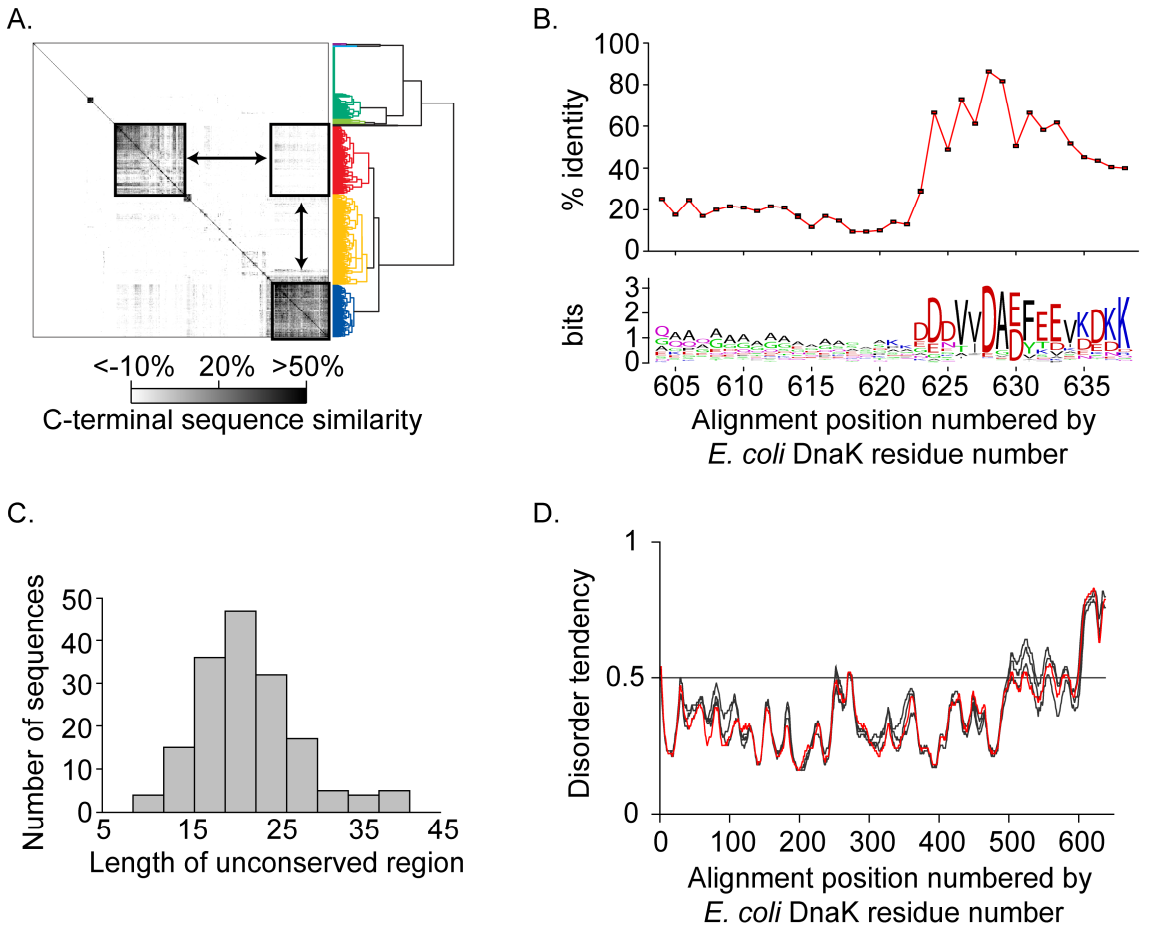

Figure 3.2. A subset of Hsp70 of bacterial origin share a novel C-terminal sequence motif.

A. A sequence similarity matrix shows pairwise conservation between aligned C-terminal Hsp70 sequences. A dendrogram illustrates the organization of patterns within the matrix and is colored according to a hierarchical division where the main clusters are visually apparent. The boxed regions indicate that two internally conserved sequence groups are dissimilar from each other. **B.** The C-terminal sequence cluster colored red in A is composed of an unconserved region followed by a conserved region. The lower panel shows relative amino acid frequencies for the sequence group (Crooks *et al.* 2004) and describes a novel motif in the conserved region. **C.** The unconserved region (corresponding to DnaK residues 604-623) is highly variable in length. **D.** The amino acid compositions of diverse sequences within the cluster, *E. coli* DnaK (highlighted in red), *B. subtilus* DnaK, *H. marismortui* DnaK and *A. thaliana* mtHsc70-1, favor structural disorder following crystallographically resolved structure (residue 603 in *E. coli* DnaK). Scores above 0.5 disorder tendency are significant predictions of disorder using a false positive threshold of 5% (Ishida and Kinoshita 2008).

but excludes its paralogs *E. coli* HscA and HscC, which have C-terminal deletions relative to DnaK. In addition, the classification of sequences by C-terminal conservation cannot be recreated using full sequence conservation. These data indicate that grouping of sequences based on this C-terminal motif is an evolutionarily widespread phenomenon and is not simply an artifact of phylogenetic subdivision.

The amino acid sequence in the red (bacterial origin) cluster shows a unique sequence motif that is conserved only for alignment positions corresponding to *E. coli* DnaK residues 624-638 (Fig. 3.2B). The motif is distinct from that of the previously characterized eukaryotic group, with an EEV tripeptide as the only feature conserved between *E. coli* DnaK and human Hsc70. This tripeptide sequence occurs as part of the conserved TPR-binding EEVD motif at the extreme C-terminus in eukaryotes, and is five residues from the C-terminus in the bacterial cluster. N-terminal to the conserved C-terminal 15 residues is a region of low sequence complexity with a broad length distribution ranging from approximately ten to forty amino acids (Fig. 3.2C). The low sequence complexity and variable length of this region suggest that it may act as flexible tether between the structured SBD lid and the more conserved region of the C-terminal tail. The length of *E. coli* DnaK 604-623 in this variable region is representative of the distribution's mean (20 residues).

The amino acid content of the flexible tether (enriched in Ala, Gly, and Gln) and the more conserved C-terminal region suggest sequence selection for intrinsic disorder (Dunker *et al.* 2002). This expectation was tested by subjecting

representative sequences from the bacterial group to disorder prediction methods based on amino acid composition. Similar results were obtained using both metaPrDOS (Ishida and Kinoshita 2008) and PONDR (Dunker and Uversky 2008) prediction algorithms. The sequences of *E. coli* DnaK (gram-negative), *B. subtilis* DnaK (gram-positive), *H. marismortui* DnaK (archaeal) and *A. thaliana* mtHsc70-1 (mitochondrial) are all highly predictive of structural disorder following residue 603 (Fig. 3.2D). Direct support for these predictions is provided by the finding that in solution the portion of *E. coli* DnaK following residue 603 has been found to lack stable structure (Bertelsen *et al.* 2009), is susceptible to proteolysis, and its removal facilitates crystallization of the SBD (Zhu *et al.* 1996).

Intriguingly, residues 624-633 are predicted to be a protein-binding region, based on their potential to gain energetically stabilizing interactions, using the program ANCHOR (Dosztányi *et al.* 2009). This region has the highest conservation of the entire sequence motif. While the C-terminus is not predicted to have a unique structure, its sequence specificity raises the possibility that it assumes a structure upon interaction with partners. Also, the eukaryotic TPR domain-binding role of the analogous region suggests that there may be an as yet undefined partner that interacts with this region in the bacterial and organellar Hsp70s. However, we made a construct of glutathione-S-transferase fused to the DnaK C-terminal peptide 619-638 and looked for interaction partners in cell lysates from several *E. coli* strains prepared under normal or stress conditions by pull down using glutathione affinity resin. No partners were found. This result may be because there is not a specific binding partner, or because the binding

mediated by the C-terminal region is weak and transient, or because cellular partners are present in low abundance.

3.2.2 The conserved *E. coli* DnaK C-terminus is not required for intramolecular or co-chaperone functions

In order to explore the role of the conserved C-terminus in *E. coli* DnaK, we constructed and purified DnaK with this region truncated (DnaK 1-631), as well as a point mutant, D628A, that substitutes the most conserved residue by Ala, and compared their performances using an array of well-established Hsp70 *in vitro* functional assays. In the ADP-bound state, Hsp70s bind tightly to their substrates. This binding has been characterized through the use of oligopeptides, such as the nine-residue substrate CLLLSAPRR (p5). As measured by fluorescence anisotropy using a FITC-conjugated derivative of p5 (fp5), wild type DnaK, D628A and 1-631 all show normal binding ability (K_d values of approximately 150 nM with overlapping confidence intervals) (Fig. 3.3A).

Interdomain allostery in Hsp70s is revealed by the activation of their basal ATPase activity upon substrate binding, and reciprocally by a reduction in substrate affinity (enhanced release kinetics) upon ATP binding. Wild type DnaK, D628A and 1-631 all show basal and p5-stimulated ATPase rates within experimental error of one another, and show approximately a six-fold stimulation by p5 (Fig. 3.3B). The substrate fp5 is a suitable probe for measuring ATP-stimulation of peptide release, as its fluorescence intensity is quenched upon

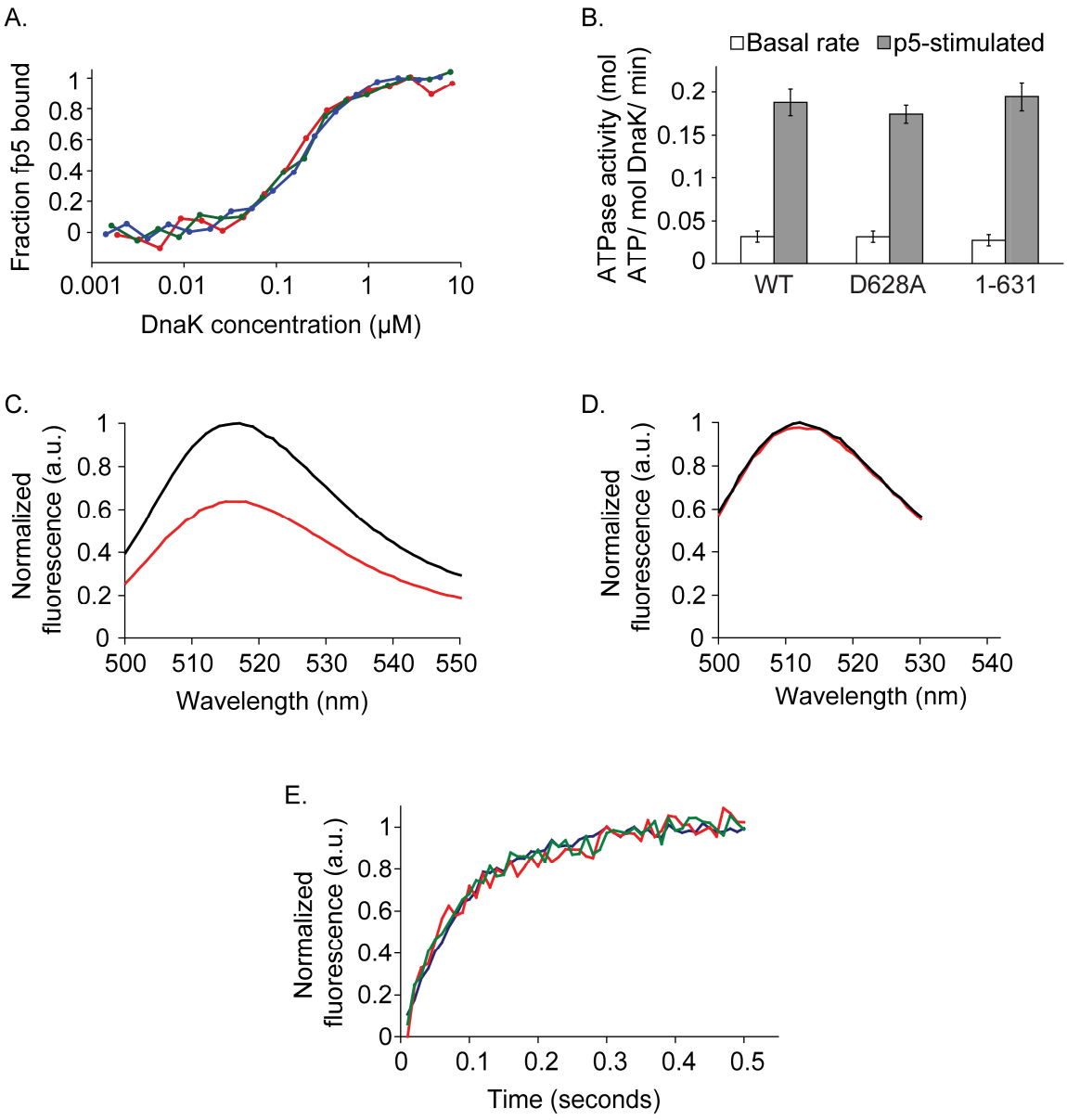
A.
Fraction fp5 bound
1
0.8
0.6
0.4
0.2
0
0.001
0.01
0.1
1
10
DnaK concentration (μM)
B.
Basal rate
p5-stimulated
ATPase activity (mol ATP/ mol DnaK/ min)
0.2
0.15
0.1
0.05
0
WT
D628A
1-631
C.
Normalized fluorescence (a.u.)
1
0.8
0.6
0.4
0.2
0
500
510
520
530
540
550
Wavelength (nm)
D.
Normalized fluorescence (a.u.)
1
0.8
0.6
0.4
0.2
0
500
510
520
530
540
Wavelength (nm)
E.
Normalized fluorescence (a.u.)
1
0.8
0.6
0.4
0.2
0
0
0.1
0.2
0.3
0.4
0.5
Time (seconds)

Figure 3.3. Allosteric signaling between the nucleotide and substrate-binding sites is normal in C-terminal mutants.

A. In an ADP-bound, high substrate affinity state, DnaK 1-631 (red) and D628A (green) bind to fp5, a substrate peptide containing a linked fluorescein group, with the same affinity as wild type DnaK (blue). **B.** The basal ATPase activity of DnaK and its allosteric stimulation by the substrate peptide p5 are normal in C-terminal mutants. **C.** The fluorescence spectrum of fp5 (black) is quenched upon addition of wild type DnaK (red). **D.** The fluorescence spectrum of free fluorescein (black) is unaffected by addition of wild type DnaK (red), indicating that the quenching of fp5 by DnaK requires specific binding to the linked peptide. **E.** Upon rapid mixing with ATP, fp5 release follows the same kinetic profile in DnaK 1-631 (red), D628A (green) and wild type DnaK (blue). In control experiments, mixing with buffer instead of ATP showed no increase in fp5 fluorescence over time, and mixing with saturating concentrations of both ATP and p5 gave the same kinetic trace as addition of ATP alone, indicating that release of fp5 was both ATP-dependent and complete.

binding by DnaK (Fig. 3.3C,D). Here again, data obtained by stopped-flow measurements of peptide release upon ATP binding showed kinetics in the range of 11.0 +/- 0.3 s^{-1} for the mutated C-terminal constructs, indistinguishable from wild type (Fig. 3.3E) and consistent with previously reported values (Theyssen *et al.* 1996).

In addition to the functional measurement of interdomain allostery by ATPase activity, a structural basis of allostery can be assessed using the sole intrinsic tryptophan of DnaK. W102 in the NBD undergoes fluorescence emission changes upon binding of ATP and in the presence of the SBD lid, and is therefore a probe for functional ATP-induced interdomain docking (Moro *et al.* 2003). The physical proximity of the disordered C-terminal tail to the structured SBD lid suggests that C-terminal mutation might alter lid interaction with the NBD. However, in wild type DnaK, D628A and 1-631, the W102 fluorescence emission maximum shifts equally from 342 to 338 nm in the presence of ATP (Fig. 3.4). The quenching of W102 fluorescence intensity upon ATP binding was also identical to that of wild type protein for each of these constructs (26% +/- 1%), providing additional support for normal interdomain packing.

By analogy with the role of the eukaryotic C-terminus in co-chaperone binding, the C-terminal region of DnaK could be involved in interaction with partner co-chaperones, DnaJ and GrpE. These co-chaperones affect DnaK function in part by increasing the steady-state ATPase activity of DnaK, and in turn, the rate of DnaK conformational cycling between high- and low-affinity substrate-binding states. GrpE catalyzes nucleotide exchange and DnaJ

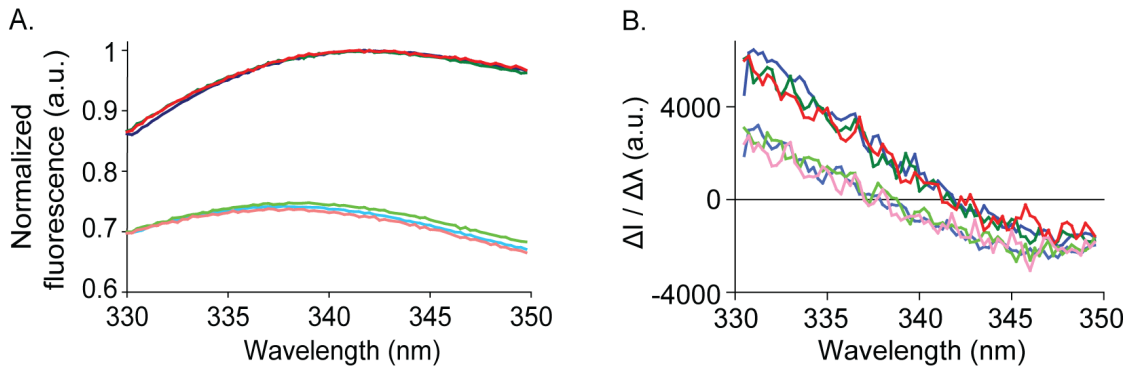

Figure 3.4. C-terminal mutants undergo normal ATP-induced interdomain docking.

A. Wild type DnaK (blue), 1-631 (red) and D628A (green) display the same characteristic blue shift and intensity quench of Trp fluorescence upon ATP binding, indicating that they undergo productive interdomain docking (Moro *et al.* 2003; Liu and Hendrickson 2007). The lower and more lightly shaded traces are from measurements after ATP addition. **B.** The derivative plots of A reveal the spectral peaks as zero crossing points for easier inspection.

stimulates ATP hydrolysis. The rate-limiting step of DnaK basal ATPase activity is ATP hydrolysis, therefore DnaJ alone causes a much larger increase in DnaK ATPase activity than GrpE alone, while the combination is synergistic (McCarty *et al.* 1995). The ATPase activity of wild type DnaK was stimulated 1.6-fold by GrpE, 12–fold by DnaJ, and 90-fold by both co-chaperones together. Importantly, no difference was observed for the C-terminally mutated DnaK D628A or 1-631 (Fig. 3.5A). These data are consistent with previous reports that a DnaK constructs with truncations exceeding the conserved region reported here have little to no effect on DnaJ binding (Gassler *et al.* 1998; Suh *et al.* 1999).

In addition to its direct association with DnaK, DnaJ can also bind misfolded substrates and deliver them to DnaK, coupling the presentation of misfolded substrates to DnaK with the conversion of DnaK to an ADP-bound, high substrate affinity state (Mayer and Bukau 2005). We found that the cooperative stimulation of DnaK ATPase activity by DnaJ and substrates could only be detected in a narrow range of molar ratios, but that even in these circumstances, there was no detectable difference between wild type DnaK and C-terminal mutants in their mutual stimulation by DnaJ and the oligopeptide substrate p5 or an unfolded protein substrate (reduced carboxymethylated α-lactalbumin, RCMLA) (Fig. 3.5B,C).

3.2.3 The C-terminal region enhances protein refolding efficiency

While many of the underlying biochemical and structural features of DnaK function show no measurable difference among C-terminal constructs, we also

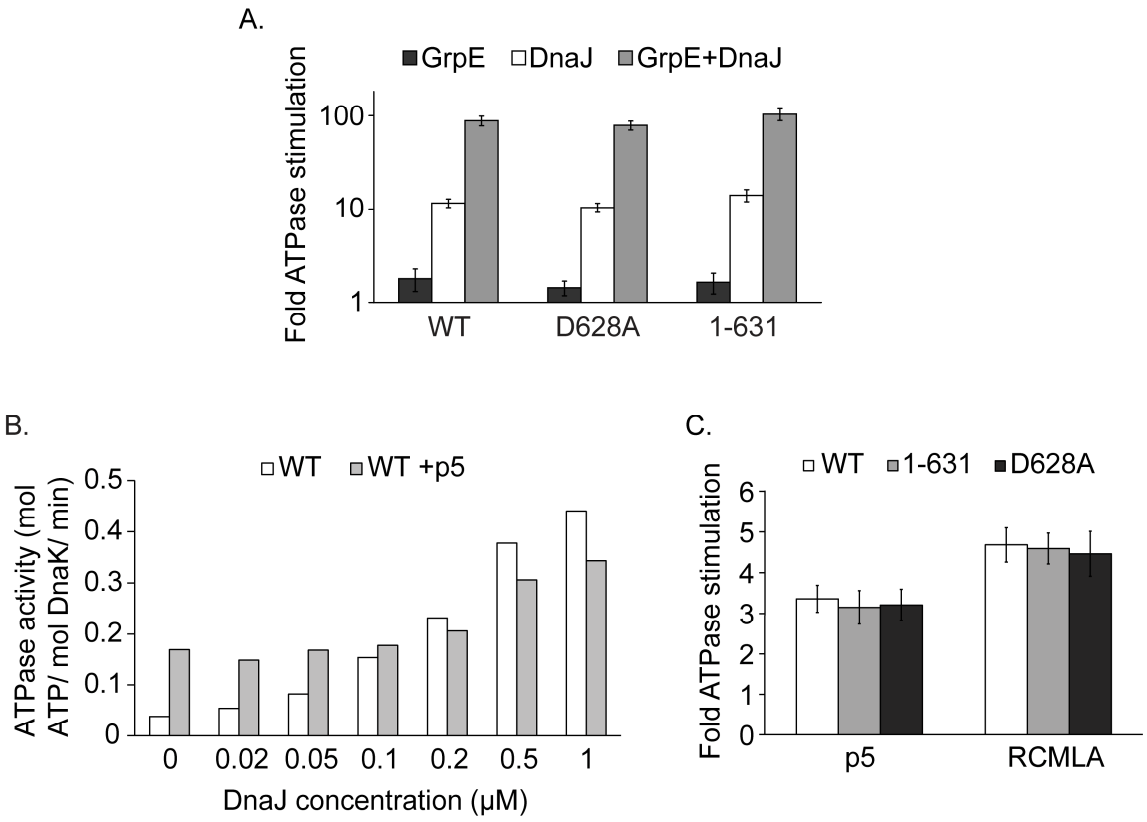

Figure 3.5. Interaction with co-chaperones GrpE and DnaJ is normal.

A. The ATPase activities of DnaK 1-631 and D628A in the presence of co-chaperones GrpE and DnaJ are indistinguishable from wild type DnaK. **B.** Conditions for measuring the joint stimulation of DnaJ and peptide p5 were investigated. In the absence of DnaJ, saturating (100 μM) p5 stimulates wild type DnaK ATPase activity. However, at high concentrations of DnaJ, addition of p5 is slightly inhibitory. Conversely, the presence of saturating p5 masks DnaK ATPase response to DnaJ at lower DnaJ concentrations. The sensitivity of DnaK ATPase activity to the combined effects of both DnaJ and substrates could be distinguished most easily at sub-saturating concentrations (DnaJ, 0.05 uM; p5, 10 μM). **C.** The ATPase activity of C-terminal mutations responds normally to joint DnaJ (0.05 μM) and substrate stimulation (10 μM p5 or 30 μM RCMLA).

tested chaperone function directly by measuring an ability to protect refolding proteins from aggregation and assist their folding to the native state. Purified DnaK variants together with co-chaperones DnaJ and GrpE were used in a reconstituted refolding assay of denatured firefly luciferase. Successful refolding of luciferase to its native state is measured by its ability to oxidize the small molecule luciferin, which leads to emission of light (Schroder *et al.* 1993).

Strikingly, we find that both DnaK 1-631 and D628A show markedly reduced ability to assist refolding of denatured luciferase (Fig. 3.6). Both the rate of appearance of folded luciferase (monitored by photon count) and the total amount of native luciferase are reduced when the C-terminal region of DnaK is truncated or mutated. After subtracting a DnaK-minus sample, mutations in the conserved DnaK C-terminal region result in an initial rate of just 25-35% of that observed with wild type DnaK rate and maximal luminescence that is 50-60% of the wild type level.

3.2.4 C-terminally truncated E. coli DnaK is functionally defective under stringent *in vivo* conditions

DnaK functions as a generalized chaperone, so we tested whether the deficiency observed for C-terminally mutated DnaK in refolding an exogenous substrate, firefly luciferase, reports on a broader, physiologically meaningful defect in *E. coli*. Two commonly used assays for *E. coli* DnaK function *in vivo* are bacteriophage λ propagation, which requires host Hsp70 to disassemble a phage protein complex to initiate DNA replication, as well as growth following protein

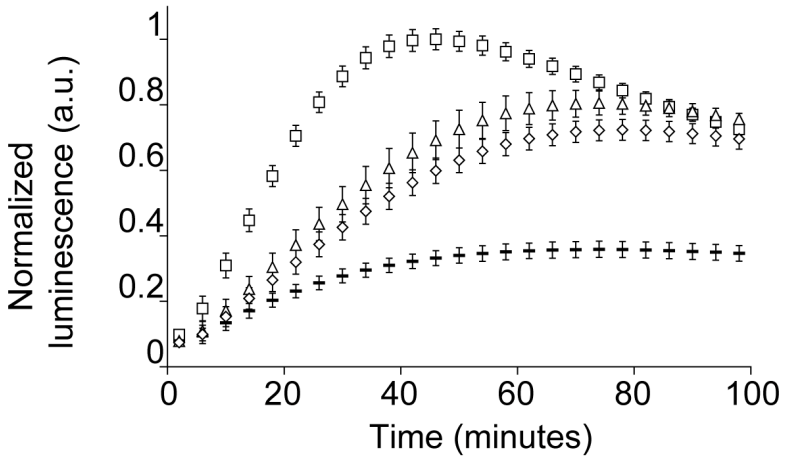

Figure 3.6. Disruption of the C-terminal motif impairs substrate refolding activity.

Wild type DnaK (squares) promotes faster formation and higher yield of native luciferase than DnaK 1-631 (diamonds), D628A (triangles) or samples omitting DnaK (dashes).

misfolding and aggregation induced by heat shock (Yochem *et al.* 1978; Bukau and Walker 1990; Osipiuk *et al.* 1993). Previous studies report no defect in phage propagation upon truncation of the *E. coli* DnaK C-terminus (Gassler *et al.* 1998; Suh *et al.* 1999; Swain *et al.* 2006), and in agreement with these studies, we confirm that a plasmid expressing DnaK with the entire disordered C-terminus removed (1-603) is fully capable of propagating λ in an *E. coli dnaK*-knockout strain at 30 ℃ (Fig. 3.7A). However, we observe sign ificant growth defects after heat shock at 43 ℃ in *E. coli* expressing DnaK with just a seven-residue truncation (1-631) (Fig. 3.7B).

Prior reports using extensive C-terminal truncations result in a range of heat shock viabilities and are apparently sensitive to the genetic background and complications arising from SBD lid destabilization. For example, it was previously reported that DnaK truncated at residue 628 or 538 supported growth at heat shock temperature in the *E. coli dnaK756* strain (Suh *et al.* 1999), while others have reported growth defects upon heat shock for DnaK 1-538 (Gassler *et al.* 1998). In the *dnaK756* strain, the chromosomal copy of DnaK harbors three point mutations and the heat shock transcription factor σ^{32} is wild type. Perturbation of DnaK likely leads to the increased expression of other heat shock proteins in a wild type σ^{32} background (Paek and Walker 1987), which may partially compensate for a DnaK defect. We also note that the effects of C-terminal truncations within the SBD lid (prior to residue 603) are difficult to interpret physiologically because of the now well-established tendency for the truncated α-helical subdomain to unwind and bind to the substrate-binding cleft in the SBD

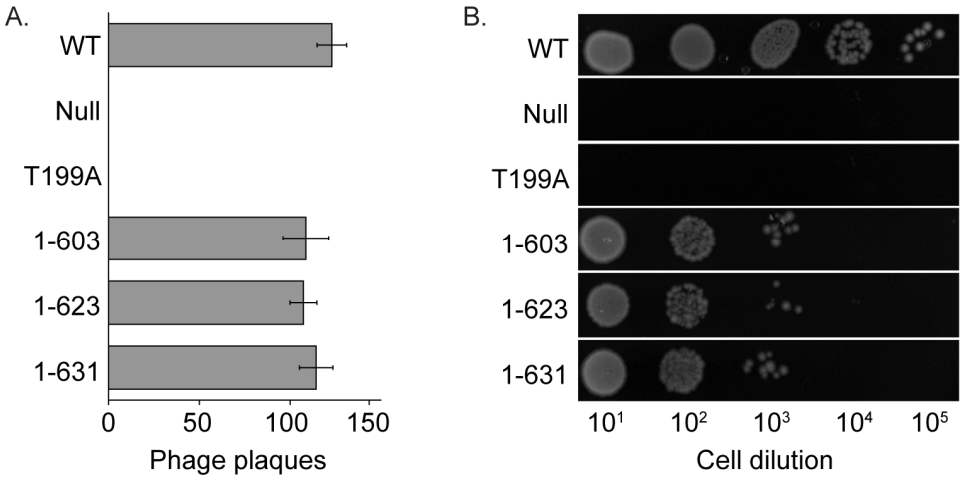

Figure 3.7. C-terminal disruption results in functional defects *in vivo*.

A. In a DnaK-dependent process, an *E. coli* Δ*dnaK* strain (BB1553) is fully capable of propagating phage λ at 30 ℃ when expressing plasmid-encoded C-terminal truncations of DnaK. **B.** The same strain shows impaired growth upon 43 ℃ heat shock when expressing C-terminal truncations o f DnaK. DnaK T199A cannot hydrolyze ATP effectively (Buchberger *et al.* 1995) and is a negative control.

(Pellecchia *et al.* 2000; Swain *et al.* 2006). In addition to *dnaK* deletion, the strain used here (BB1553) has reduced levels of co-operonic *dnaJ* expression and harbors a secondary σ^{32} mutation that suppresses high level expression of genes under control of the heat shock promoter (Bukau and Walker 1990).

Revealing defects in DnaK function is complicated by the partially overlapping functions among different chaperones. To gain greater insight into the functions of the DnaK C-terminus, we took advantage of the synthetic lethality recently reported for *E. coli* Δ*secB* Δ*dnaK* (Ullers *et al.* 2007). Starting with an *E. coli* Δ*secB* strain, we introduced a truncation into the chromosomal *dnaK* gene after residue 631 by homologous recombination. At 30 ℃, the resulting strain shows a slightly reduced growth rate (Fig. 3.8A) and a highly filamentous morphology (Fig. 3.8B), revealing a requirement for the intact DnaK C-terminus at unstressed temperature in the absence of the chaperone SecB.

Ullers and colleagues previously developed *E. coli* strain GP502, with a *secB* knockout and the co-operonic *dnaK* and *dnaJ* genes placed under control of the P_{BAD} promoter. When grown in media supplemented with 0.5% arabinose, these cells have normal (not stress-induced) levels of DnaK and DnaJ. Furthermore, growth in media supplemented with glucose to inhibit P_{BAD} expression results in significant protein aggregation (Ullers *et al.* 2007). Given the observed influence of the DnaK C-terminus on cellular health in the absence of SecB, and the partially overlapping capabilities of SecB and DnaJ to bind misfolded proteins and maintain their solubility (Langer *et al.* 1992; Ullers *et al.* 2004), we hypothesized that cells depleted of both SecB and DnaJ would be

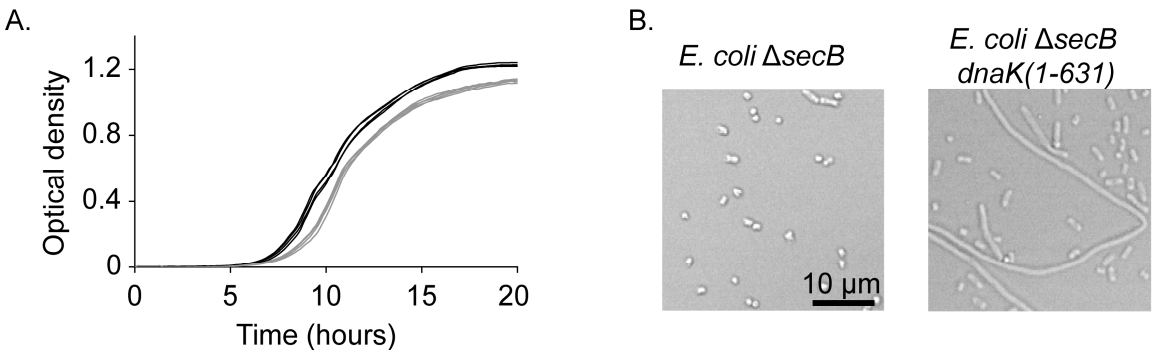

Figure 3.8. *E. coli* is impaired by C-terminal truncation of DnaK in the absence of SecB.

A. *E. coli* Δ*secB dnaK(1-631)* grows at a slightly reduced rate (gray traces) relative to *E. coli* Δ*secB* (encoding full-length *dnaK*, black traces). Cultures are grown at 30 ℃. **B.** *E. coli* Δ*secB* cells have a normal rod-shaped morphology while *E. coli* Δ*secB dnaK(1-631)* cells are prone to filamentation.

even more sensitive to the influence of the DnaK C-terminus. This strain was transformed with a plasmid that encodes a *dnaK* gene with an IPTG-inducible promoter, allowing controlled expression of chromosomal wild type *dnaK* and *dnaJ* genes as well as plasmid-encoded *dnaK* variants (Fig. 3.9A). Three growth conditions provide an internal positive control (grown with arabinose; chromosomal wild type *dnaK* and *dnaJ* expression), internal negative control (grown with glucose; repression of all *dnaK* and *dnaJ* expression), and test condition (grown with glucose plus IPTG; plasmid-encoded *dnaK* expression in the absence of *dnaJ* expression) (Fig. 3.9B).

Accordingly, we find that strain GP502 transformed with plasmid-encoded wild type *dnaK* shows a high growth rate in arabinose media, a very low growth rate in the presence of glucose, and rescued growth in the presence of glucose plus IPTG (Fig. 3.9C). The trends in growth rate are mirrored by features of cell morphology, with a filamentous phenotype observed only in the presence of glucose (Fig. 3.9F). In contrast, growth is not rescued with IPTG in cells carrying an empty vector or carrying a vector that expresses non-functional DnaK with mutation at T199A or K414I (Buchberger *et al.* 1995; Montgomery *et al.* 1999). C-terminal truncation constructs 1-603, 1-623 and 1-631 behave normally in control conditions (Fig. 3.9D), but when grown in the IPTG rescue condition, all show severe defects with little or no growth benefit relative to the empty vector (Fig. 3.9E) and have a filamentous morphology (Fig. 3.9F). Therefore, depleting DnaJ exacerbates the C-terminal truncation phenotype observed in the absence of

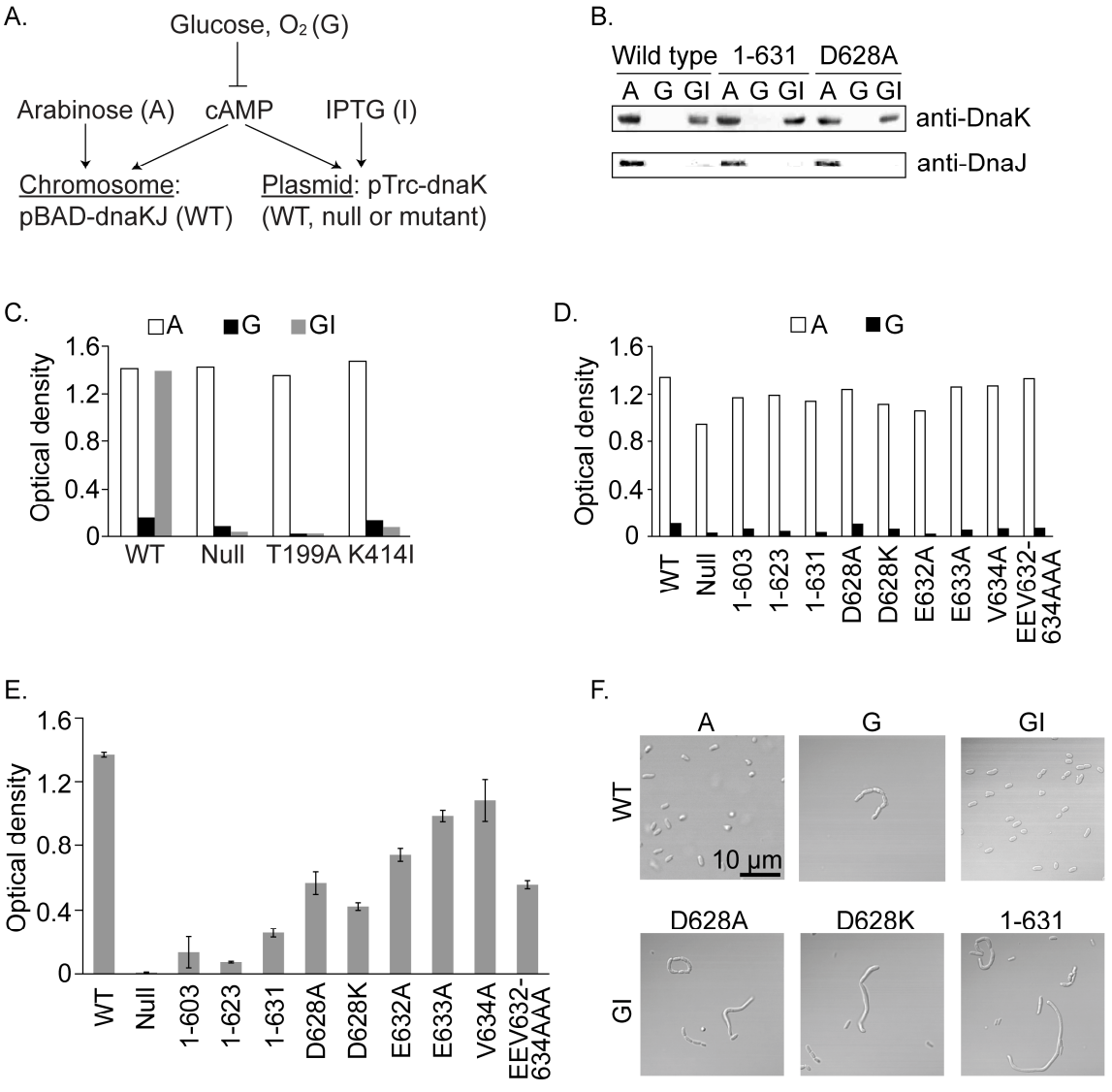
A.
Glucose, O2 (G)
Arabinose (A)
cAMP
IPTG (I)
Chromosome:
pBAD-dnaKJ (WT)
Plasmid: pTrc-dnaK
(WT, null or mutant)
B.
Wild type
1-631
D628A
A G GI A G GI A G GI
anti-DnaK
anti-DnaJ
C.
A
G
GI
Optical density
1.6
1.2
0.8
0.4
0
WT
Null
T199A
K414I
D.
A
G
Optical density
1.6
1.2
0.8
0.4
0
WT
Null
1-603
1-623
1-631
D628A
D628K
E632A
E633A
V634A
EEV632-634AAA
E.
Optical density
1.6
1.2
0.8
0.4
0
WT
Null
1-603
1-623
1-631
D628A
D628K
E632A
E633A
V634A
EEV632-634AAA
F.
A
G
GI
WT
10 μm
D628A
D628K
1-631
GI

Fig. 3.9. DnaK C-terminal function is sensitive to chaperone depletion and site-specific disruption of conserved sites.

A. *E. coli* strain GP502 (Δ*secB*, P_{BAD}-*dnaKdnaJ*) (Ullers *et al.* 2007) transformed with a plasmid carrying an IPTG-inducible *dnaK* gene allows for the controlled expression of wild type *dnaK* and *dnaJ*, and separately, a *dnaK* variant. **B.** Western blots of *E. coli* GP502 transformed with different *dnaK* variants show the expected expression profiles when grown in the presence of different metabolites (as depicted in A). **C.** Control experiments demonstrate that all transformants grown at 30 ℃ show high and low levels of growth in media supplemented with arabinose and glucose, respectively, but growth rescue in media supplemented with glucose and IPTG is dependent on DnaK function. DnaK T199A (Buchberger *et al.* 1995) and DnaK K414I (Montgomery *et al.* 1999) are non-functional DnaK constructs. **D.** In the same assay, C-terminal mutants of DnaK behave normally in control conditions, **E.** but are unable to support normal growth when wild type DnaK and DnaJ are depleted in media supplemented with glucose and IPTG. Importantly, site-specific mutations result in growth defect severity following the observed conservational pattern (Fig. 3.2). **F.** Growth rate defects coincide with cellular filamentation.

SecB alone and provides an even more stringent assay for DnaK C-terminal function.

3.2.5 Site-specific contributions of the C-terminal tail to cellular function

Cells depleted of SecB and DnaJ reveal a severe phenotype upon DnaK C-terminal truncation and allow us to explore specific sequence requirements within the conserved C-terminal tail in greater detail. Alanine substitutions were made in DnaK at the most highly conserved position in the bacterial C-terminal sequence cluster, D628, and within the ^{632}EEV^{634} motif shared with eukaryotic Hsp70s. Cells expressing plasmid-encoded C-terminal mutation constructs show a high rate of growth in the presence of arabinose (positive control) and a low rate of growth in the presence of glucose (negative control), as expected (Fig. 3.9D). All of these site-specific mutations result in significant growth defects in the test condition (Fig. 3.9E,F), as would be anticipated if the pattern of conservation arose from specific residue roles (Fig. 3.2), with mutation at D628 having the greatest effect. Charge reversal (D628K) is more deleterious than removal of charge (D628A). Single Ala substitutions within ^{632}EEV^{634} show significant defects that are less pronounced than at D628, in agreement with their lesser degree of conservation. A triple Ala mutation at the ^{632}EEV^{634} motif is defective to a similar extent as the single D628 mutations, but none of these changes cause as large a defect as the truncation of seven residues in 1-631, indicating that there are multiple determinants for function dispersed throughout the conserved region.

3.3 Discussion

The conserved sequence character of the entire C-terminal region predicts that it is disordered, which raises interesting questions regarding its function, given the roles of structural disorder in other proteins and chaperone systems (Dyson and Wright 2005). Intrinsically disordered regions (IDRs) can act as scaffolds for larger molecular assemblies. Interestingly, *E. coli* YbbN has been identified as a DnaK-interacting chaperone with an atypical TPR domain (Kthiri *et al.* 2008; Lin and Wilson 2011) and DnaK cooperates with the *E. coli* Hsp90, HtpG (Genest *et al.* 2011), in possible analogy to assemblies with eukaryotic cytoplasmic Hsp70s. Nonetheless, a C-terminal defect is manifested in a simple reconstituted refolding assay with only DnaJ and GrpE present, and we observe no deficiency in assays that probe functional interactions between DnaK and these co-chaperones upon C-terminal mutation. Thus, a scaffolding role for the C-terminus is not directly supported by the current data.

Another possibility is that the C-terminal tail shields misfolded regions of bound substrate and blocks aggregation, by analogy to small heat shock proteins (Tompa and Csermely 2004). However, the conservation of specific residues in the C-terminus and functional defect upon charge reversal at a single site in DnaK D628K demonstrate that that the C-terminal tail's function cannot be ascribed to occlusion based on chain length and disorder alone.

As a potential protein-binding region, the more highly conserved extreme C-terminal sequence (residues 624 to 638 of DnaK) may present a weak, auxiliary binding site for denatured substrate proteins. Upon ATP-induced

release from the canonical substrate-binding site, weak association with the C-terminal IDR would retain the substrate in the vicinity of the chaperone and thus promote subsequent cycles of binding and release (Fig. 3.10A). This model is consistent with the finding that IDRs frequently act as flexible tethers and present molecular recognition features; the resulting flexibility enables the same regions to engage in multiple modes of binding (Smock and Gierasch 2009), as suggested here. For example, the N-terminal disordered region of the small heat shock protein PsHsp18.1 allows the molecule to recognize a variety of misfolded substrates (Jaya *et al.* 2009), as do the N- and C-terminal disordered regions of the ubiquitin ligase San1 (Rosenbaum *et al.* 2011).

An additional outcome of substrate binding to the disordered C-terminal tail can be envisioned: The binding itself can transiently disrupt non-native local structure of misfolded substrates while the substrate is bound to the higher affinity site in the β-sandwich binding pocket. Local unfolding would facilitate conformational search of the substrate chain, which may fold spontaneously upon release or be subjected to ongoing cycles of interaction (Fig. 3.10B). This idea is similar to an iterative annealing model for GroEL function (Todd *et al.* 1996). The general idea that chaperones may exploit weak multivalent binding events offered by an IDR at a secondary proximal site was suggested by Tompa and Csermely, who termed it entropy transfer (Tompa and Csermely 2004). In the cell, multivalent molecular recognition by the DnaK C-terminus may accommodate the chaperoning of diverse cellular substrates, as opposed to other *E. coli* Hsp70s lacking the C-terminal motif that have specialized functional

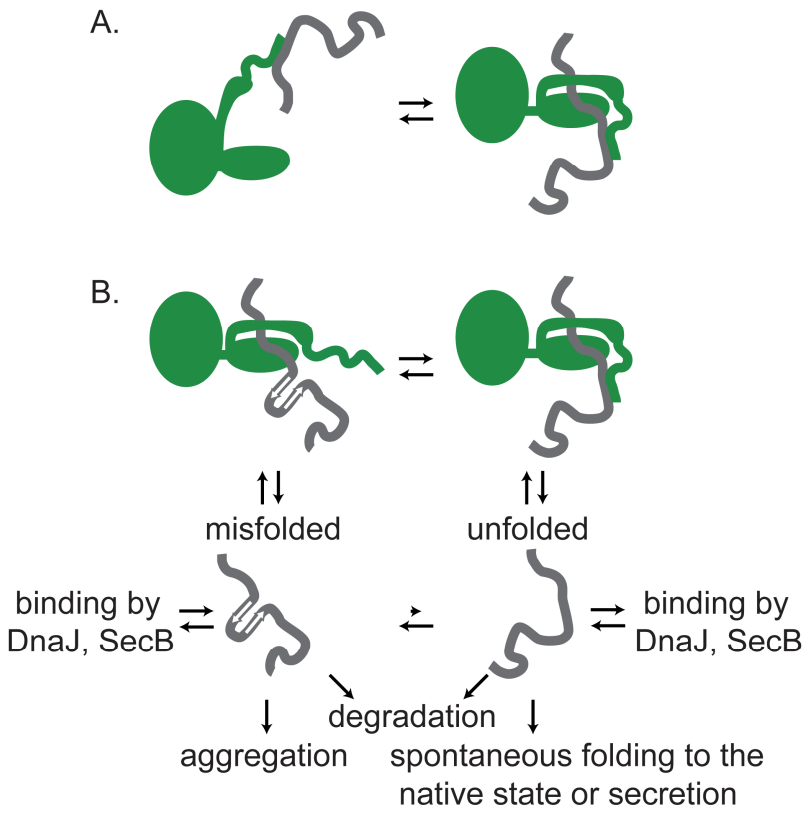

Figure 3.10. A model for Hsp70 C-terminal function.

The disordered C-terminal region of Hsp70 (green) acts as a flexible, multivalent binding site for direct interaction with substrate proteins (gray), which may contribute to the chaperone mechanism in different ways. **A.** Enhanced kinetic partitioning. After ATP-induced lid opening and release of the substrate to solvent, the substrate may fold to the native state or repeatedly rebind the chaperone. The C-terminal tail weakly interacts with substrates even after release, increasing the probability of reforming a stable substrate-chaperone complex. **B.** Local unfolding. Unfolded proteins may spontaneously fold to the native state or become kinetically trapped in a misfolded state, leading to possible aggregation. Upon binding to Hsp70, the C-terminal tail disrupts locally misfolded structure through transient multivalent binding, thereby relieving the kinetic trap and allowing the release of unfolded substrate. ATP-mediated substrate release in DnaK as well as DnaJ-mediated delivery of substrates to DnaK are not shown for simplicity.

roles. For example, *E. coli* HscA specifically assists in the maturation of iron-sulfur cluster assembly protein IscA (Vickery and Cupp-Vickery 2007) and has a natural C-terminal deletion.

The chaperones SecB and DnaJ, which are functionally linked to the C-terminal role of DnaK, play related roles in the protection of misfolded states from aggregation and degradation as well in the maintenance of the unfolded state for delivery to the secretory pathway (Langer *et al.* 1992; Randall and Hardy 2002; Ullers *et al.* 2004; Sakr *et al.* 2010).

A weak, multivalent binding of the C-terminal IDR and the heterogeneity of target substrates suggest that other sequence variations might functionally substitute in some cases. Indeed, complete truncation of the DnaK C-terminal region after the structured SBD lid and its replacement with a His tag actually enhances the refolding of exogenous firefly luciferase substrate (Aponte *et al.* 2010). Given the evolutionary selection of the C-terminal motif (Fig. 3.2), one might suspect that it is optimized for substrates of bacterial origin.

We note that the mechanism we propose might also be utilized by eukaryotic cytoplasmic Hsp70s. If so, evolution has overlaid more specialized functionality on the basic function by selecting C-terminal motifs that also promote assembly with other chaperone and degradation machinery. Thus, their C-terminal regions may play multiple roles.

3.4 Methods

3.4.1 Sequence analysis

Pairwise sequence similarity was scored using a BLOSUM 62 substitution matrix (Henikoff and Henikoff 1992) and the resulting sequence similarity matrix was hierarchically clustered using a city block metric and single linkage in Matlab.

3.4.2 Plasmids and strains

The wild type *dnaK* gene is carried on vector pMS119 (p*tac*, pBR322 *ori,* Amp^R) (Montgomery *et al.* 1999). All *dnaK* mutations were generated by site-directed mutagenesis of pMS119-*dnaK* by introducing Ala, Lys, or stop codons. *E. coli* strains BB1553 (MC4100 Δ*dnak52*::Cam^R, *sidB1*) (Bukau and Walker 1990) and GP502 (Mph42 Δ*secB*, P_{BAD}-*dnaKJ*, Kan^R, *araC*) (Ullers *et al.* 2007) are used for *in vivo* assays and DnaK purifications.

The *E. coli* Δ*secB dnaK(1-631)* strain was created by λ *red*-mediated homologous combination of the chromosomal *dnaK* gene. The bacteriophage genes were expressed from plasmid pSIM9 (Sharan *et al.* 2009) in *E. coli* strain MC4100. Cells were then electroporated with a linear DNA recombination substrate composed of a *dnaK* C-terminal homology arm with the codon at residue 632 mutated to TAA, an FRT site, a Tet^R gene taken from plasmd pGPMTet5 (George Munson, University of Miami), a second FRT site, and a *dnaK-dnaJ* intergenic homology arm. Following recombination, selected clones were used to transfer tetracycline resistance and the linked *dnaK* truncation to *E.*

coli Keio knockout strain JW3584-1 (Δ*secB*::kan^{R}) (Baba *et al.* 2006) by P1 phage transduction. Chromosomal antibiotic resistance genes were removed by expression of FLP recombinase using plasmid pFT-A (Posfai *et al.* 1997). The plasmid was removed during growth at 37 ℃ and the cor rect losses of antibiotic resistance were verified by plating experiments. The correct chromosomal replacement of the *dnaK* C-terminus and deletion of *secB* were verified by PCR and DNA sequencing.

3.4.3 Purification of proteins and peptides

E. coli DnaK was prepared as previously described (Smock *et al.* 2010). *E. coli* GrpE was purified by utilizing a cleavable His tag (Chang *et al.* 2010), which was removed by treatment with His-tagged S219V TEV protease (Kapust *et al.* 2001). Concentration was estimated in a Bradford assay using BSA as a standard. *E. coli* DnaJ was purified using plasmid p29SEN-*dnaJ* (p*trc*, Amp^{R}) (Ullers *et al.* 2007) following established protocols (Chang *et al.* 2008). Concentration was calculated from absorbance at 280 nm using an extinction coefficient of 13,400 M^{-1} cm^{-1}. The peptide p5 (CLLLSAPRR) was prepared as previously described (Smock *et al.* 2010). A purified derivative of p5 (ala-p5, ALLLSAPRR) N-terminally conjugated with FITC (fp5) was purchased from Genscript. Concentration was determined by absorbance at 494 nm in phosphate buffer at pH 8.0 using an extinction coefficient of 72,000 $M^{-1}cm^{-1}$. RCMLA was prepared (Siegenthaler and Christen 2006) and N-terminally conjugated with FITC (Tanaka *et al.* 2002) as previously described.

3.4.4 *In vitro* measurements

All absorbance and luminescence measurements were made using a Biotek Synergy2 microplate reader. Steady-state ATPase activity was measured at 30°C using an enzyme-coupled system as previously described (Smock *et al.* 2010). One µM DnaK was used for all ATPase measurements in the presence or absence of 100 µM p5, 1 µM DnaJ and 1 µM GrpE dimer. The refolding of denatured firefly luciferase (Promega) was measured as previously described (Aponte *et al.* 2010) using Steady Glo reagent (Promega) as a source of luciferin, 80 nM denatured luciferase, 3.2 µM DnaK, 0.8 µM DnaJ and 0.4 µM GrpE.

All fluorescence measurements were taken on a Photon Technology International fluorometer. DnaK W102 fluorescence emission spectra were measured as previously described (Smock *et al.* 2010), in which multiple measurements are averaged resulting in an extra 7% W102 intensity quench due to photobleaching. Fluoresence anisotropy measurements of fp5 incubated with DnaK were made essentially as described previously (Montgomery *et al.* 1999) but in the presence of 1 mM ADP. Kinetic measurements were made using a Bio-Logic stopped-flow device similarly as previously described (Theyssen *et al.* 1996) except that fp5 was used as a substrate with excitation at 494 nm and detection at 517 nm.

3.4.5 *In vivo* assays

Phage propagation and heat shock assays were performed as previously described using *E. coli* strain BB1553 transformed with plasmid pMS119-*dnaK*

(Bukau and Walker 1990; Clérico *et al.* 2010). In the chaperone depletion assay with a *secB* knockout and controllable *dnaKdnaJ* expression, *E. coli* strain GP502 was transformed with pMS119-*dnaK* plasmids. Resulting colonies were used to inoculate overnight cultures grown at 30 ℃ in LB media supplemented with 0.5% arabinose, 50 mg/L ampicillin and 50 mg/L kanamycin. Optical density at 600 nm (OD_{600}) was normalized to 0.4 by dilution with LB media for each growth. 10 µL of each growth was diluted in 1 ml of water and 5 µl was used to inoculate 1.5 ml of LB media supplemented with antibiotics and either 0.5% arabinose, 0.5% glucose or 0.5% glucose plus 50 µM IPTG. This concentration of IPTG gave maximal growth recovery of strain GP502 transformed with pMS119-*dnaK* (wild type) in the presence of 0.5% glucose. Growths were carried out at 30 ℃ in 15 ml falcon tubes secured on their sid es in a shaker to ensure efficient aeration. OD_{600} was measured after twenty hours. Samples for Western blots and microscopy were prepared in the same way except that optical density was normalized for Western blot samples. Blots were probed using mouse anti-DnaK and rabbit anti-DnaJ antibodies (Enzo Life Sciences).

Pulldown assays were performed using a GST fusion construct with the DnaK C-terminus. The fusion was made by the in-frame ligation of sequence encoding DnaK residues 619-638 into the C-terminal linker of GST in plasmid pGEX6p1 (Amersham). GST and GST-DnaK(619-638) were purified using glutathione-sepharose 4B resin. Cell lysates of unstressed (DH5α) and stressed *E. coli* cells (BB1553 and GP502) were prepared and putative interaction partners were pulled down with GST or GST-DnaK(619-638) as previously

described (Classic Protocol 2004). Additionally, the plasmids pGEX6p1 and pGEX6p1-DnaK(619-638) were transformed into DH5α cells, induced for protein expression, subjected to heat shock treatment or not, and crosslinked with 0.1% or 1% formaldehyde prior to lysis and pulldown. No stable, abundant interaction partners with GST-DnaK(619-638) could be visualized by Coomassie or silver staining above the background observed for GST alone.

CHAPTER 4

MECHANISTIC FEATURES OF LIGAND AND CO-CHAPERONE BINDING THAT REMODEL DNAK DOMAIN ASSEMBLY

Parts of this chapter are the result of collaboration with Eugenia Clérico, Anastasia Zhuravleva and Lila Gierasch (Smock and Gierasch 2009, Publication 3; Clérico *et al.* 2010, Publication 4).

4.1 Introduction

The transmission of signals between proteins is essential to life. In recent years, we have gained tremendous knowledge of the interacting networks that act as communication pathways for cellular signaling (*e.g.*, Yu *et al.* 2008), yet we know far less about how signals are passed from one component of a network to another.

The modular linkage of domains underlies the signaling networks of many protein systems (Pawson and Nash 2003). For example, the first PDZ domain of mouse phosphotyrosine phosphatase BL (which contains five PDZ domains) modulates the peptide affinity and specificity of its second PDZ domain through an interdomain interaction site (van den Berk *et al.* 2007). Similarly, the *Drosophila* cell polarity protein Par-6 is regulated by the Rho guanosine triphosphatase (GTPase) Cdc42 via binding of Cdc42 to a CRIB domain adjacent to the Par-6 PDZ domain. Cdc42-CRIB activation of the PDZ domain occurs with pronounced rigidification of CRIB and CRIB-PDZ contacts on the opposite side from the PDZ ligand-binding site, consistent with the signaling pathways observed for isolated PDZ domains. This indirect domain-domain rearrangement

changes the conformation of the PDZ domain, increasing its affinity for peptide and triggering subsequent cell polarity signaling (Peterson *et al.* 2004). Incoming signals can also promote assembly and disassembly of multiple domains. Src tyrosine kinase SH2 and SH3 domains undergo reversible assembly and disassembly in a dynamic equilibrium that is dependent in part on binding of a phosphotyrosine from the adjacent kinase domain. Binding of the phosphorylated kinase segment to the SH2 domain favors SH2-SH3 assembly and inactivates the kinase. When the kinase is activated (and the tyrosine is not phosphorylated), the SH2-SH3 tandem domains are uncoupled from one another. They nonetheless fluctuate between the assembled and disassembled arrangements (Faraldo-Gómez and Roux 2007).

Domain assembly and disassembly of Hsp70 proteins is captured to modulate interaction with numerous substrates. This mechanistic feature enables Hsp70 to remodel substrate conformation and activity, such as in the assistance of protein folding and regulation of factors that control cellular stress response and apoptosis (Ravagnan *et al.* 2001; Mayer and Bukau 2005). The Hsp70 SBD exhibits some chaperone activity as an isolated domain (Tanaka *et al.* 2002). A linked actin-like NBD and co-chaperones impart tunable substrate binding affinity through allosterically controlled cycles of interdomain interaction (Mayer and Bukau 2005).

A more complete picture of Hsp70 interdomain arrangement has begun to emerge in recent years using NMR. This method provides information on the chemical environment and dynamic nature of each residue in a protein

simultaneously and is therefore a powerful experimental technique for measuring conformational transitions. Analysis of residual dipolar coupling has allowed initial characterization of average domain orientations along two axes of rotation in the ADP-bound state of *T. thermophilus* DnaK (Revington *et al.* 2005). In the ADP-bound state of *E. coli* DnaK, the spectra of a two-domain construct overlaps with those of the individual domains (Swain *et al.* 2007), which facilitates transfer of chemical shift assignments to the two-domain protein (Bertelsen *et al.* 2009). Domains are observed to be flexibly independent modules in the ADP-bound state of DnaK and behave similarly in the absence of nucleotide (Fig. 4.1A) (Swain *et al.* 2007; Bertelsen *et al.* 2009). In contrast, an interdomain association is indicated in the absence of nucleotide for bovine Hsc70 (Jiang *et al.* 2007).

Prior work in our lab has described the ATP-induced conformational changes in the NBD as an isolated domain and in preliminary studies of a two-domain DnaK construct (Swain *et al.* 2007; Zhuravleva and Gierasch 2011). In particular, the ATP-bound NBD is required to elicit a complete conformational transition of the SBD, underscoring the importance of studying the linked properties of DnaK domains in this allosteric state. A crystal structure of a homologous Hsp110 protein bound to ATP likely informs on several structural features relevant to the ATP-bound, domain-docked state of Hsp70 (Fig. 4.1C) (Liu and Hendrickson 2007), while their functional divergence and a mechanistic basis for ATP-induced allostery (which is not present in Hsp110 proteins) are less clear. In Chapter 2, we describe a co-evolving network of residues that defines the Hsp70/110 separation and mediates allostery in Hsp70 proteins. The

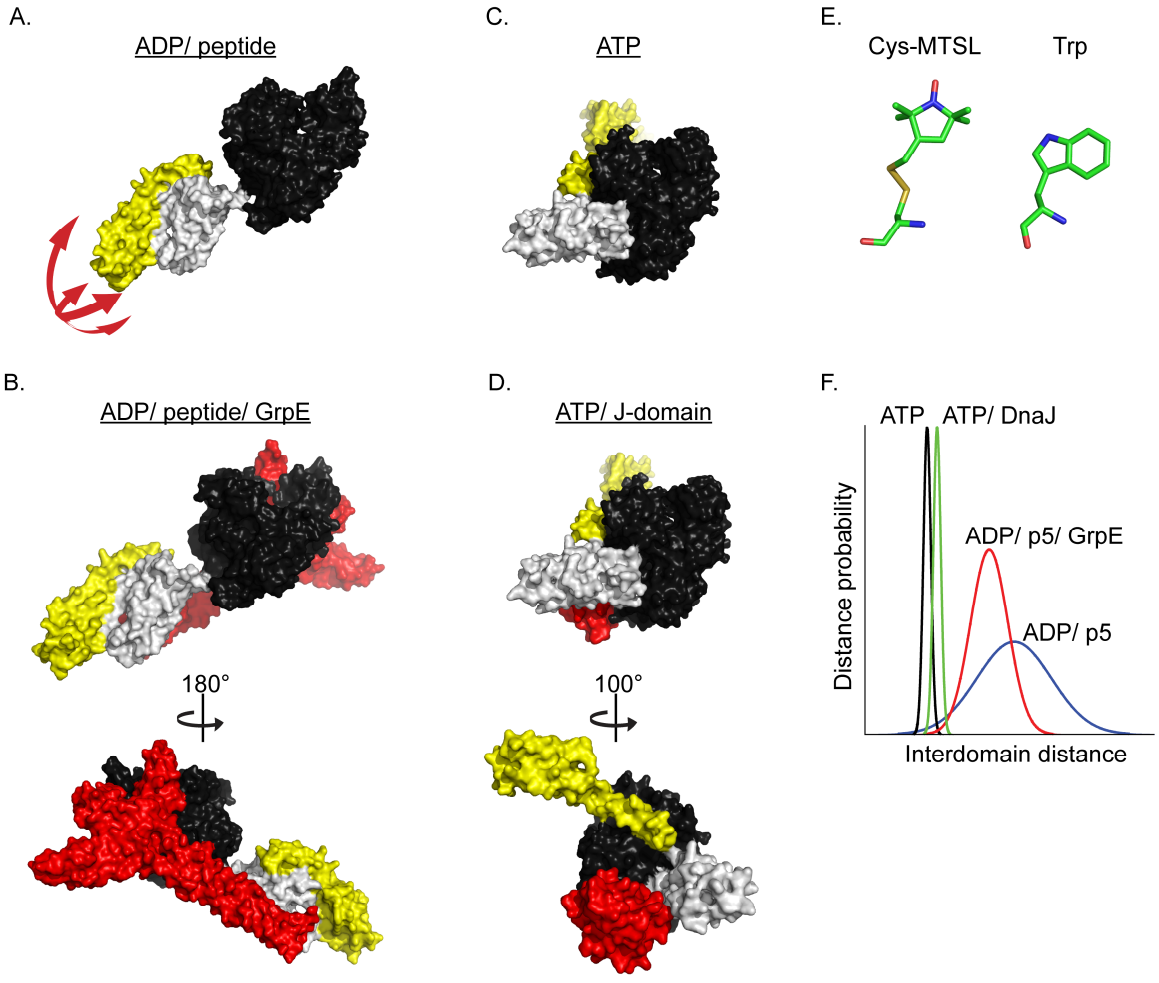
A.
ADP/ peptide
C.
ATP
E.
Cys-MTSL
Trp
B.
ADP/ peptide/ GrpE
180°
D.
ATP/ J-domain
100°
F.
ATP
ATP/ DnaJ
ADP/ p5/ GrpE
ADP/ p5
Distance probability
Interdomain distance

Figure 4.1. Structural models illustrate potential roles for ligands and co-chaperones in re-shaping DnaK domain arrangements.

A. Bound to ADP and peptide, the DnaK NBD (black) and SBD (white, β-sandwich subdomain; yellow, α-helical lid subdomain) are undocked and flexible with respect to one another (PDB 2KHO, an average DnaK domain orientation model based on residual dipolar couplings of the two-domain protein and isolated domain crystal structures PDB 1DKX and 1DKZ) (Zhu *et al.* 1996; Harrison *et al.* 1997; Bertelsen *et al.* 2009). **B.** Complex formation between the ADP/ peptide-bound NBD and the co-chaperone GrpE (red) promotes nucleotide exchange. This superposition of different NBD structural complexes suggests a restriction of SBD rotational freedom by GrpE. Additional interactions of GrpE with the SBD have been previously indicated and may further assist DnaK in its conversion to the ATP-bound state (superposition of PDB 2KHO, described in A, with PDB 1DKG, a crystal structure of the DnaK NBD bound to GrpE) (Harrison *et al.* 1997). **C.** ATP binding induces widespread conformational changes including interdomain docking, lid opening and substrate release (PDB 2QXL, a crystal structure of the homologous yeast Hsp110 protein Sse1 bound to ATP) (Liu and Hendrickson 2007). **D.** Complex formation of the ATP-bound NBD with a J-domain co-chaperone (red) stimulates ATP hydrolysis. In this superposition of different NBD structural complexes, the J-domain and SBD β-sandwich partially overlap, indicating a requirement for SBD/ J-domain re-positioning with tight physical association. Furthermore, the co-chaperone DnaJ encodes a J domain as well as a peptide binding domain and has been previously indicated to interact with the SBD. These data suggest that DnaJ may directly re-shape DnaK interdomain assembly to further assist in its converstion to the ADP/ peptide bound state (superposition of PDB 2QXL, described in C, with PDB 2QWR, a crystal structure of the bovine Hsc70 NBD bound to the J-domain of auxilin) (Jiang *et al.* 2007). **E.** The nitroxide derivative MTSL can be specifically attached to cysteine residues in proteins, giving an EPR-sensitive probe that is of the same size range as natural amino acids. **F.** With a Cys-MTSL residue in the NBD and another in the SBD β-sandwich, pulsed EPR data can characterize the reshaping of DnaK interdomain assemblies by ligands and co-chaperones as described in A-D, as well as describing sub-populations within these states (only single distributions are shown in these predictions).

structural and dynamic solution properties of domain-docked, ATP-bound Hsp70 can be characterized using NMR if chemical shifts can be assigned to specific residues. However, the large number and broadening of chemical shifts in this state (Swain *et al.* 2007) create significant overlap among them and ambiguity of assignment. To help overcome this technical challenge, a segmental isotopic labeling strategy is employed in this chapter to simplify NMR spectra and facilitate chemical shift assignment.

The arrangement of Hsp70 domains in various functional states can be observed directly using methods that characterize distance distributions between the domains. Interdomain distance distributions can be measured using single molecule Förster resonance energy transfer. In a recent study utilizing this technique, the yeast mitochondrial Hsp70, Sse1, was found to exist in a uniform, closed conformation in the ATP-bound state (Mapa *et al.* 2010). In the ADP-bound form, Sse1 domains were separated with a broad distance distribution in at least two conformations. Addition of the J-domain co-chaperone Mdj1 converted Sse1 to the ADP-bound state, presumably due to stimulated ATP hydrolysis during the measurement. Interestingly, the combined addition of Mdj1 with a peptide substrate allowed a more complete undocking of domains. Interdomain distances were also measured in *E. coli* DnaK, which contains a single intrinsic cysteine. Labeled double cysteine constructs were made in the context of a C15A mutation and different properties of the conformational ensembles of DnaK and Sse1 were interpreted as a species-specific tuning of chaperone function. However, DnaK constructs were not tested for their

maintenance of physiological function. In other reports, DnaK C15A exhibited a basal ATPase activity that is ~4-fold greater than wild-type DnaK (Linke 2005) and altered the redox-sensitive properties of the chaperone (Winter *et al.* 2005).

Alternatively, distance distributions can be measured between paramagnetic centers on protein domains, which may be introduced through the covalent attachment of a nitroxide spin label to cysteine residues (Fig. 4.1E) (Berliner 2000). Labeling proteins with nitroxides gives a probe that is sensitive to electron paramagnetic resonance (EPR) phenomena. Continuous wave EPR has captured the flexible nature of SBD subdomains in *E. coli* DnaK to accommodate substrates of various sizes (Schlecht *et al.* 2011). In pulsed EPR, carefully designed microwave pulse sequences extract the effect of interspin coupling in magnetic field re-alignment, which is proportional to the inverse cube of the distance between spin labels (Berliner 2000). These measurements provide the centers and widths of one or multiple distance distributions over a large distance range (Blackburn *et al.* 2009). In this chapter, a pulsed EPR investigation of domain assembly is initiated by creating a DnaK construct that is amenable to these measurements and retains proper function. This will allow the examination of conformational ensembles in different nucleotide states in a physiologically relevant context (Fig. 4.1A,C,F).

Moreover, the functional remodeling of DnaK domain architecture by co-chaperones has been suggested in a number of studies but is a poorly understood process. Nucleotide exchange factors and J-domain proteins have well-defined roles in their binding to the Hsp70 NBD and effect the nucleotide

state of the isolated domain, which in turn can modulate substrate binding in the SBD (Mayer and Bukau 2005). However, direct interactions of co-chaperones with the SBD have also been indicated. The nucleotide exchange factor of DnaK, GrpE, was observed to bind to the isolated DnaK SBD β-sandwich subdomain, with further involvement of the SBD α-helical subdomain in full-length protein. Truncation experiments indicated that this binding is mediated by an N-terminal segment in the helical tail of GrpE (Fig. 4.1B) (Chesnokova *et al.* 2003). Another study showed that the N-terminal tail of GrpE reduces proteolytic accessibility of the allosterically important interdomain linker of DnaK and suggested a role for the physical approximation of DnaK domains by GrpE (Moro *et al.* 2007).

The other major co-chaperone of DnaK, DnaJ, contains four domains. Only two domains (the J-domain and Gly/Phe-rich region) are essential for stimulation of NBD ATPase activity (Karzai and McMacken 1996). In a bovine chaperone system, the J-domain binding site on the NBD partially overlaps with the NBD-SBD interface and is suggested to disrupt Hsp70 interdomain assembly (Fig. 4.1D) (Jiang *et al.* 2007). Direct binding of DnaJ to the DnaK SBD has been suggested by several studies but remains poorly defined (Gassler *et al.* 1998; Suh *et al.* 1998; Suh *et al.* 1999). A zinc-binding domain of DnaJ binds to protein substrates, which is itself a chaperone function and further allows substrate transfer to DnaK. When a particular mutation is made in the zinc-binding domain, DnaJ retains substrate binding and DnaK ATPase stimulation but is impaired in its ability to promote DnaK state transition (Linke *et al.* 2003). These studies demonstrate that while GrpE and DnaJ perform criticial functions on the DnaK

NBD, their full functional profile must also address roles in direct SBD interaction and DnaK interdomain assembly.

The dynamic interaction between Hsp70 domains and their re-modelling by co-chaperones, as well as the size of the overall assemblies, impair their characterization by crystallography and NMR. Pulsed EPR is an excellent tool to describe DnaK domain architecture because it is not limited by the size of the system studied and describes full conformational ensembles. Moreover, the capability for rapid mixing and freezing of samples for pulsed EPR measurement will enable the novel characterization of transitional forms, such as encounter complexes formed with co-chaperones DnaJ and GrpE and direct roles for their functional re-distribution of DnaK domain assembly (Fig. 4.1F).

4.2 Results

4.2.1 Tolerance for the addtition of a polyhistidine tag and substitution of an intrinsic cysteine residue in DnaK

Developing an experimental capability to measure properties of interdomain assembly can be simplified using modified DnaK constructs. In particular, the use of a polyhistidine tag would facilitate purification of DnaK constructs, and C-terminally His-tagged construct of DnaK (C-His) has been utilized in previous reports (*e.g.*, Suh *et al.* 1998). Furthermoe, the sole intrinsic cysteine residue in DnaK, C15, may be accessible to the attachment of thiol-reactive labels in an undesirable side reaction. An effort was made to resolve this

potential complication by mutating the thiol moiety to a hydroxyl using the substitution C15S.

The functional properties of DnaK C-His and C15S were first evaluated using DnaK-dependent cellular assays. The requirement for DnaK in propagating the bacteriophage λ (Yochem *et al.* 1978; Osipiuk *et al.* 1993) can be observed using an *E. coli dnaK*-knockout strain that expresses plasmid-encoded wild type DnaK, the hydrolysis-defective mutant DnaK T199A (Buchberger *et al.* 1995), or no DnaK using an empty vector (Fig. 4.2A). Both DnaK C-His and C15S retain full phage propagation activity. In a second cellular assay, DnaK is required for *E. coli* to recover after being subjected to heat shock at 43 ℃ (Bukau and Walker 1990). In this assay, both DnaK C-His and C15S are unable to rescue *E. coli* growth defects and perform similarly to the DnaK null and T199A constructs (Fig. 4.2B). Finally, a third assay measures the requirement for functional DnaK upon depletion of the chaperones SecB and DnaJ (Fig. 3.9). DnaK C-His shows anomalous growth in an internal negative control, which renders interpretation of its growth in a test condition meaningless (Fig. 4.2C). DnaK C15S behaves normally in internal controls but is observed to be defective to a similar extent as DnaK null and T199A in its test condition. Together, these data indicate that both DnaK C-His and C15S deviate from the physiological function of the wild type molecule.

DnaK C-His and C15S were purified in order to examine their biophysical properties in more detail. The fluorescence of the sole intrinsic tryptophan residue in DnaK, W102, is diagnostic for ATP-dependent interdomain docking

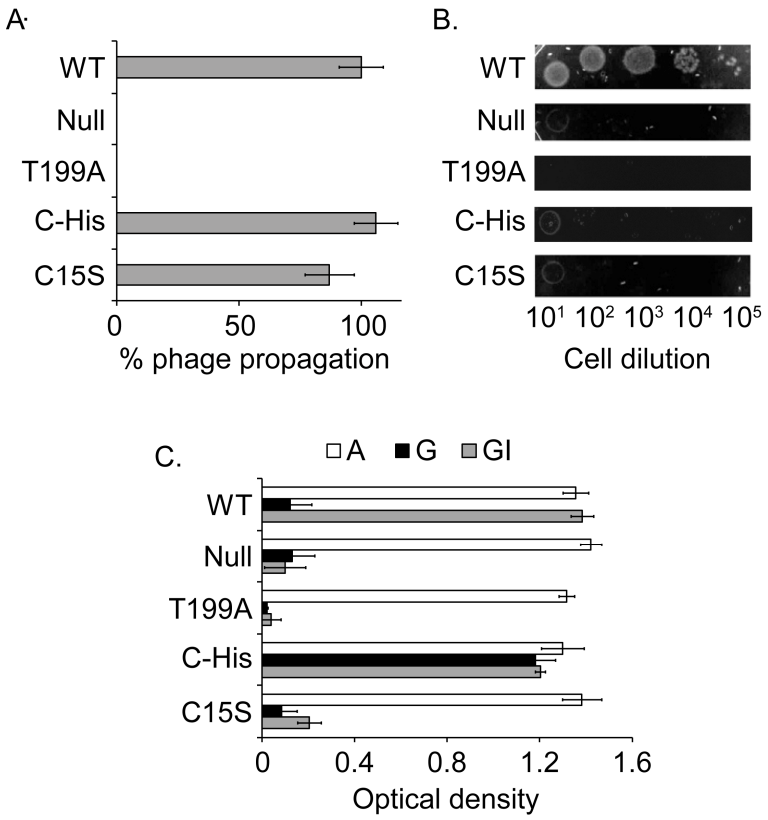

Figure 4.2. Intolerance for the addition of a polyhistidine tag to DnaK or substitution of an intrinsic cysteine in stringent cellular assays.

A variety of DnaK-dependent cellular assays indicate the tolerance of *E. coli* to DnaK mutation. DnaK T199A is a non-functional construct that cannot hydrolyze ATP effectively (Buchberger *et al.* 1995) and acts as a negative control. **A.** C-terminally His-tagged DnaK construct (C-His) and DnaK C15S are fully active in the propagation of phage λ at 30 ℃. **B.** Cells expressing either DnaK C-His or C15S are unable to recover after heat shock at 43 ℃. **C.** In a chaperone depletion assay, constructs are expected to show robust growth in media supplemented with arabinose (A, internal positive control), poor growth in media supplemented with glucose (G, internal negative control), and the rescue of growth by the indicated construct is tested in media supplemented with glucose and IPTG (GI) (Fig. 3.9). DnaK C-His shows reproducibly robust growth in its internal negative control, prohibiting interpretation of its functional effects in this assay. DnaK C15S shows normal behavior in the A and G control conditions and is unable to rescue growth in the experimental GI condition, indicating a functional deficiency.

(Moro *et al.* 2003). The wavelength of maximum W102 fluorescence for wild type DnaK undergoes a 4 nm transition upon addition of ATP (Table 4.1). However, the spectral peak of DnaK C-His in the absence of ATP is intermediate to the peaks observed in the wild type transition and undergoes little change upon addition of ATP. Therefore, the interdomain arrangement of DnaK C-His is perturbed and ATP-induced allostery is compromised. In contrast, the spectral peaks of DnaK C15S are equivalent to wild type in both the presence and absence of ATP, indicating normal ATP-induced interdomain allostery for this construct.

The ATPase activity of DnaK provides another measure of the functional properties of mutants. Wild type DnaK hydrolyzes ATP at an intrinsically low rate which is stimulated ~6-fold by the peptide substrate p5 (Fig. 4.3). The basal ATPase activity of DnaK C-His is elevated ~3-fold relative to the basal rate of the wild type protein and is not stimulated further by p5. DnaK C15S shows a basal ATPase rate that is reduced ~4-fold relative to the basal rate of wild type DnaK. The p5-stimulated rate of DnaK C15S is reduced by a similar extent, resulting in a fold stimulation that is within experimental error of the wild type protein. This indicates that while the intrinsic ATPase activity of DnaK C15S is perturbed, allosteric signal is communicated normally from the peptide-binding site in the SBD to the nucleotide-binding site in the NBD.

It can be concluded that DnaK C-His is an allosterically non-functional construct and should not be used for the study of physiologically relevant interdomain assembly of DnaK. This result is unexpected since truncation of the

	λ max (nm)		
	-ATP	+ATP	Intensity quench
Wild type	342	338	19%
C-His	340	339	n.d.
C15S	342	338	n.d.
T383C	342	338	19%

Table 4.1. DnaK W102 fluorescence as a probe for ATP-induced domain docking in various constructs.

The wavelength of maximum W102 fluorescence intensity plus or minus ATP describes a 4 nm blueshift in all constructs except C-His. Addition of ATP also quenches W102 intensity normally for T38C. n.d.= not determined.

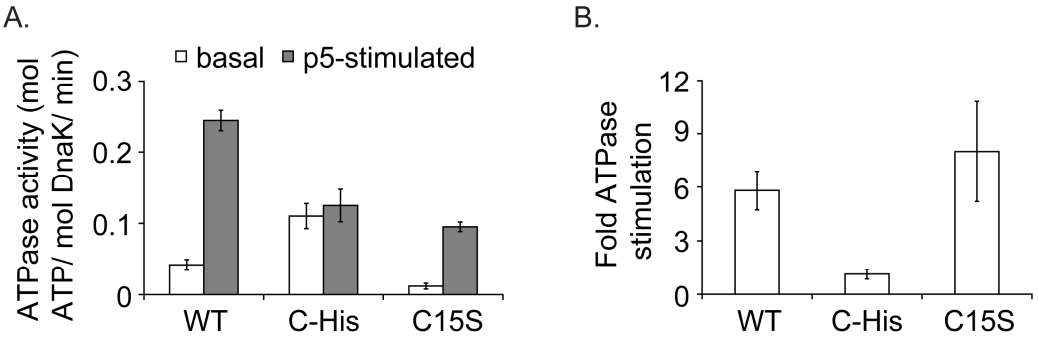

Figure 4.3. Basal and peptide-stimulated ATPase activities of C-terminally His-tagged DnaK and DnaK C15S.

A. The basal ATPase activity of DnaK C-His is elevated ~3-fold relative to wild type DnaK and that of DnaK C15S is reduced ~4-fold. **A,B.** The ATPase activity of DnaK C-His is not allosterically stimulated by the peptide substrate p5 while DnaK C15S shows normal allosteric function.

C-terminus does not alter intramolecular functions of DnaK (Chapter 3). The flexibility of the natural C-terminus and additional 15 residues present on DnaK C-His may allow for the extended sequence to interact with the NBD and stabilize interdomain interaction in a perturbed structural state. Attachment of a polyhistidine tag to the N-terminus of DnaK was not pursued because this region of the protein is adjacent to the interdomain interface and modification might disrupt allostery. Substitution of the single native Cys in DnaK to Ser preserves normal allosteric properties but results in a reduced basal ATPase activity and a functional defect *in vivo*. This effect might be rationalized by the absence of C15 from an evolutionary network associated with Hsp70 allostery (Chapter 2) while also having close physical proximity to the ATP binding site. A slight change in side chain packing around C15 is expected to alter the orientation of catalytic residues relative to ATP.

4.2.2 A functional cysteine mutation in the interdomain linker for segmental isotopic labeling of DnaK domains

The assignment of NMR chemical shifts to specific residues presents a significant challenge for the study of the ATP-bound, domain-docked conformation of DnaK. A complete NMR spectrum of a two-domain construct contains more chemical shifts than single domain constructs. Moreover, the intact molecule tumbles more slowly and signals are broadened, resulting in the overlap and ambiguous identification of chemical shifts.

Assignment of chemical shifts in a two-domain construct can be simplified by segmental isotopic labeling. In this approach, one domain is labeled with NMR-sensitive isotopes and the other without. The domains are then ligated so that the NMR signals of a single domain can be measured in the context of a two-domain protein. This treatment reduces the number of chemical shifts in the spectrum and the degree of overlap among them in order to facilitate their assignment to specific residues. Ligation is achieved by expressing the N-terminal domain as a fusion with an intein, a protein that naturally functions in protein splicing. Self-cleavage of the intein can then be induced to generate a reactive α-thioester on the domain's C-terminus (Xu *et al.* 1999). The thioester selectively reacts with an N-terminal cysteine of the second domain. The resulting linkage undergoes spontaneous rearrangement with the cysteine's α-amino group to generate a native peptide bond (Dawson *et al.* 1994).

Application of this technique to the study of allostery in DnaK depends on the identification of a functional cysteine substitution within the solvent-exposed interdomain linker. However, the interdomain linker is highly conserved in Hsp70 proteins and plays a critical role in allosteric signaling (Vogel *et al.* 2006; Swain *et al.* 2007). Candidate sites for cysteine mutation in the linker were selected based on their relatively low conservation, and wherever possible, natural tolerance for substitution by cysteine or serine in an Hsp70 multiple sequence alignment.

Linker cysteine mutations were initially introduced into DnaK C-His (prior to the characterization of its allosteric defects) and the resulting constructs were

subjected to a phage λ propagation assay (Fig. 4.4A). Two of the five linker cysteine mutations (DnaK T383C C-His and DnaK G379C C-His) support λ replication at comparable levels to wild type protein. The three sites that do not allow normal λ replication (G380C C-His, D385C C-His and K387 C-His) show no natural frequency for Cys substitution and very little Ser in the sequence alignment. In contrast, the amino acid distribution at site G379 includes 5% Cys. However, this site is partially embedded in NBD structure, suggesting that its use in domain linkage could be problematic. At position T383, serine is the most conserved amino acid (33% serine), indicating a high likelihood for tolerance of cysteine at this site in more stringent functional tests. Therefore, the T383C mutation was introduced in a wild type background and subjected to additional functional screens. DnaK T383C behaves identically to wild type protein in both a phage propagation assay (Fig. 4.4A) and heat shock assay (Fig. 4.4B).

DnaK T383C was purified in order to characterize its capacity for allosteric signaling. The W102 fluorescence spectrum of DnaK T383C undergoes a 4 nm blue shift and 19% intensity quench upon addition of ATP that is characteristic of wild type protein (Table 4.1). Furthermore, its basal and p5-stimulated ATPase activities are indistinguishable from wild type DnaK (Fig. 4.5). Together, these data demonstrate that the T383C mutation does not interfere with allosteric conformational transitions in DnaK. Initial work labeling T383C with the fluorophor IAEDANS indicated that fluorophor attachment perturbed allostery (data not shown).

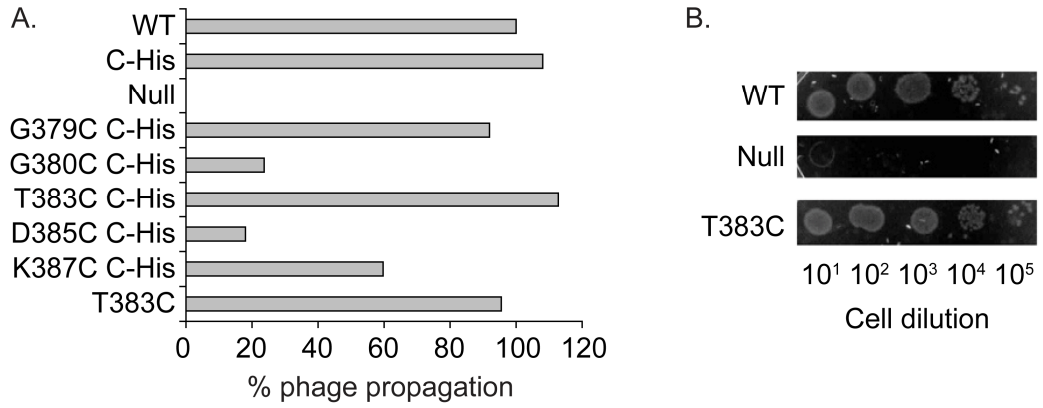

Figure 4.4. An *in vivo* screen for functional cysteine substitutions in the DnaK interdomain linker.

A. Cysteine mutations in the interdomain linker were initially introduced in a C-terminally His-tagged DnaK construct. *E. coli* cells expressing DnaK G379C and T383C were able to propagate phage λ to a similar extent as wild type protein. The T383C mutation was then introduced into an untagged DnaK construct and was determined to be fully functional by both the phage propagation and **B**. heat shock assays.

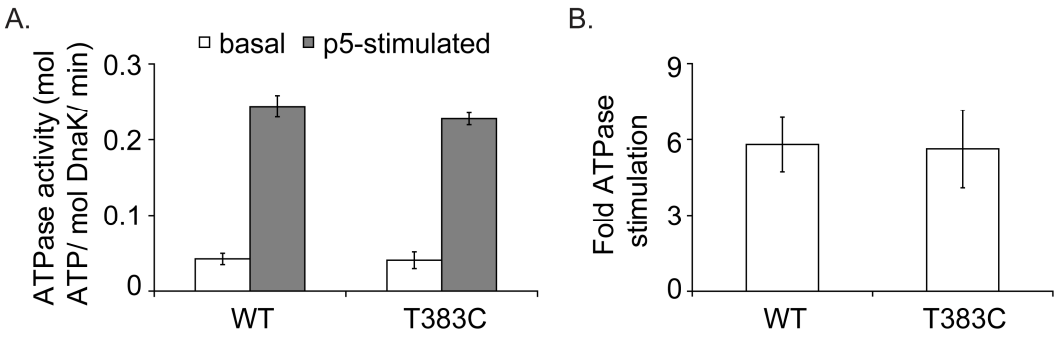

Figure 4.5. DnaK T383C has normal ATPase activity and allosteric stimulation by peptide.

A. The ATPase activities of wild type DnaK and DnaK T383C are within experimental error of one another. **A,B.** The allosteric stimulation of ATPase activity by the peptide substrate p5 is normal in DnaK T383C.

Eugenia Clérico and Anastasia Zhuravleva, postdoctoral fellows in our lab, continued to develop the application of this cysteine substitution to segmental isotopic labeling. Prior work demonstrated that a two-domain DnaK construct, DnaK(1–552)ye T199A, is amenable to the measurement of allosteric states using NMR (Swain *et al.* 2007). Briefly, efforts to ligate DnaK 1-382 T199A (the NBD) and DnaK (383-552)ye T383C (the SBD) using the intein fusion strategy indicated incomplete accessibility of the generated thioester to solvent. Ligation efficiency was optimized using a low concentration of urea to more readily expose the thioester to solvent as well as using high concentrations of domains to increase the probability of their intermolecular collision. The successfully ligated product retained allosteric function, as determined by W102 fluorescence and peptide-stimulated ATPase activity. Finally, ligation of the isotopically labeled NBD with the unlabeled SBD provided significant simplification of NMR spectra compared to a fully labeled two-domain construct. This treatment facilitated the assignment of chemical shifts to specific residues and will assist in the experimentally difficult characterization of the ATP-bound, domain-docked state of DnaK.

4.2.3 Suitability of DnaK for pulsed EPR studies

Pulsed EPR studies are typically performed on protein samples containing 30% glycerol (*e.g.*, Blackburn *et al.* 2009). Glycerol acts as a cryoprotectant and extends the transverse magnetization decay of EPR signals to a more easily measured time regime. However, a high concentration of glycerol may alter the

assembly of DnaK domains and the interpretation of pulsed EPR data. The effect of pulsed EPR sample conditions on DnaK domain assembly was investigated by measuring W102 fluorescence in the presence of high concentrations of glycerol. The spectra of wild type DnaK in the presence or absence of ATP overlap when 0%, 10% or 20% glycerol is included in sample buffer (Fig. 4.6A). Initial measurements of W102 fluorescence in the presence of 30% glycerol indicated a normal spectrum in the absence of ATP but only a partial transition to the ATP state (~2 nm blue shift and incomplete intensity quench). However, ongoing blue shift and intensity quench were observed as a function of time, with an additional ~2 nm blue shift and full conversion to the ATP-bound state after approximately 20 minutes (Fig. 4.6B). These data suggest that high concentrations of glycerol can affect the timescale of interstate conversion but do not affect the conformations of allosteric end states in DnaK. Therefore, the use of 30% glycerol is compatible with the physiologically meaningful conformational ensembles of DnaK in different allosteric states as long as samples are given sufficient time between the addition of ATP and sample freezing.

The application of pulsed EPR to DnaK requires selective labeling of nitroxide derivatives to cysteine residues at designed sites. However, DnaK contains a single intrinsic cysteine, C15, whose mutation alters basal properties of DnaK without affecting the efficiency of allosteric signal transmission (Table 4.1 and Fig. 4.3). An alternative strategy was considered to ensure that spurious data of a functionally compromised construct of DnaK would not be mis-interpreted. Since C15 is largely buried in NBD structure, its accessibility to

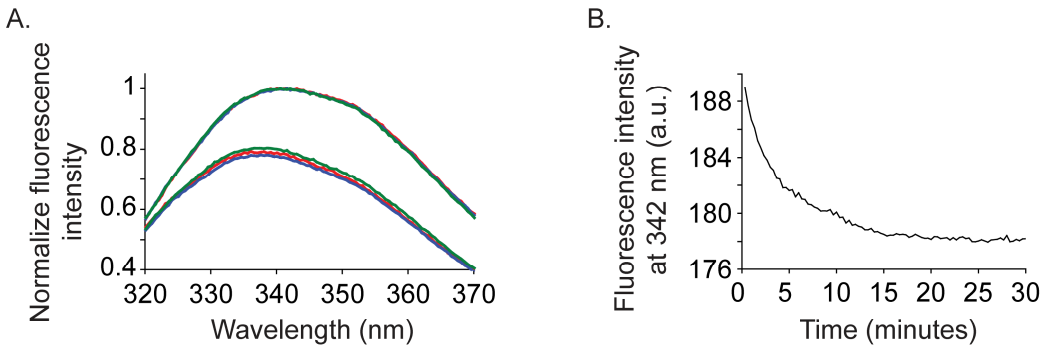

Figure 4.6. The presence of glycerol does not disrupt interdomain arrangements in DnaK.

A. The W102 fluorescence spectra of wild type DnaK in the absence of glycerol (red) or in the presence of 10% (blue) or 20% (green) glycerol overlap in either the absence (upper traces) or presence (lower traces) of ATP. **B.** In the presence of 30% glycerol, the W102 spectrum was normal in the absence of ATP but showed only a partial transition to the ATP state. However, a full conversion was observed after approximately 20 minutes, as indicated by the ATP-induced quench of W102 fluorescence intensity plotted as a function of time.

reactive labels in the solvent was tested. Wild type DnaK was incubated with a 5 kDa derivative of poly-ethylene glycol (PEG) linked to a thiol-reactive maleimide. The size-dependent mobility of proteins on a polyacrylamide gel indicates that DnaK is partially labeled by PEG-maleimide. However, no labeling is observed in the presence of ATP (Fig. 4.7). These data indicate that conformational transitions of DnaK alter the exposure of C15 to solvent and that the labeling of this site can be blocked in the ATP-bound state.

In contrast, cysteine residues introduced on domain surfaces are expected to be fully accessible to reactive labels in the solvent. In order to capture physiologically relevant events with pulsed EPR, cysteine mutation must preserve the structural and functional properties of DnaK. Candidate sites for mutation were designed based on their relatively low conservation in the Hsp70 family, their lack of involvement in evolutionary sectors (Chapter 2), and their predicted solvent accessibility in both the ADP-bound and ATP-bound conformations of DnaK using available crystal structures. A variety of sites that meet these criteria were used to create single cysteine mutation constructs (Fig. 4.8). Collectively, they are distributed across domain surfaces to allow for flexibility in the design of double cysteine mutation constructs.

Single cysteine mutation constructs were screened for function using cellular assays. Initially, cysteines were introduced into DnaK C15S and tested for their ability to propagate phage λ. This assay is not sensitive to the C15S mutation (Fig. 4.2). Seven cysteine mutations reduce λ replication efficiency

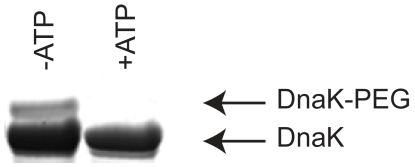

Figure 4.7. The sole intrinsic cysteine residue of DnaK becomes inaccessible to covalent modification in the presence of ATP.

A 5 kDa polyethylene glycol (PEG) derivative linked to a thiol-reactive maleimide group can be covalently attached to solvent-accessible cysteine residues of proteins. Attachment is observed as a shift in protein mobility in SDS-PAGE. The reaction of PEG with the native C15 residue of wild type DnaK indicates slight labeling in the absence of ATP and no labeling in the presence of ATP.

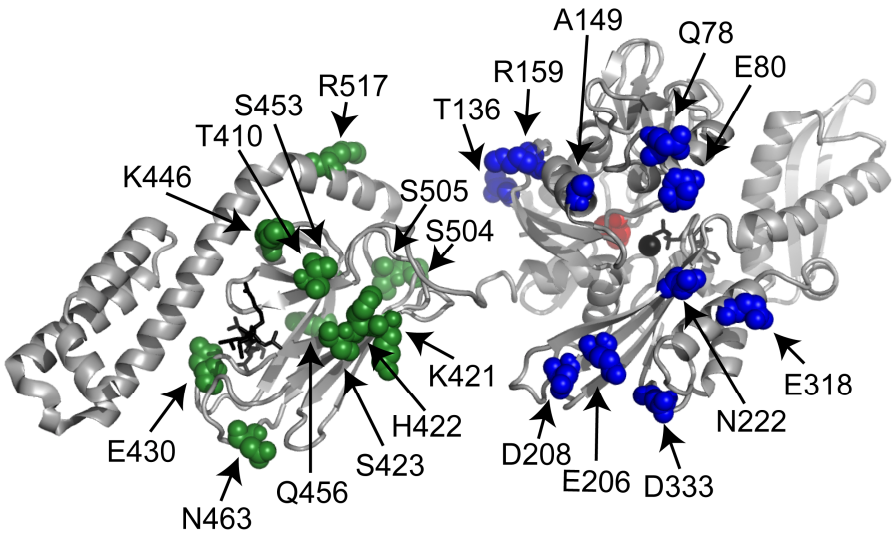

Figure 4.8. Structural locations of cysteine mutation sites.

A number of DnaK constructs were created with single cysteine mutations at sites distributed over DnaK domain surfaces.

while eight cysteine mutations are functional (Table 4.2). In subsequent work, cysteine mutations were introduced into a wild type DnaK background and screened for function using heat shock and chaperone depletion assays. In these more stringent assays for DnaK function, all but one of the designed constructs are fully functional (Fig. 4.9). In total, eleven unique cysteine mutations are functional in the heat shock and chaperone depletion assays and an additional two unique cysteine mutations are functional in the phage assay.

4.2.4 Selective nitroxide labeling of a double cysteine mutation construct

Functional cysteine mutations from each domain were combined to create a double cysteine mutation construct, DnaK T136C K421C. The distance between the β-carbons of T136C and K421C are predicted to be 33 Å in the ADP-bound state (using PDB 2KHO) and 27 Å in the ATP-bound state (using a homology model based on PDB 2QXL). The difference in distances between attached nitroxide labels in these states as well as their full dynamical distributions can be resolved by pulsed EPR in this distance regime (Berliner 2000).

A prior experiment indicated that attachment of the native C15 residue to a 5 kDa PEG-maleimide label could be blocked in the presence of ATP (Fig. 4.7). This premise was tested using the smaller label MTSL, an EPR-sensitive nitroxide linked to a thiol-reactive methanesulfonothioate group. Wild type DnaK was reacted with MTSL in both the presence and absence of ATP. After

Phage λ Propagation		
+++ (>75%)	**++ (25-75%)**	**+ (<25%)**
WT	Q78C C15S	Null
C15S	R159C C15S	E206C C15S
A149C C15S	S453C C15S	D208C C15S
D333C C15S		N222C C15S
K421C C15S		S505C C15S
H422C C15S		
E430C C15S		
K446C C15S		
S504C C15S		
R517C C15S		

Table 4.2. Activity of cysteine-substituted DnaK constructs in a phage propagation assay.

Cysteine mutations were initially introduced in a C15S background and screened for their ability to propagate phage λ in *E. coli*.

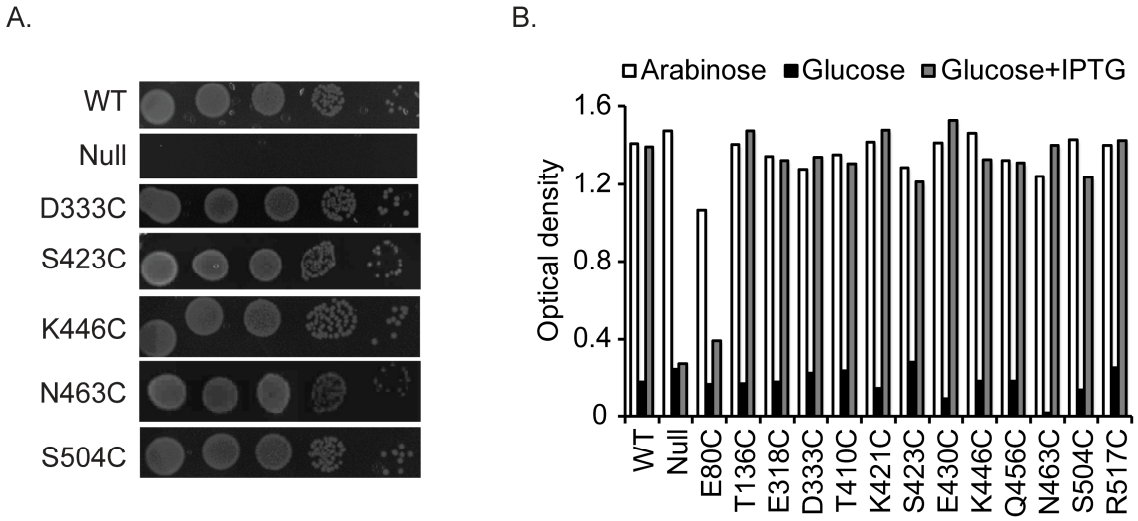

Figure 4.9. Functional screens for single cysteine mutation constructs *in vivo*.

Cysteine mutation constructs in a native C15 background are screened by **A,** a heat shock assay, and **B,** a chaperone depletion assay. With the exception of DnaK E80C, all of the constructs tested are fully functional.

removing unreacted MTSL, the labeling efficiency can be estimated by measuring an EPR spectrum. EPR signal intensities were calibrated to nitroxide concentration by measuring spectra of the free nitroxide TEMPO at known concentrations (Fig. 4.10A). The line shape of free nitroxide is composed of three symmetric and narrow peaks. Calibration is performed by calculating the double integral of fit EPR spectra, which is linearly proportional to the concentration of nitroxide. Intensity is calibrated in this way because line shape is sensitive to the chemical environment of the paramagnetic center as well as its rate of molecular tumbling, and is therefore expected to change upon attachment to DnaK. The EPR spectrum of wild type DnaK labeled in the absence of ATP shows a small but measurable signal (Fig. 4.10B) with labeling efficiency estimated at 0.06 MTSL:DnaK (molar ratio). Labeling in the presence of ATP produces a spectrum that falls within the noise of signal intensity with labeling efficiency of less than 0.02 MTSL:DnaK.

These data corroborate preliminary findings from the PEG-maleimide experiment and indicate that C15 labeling by MTSL is effectively blocked in the ATP-bound state of DnaK. DnaK T136C K421C was labeled under the same conditions in the presence of ATP. The line shape of its EPR spectrum is composed of peaks that are asymmetric from one another and are significantly broadened (Fig. 4.10C) relative to the line shape of free nitroxide (Fig. 4.10A), ensuring that the observed signal results from nitroxide that is attached to protein. Labeling efficiency is estimated at 2.2 MTSL:DnaK. Given the

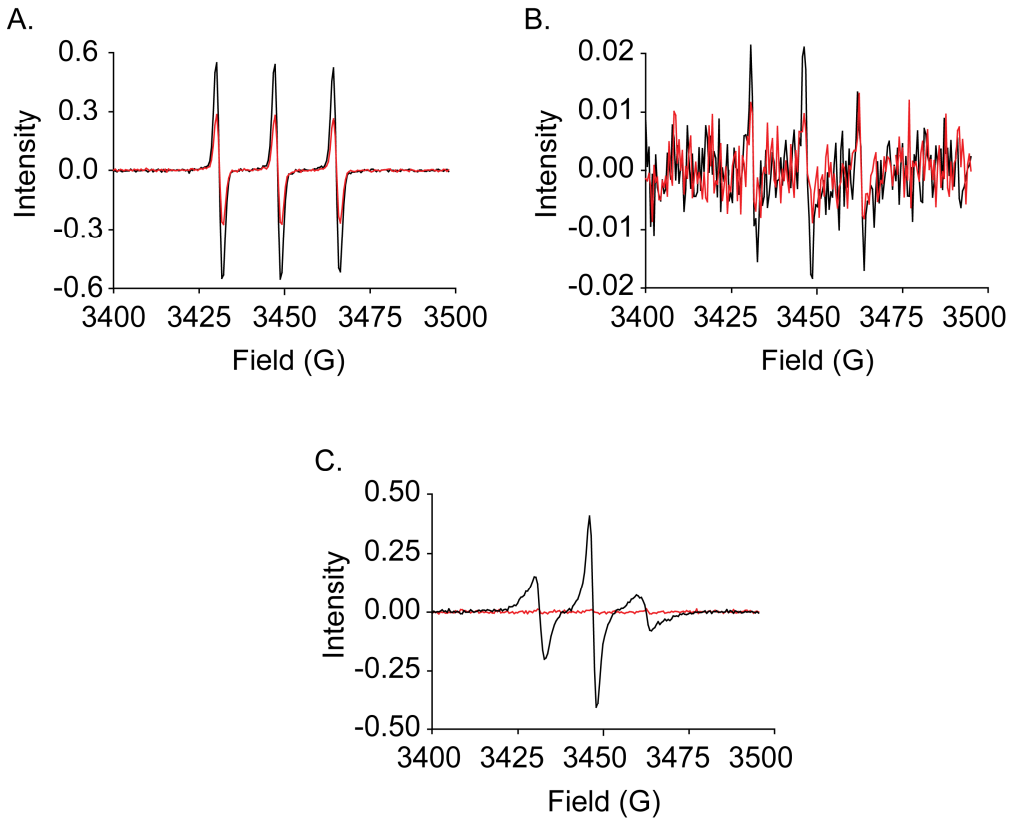

Figure 4.10. Efficient and selective labeling of DnaK T136C K421C with the nitroxide derivative MTSL.

A. EPR spectra of the free nitroxide radical TEMPO are shown at 25 μM (red) and 50 μM (black). Signal intensity is linearly proportional to concentration and allows for the calibration of signal intensities of samples at unknown nitroxide concentrations. **B.** EPR spectra of wild type DnaK (containing a single intrinsic cysteine at residue C15) are shown for samples labeled with the thiol-reactive nitroxide derivative MTSL in the presence (red) or absence (black) of ATP. The vertical axis is shown on a much smaller scale relative to the other panels. **C.** An EPR spectrum is shown for DnaK T136C K421C reacted with MTSL in the presence of ATP (black). The lineshape of this spectrum can be clearly distinguished from the EPR spectra of free nitroxide in A, indicating attachment of the label to DnaK. The EPR spectrum of wild type DnaK labeled under the same conditions is also shown (red) and is the same spectrum shown in B for comparison on this vertical axis scale.

experimental errors associated with estimating nitroxide and protein concentrations (perhaps 20% total error), as well as the labeling of this sample under conditions that block access to C15, it can be concluded that the reaction of MTSL with ATP-bound DnaK T136C K421C is both selective and essentially complete. Additional tests for labeling efficiency can also be employed in future efforts, such as sample denaturation followed by PEGylation. This analysis is expected to give a single high intensity band in SDS-PAGE corresponding to the single attachment of PEG to DnaK C15. Nonetheless, absolute completion of MTSL labeling is not necessary to obtain pulsed EPR data.

Biophysical properties of DnaK T136C K421C $MTSL_2$ were measured to ensure that the double cysteine mutation and attached labels do not interfere with the proper function of the molecule. The W102 fluorescence spectra of unlabeled (Fig. 4.11A) and labeled (Fig. 4.11B) mutation construct overlap with wild type DnaK in the presence and absence of ATP, indicating the correct interdomain arrangement in different allosteric states. The basal and p5-stimulated ATPase activities of these constructs are also indistinguishable from wild type DnaK (Fig. 4.11C), demonstrating that mutation and labeling do not interfere with intrinsic or ligand-induced allosteric functions. Similarly, the stimulation of ATPase activity by co-chaperones DnaJ and GrpE observed for wild type is fully preserved in DnaK T136C K421C $MTSL_2$ (Fig. 4.11D). These data underscore the utility of this construct for characterizing the conformational ensembles of DnaK domains in various co-chaperone and ligand-bound states.

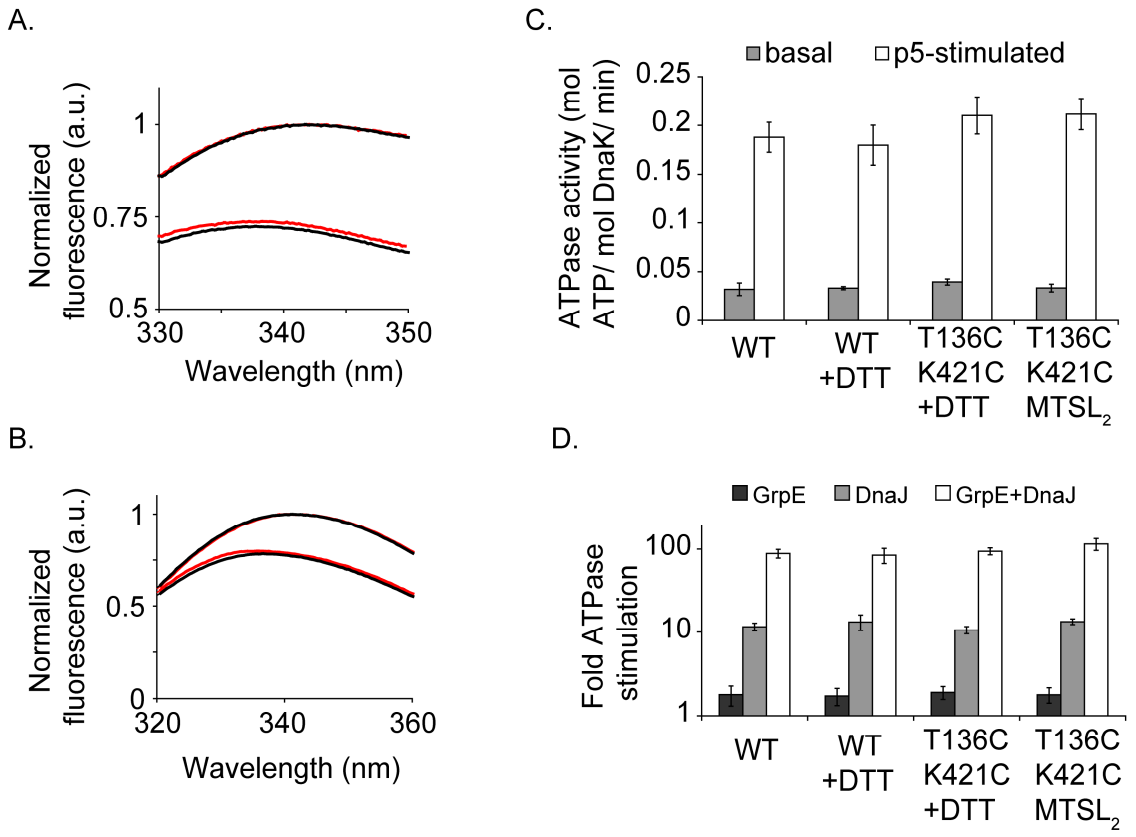

Figure 4.11. DnaK T136C K421C $MTSL_2$ is fully functional in interdomain allostery and co-chaperone interaction.

A. The W102 fluorescence spectra of DnaK T136C K421C (red) is indistinguishable from wild type DnaK (black) in the absence (upper traces) and presence (lower traces) of ATP. **B.** W102 fluorescence spectra are also normal for DnaK T136C K421C $MTSL_2$ (red) relative to wild type DnaK (black). **C.** The labeled and unlabeled constructs of DnaK T136C K421C show normal basal ATPase activity and allosteric stimulation by p5. **D.** The labeled and unlabeled constructs of DnaK T136C K421C undergo normal ATPase stimulation by the co-chaperones GrpE and DnaJ.

4.3 Discussion

In this chapter, domain labeling strategies were employed using constructs that retain physiological function of DnaK. It was found that in cellular assays, *E. coli* does not tolerate substitution in DnaK for a native cysteine residue or addition of a polyhistidine tag. Purified proteins also had compromised functional properties *in vitro*. Therefore, subsequent cysteine mutations were introduced into a native background, and numerous sites that tolerate substitution to cysteine were identified using stringent cellular assays for DnaK function (Table 4.3). These cysteine mutation sites identified here can be used for a variety of purposes.

Segmentally labeled DnaK domains and double nitroxide contructs are being used in ongoing NMR and EPR studies to bolster our understanding of DnaK conformational transitions and the mechanism of chaperone function. Extending our knowledge of molecular biophysics to the cell can include increasingly complex systems such as chaperone assemblies. Pulsed EPR will assess a capacity for GrpE and DnaJ to directly remodel DnaK assembly. Looking beyond a two-state model of Hsp70 function, this remodeling can serve to assist DnaK domains in their transitions between ATP-induced docking and ADP-induced undocking events by providing partial domain transition prior to ATP hydrolysis or nucleotide exchange (Fig. 4.1F). Initial encounter complexes between DnaK and GrpE or DnaJ can be captured through the use of a T199A mutation, saturating ligand, and a rapid mixing/ freezing device. Pulsed EPR data will also describe sub-populations within each state. A recent computational

	Assay:	**Phage λ**	**Heat shock**	**Chaperone depletion**
	Stringency:	+	++	+++
	Genetic background:	C15S	WT	WT
	Mutation			
Nucleotide Binding Domain	Q78C	-		
	E80C			-
	T136C			+
	A149C	+		
	R159C	-		
	E206C	-		
	D208C	-		
	N222C	-		
	E318C			+
	D333C	+	+	+
Substrate Binding Domain	T410C			+
	K421C	+		+
	H422C	+		
	S423C		+	+
	E430C	+		
	K446C	+	+	+
	S453C	-		
	Q456C			+
	N463C		+	+
	S504C	+	+	+
	S505C	-		
	R517C	+		+

Table 4.3. Summary of the *in vivo* function of cysteine-substituted DnaK constructs.

Cellular assay screens for cysteine mutations distributed on domain surfaces are summarized (Table 4.2, Fig. 4.9). A plus sign indicates full activity relative to wild type DnaK and a minus sign symbolizes any measurable defect.

study has predicted that intrinsic motions of the NBD provide a conformation that GrpE can bind and stabilize (Liu *et al.* 2010). Similarly, it will be interesting to see if a given state of domain assembly includes sub-populations that facilitate conversion to alternate allosteric states.

4.4 Methods

4.4.1 *In vivo* assays

Phage propagation and heat shock assays were performed as previously described using *E. coli* strain BB1553 transformed with plasmid pMS119-*dnaK* or pWCS57-*dnaK(C-His)* (Bukau and Walker 1990; Clérico *et al.* 2010). In the chaperone depletion assay, *E. coli* strain GP502 (Mph42 Δ*secB*, P_{BAD}-*dnaKJ*, Kan^R, *araC*) (Ullers *et al.* 2007) was transformed with plasmid pMS119-*dnaK*. Resulting colonies were used to inoculate overnight cultures grown at 30 ℃ in LB media supplemented with 0.5% arabinose, 50 mg/L ampicillin and 50 mg/L kanamycin. Optical density at 600 nm (OD_{600}) was normalized to 0.4 by dilution with LB media for each growth. 10 µL of each growth was diluted in 1 ml of water and 5 µl was used to inoculate 1.5 ml of LB media supplemented with antibiotics and either 0.5% arabinose, 0.5% glucose or 0.5% glucose plus 50 µM IPTG. This concentration of IPTG gave maximal growth recovery of strain GP502 transformed with pMS119-*dnaK* (wild type) in the presence of 0.5% glucose. Growths were carried out at 30 ℃ in 15 ml falcon tube s secured on their sides in a shaker to ensure efficient aeration. OD_{600} was measured after twenty hours.

4.4.2 Purification of proteins and peptides

E. coli DnaK was prepared as previously described (Smock *et al.* 2010) using *E. coli* strain BB1553 (MC4100 Δ*dnak52*::Cam^R, *sidB1*) (Bukau and Walker 1990) transformed with plasmid pMS119-*dnaK* (p*tac*, pBR322 *ori,* Amp^R) (Montgomery *et al.* 1999). Plasmids pMS119-*dnaK(S423C)* and pMS119-*dnaK(N463C)* were provided by Sigurd Wilbanks (University of Otago). Proteins containing solvent-exposed cysteine residues were stored in the presence of 100 µM TCEP. DnaK with an eleven-residue C-terminal sequence extension, RRASVHHHHHH, was carried on plasmid pWCS57 (p*trc*, colE1 *ori,* Amp^R) and was purified as previously described using *E. coli* strain BL21(DE3) (Suh *et al.* 1999). Compared to plasmid pMS119 used in other cellular assays, pWCS57 has a similar low copy colE1-like origin of replication and expresses protein at similar levels with a closely related IPTG-inducible promoter. *E. coli* GrpE was purified by utilizing a cleavable His tag (Chang *et al.* 2010), which was removed by treatment with His-tagged S219V TEV protease (Kapust *et al.* 2001). Concentration was estimated in a Bradford assay using BSA as a standard. *E. coli* DnaJ was purified using plasmid p29SEN-*dnaJ* (p*trc*, Amp^R) (Ullers *et al.* 2007) following established protocols (Chang *et al.* 2008). Concentration was calculated from absorbance at 280 nm using an extinction coefficient of 13,400 M^{-1} cm^{-1}. The peptide p5 (CLLLSAPRR) was prepared as previously described (Smock *et al.* 2010).

4.4.3 *In vitro* measurements

Steady-state ATPase activity was measured at 30 ℃ using an enzyme-coupled system and a Biotek Synergy2 microplate reader as previously described (Smock *et al.* 2010). One µM DnaK was used for all ATPase measurements in the presence or absence of 100 µM p5, 1 µM DnaJ and 1 µM GrpE dimer. DnaK W102 fluorescence emission spectra were measured as previously described using a Photon Technology International fluorometer (Smock *et al.* 2010). The contribution of photobleaching to ATP-induced quenching of W102 fluorescence varied between some measurements due to differences in the scan parameters utilized. However, scan parameters were identical among spectra shown on the same plot.

Wild type DnaK was labeled with PEG-maleimide (Nektar) by first mixing 10 µM DnaK and 50 µM DTT, with or without 5 mM ATP, in buffer containing HMK buffer (20 mM Hepes, 10 mM $MgCl_2$ and 100 mM KCl) at pH 7. PEG-maleimide was then added to a final concentration of 2 mM. Samples were incubated for a half hour at room temperature and quenched with excess DTT. Wild type DnaK and DnaK T136C K421C were labeled with MTSL (Toronto Research Chemicals) by first incubating 50 µM DnaK and 10 mM β-mercaptoethanol in HMK buffer at pH 7.6 for an hour at 4 ℃. Protein samples were then desalted with four cycles of 30-fold dilution in HMK buffer at pH 7.6 and re-concentration. ATP was added to a concentration of 5 mM or omitted, followed by the addition of 1 mM MTSL. Samples were incubated at room temperature for four hours and desalted with seven cycles of five-fold dilution in

HMK buffer at pH 7.6 and re-concentration. The concentrations of labeled DnaK samples were estimated by their SDS-PAGE band intensities and the use of a calibration curve constructed from the band intensities of known concentrations of unlabeled DnaK on the same gel. EPR was measured using a Bruker ELEXSYS E-500 X-band spectrometer. Each sample was measured using the same scan parameters with spectral averaging over 20 scans. Total signal intensities were estimated by fitting simulated spectra to EPR data (software provided by Christian Altenbach, UCLA) and calculating integrals of the fit curves (Origin software).

CHAPTER 5

CONCLUSIONS

5.1 Summary

Allosteric signal transmission between structural domains forms an essential component of Hsp70 chaperone function. A network of Hsp70 residues that mediates allostery has been identified by performing statistical analysis of genetic data. In this analysis, position and sequence correlations were simultaneously evaluated to extract sites that define the divergence of the allosteric Hsp70 and non-allosteric Hsp110 sub-families. The resulting evolutionary sector was observed to join the distal substrate and nucleotide-binding sites through domain interaction surfaces. A basis for the sector in interdomain allostery was further validated through molecular dynamics-based homology modeling, site-directed mutagenesis, cellular fitness assays, and biophysical and enzymological experimentation. This work sheds mechanistic insight into features of allostery that are shared throughout the Hsp70 family. The evolutionary constraints shared between Hsp70 domains also highlights the potential for observing similar interactions across protein-protein interfaces.

The variable and disordered C-terminus is not an evolutionary conserved sequence feature of Hsp70s, yet it exists in some form throughout the protein family. Surprisingly, a conserved C-terminal sequence motif was discovered in a large and diverse subset of Hsp70s of bacterial origin, including *E. coli* DnaK. A cellular assay with greater dependence on DnaK function was developed by

depleting compensatory chaperones. This assay was utilized to reveal a site-specific correspondence of DnaK function *in vivo* to the identified conservational pattern. Consistent with our prior results, the C-terminus does not affect interdomain allostery, as determined by the characterization of purified mutation constructs. The C-terminus is also dispensable for interaction with co-chaperones, but does enhance the refolding of a denatured DnaK substrate, indicating a previously unidentified mode of chaperone action. We therefore propose a model in which the C-terminus transiently interacts with client proteins to maintain a high local concentration and disrupt misfolded structures.

Ligand-induced allostery and co-chaperone binding are interacting components of the Hsp70 chaperone machine. Several domain labeling studies were initiated to investigate the influences of ligands and co-chaperones on conformational transitions in DnaK. Using stringent cellular assays as a reliable functional screen, a variety of sites were identified that tolerate mutation to cysteine. One functional cysteine mutation in the interdomain linker established the groundwork for segmental isotopic labeling of DnaK domains, which has facilitated NMR chemical shift assignments of DnaK in its ATP-bound, domain-docked state. Furthermore, a double cysteine construct labeled with an EPR-sensitive probe was found to retain physiological functions. The use of this construct in pulsed EPR experiments will allow the measurement of interdomain distance distributions in various states. In particular, the capacity of co-chaperones to directly remodel DnaK domain assembly can be tested using this strategy.

5.2 Implications for future research

The allosteric network identified in our analysis of position correlations can be systematically assessed by analyzing NMR chemical shift perturbations in different ligand-bound states of two-domain DnaK. Anastasia Zhuravleva, a postdoctoral fellow in our lab, is performing these analyses, and has successfully assigned chemical shifts for the ADP-bound state of DnaK. Chemical shift assignments for the ATP-bound state of DnaK are underway. Assignments have been facilitated using the isotopically labeled ligated domain construct that was created by Eugenia Clérico, another postdoctoral fellow, and was made with the functional T383C mutation reported in this thesis. This work will provide the first protein-wide and site-specific experimental characterization of an Hsp70 protein in its domain-docked allosteric state. It will be interesting to examine the basis of the Hsp70 interdomain evolutionary sector with regard to specific roles in nucleotide-induced roles conformational changes and dynamic processes in solution. Furthermore, our collaborators at the University of Texas Southwestern Medical Center, who developed the correlated mutations analysis, are currently examining a role for sectors shared across protein-protein interfaces.

Mandy Blackburn, a postdoctoral fellow in our lab, is performing an EPR study to test our model for DnaK C-terminal substrate interaction. Her results indicate a loss of mobility for the C-terminus upon binding to a misfolded protein substrate, in agreement with our model (Fig. 3.10). A more comprehensive evaluation of substrate interaction could be performed using differentially labeled

complexes of misfolded susbtrate and the DnaK SBD, which on its own is sufficient to bind substrates with high affinity (Tanaka *et al.* 2002). For example, the isotopically labeled SBD (*e.g.*, residues 395-638) is expected to give especially sharp, high intensity peaks for sites in the disordered C-terminus. Peaks that broaden or shift upon binding to a misfolded protein substrate could indicate a role for specific C-terminal residues in substrate binding. Performing the same experiment with a peptide substrate would control for chemical shift perturbations associated with binding to just the canonical SBD β-sandwich binding pocket, as would the binding of a misfolded substrate to a C-terminally truncated SBD.

The SBD β-sandwich is able to bind to misfolded substrates in a general way through non-specific hydrophobicity and charge recognition (Rudiger *et al.* 1997). It follows that the C-terminus may also interact with substrates through a similarly generic mechanism and disorder in this segment may provide the capacity for variable multivalent binding (Smock and Gierasch 2009). This hypothesis could be tested by using a variety of misfolded substrates to determine any differences in the mode of C-terminal binding. In a converse labeling scheme, a substrate can be isotopically labeled while the SBD is not in order to identify binding sites on the substrate with which the C-terminus interacts.

In order to further determine the role of the C-terminus in substrate folding efficiency, an overarching experimental design can be envisioned in which a relatively small protein or protein fragment is first selected on the basis of

extensive prior knowledge of its folding mechanism. The substrate is then subjected to a refolding assay in the presence or absence of the DnaK chaperone system with or without C-terminal mutation. A candidate substrate for further study would be observed to show enhanced refolding in the presence of the DnaK C-terminus, as was observed for denatured luciferase in this thesis. However, the experimental determination of folding status would need to be tailored to the substrate studied in this experiment. For example, the tryptophan fluorescence signature of a substrate's folding status might be extracted from background signals in a refolding reaction, since the entire DnaK-DnaJ-GrpE chaperone system has only two tryptophan residues in total. The substrate could then be isotopically labeled for NMR analysis and its chemical shift perturbations analyzed in the presence or absence of SBD C-terminal modification. One might then relate the binding interactions by the DnaK C-terminus to the observed effect on refolding efficiency through prior knowledge of its folding mechanism. Our current model predicts that the C-terminus transiently interacts with substrate residues responsible for the formation of an off-pathway, kinetically trapped intermediate.

A role for C-terminal disruption of misfolded structure may also affect the ability of DnaK to cooperate with chaperone networks in the cell. A requirement for DnaK has been implicated in an early stage of the disaggregation mechanism of the Hsp100 ClpB (Weibezahn *et al.* 2004). Our model for DnaK C-terminal function suggests that mutation of this region could impair the DnaK-mediated presentation of unfolded substrates to ClpB.

The diversification of C-terminal motifs in bacteria and eukaryotes may be incomplete and retain vestiges of functional similarity. While the C-terminal sequences of *E. coli* DnaK and human Hsp70 fall into different conservation groups, they do share an EEV tripeptide with each other, suggesting that their divergence is not absolute. The C-terminus of human Hsp70 motif mediates binding with TPR domain proteins and also plays a role in the intramolecular functions of the protein (Freeman *et al.* 1995). In DnaK, mutation of the C-terminus does not affect intramolecular function and we did not observe any obvious interaction partners in pulldown assays using a fusion of GST with a DnaK C-terminal peptide. However, *E. coli* YbbN has been identified as a DnaK-interacting chaperone with an atypical TPR domain (Kthiri *et al.* 2008; Lin and Wilson 2011). Pulldown assays can be performed again but probed for this specific interaction using an antibody directed against YbbN, which may be present in low abudance.

In addition to examining a C-terminal TPR-binding function in *E. coli* DnaK, our model for C-terminal DnaK interaction with substrates might be extended to human Hsp70. Other roles for the human Hsp70 C-terminus complicate, but do not exclude, the possibility that our DnaK C-terminal model may also apply to the human protein. This premise can be tested with a C-terminal sequence swapping experiment. A truncation of the entire DnaK C-terminus is an appropriate control, and the Hsp70 C-terminus can be appended onto this construct. Relative to the truncated construct, the C-terminally swapped variant could show enhanced luciferase refolding activity that is comparable to

native DnaK. This result would indicate that the human Hsp70 C-terminus can play a similar role as the DnaK C-terminus, in addition to its more specialized functions.

Cellular screens have identified a number of cysteine substitutions that are tolerated on DnaK domain surfaces. These may be labeled with fluorophores, nitroxide derivatives, crosslinkers or any other thiol-reactive chemical groups for a variety of applications. The labeling of cysteines with nitroxide for pulsed EPR provides a powerful general tool to measure distance distributions between the NBD, SBD β-sandwich and SBD α-helical lid using a variety of cysteine combinations and the labeling strategy described in this thesis. Moreover, the states of domain assembly being studied can be expanded. For example, at elevated temperature, the oxidation-sensitive role of DnaK C15 on refolding activity and its alteration by mutation at this site (Winter *et al.* 2005) can be examined in the context of interdomain assembly. A general redox-sensitive role for cysteine residues in proteins (Barford 2004) may manifest in DnaK as a response to oxidative heat stress through NBD deactivation (Winter *et al.* 2005) and domain uncoupling to promote a persistent high substrate affinity state. This internal regulatory component would allow DnaK to favor its role in substrate protection from degradation under conditions that may not permit spontaneous refolding upon release from the chaperone.

APPENDIX

HSP70 SECTORS AND PSEUDO-SECTORS

Groups of residues in proteins can evolve cooperatively to form functional units called sectors. In Chapter 2, a single sector is identified that mediates interdomain allostery in Hsp70, based on the joint analysis of differential sequence and position correlations among allosteric Hsp70 and non-allosteric Hsp110 proteins. While this allows unambiguous assignment of the sector to a particular functional role, the same group of residues can also be identified using prior versions of SCA (Estabrook *et al.* 2005; Russ *et al.* 2005; Socolich *et al.* 2005). Here, SCA is applied to 739 Hsp70 sequences (and no Hsp110 sequences) and all of the resulting position correlation groups are investigated. This approach identifies the same interdomain allosteric sector as well as an intradomain sector within just the NBD. In addition, a sector is identified that results from sequence differences in the phlogenetic divergence of eukaryotes and prokaryotes and does not represent unifying aspects of Hsp70 structure or function.

Sectors are identified in Hsp70 through multiple steps. Raw correlated mutations are first calculated, indicating that they are sparse and distributed throughout Hsp70 primary sequence (Fig. A.1A). The position correlation matrix of Hsp70 is cleaned to remove phylogenetic noise (Fig. A.1B) (Halabi *et al.* 2009). The matrix is then subjected to principal component analysis and the highest eigenvalue components are plotted on orthogonal axes to separate

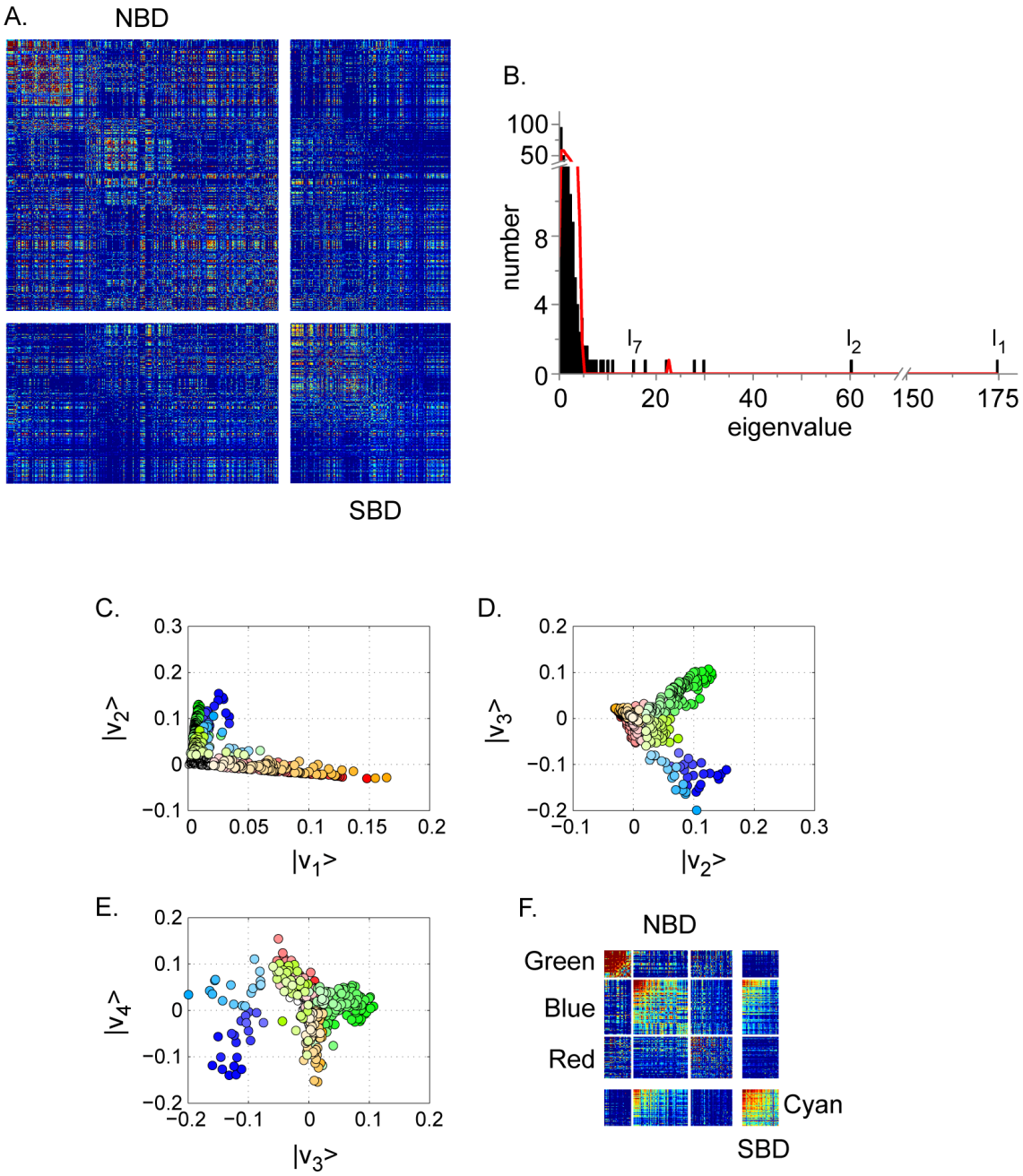
A.
NBD
SBD
B.
100
50
8
4
0
number
l_7
l_2
l_1
0
20
40
60
150
175
eigenvalue
C.
$|v_2\rangle$
$|v_1\rangle$
D.
$|v_3\rangle$
$|v_2\rangle$
E.
$|v_4\rangle$
$|v_3\rangle$
F.
NBD
Green
Blue
Red
Cyan
SBD

Figure A1. Sector identification in an Hsp70 sequence alignment.

A. The preliminary, unordered SCA matrix. Both axes are represented by Hsp70 positions in linear order from the N- to C-terminus of the protein. Each pixel represents the correlation between two positions with larger magnitudes mapped by hotter colors. **B.** Eigenvalue histogram of the SCA matrix for the NBD (black) shows that the bulk of eigenvalues are small and within random expectation. The red curve is the eigenvalue spectrum for 100 trials of randomizing the alignment. Thus, these small eigenvalues could be accounted for by just the finite size of the alignment. The SCA matrix is cleaned by throwing out all but eigenmodes 2-7, recomputing the SCA matrix for just these modes, and then recalculating eigenvalues and eigenvectors. **C.** Residue weights along the top few eigenvectors of the cleaned matrix. This shows that there is a Red sector that comes out in the top mode, that the Blue/Cyan and Green sectors emerge together along the second mode, **D.** the Blue/Cyan and Green sectors are split by the third mode, **E.** and the Blue and Cyan sectors can be separated by the fourth mode. **F.** The SCA matrix from A is ordered by the identified sectors.

coupling signals and define sectors, giving essentially the same result as that obtained from hierarchical clustering (Fig. A.1C-F). Three sectors in the NBD are labeled Blue, Green and Red. A single sector is found in the SBD and is labeled Cyan.

Phylogenetic bias in sectors is determined by first performing principal component analysis on sequence correlations to separate the alignment into sequence sub-families. The contribution of each amino acid in each sequence to position correlations is then displayed in a matrix that is ordered by sequence correlation groups (phylogenetic sub-families) along one axis and position correlations groups (sectors) along the other axis. The Red NBD sector shows coupling that originates from non-uniform sampling among these sequence sub-families (Fig. A.2). The origins of the Red sector are obscure. For a subset of residues in this sector, a function can be attributed to interaction with non-homologous NEFs. However, most residues are distributed throughout structure in a seemingly random way. The Green, Blue and Cyan sectors are different as they are composed of correlated mutations that are sampled throughout phylogeny.

The Green and Blue sectors in the NBD comprise 26% and 10% of the domain's residues, respectively. The Cyan sector in the SBD comprises 20% of this domain's residues. Strikingly, the Blue and Cyan sectors display strong coupling, which reflects evolutionary linkages of residues between the two domains. Painting these three sectors structures of individual Hsp70 domains reveals that each sector forms a structurally contiguous unit. Moreover, each

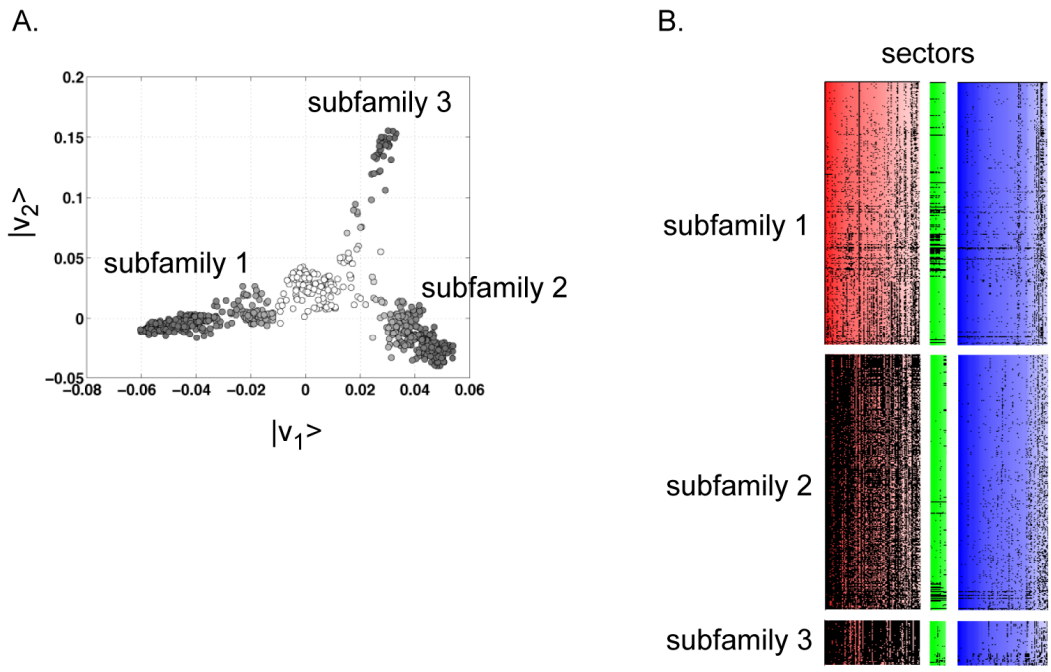

Figure A2. Real sectors and pseudo-sectors.

A. The two top principal components of sequence correlations in the Hsp70 alignment. This illustrates the presence of a very strong phylogenetic classification of sequences into approximately three subfamilies. The two main subfamilies represent the prokaryotes and eukaryotes. **B.** Recognition of the role of historical inheritance in the Red sector with non-uniform sampling in the alignment. Along the vertical axis, sequences are ordered by the three subfamilies. Along the horizontal axis, positions are ordered by sectors. A pixel is shaded according to how the corresponding amino acid from the alignment strengthens correlations in the sector.

connects a ligand-binding site to a distal region of the domain (Fig. A.3). The Blue sector in the NBD and Cyan sector in the SBD each terminate at large surface-exposed regions near the point of attachment of the interdomain linker—they are essentially equivalent to the single interdomain sector identified in Chapter 2 that underlies interdomain allostery. The two sectors in the NBD (Blue and Green), which are weakly coupled to each other, are linked by their common connection to the nucleotide-binding site. A more complete functional interpretation of the Green sector is not yet clear from existing experimental evidence.

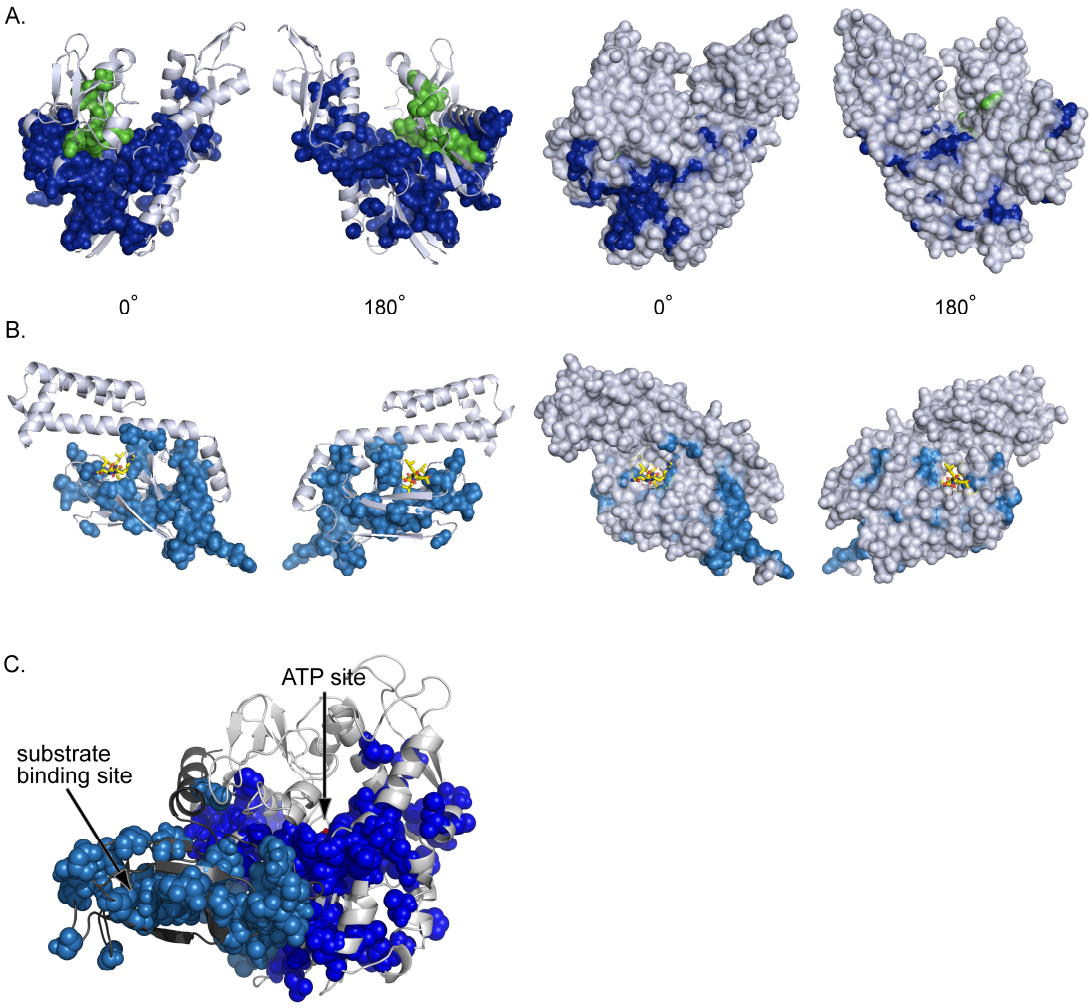

Figure A3. Sectors form structural networks in Hsp70 domains.

A. The Blue and Green sectors each form physically connected networks in the NBD. The Green sector is fully buried in subdomain 1B, and the Blue sector is made up of residues in all subdomains (mostly 1A) that connect the ATP binding site to a surface patch. **B.** The Cyan sector forms a physically connected network in the SBD that links the substrate binding site to the interdomain linker through residues in the β sheet. **C.** Surface patches presented by the Blue and Cyan sectors are joined through interdomain packing.

BIBLIOGRAPHY

Publication

1 Smock RG and Gierasch LM (2005). Finding the fittest fold: using the evolutionary record to design new proteins. *Cell* 122: 832-4.

2 Marcelino AC, Smock RG and Gierasch LM (2006). Evolutionary coupling of structural and functional sequence information in the intracellular lipid-binding protein family. *Proteins* 63: 373-84.

3 Smock RG and Gierasch LM (2009). Sending signals dynamically. *Science* 324: 198-203.

4 Clérico EM, Zhuravleva A, Smock RG and Gierasch LM (2010). Segmental isotopic labeling of the Hsp70 molecular chaperone DnaK using expressed protein ligation. *Biopolymers* 94: 742-52.

5 Smock RG, Rivoire O, Russ WP, Swain JF, Leibler S, Ranganathan R and Gierasch LM (2010). An interdomain sector mediating allostery in Hsp70 molecular chaperones. *Mol Sys Bio* 6: 414.

6 Smock RG, Blackburn ME and Gierasch LM (2011). The conserved, disordered C terminus of DnaK enhances cellular survival upon stress and *in vitro* chaperone activity. *J Biol Chem*, in press.

7 Buchner J, Smock RG, Gierasch LM, Saibil HR and Horwich AL (2011). Chaperones and protein folding (book chapter) in Comprehensive Biophysics, ed. Daggett V, in press.

Altman E, Kumamoto CA and Emr SD (1991). Heat-shock proteins can substitute for SecB function during protein export in Escherichia coli. *EMBO J* **10**: 239-45.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W and Lipman DJ (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-402.

Andreasson C, Fiaux J, Rampelt H, Druffel-Augustin S and Bukau B (2008). Insights into the structural dynamics of the Hsp110-Hsp70 interaction reveal the mechanism for nucleotide exchange activity. *Proc Natl Acad Sci U S A* **105**: 16519-24.

Anfinsen CB (1973). Principles that govern the folding of protein chains. *Science* **181**: 223-30.

Aponte RA, Zimmermann S and Reinstein J (2010). Directed evolution of the DnaK chaperone: mutations in the lid domain result in enhanced chaperone activity. *J Mol Biol* **399**: 154-67.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL and Mori H (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**: 2006 0008.

Barford D (2004). The role of cysteine residues as redox-sensitive regulatory switches. *Curr Opin Struct Biol* **14**: 679-86.

Berliner LJ, Eaton SS and Eaton GR, eds. (2000). *Distance measurements in biological systems by EPR*. New York, Kluwer Academic/ Plenum Publishers.

Bertelsen EB, Chang L, Gestwicki JE and Zuiderweg ERP (2009). Solution conformation of wild-type E. coli Hsp70 (DnaK) chaperone complexed with ADP and substrate. *Proc Natl Acad Sci U S A* **106**: 8471-6.

Bhattacharya A, Kurochkin AV, Yip GNB, Zhang Y, Bertelsen EB and Zuiderweg ERP (2009). Allostery in Hsp70 chaperones is transduced by subdomain rotations. *J Mol Biol* **388**: 475-90.

Blackburn ME, Veloro AM and Fanucci GE (2009). Monitoring inhibitor-induced conformational population shifts in HIV-1 protease by pulsed EPR spectroscopy. *Biochemistry* **48**: 8765-7.

Bork P, Sander C and Valencia A (1992). An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. *Proc Natl Acad Sci U S A* **89**: 7290-4.

Brehmer D, Rudiger S, Gassler CS, Klostermeier D, Packschies L, Reinstein J, Mayer MP and Bukau B (2001). Tuning of chaperone activity of Hsp70 proteins by modulation of nucleotide exchange. *Nat Struct Biol* **8**: 427-32.

Buchberger A, Theyssen H, Schroder H, McCarty JS, Virgallita G, Milkereit P, Reinstein J and Bukau B (1995). Nucleotide-induced conformational changes in the ATPase and substrate binding domains of the DnaK chaperone provide evidence for interdomain communication. *J Biol Chem* **270**: 16903-10.

Buchner J, Smock RG, Gierasch LM, Saibil HR and Horwich AL (in press). Chaperones and protein folding. *Comprehensive Biophysics*, Daggett V, Elsevier.

Bukau B, Deuerling E, Pfund C and Craig E (2000). Getting newly synthesized proteins into shape. *Cell* **101**: 119-22.

Bukau B and Walker GC (1990). Mutations altering heat shock specific subunit of RNA polymerase suppress major cellular defects of E. coli mutants lacking the DnaK chaperone. *EMBO J* **9**: 4027-36.

Burkholder WF, Zhao X, Zhu X, Hendrickson WA, Gragerov A and Gottesman ME (1996). Mutations in the C-terminal fragment of DnaK affecting peptide binding. *Proc Natl Acad Sci U S A* **93**: 10632-7.

Chang L, Bertelsen EB, Wisen S, Larsen EM, Zuiderweg ERP and Gestwicki JE (2008). High-throughput screen for small molecules that modulate the ATPase activity of the molecular chaperone DnaK. *Anal Biochem* **372**: 167-76.

Chang L, Thompson AD, Ung P, Carlson HA and Gestwicki JE (2010). Mutagenesis reveals the complex relationships between ATPase rate and the chaperone activities of Escherichia coli heat shock protein 70 (Hsp70/DnaK). *J Biol Chem* **285**: 21282-91.

Chang Y-W, Sun Y-J, Wang C and Hsiao C-D (2008). Crystal structures of the 70-kDa heat shock proteins in domain disjoining conformation. *J Biol Chem* **283**: 15502-11.

Chesnokova LS, Slepenkov SV, Protasevich II, Sehorn MG, Brouillette CG and Witt SN (2003). Deletion of DnaK's lid strengthens binding to the nucleotide exchange factor, GrpE: A kinetic and thermodynamic analysis. *Biochemistry* **42**: 9028-40.

Clark PL, Liu ZP, Rizo J and Gierasch LM (1997). Cavity formation before stable hydrogen bonding in the folding of a beta-clam protein. *Nat Struct Biol* **11**: 883-6.

Classic Protocol (2004). Detection of protein-protein interactions using the GST fusion protein pull-down technique. *Nat Methods* **1**: 275-6.

Clérico EM, Zhuravleva A, Smock RG and Gierasch LM (2010). Segmental isotopic labeling of the Hsp70 molecular chaperone DnaK using expressed protein ligation. *Biopolymers* **94**: 742-52.

Colombo G, Morra G, Meli M and Verkhivker G (2008). Understanding ligand-based modulation of the Hsp90 molecular chaperone dynamics at atomic resolution. *Proc Natl Acad Sci U S A* **105**: 7976-81.

Crick F (1970). Central dogma of molecular biology. *Nature* **227**: 561-3.

Crooks G, Hon G, Chandonia J and Brenner S (2004). WebLogo: a sequence logo generator. *Genome Res* **14**: 1188-90.

Cupp-Vickery JR, Peterson JC, Ta DT and Vickery LE (2004). Crystal structure of the molecular chaperone HscA substrate binding domain complexed with the IscU recognition peptide ELPPVKIHC. *J Mol Biol* **342**: 1265-78.

Cyr DM, Lu XY and Douglas MG (1992). Regulation of Hsp70 function by a eukaryotic DnaJ homolog. *J Biol Chem* **267**: 20927-31.

Davis JE, Voisine C and Craig EA (1999). Intragenic suppressors of Hsp70 mutants: interplay between the ATPase- and peptide-binding domains. *Proc Natl Acad Sci U S A* **96**: 9269-76.

Dawson P, Muir T, Clark-Lewis I and Kent S (1994). Synthesis of proteins by native chemical ligation. *Science* **266**: 776-9.

De Los Rios P, Ben-Zvi A, Slutsky O, Azem A and Goloubinoff P (2006). Hsp70 chaperones accelerate protein translocation and the unfolding of stable protein aggregates by entropic pulling. *Proc Natl Acad Sci U S A* **103**: 6166-71.

Deuerling E, Schulze-Specking A, Tomoyasu T, Mogk A and Bukau B (1999). Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature* **400**: 693-6.

Dominguez R (2004). Actin-binding proteins - a unifying hypothesis. *Trends Biochem Sci* **29**: 572-8.

Doolittle RF, Abelson JN and Simon MI (1996). Computer methods for macromolecular sequence analysis *Methods Enzymol*, ed. San Diego, Academic Press. **266:** 497-598.

Dosztányi Z, Mészáros B and Simon I (2009). ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* **25**: 2745-6.

Dunker A, Brown C and Obradovic Z (2002). Identification and functions of usefully disordered proteins. *Adv Protein Chem* **62**: 25-49.

Dunker A and Uversky V (2008). Signal transduction via unstructured protein conduits. *Nat Chem Biol* **4**: 229-30.

Dyson HJ and Wright PE (2005). Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**: 197-208.

Easton DP, Kaneko Y and Subjeck JR (2000). The Hsp110 and Grp170 stress proteins: newly recognized relatives of the Hsp70s. *Cell Stress Chaperones* **5**: 276-90.

Estabrook RA, Luo J, Purdy MM, Sharma V, Weakliem P, Bruice TC and Reich NO (2005). Statistical coevolution analysis and molecular dynamics: Identification of amino acid pairs essential for catalysis. *Proc Natl Acad Sci U S A* **102**: 994-9.

Faraldo-Gómez JD and Roux B (2007). On the importance of a funneled energy landscape for the assembly and regulation of multidomain Src tyrosine kinases. *Proc Natl Acad Sc U S A* **104**: 13643-8.

Fedorov AN and Baldwin TO (1997). GroE modulates kinetic partitioning of folding intermediates between alternative states to maximize the yield of biologically active protein. *J Mol Biol* **268**: 712-23.

Flaherty KM, DeLuca-Flaherty C and McKay DB (1990). Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. *Nature* **346**: 623-8.

Flaherty KM, McKay DB, Kabsch W and Holmes KC (1991). Similarity of the three-dimensional structures of actin and the ATPase fragment of a 70-kDa heat shock cognate protein. *Proc Natl Acad Sci U S A* **88**: 5041-5.

Flaherty KM, Wilbanks SM, DeLuca-Flaherty C and McKay DB (1994). Structural basis of the 70-kilodalton heat shock cognate protein ATP hydrolytic activity. II. Structure of the active site with ADP or ATP bound to wild type and mutant ATPase fragment. *J Biol Chem* **269**: 12899-907.

Flynn GC, Chappell TG and Rothman JE (1989). Peptide binding and release by proteins implicated as catalysts of protein assembly. *Science* **245**: 385-90.

Freeman BC, Myers MP, Schumacher R and Morimoto RI (1995). Identification of a regulatory motif in Hsp70 that affects ATPase activity, substrate binding and interaction with HDJ-1. *EMBO J* **14**: 2281-92.

Gamer J, Multhaup G, Tomoyasu T, McCarty JS, Rüdiger S, Schönfeld HJ, Schirra C, Bujard H and Bukau B (1996). A cycle of binding and release of the DnaK, DnaJ and GrpE chaperones regulates activity of the Escherichia coli heat shock transcription factor sigma32. *EMBO J* **15**: 607-17.

Gassler CS, Buchberger A, Laufen T, Mayer MP, Schroder H, Valencia A and Bukau B (1998). Mutations in the DnaK chaperone affecting interaction with the DnaJ cochaperone. *Proc Natl Acad Sci U S A* **95**: 15229-34.

Genest O, Hoskins JR, Camberg JL, Doyle SM and Wickner S (2011). Heat shock protein 90 from Escherichia coli collaborates with the DnaK chaperone system in client protein remodeling. *Proc Natl Acad Sc U S A* in press.

Gershenson A and Gierasch LM (2011). Protein folding in the cell: challenges and progress. *Curr Opin Struct Biol* **21**: 32-41.

Goloubinoff P and De Los Rios P (2007). The mechanism of Hsp70 chaperones: (entropic) pulling the models together. *Trends Biochem Sci* **32**: 372-80.

Gragerov A, Nudler E, Komissarova N, Gaitanaris GA, Gottesman ME and Nikiforov V (1992). Cooperation of GroEL/GroES and DnaK/DnaJ heat shock proteins in preventing protein misfolding in Escherichia coli. *Proc Natl Acad Sci U S A* **89**: 10341-4.

Gunasekaran K, Hagler AT and Gierasch LM (2004). Sequence and structural analysis of cellular retinoic acid-binding proteins reveals a network of conserved hydrophobic interactions. *Proteins* **54**: 179-94.

Ha JH and Mckay DB (1995). Kinetics of nucleotide-induced changes in the tryptophan fluorescence of the molecular chaperone Hsc70 and its subfragments suggest the ATP-induced conformational change follows initial ATP binding. *Biochemistry* **34**: 11635-44.

Halabi N, Rivoire O, Leibler S and Ranganathan R (2009). Protein sectors: evolutionary units of three-dimensional structure. *Cell* **138**: 774-86.

Hardy SJ and Randall LL (1991). A kinetic partitioning model of selective binding of nonnative proteins by the bacterial chaperone SecB. *Science* **251**: 439-43.

Harrison CJ, Hayer-Hartl M, Liberto MD, Hartl F-U and Kuriyan J (1997). Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. *Science* **276**: 431-5.

Hartl FU and Hayer-Hartl M (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295**: 1852-8.

Henikoff S and Henikoff JG (1992). Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A* **89**: 10915-9.

Hess B (2002). Convergence of sampling in protein simulations. *Phys Rev E Stat Nonlin Soft Matter Phys* **65**: 031910.

Hess B, Kutzner C, van der Spoel D and Lindahl E (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* **4**: 435-47.

Ishida T and Kinoshita K (2008). Prediction of disordered regions in proteins based on the meta approach. *Bioinformatics* **24**: 1344-8.

Jakob U, Gaestel M, Engel K and Buchner J (1993). Small heat shock proteins are molecular chaperones. *J Biol Chem* **268**: 1517-20.

Jaya N, Garcia V and Vierling E (2009). Substrate binding site flexibility of the small heat shock protein molecular chaperones. *Proc Natl Acad Sci U S A* **106**: 15604-9.

Jiang J, Maes EG, Taylor AB, Wang L, Hinck AP, Lafer EM and Sousa R (2007). Structural basis of J cochaperone binding and regulation of Hsp70. *Mol Cell* **28**: 422-33.

Jiang JW, Prasad K, Lafer EM and Sousa R (2005). Structural basis of interdomain communication in the Hsc70 chaperone. *Mol Cell* **20**: 513-24.

Johnson ER and McKay DB (1999). Mapping the role of active site residues for transducing an ATP-induced conformational change in the bovine 70-kDa heat shock cognate protein. *Biochemistry* **38**: 10823-30.

Kampinga HH and Craig EA (2010). The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nat Rev Mol Cell Biol* **11**: 579-92.

Kapust RB, Tozser J, Fox JD, Anderson DE, Cherry S, Copeland TD and Waugh DS (2001). Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng* **14**: 993-1000.

Karzai AW and McMacken R (1996). A bipartite signaling mechanism involved in DnaJ-mediated activation of the Escherichia coli DnaK protein. *J Biol Chem* **271**: 11236-46.

Kass I and Horovitz A (2002). Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations. *Proteins* **48**: 611-7.

Kthiri F, Le H-T, Tagourti J, Kern R, Malki A, Caldas T, Abdallah J, Landoulsi A and Richarme G (2008). The thioredoxin homolog YbbN functions as a chaperone rather than as an oxidoreductase. *Biochem Biophys Res Commun* **374**: 668-72.

Langer T, Lu C, Echols H, Flanagan J, Hayer MK and Hartl FU (1992). Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature* **356**: 683-9.

Larson SM, Di Nardo AA and Davidson AR (2000). Analysis of covariation in an SH3 domain sequence alignment: applications in tertiary contact prediction and the design of compensating hydrophobic core substitutions. *J Mol Biol* **303**: 433-46.

Laufen T, Mayer MP, Beisel C, Klostermeier D, Mogk A, Reinstein J and Bukau B (1999). Mechanism of regulation of Hsp70 chaperones by DnaJ cochaperones. *Proc Natl Acad Sci U S A* **96**: 5452-7.

Lee GJ and Vierling E (2000). A small heat shock protein cooperates with heat shock protein 70 systems to reactivate a heat-denatured protein. *Plant Physiol* **122**: 189-98.

Lee J, Natarajan M, Nashine VC, Socolich M, Vo T, Russ WP, Benkovic SJ and Ranganathan R (2008). Surface sites for engineering allosteric control in proteins. *Science* **322**: 438-42.

Liberek K, Skowyra D, Zylicz M, Johnson C and Georgopoulos C (1991). The Escherichia coli Dnak chaperone, the 70-kDa heat shock protein eukaryotic equivalent, changes conformation upon ATP hydrolysis, thus triggering its dissociation from a bound target protein. *J Biol Chem* **266**: 14491-6.

Lichtarge O, Bourne HR and Cohen FE (1996). An evolutionary trace method defines binding surfaces common to protein families. *J Mol Biol* **257**: 342-58.

Lin J and Wilson MA (2011). Escherichia coli thioredoxin-like protein YbbN contains an atypical tetratricopeptide repeat motif and is a negative regulator of GroEL. *J Biol Chem* **286**: 19459-69.

Lindorff-Larsen K, Best RB, Depristo MA, Dobson CM and Vendruscolo M (2005). Simultaneous determination of protein structure and dynamics. *Nature* **433**: 128-32.

Linke K (2005). Characterization of the DnaK-DnaJ-GrpE system under oxidative heat stress *Ph.D. disseration*, Technical University of Munich.

Linke K, Wolfram T, Bussemer J and Jakob U (2003). The roles of the two zinc binding sites in DnaJ. *J Biol Chem* **278**: 44457-66.

Liu Q and Hendrickson WA (2007). Insights into Hsp70 chaperone activity from a crystal structure of the yeast Hsp110 Sse1. *Cell* **131**: 106-20.

Liu Y, Gierasch LM and Bahar I (2010). Role of Hsp70 ATPase domain intrinsic dynamics and sequence evolution in enabling its functional interactions with NEFs. *PLoS Comput Biol* **6**: e1000931.

Lockless SW and Ranganathan R (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**: 295-9.

Mapa K, Sikor M, Kudryavtsev V, Waegemann K, Kalinin S, Seidel CAM, Neupert W, Lamb DC and Mokranjac D (2010). The conformational dynamics of the mitochondrial Hsp70 chaperone. *Mol Cell* **38**: 89-100.

Marcelino AMC, Smock RG and Gierasch LM (2006). Evolutionary coupling of structural and functional sequence information in the intracellular lipid-binding protein family. *Proteins* **63**: 373-84.

Mayer M and Bukau B (2005). Hsp70 chaperones: Cellular functions and molecular mechanism. *Cell Mol Life Sci* **62**: 670-84.

Mayer MP, Schroder H, Rudiger S, Paal K, Laufen T and Bukau B (2000). Multistep mechanism of substrate binding determines chaperone activity of Hsp70. *Nat Struct Biol* **7**: 586-93.

McCarty JS, Buchberger A, Reinstein J and Bukau B (1995). The role of ATP in the functional cycle of the DnaK chaperone system. *J Mol Biol* **249**: 126-37.

Minami Y, Hohfeld J, Ohtsuka K and Hartl FU (1996). Regulation of the heat-shock protein 70 reaction cycle by the mammalian DnaJ homolog, Hsp40. *J Biol Chem* **271**: 19617-24.

Montgomery DL, Morimoto RI and Gierasch LM (1999). Mutations in the substrate binding domain of the Escherichia coli 70 kDa molecular chaperone, DnaK, which alter substrate affinity or interdomain coupling. *J Mol Biol* **286**: 915-32.

Morimoto RI (1993). Cells in stress: transcriptional activation of heat shock genes. *Science* **259**: 1409-10.

Moro F, Fernandez V and Muga A (2003). Interdomain interaction through helices A and B of DnaK peptide binding domain. *FEBS Lett* **533**: 119-23.

Moro F, Taneva SG, Velázquez-Campoy A and Muga A (2007). GrpE N-terminal domain contributes to the interaction with DnaK and modulates the dynamics of the chaperone substrate binding domain. *J Mol Biol* **374**: 1054-64.

Morshauser RC, Hu W, Wang H, Pang Y, Flynn GC and Zuiderweg ER (1999). High-resolution solution structure of the 18 kDa substrate-binding domain of the mammalian chaperone protein Hsc70. *J Mol Biol* **289**: 1387-403.

Neher E (1994). How frequent are correlated changes in families of protein sequences? *Proc Natl Acad Sci U S A* **91**: 98-102.

Noivirt O, Eisenstein M and Horovitz A (2005). Detection and reduction of evolutionary noise in correlated mutation analysis. *Protein Eng Des Sel* **18**: 247-53.

Osipiuk J, Georgopoulos C and Zylicz M (1993). Initiation of lambda DNA replication. The Escherichia coli small heat shock proteins, DnaJ and GrpE, increase DnaK's affinity for the lambda P protein. *J Biol Chem* **268**: 4821-7.

Paek KH and Walker GC (1987). Escherichia coli dnaK null mutants are inviable at high temperature. *J Bacteriol* **169**: 283-90.

Palleros DR, Reid KL, Shi L, Welch WJ and Fink AL (1993). ATP-induced protein Hsp70 complex dissociation requires K+ but not ATP hydrolysis. *Nature* **365**: 664-6.

Patury S, Miyata Y and Gestwicki JE (2009). Pharmacological targeting of the Hsp70 chaperone. *Curr Top Med Chem* **9**: 1337-51.

Pawson T and Nash P (2003). Assembly of cell regulatory systems through protein interaction domains. *Science* **300**: 445-52.

Pellecchia M, Montgomery D, Stevens D, Vander Kooi C, Feng H and Gierasch LZ, ERP (2000). Structural insights into substrate binding by the molecular chaperone DnaK. *Nat Struct Mol Bio* **7**: 298-303.

Pellecchia M, Montgomery DL, Stevens SY, Vander Kooi CW, Feng HP, Gierasch LM and Zuiderweg ER (2000). Structural insights into substrate binding by the molecular chaperone DnaK. *Nat Struct Biol* **7**: 298-303.

Peterson FC, Penkert RR, Volkman BF and Prehoda KE (2004). Cdc42 regulates the Par-6 PDZ domain through an allosteric CRIB-PDZ transition. *Mol Cell* **13**: 665-76.

Polier S, Dragovic Z, Hartl FU and Bracher A (2008). Structural basis for the cooperation of Hsp70 and Hsp110 chaperones in protein folding. *Cell* **133**: 1068-79.

Posfai G, Koob MD, Kirkpatrick HA and Blattner FR (1997). Versatile insertion plasmids for targeted genome manipulations in bacteria: isolation, deletion, and rescue of the pathogenicity island LEE of the Escherichia coli O157:H7 genome. *J Bacteriol* **179**: 4426-8.

Randall LL and Hardy SJS (2002). SecB, one small chaperone in the complex milieu of the cell. *Cell Mol Life Sci* **59**: 1617-23.

Ravagnan L, Gurbuxani S, Susin SA, Maisse C, Daugas E, Zamzami N, Mak T, Jaattela M, Penninger JM, Garrido C and Kroemer G (2001). Heat-shock protein 70 antagonizes apoptosis-inducing factor. *Nat Cell Biol* **3**: 839-43.

Revington M, Holder TM and Zuiderweg ERP (2004). NMR study of nucleotide-induced changes in the nucleotide binding domain of Thermus thermophilus Hsp70 chaperone DnaK - Implications for the allosteric mechanism. *J Biol Chem* **279**: 33958-67.

Revington M, Zhang Y, Yip GN, Kurochkin AV and Zuiderweg ER (2005). NMR investigations of allosteric processes in a two-domain Thermus thermophilus Hsp70 molecular chaperone. *J Mol Biol* **349**: 163-83.

Rist W, Graf C, Bukau B and Mayer MP (2006). Amide hydrogen exchange reveals conformational changes in Hsp70 chaperones important for allosteric regulation. *J Biol Chem* **281**: 16493-501.

Rosenbaum JC, Fredrickson EK, Oeser ML, Garrett-Engele CM, Locke MN, Richardson LA, Nelson ZW, Hetrick ED, Milac TI, Gottschling DE and Gardner RG (2011). Disorder targets misorder in nuclear quality control degradation: a disordered ubiquitin ligase directly recognizes its misfolded substrates. *Mol Cell* **41**: 93-106.

Rudiger S, Germeroth L, Schneider-Mergener J and Bukau B (1997). Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *EMBO J* **16**: 1501-7.

Russ WP, Lowery DM, Mishra P, Yaffe MB and Ranganathan R (2005). Natural-like function in artificial WW domains. *Nature* **437**: 579-83.

Sakr S, Cirinesi AM, Ullers RS, Schwager F, Georgopoulos C and Genevaux P (2010). Lon protease quality control of presecretory proteins in Escherichia coli and its dependence on the SecB and DnaJ (Hsp40) chaperones. *J Biol Chem* **285**: 23506-14.

Sali A and Blundell TL (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* **234**: 779-815.

Schlecht R, Erbse AH, Bukau B and Mayer MP (2011). Mechanics of Hsp70 chaperones enables differential interaction with client proteins. *Nat Struct Mol Biol* **18**: 345-51.

Schmid D, Baici A, Gehring H and Christen P (1994). Kinetics of molecular chaperone action. *Science* **263**: 971-3.

Schroder H, Langer T, Hartl FU and Bukau B (1993). DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage. *EMBO J* **12**: 4137-44.

Schuermann JP, Jiang JW, Cuellar J, Llorca O, Wang LP, Gimenez LE*, et al.* (2008). Structure of the Hsp110 : Hsc70 nucleotide exchange machine. *Mol Cell* **31**: 232-43.

Schuler H (2001). ATPase activity and conformational changes in the regulation of actin. *Biochim Biophys Acta* **1549**: 137-47.

Shaner L and Morano KA (2007). All in the family: atypical Hsp70 chaperones are conserved modulators of Hsp70 activity. *Cell Stress Chaperones* **12**: 1-8.

Sharan SK, Thomason LC, Kuznetsov SG and Court DL (2009). Recombineering: a homologous recombination-based method of genetic engineering. *Nat Protoc* **4**: 206-23.

Sharma S, De Los Rios P, Christen P, Lustig A and Goloubinoff P (2010). The kinetic parameters and energy cost of the Hsp70 chaperone as a polypeptide unfoldase. *Nat Chem Bio* **6**: 914-20.

Shomura Y, Dragovic Z, Chang HC, Tzvetkov N, Young JC, Brodsky JL, Guerriero V, Hartl FU and Bracher A (2005). Regulation of Hsp70 function by HspBP1: Structural analysis reveals an alternate mechanism for Hsp70 nucleotide exchange. *Mol Cell* **17**: 367-79.

Siegenthaler RK and Christen P (2006). Tuning of DnaK chaperone action by nonnative protein sensor DnaJ and thermosensor GrpE. *J Biol Chem* **281**: 34448-56.

Slepenkov SV and Witt SN (2002). The unfolding story of the Escherichia coli Hsp70 DnaK: is DnaK a holdase or an unfoldase? *Mol Microbiol* **45**: 1197-206.

Smock RG, Blackburn ME and Gierasch LM (2011). The conserved, disordered C terminus of DnaK enhances cellular survival upon stress and *in vitro* chaperone activity. *J BIol Chem* in press.

Smock RG and Gierasch LM (2005). Finding the fittest fold: using the evolutionary record to design new proteins. *Cell* **122**: 832-4.

Smock RG and Gierasch LM (2009). Sending signals dynamically. *Science* **324**: 198-203.

Smock RG, Rivoire O, Russ WP, Swain JF, Leibler S, Ranganathan R and Gierasch LM (2010). An interdomain sector mediating allostery in Hsp70 molecular chaperones. *Mol Syst Biol* **6**: 414.

Socolich M, Lockless SW, Russ WP, Lee H, Gardner KH and Ranganathan R (2005). Evolutionary information for specifying a protein fold. *Nature* **437**: 512-8.

Sondermann H, Scheufler C, Schneider C, Hohfeld J, Hartl FU and Moarefi I (2001). Structure of a Bag/Hsc70 complex: Convergent functional evolution of Hsp70 nucleotide exchange factors. *Science* **291**: 1553-7.

Suh W-C, Burkholder WF, Lu CZ, Zhao X, Gottesman ME and Gross CA (1998). Interaction of the Hsp70 molecular chaperone, DnaK, with its cochaperone DnaJ. *Proc Natl Acad Sci U S A* **95**: 15223-8.

Suh WC, Lu CZ and Gross CA (1999). Structural features required for the interaction of the Hsp70 molecular chaperone DnaK with its cochaperone DnaJ. *J Biol Chem* **274**: 30534-9.

Swain JF, Dinler G, Sivendran R, Montgomery DL, Stotz M and Gierasch LM (2007). Hsp70 chaperone ligands control domain association via an allosteric mechanism mediated by the interdomain linker. *Mol Cell* **26**: 27-39.

Swain JF, Schulz EG and Gierasch LM (2006). Direct comparison of a stable isolated Hsp70 substrate-binding domain in the empty and substrate-bound states. *J Biol Chem* **281**: 1605-11.

Szabo A, Langer T, Schroder H, Flanagan J, Bukau B and Hartl FU (1994). The ATP hydrolysis-dependent reaction cycle of the Escherichia coli Hsp70 system DnaK, DnaJ, and GrpE. *Proc Natl Acad Sci U S A* **91**: 10345-9.

Tanaka N, Nakao S, Wadai H, Ikeda S, Chatellier J and Kunugi S (2002). The substrate binding domain of DnaK facilitates slow protein refolding. *Proc Natl Acad Sci U S A* **99**: 15398-403.

Tapley TL and Vickery LE (2004). Preferential substrate binding orientation by the molecular chaperone HscA. *J Biol Chem* **279**: 28435-42.

Theyssen H, Schuster HP, Packschies L, Bukau B and Reinstein J (1996). The second step of ATP binding to DnaK induces peptide release. *J Mol Biol* **263**: 657-70.

Thompson JD, Higgins DG and Gibson TJ (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673-80.

Todd MJ, Lorimer GH and Thirumalai D (1996). Chaperonin-facilitated protein folding: optimization of rate and yield by an iterative annealing mechanism. *Proc Natl Acad Sci U S A* **93**: 4030-5.

Tompa P and Csermely P (2004). The role of structural disorder in the function of RNA and protein chaperones. *FASEB J* **18**: 1169-75.

Ullers RS, Ang D, Schwager F, Georgopoulos C and Genevaux P (2007). Trigger Factor can antagonize both SecB and DnaK/DnaJ chaperone functions in Escherichia coli. *Proc Natl Acad Sci U S A* **104**: 3101-6.

Ullers RS, Luirink J, Harms N, Schwager F, Georgopoulos C and Genevaux P (2004). SecB is a bona fide generalized chaperone in Escherichia coli. *Proc Natl Acad Sci U S A* **101**: 7583-8.

van den Berk LCJ, Landi E, Walma T, Vuister GW, Dente L and Hendriks WJAJ (2007). An allosteric intramolecular PDZ−PDZ interaction modulates PTP-BL PDZ2 binding specificity. *Biochemistry* **46**: 13629-37.

Vickery L and Cupp-Vickery J (2007). Molecular chaperones HscA/Ssq1 and HscB/Jac1 and their roles in iron-sulfur protein maturation. *Crit Rev Biochem Mol Biol* **42**: 95-111.

Vogel M, Bukau B and Mayer MP (2006). Allosteric regulation of Hsp70 chaperones by a proline switch. *Mol Cell* **21**: 359-67.

Vogel M, Mayer MP and Bukau B (2006). Allosteric regulation of Hsp70 chaperones involves a conserved interdomain linker. *J Biol Chem* **281**: 38705-11.

Weibezahn J, Tessarz P, Schlieker C, Zahn R, Maglica Z, Lee S*, et al.* (2004). Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. *Cell* **119**: 653-65.

Wilbanks SM, Chen L, Tsuruta H, Hodgson KO and McKay DB (1995). Solution small-angle X-ray scattering study of the molecular chaperone Hsc70 and its subfragments. *Biochemistry* **34**: 12095-106.

Wilbanks SM, DeLuca-Flaherty C and McKay DB (1994). Structural basis of the 70-kilodalton heat shock cognate protein ATP hydrolytic activity. I. Kinetic analyses of active site mutants. *J Biol Chem* **269**: 12893-8.

Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD and Wishart DS (2003). VADAR: a web server for quantitative evaluation of protein structure quality. *Nucleic Acids Res* **31**: 3316-9.

Winter J, Linke K, Jatzek A and Jakob U (2005). Severe oxidative stress causes inactivation of DnaK and activation of the redox-regulated chaperone Hsp33. *Mol Cell* **17**: 381-92.

Xu R, Ayers B, Cowburn D and Muir TW (1999). Chemical ligation of folded recombinant proteins: Segmental isotopic labeling of domains for NMR studies. *Proc Natl Acad Sc U S A* **96**: 388-93.

Yochem J, Uchida H, Sunshine M, Saito H, Georgopoulos CP and Feiss M (1978). Genetic analysis of two genes, dnaJ and dnaK, necessary for Escherichia coli and bacteriophage lambda DNA replication. *Mol Gen Genet* **164**: 9-14.

Yu H, Braun P, Yıldırım MA, Lemmens I, Venkatesan K, Sahalie J*, et al.* (2008). High-quality binary protein interaction map of the yeast interactome network. *Science* **322**: 104-10.

Zhang Y and Zuiderweg ERP (2004). The 70-kDa heat shock protein chaperone nucleotide-binding domain in solution unveiled as a molecular machine that can reorient its functional subdomains. *Proc Natl Acad Sci U S A* **101**: 10272-7.

Zhu X, Zhao X, Burkholder WF, Gragerov A, Ogata CM, Gottesman ME and Hendrickson WA (1996). Structural analysis of substrate binding by the molecular chaperone DnaK. *Science* **272**: 1606-14.

Zhuravleva A and Gierasch LM (2011). Allosteric signal transmission in the nucleotide-binding domain of 70-kDa heat shock protein (Hsp70) molecular chaperones. *Proc Natl Acad Sc U S A* **108**: 6987-92.