Published in final edited form as:

Biopolymers. 2010 ; 94(6): 742–752. doi:10.1002/bip.21426.

Segmental isotopic labeling of the Hsp70 molecular chaperone DnaK using expressed protein ligation

Eugenia M. Clerico, Anastasia Zhuravleva, Robert G. Smock, and Lila M. Gierasch*

Department of Biochemistry & Molecular Biology and Department of Chemistry, University of Massachusetts, Amherst MA 01003, U.S.A

Abstract

Introducing biophysical labels into specific regions of large and dynamic multidomain proteins greatly facilitates mechanistic analysis. Ligation of expressed domains that are labeled in a desired manner before assembly of the intact molecular machine provides such a strategy. We have elaborated an experimental route using expressed protein ligation (EPL) to create an Hsp70 molecular chaperone (in this case the *E. coli* Hsp70, DnaK) where only one of the two constituent domains was labeled, in this case with NMR active isotopes, allowing visualization of the single domain in the context of the two domain protein. Several technical obstacles were overcome, including choice of site for ligation with retention of function, optimization of ligation yield, and purification from unreacted domains. Ligated semi-labeled DnaK was successfully produced with a Cys residue at position 383, and the ligated product harboring the Cys mutation was confirmed to be functional and identical to an expressed Cys-containing two-domain construct. The NMR spectrum of the segmentally labeled protein was considerably simplified, enabling unequivocal assignment and enhanced analysis of dynamics, as a prelude to exploring the energy landscape for allostery in the Hsp70 family.

Keywords

Hsp70 molecular chaperone; DnaK; segmental labeling; expressed protein ligation; intein-mediated cleavage; NMR

Introduction

The ability to engineer biophysical probes into a large protein is empowering: Many key mechanistic questions require analysis of domain rearrangements and dynamical sampling of conformational ensembles.¹ Among the methods most suited to such questions are fluorescence, especially Forster resonance energy transfer, NMR, and EPR. In all these methods, regiospecific labels are needed to probe the nature of conformational transitions. Typically, complex strategies must be invoked to accomplish this specific labeling such as orthogonal labeling reactions^{2,3} or genetic incorporation of non-native amino acids.^{4,5} Additionally, purification hurdles must be overcome to isolate homogeneous labeled samples. The development of methods to link whole protein domains and fragments by direct ligation^{6–8} opened the door to new routes to achieve segmental regiospecific biophysical labeling. Yet the number of examples of large, multidomain proteins where this strategy has been successfully applied remains small.^{9–12} The reasons for this include the

*Correspondence to Department of Biochemistry & Molecular Biology, University of Massachusetts-Amherst, Amherst, MA 01003, USA; tel: 1-413-545-6094; fax: 1-413-545-1289; gierasch@biochem.umass.edu.

need for case-specific design of constructs, the challenge of optimizing ligation reactions, and the requirement for retention of function in the ligated product.

An important class of multidomain allosteric molecular machines, the Hsp70 molecular chaperones, under intensive study in our laboratory, presents an apt opportunity to apply protein ligation strategies to make powerful biophysical experiments possible. Hsp70s comprise a family of molecular chaperones present in organisms from bacteria to humans, and found in almost all cellular compartments.¹³ They act in the cell under normal physiological conditions by performing functions like chaperoning nascent chains, dissociating macromolecular complexes, and assisting protein import into organelles. They also enable cells to respond to stress conditions such as heat shock by binding partially misfolded proteins that accumulate during stress and preventing aggregation, and helping with the dissociation of protein aggregates. In the cell, Hsp70s interact with other chaperones and with co-chaperones to form complex and interactive networks.^{14–19}

All Hsp70s bind through their C-terminal substrate-binding domain (SBD) to extended, hydrophobic stretches of substrate proteins, and this interaction is modulated by ADP or ATP binding to the N-terminal Hsp70 nucleotide binding or ATPase domain. The N-terminal ATPase domain has a bi-lobed actin-like fold with each lobe made up of two sub-domains. The two lobes are connected by a pair of crossing helices at the bottom of the cleft where nucleotide binds. The SBD is comprised of a β -sandwich of two four-stranded β -sheets followed by a helical lid sub-domain and a helical bundle. The substrate-binding groove has a hydrophobic pocket to accommodate a central hydrophobic residue within the bound peptide substrate.

When ADP is bound to an Hsp70, substrate binds with high affinity to the SBD, but upon nucleotide exchange to ATP, substrate affinity is greatly decreased. In turn, ATP binding leads to lower affinity for substrate. The two domains of Hsp70s are connected by a highly conserved, hydrophobic linker that is crucial for this interdomain communication.^{20–23}

Based on combination of NMR and other biophysical results, we proposed a model for the allosteric cycle of Hsp70s²⁴: In the ADP-bound state, the two domains are largely independently tumbling and connected by a flexible linker with no stable interdomain interactions. Upon ATP binding, both domains undergo a significant conformational change and dock onto each other, sequestering the connecting linker from the solvent. In this state, portions of the SBD are stabilized while the regions that form the substrate-binding pocket become dynamic.

A wide array of biochemical and structural data support the major features of this model: Other NMR analysis of the ADP-bound form of Hsp70s showed that the two domains are not docked.^{25,26} ATP binding leads to enhanced proteolytic lability of the SBD, stabilization of the ATPase domain, protection of the interdomain linker^{27,28}, solvent inaccessibility of Trp102 in DnaK ^{27,29,30}, and compaction of the overall molecule as signaled by a decrease in the radius of gyration by 15 Å.³¹ The ATP-bound state of Hsp70 has proven refractory to study by crystallography. A recently published structure of a yeast ATP-bound Sse1 protein, a member of the Hsp110 family which is a distant relative of Hsp70s and shares their domain arrangements, provides a plausible model for the mode of interdomain packing in ATP-bound Hsp70s.³² The global architecture of ATP-bound Sse1 agrees with all the previous biochemical data for the analogous Hsp70 state, and fits the allosteric model proposed before.²⁴ From this model emerges a fundamental puzzle: what is the mechanism in Hsp70s whereby ligands modulate the mode of interdomain packing and the affinity of their partner domains?

In order to address this question, we are investigating the structure and dynamical properties of all the species participating in the allosteric cycle of the Hsp70 DnaK using a wide array of biophysical methods, including NMR. Our earlier work indicated that the dynamic nature of different nucleotide states along the allosteric sequence of DnaK could be addressed via NMR.²⁴ However, both the size of the overall molecule and the complexities that arise upon ATP binding made it clear that NMR analysis would be challenging. Note that the NMR spectrum of the ATP-bound state of DnaK shows global chemical shift changes as a consequence of the large changes in structure and in dynamics all through the protein. While signals were sharp for the ADP-bound state when the domains tumbled as smaller entities, the quality of the spectrum for the 552-amino acid construct we designed for NMR studies²⁴ decreased in the ATP-bound state, as the domains dock and the protein tumbles as a 60 kDa species. Also, the range of line widths observed points to differential dynamics and conformational exchange in different parts of the molecule. In order to successfully study this complex docked state of DnaK by NMR, spectral simplification is required.

Thus, our plan to apply NMR to this system presents an opportunity to exploit expressed protein ligation (EPL) strategies to produce segmentally labeled Hsp70s. This approach has the potential to simplify and improve the assignment of the protein by resolving peak overlaps and reducing the number of spin systems of the two domain constructs of DnaK. We have isotopically labeled only one domain, then reassembled the two-domain protein using EPL, and analyzed the NMR spectrum for the single domain in the context of two-domain protein. One limitation of this method is that the ligation procedure leaves a Cys residue at the ligation point. An obvious potential attachment point is the interdomain linker, but this region is known to be refractory to mutations.²⁰ A Cys-scanning mutagenesis study was performed on DnaK in order to find sites that tolerate Cys replacement near the interdomain linker with retention of full allosteric function.

Our optimized protocol enabled us to express the ATPase domain of DnaK as a ^{15}N , ^{13}C , ^2H labeled protein and ligate it to the unlabeled SBD, yielding an active, properly folded protein suitable for NMR studies. This strategy now makes possible the analysis of a segmentally labeled DnaK molecule with a reduced number of overlapping spin systems in the elusive ATP-bound docked state and to relate the structure and dynamics of this state to others in the allosteric cycle. In order to observe the ATP-bound state, a mutation that blocks ATP hydrolysis (T199A²⁷) is incorporated into the ligated product. In addition to the utility of the ligation method for NMR, a parallel strategy can now be deployed in order to introduce domain-specific labels for fluorescence or EPR measurements.

Materials and Methods

Reagents

Primers were obtained from Integrated DNA Technologies (Coralville, IA). Restriction enzymes, ImpactTM kit and chitin-binding (CB) beads were purchased from New England Biolabs (NEB, Ipswich, MA). Pfu polymerase was from Stratagene (La Jolla, CA). Guanidine hydrochloride (GuHCl) was purchased from USB (Cleveland, OH). Isotopes were from Cambridge Isotopes (Andover, MA). The model substrate peptide, p5 (CLLLSAPRR), was synthesized by Genscript (Piscataway, NJ) and purified as described. 33 Reagents used for the coupled enzyme ATPase activity assay were obtained from Sigma (St. Louis, MO).

Cloning

The plasmid encoding DnaK(1–552)_{ye} T199A was obtained as previously described. DnaK wild type and DnaK(1–552)_{ye} T199A encoding genes were mutagenized by QuikChange

(Stratagene, La Jolla, CA) to obtain DnaK T383C and DnaK(1–552)_{ye} T199A T383C respectively. Various DnaK Cys mutants were prepared as C-terminal His-tagged proteins by QuikChange using the pWCS plasmid carrying DnaK cloned into derivatives of the pTrc HisA plasmid (pWCS was obtained from Carol Gross' laboratory³⁴).

For EPL, the DNA encoding the ATPase domain of DnaK (1–382 T199A) was amplified by PCR using the plasmid containing DnaK(1–552)_{ye} T199A as a template and cloned into pTwin1 between the NdeI and SapI sites, upstream of the *Mycobacterium xenopi* gyrA gene-derived intein (Mxe) and the chitin-binding domain (CBD). The pTwin plasmid containing the DnaK (1–382 T199A) gene was used as template to generate the ATPase domain (1–382) without the T199A mutation.

To create the intein fusion of the SBD (383C–552)_{ye}, the plasmid containing DnaK(1–552)_{ye} T199A was mutagenized to remove an internal SapI site, the SBD was amplified and cloned into pTwin1 between the SapI and BamHI restriction sites, downstream of the *Synechocystis* sp DnaB intein (*Ssp*) and the CBD (see Figure 1). Cloned genes were sequenced by Genewiz (South Plainfield, NJ).

ASA calculation

The solvent-accessible surface areas (ASA) of DnaK residues in the ADP-bound state were calculated using VADAR35 and the PDB entry 2KHO25 as input.

Protein expression and purification

DnaK wild type, DnaK T383C, and DnaK(1–552)_{ye} T199A T383C were expressed from pMS plasmids and purified as described³⁶, except that *E. coli* BL21(DE3) cells were used. In order to prepare samples of protein suitable to directly compare to that obtained by EPL, all proteins were denatured after purification in 6 M GuHCl for 1 h at 37 °C, refolded by 30X dilution in native buffer (either HMK (20 mM HEPES pH 7.6, 100 mM KCl, 5 mM MgCl₂) or NMR buffer (10 mM potassium phosphate pH 7.3, 10 mM KCl, 5 mM MgCl₂, 1 mM DTT) with constant slow stirring at 4 °C for 2 h, and concentrated using Amicon Ultra Centrifugal Filter Units (Billerica, MA).

Isotopic labeling and preparation of NMR samples

Isotopically labeled samples were prepared following established protocols.³⁷ Cells for the expression of uniformly ¹⁵N, ²H and ¹³C labeled proteins (DnaK(1–552)_{ye} T199A T383C or ATPase domain 1–382 T199A fused to the intein) were grown in M9 minimal medium in D₂O containing ¹⁵N-ammonium chloride and ¹³C (deuterated) glucose. After purification, proteins were concentrated to ≈ 0.3 mM in NMR buffer, 5% D₂O, 0.02% sodium azide and placed in a Shigemi NMR tube. These samples were directly measured as “nucleotide-free” samples; subsequently, ATP (pH 7.3) was added up to a 5 mM concentration to obtain the ATP-bound state.

Expression of the intein fusions

ATPase domain (1–382 T199A, from now on referred to as “ATPase domain”), ATPase domain 1–382, and SBD ((383C–552)_{ye} from now on “SBD”) cloned as intein fusions were expressed in *E. coli* BL21(DE3) until OD ≈ 0.8. Protein expression was induced by addition of 1 mM IPTG for 2 h at 37 °C. Cells were collected, re-suspended in buffer Impact 1 (20 mM HEPES pH 8.5, 500 mM NaCl, 1 mM EDTA), and broken by passage through a Microfluidics M-110L pneumatic cell disruptor (Newton, MA). Cell lysates were spun down at 48000 × g for 20 min at 4 °C, and the cleaved proteins were obtained from the supernatant as described below.

For optimization of the EPL protocol, fluorescence, and ATPase activity assays, cells were grown in LB. For NMR, the ATPase domain was produced in triply labeled cultures (^2H , ^{13}C , and ^{15}N) as described above. In both cases the same purification and ligation protocols were followed.

Cleavage of ATPase domain and SBD from the intein fusion

The supernatants obtained after spinning the cell lysates were passed three times through a 10 ml column packed with CB beads at 4 °C. Columns were washed with 40 ml of buffer Impact 1 and then with 20 ml of buffer Impact 1 + 50 mM sodium mercaptoethane sulfonate (MESNA) for ATPase domain cleavage, or buffer Impact 2 (20 mM HEPES pH 6.0, 500 mM NaCl, 1 mM EDTA) for SBD cleavage. Columns were incubated at 37 °C for 16 h.

Cleaved products (ATPase domain with a C-terminal α -thioester and SBD with a free N-terminal Cys) were eluted from the CB beads, concentrated to the mM range and kept at 4 °C until ligation. Cleaved products could be frozen before ligation.

Expressed protein ligation

Ligation reactions were set up by mixing the cleaved products (ATPase 1–382 with a C-terminal α -thioester and SBD with a free N-terminal Cys) in the appropriate conditions. Several reactions were first run in order to determine the conditions for optimum ligation; different pHs in the 7.5–8.5 range were tested, as well as molar ratios of ATPase domain:SBD and several different urea concentrations (0.0–8.0 M) (see Results and Discussion section).

Optimal ligation conditions were explored by mixing the ATPase domain and SBD in molar ratio of 1:7 (the ratios were qualitatively determined by estimating the relative protein amount by band intensities on Coomassie-stained SDS-PAGE, assuming that the ATPase domain and SBD do not differ in Coomassie binding per mass unit) in buffer containing 20 mM HEPES pH 8.0, 500 mM NaCl, 2 M urea, 50 mM MESNA, 10 mM glycine, and incubated overnight at 37 °C. Unreacted ATPase domain and SBD were separated from the ligated protein by diluting the ligation reaction 5X with 10 mM HEPES pH 7.6, 1 mM EDTA, and running it on a MonoQ column 5/50 GL coupled to an AKTA Purifier 10 (GE healthcare, Piscataway, NJ) using a NaCl gradient. Fractions containing the ligated product were pooled and concentrated; in some cases these fractions were further purified by re-running on the MonoQ column. Proteins in the mixture were denatured by adding GuHCl to 6.0 M and DTT to 1 mM and loaded in a 60 cm Superdex S200 column equilibrated in 20 mM HEPES pH 7.6, 6.0 M GuHCl, at 1 ml/min. Collected fractions (1 ml) TCA precipitated and analyzed by SDS-PAGE. Fractions containing pure ligated protein were pooled and refolded as described above. The mass of the purified ligated protein was confirmed by mass spectrometry at the University of Massachusetts Amherst Mass Spectrometry Core Facility.

In vivo activity assays

Heat shock assay—To assay the ability of DnaK mutants to support bacterial growth under heat shock stress, DnaK wild type, DnaK T199A and DnaK T383C containing plasmids were transformed into *E. coli* BB1553 cells (MC4100 $\Delta dnaK52::Cm^r sidB1$, a gift from B. Bukau³⁸). Single colonies were grown overnight in liquid media at 30 °C and the OD₆₀₀ of the cultures normalized to 0.2. 10-fold serial dilutions of all cultures were equilibrated at 43 °C, spotted on LB plates pre-equilibrated at 43 °C and the plates incubated at the same temperature for 16 h. Basal expression of the *dnaK* wild type gene in pMS under the control of the *tac* promoter was sufficient to achieve nearly optimal growth rescue (no addition of IPTG needed).

Bacteriophage λ propagation—To test the functionality of various DnaK Cys mutants (see Results and Discussion) the λ bacteriophage propagation assay was performed in *dnaK*[−] *E. coli* cells BB1553 as described³⁶, except that no IPTG was used and all cultures were normalized to OD₆₀₀=1.0 prior to incubation with the phage.

Fluorescence assay

Measurements of the ATP-induced blue shift of W102 fluorescence were made in an Photon Technology International Alpha Scan Fluorometer (Birmingham, NJ) as described³⁶ for DnaK wild type, DnaK T383C, DnaK (1–552)_{ye} T199A T383C and ligated DnaK (1–552)_{ye} T199A T383C.

ATPase assay

Steady-state ATPase rate was measured using a Biotek Synergy 2 Multi-Mode Microplate Reader (Winooski, VT) at 30 °C using a coupled enzyme assay³⁹, as previously described for DnaK³⁶, except that 1 mM DTT was added to the reactions. For substrate-stimulated activity, 200 μ M p5 peptide were pre-incubated with the protein for 15 min at room temperature prior to measurement. DnaK wild type, DnaK T383C and ligated DnaK (1–552)_{ye} were tested.

NMR spectroscopy

All NMR experiments were performed at 25 °C on a 600 MHz Bruker Avance spectrometer using a TCI cryoprobe. ¹H-¹⁵N TROSY and 3D HNCQ spectra were obtained for ²H, ¹³C, ¹⁵N uniformly or segmentally (residues 1–382) labeled DnaK(1–552)_{ye} T199A T383C. To transfer the backbone assignments from the isolated domains to the two-domain DnaK, a 3D HNCA spectrum was also recorded for the uniformly labeled protein, again using a standard TopSpin 1.3 pulse sequence. All NMR experiments were performed using standard pulse sequences from Bruker TopSpin 1.3 software. Data were processed using NMRPIPE⁴⁰ and analyzed using CARA.⁴¹

Results and Discussion

Strategy

The overall goal of this work was to develop a strategy to introduce specific labels into an Hsp70 molecular chaperone so as to enable dissection of states sampled during the allosteric cycle modulated by nucleotide and substrate binding to each of the two domains. In particular, we focused on segmentally labeling a DnaK molecule with NMR active isotopes in order to make its NMR spectra tractable for detailed structural and dynamic analysis. The method developed would be of more general use when differential modification of Hsp70 domains is desired—for example, for fluorescence resonance energy transfer experiments. EPL6⁸ was the approach chosen for coupling of the differentially labeled domains, so that we could obtain a product with only one of the two domains of DnaK fully ¹⁵N, ¹³C and ²H labeled in the context of the two domain protein.

Figure 1 outlines the strategy used to obtain natively folded, allosterically active segmentally labeled DnaK: Briefly, the ATPase domain of DnaK was expressed as an N-terminal fusion of the *Mxe* intein, which upon cleavage generates a domain with a C-terminal α -thioester.⁴² The SBD was expressed as a C-terminal fusion of the *Ssp* intein, which upon cleavage generates a SBD with an N-terminal free Cys. Both intein fusions contain a CBD to facilitate the purification of the fusion protein and separation of the intein after cleavage. Once the cleaved products were purified, EPL conditions were established: The first step of the ligation involves the chemoselective reaction between an N-terminal Cys of one peptide

fragment and the C-terminal α -thioester of a second peptide fragment (in this case, SBD and ATPase domain, respectively). The initial transthioesterification reaction is followed by a spontaneous intramolecular S $\rightarrow$ N acyl shift to generate an amide bond at the ligation junction.⁴² This strategy leaves behind a Cys residue at the point of ligation, but with normal amide bond linkages in the backbone. In the case of DnaK, an obvious candidate for the junction point would appear to be the flexible interdomain linker, but as pointed out above, this linker is highly conserved and important for allosteric communication in DnaK.^{22,23} The junction site was chosen by Cys-scanning through the linker and neighboring sites, as described below, and the impact of the various Cys mutations on the function of DnaK was evaluated by making and studying the DnaK Cys mutant prior to carrying out the ligation.

DnaK T383C retains wild type activity in *in-vivo* assays

In an effort to find DnaK mutants that tolerate Cys replacements, several positions around the interdomain linker were mutated to Cys, and the functionality of the resulting mutants was tested. DnaK is essential for λ phage propagation in *E. coli*, and thus the ability to support λ phage propagation was used to assess *in vivo* function.^{43,44} Initially, the Cys mutants were constructed in a C-terminal His-tagged protein background. DnaK variants were tested for bacteriophage λ replication in BB155338, which lack the chromosomal copy of DnaK. In the presence of a plasmid producing active DnaK the cells can support λ propagation.^{45,46} Figure 2a shows the number of λ plaques formed due to the presence of DnaK-expressed from pMS or pWCS plasmids in BB1553. In the absence of DnaK expression plaque formation is not observed. Two of the five linker Cys mutations (T383C and G379C) supported λ replication at comparable levels to wild type protein.

Sequence analysis shows that T383 is one of the least conserved sites around the interdomain linker region and therefore highly likely to tolerate a Cys substitution. Also, Ser is the most conserved amino acid at position 383 within the Hsp70 family (33%), which presages good tolerance of a Cys mutation at that site. For this reason, as well as visual inspection of the greater accessibility of this site versus G379, we focused on T383C as the preferred site of ligation. To further assess the function of the DnaK T383C mutant, its ability to support cell growth under heat shock was tested (in this assay, a non-His-tagged DnaK background was utilized). As mentioned, the *dnaK*-deficient *E. coli* strain BB1553 is not viable when grown at 43 °C, but exogenous expression of a functional DnaK from the pMS plasmid can restore cell growth. Figure 2b shows that expression of DnaK T383C from the pMS plasmid supports cell growth at 43 °C to the same extent as comparable expression of wild type DnaK, while the BB1553 strain by itself fails to propagate, as does the ATP hydrolysis-deficient DnaK T199A, used as negative control. Based on the functionality of this mutant in the *in vivo* tests, we selected residue T383 to be the junction point for the EPL.

A C-terminally truncated, 552-amino acid version of DnaK was used in this study as a model for DnaK wild type: DnaK(1–552)_{ye} T199A. This construct retains the ability to undergo ATP-induced conformational changes, and its reduced size makes it a more tractable target for NMR studies.³³ Two adjacent Leu residues at positions 542 and 543 were mutated to Tyr and Glu respectively to avoid intramolecular binding of the C-terminal helix in the substrate-binding pocket, which was observed when the Leu residues were retained and is a widespread artifact in truncated Hsp70 constructs.³³ Previous functional tests for allostery showed that DnaK(1–552)_{ye} behaves similarly to the wild-type protein.³³ The T199A mutation leads to impaired ATP hydrolysis; this mutant has been extensively characterized and was used here to enable study of the ATP-bound state of DnaK over the time scale required for NMR.^{24,27,30} DnaK(1–552)_{ye} T199A T383C was constructed, over-expressed in *E. coli*, and purified as described. The purified mutant protein showed the expected characteristic blue shift and quenching of its Trp fluorescence upon interaction

with ATP, analogous to wild type protein (data not shown). The NMR spectrum of the isotopically labeled DnaK(1–552)_{ye} T199A T383C overlaps with the spectra of DnaK(1–552)_{ye} T199A (except for residue 383), both in the nucleotide-free and ATP-bound form (not shown), confirming that the T383C mutant can undergo the structural changes required for DnaK allosteric function.

The individual domains of DnaK can be efficiently expressed as intein fusions, purified and cleaved from the fused intein

The individual domains of DnaK were expressed as intein fusions (Figure 3) (see Materials and Methods section). Both fusion proteins were largely soluble when over-expressed in *E. coli* BL21(DE3) at 37 °C. Typically, production of the ATPase domain intein fusion was more efficient than production of its SBD counterpart, as some cleavage of the SBD fusion occurs in the cell. Reduction of the induction temperature decreased this *in vivo* cleavage but also decreased overall protein production. We found that a suitable balance with a good yield of the SBD fusion was obtained when cells were grown and induced at 37 °C for 2 h. Essentially complete release of the ATPase domain from the fusion protein bound to the affinity column was achieved with a fairly long incubation 16 h. We found that binding the SBD to the affinity column at low temperature and pH 8.5 minimized premature cleavage. Lowering the pH and raising the temperature led to cleavage in acceptable yield. Figure 3 shows the purity of the cleaved domain products.

DnaK(1–552)_{ye} T199A T383C can be generated by chemical ligation of the individual domains

Many conditions were tested to optimize ligation of the two domains of DnaK; pH, temperature, time, concentrations of reacting domains, relative ratios of reacting domains, and concentration of added urea were all varied. The ligation reaction efficiency was monitored by SDS-PAGE as a 60 kDa band corresponding to the ligated product molecular mass. [From now on, “ligated protein” will refer to the DnaK(1–552)_{ye} T199A T383C that resulted from the ligation reaction.]

A factor that may govern ligation efficiency is the accessibility of the reactive termini. L382 is the residue at the C-terminus of the ATPase domain that contains the α -thioester to be attacked by the reactive group of the N-terminal cysteine of the SBD in the ligation reaction. By visual inspection of the structure of ADP-bound DnaK (Figure 4, PDB code: 2KHO25), it is evident that the side chains of L382 and T383 are not completely accessible to the solvent. Their side chain ASAs in the context of the full-length protein are 33 and 14% (for L382 and T383, respectively, calculated using VADAR35). We tested the hypothesis that a small amount of a chemical denaturant (here, urea) might cause ‘fraying’ of the termini and enhance the accessibility of the reactive group. At micromolar concentrations of reacting domains and using an approximate molar ratio of 1:7 ATPase domain:SBD, we found that a urea concentration of 2.0 M enhanced ligation efficiency (Figure 5a and b). We interpret the diminution in ligation efficiency at higher urea concentrations as arising from disruption of a weak but specific domain-domain association. At lower urea concentrations (<2.0 M), the favorable interaction between the domains may occur, but the poor accessibility of the area around the reactive groups (383 in SBD and 382 in the ATPase domain) decreases the ligation yield. We are currently exploring to what extent the two domains of DnaK interact when they are not covalently connected.

As the optimization of the ligation continued, it was found that in order to obtain reproducible high yields of ligated product, it was essential to have both ATPase domain and SBD at high (mM) concentrations, and an excess of the SBD domain over the ATPase domain (typically an approximate molar ratio of 1:7 ATPase domain:SBD was used). Under

these conditions, the addition of urea did not significantly improve the ligation (not shown), presumably because the reaction was no longer limited by accessibility of the termini when concentrations are high enough to drive interdomain association.

As can be observed in Figure 5c, even after 16 h of ligation an excess of unreacted domains remains in the mixture. Purification of the ligated product from the unreacted domains was challenging, as the pI of the species in the mixture is very similar, and at such high concentrations they form heterogeneous, high molecular weight complexes that are difficult to disrupt. Ultimately, purification of the ligated product was effected using an anion exchange step and a gel filtration separation under harsh denaturing conditions (6.0 M GuHCl). After refolding, the ligated product could be used directly for functional studies (unlabeled ligated protein) or NMR (semi-labeled protein). Mass spectrometry analysis of the purified ligated DnaK confirmed the identity of the desired product (observed mass 59870.2 Da; average mass calculated for the ligated protein lacking the initial Met = 59870.5).

Typically, the amount of ATPase domain obtained from 1 L of triply labeled culture and SBD obtained from 6 L of unlabeled culture yielded ~300 μ l of an ~200 μ M ligated protein sample.

Ligated DnaK is allosterically functional

To insure that the chemical ligation and purification/refolding steps had not compromised the DnaK molecule, we put the ligated product through a battery of functional and structural tests. DnaK contains only one Trp residue (W102) located in the ATPase domain (see Figure 4a). The interdomain docking accompanying ATP binding (and not hydrolysis^{27,47}) leads to sequestration of W102 from the solvent with an associated blue shift and intensity drop. ^{30,48} Figure 6 shows that the ligated DnaK displays fluorescence changes upon ATP addition that are comparable to those of DnaK wild type or DnaK T383C (to assess the effect of the Cys mutation in the purified protein). This result indicates that the ligated DnaK is able to bind ATP and undergo the conformational changes associated with binding. Also, it shows that the T383C mutation does not have a significant effect on ATP binding.

The ATPase activity of the ligated protein was tested using a coupled enzyme colorimetric assay.³⁹ DnaK wild type displays a steady-state activity 0.05 ($\pm$ 0.027) mol ATP $\times$ DnaK⁻¹ $\times$ min⁻¹ at 30 °C while DnaK T383C shows an activity of 0.07 ($\pm$ 0.002) mol ATP $\times$ DnaK⁻¹ $\times$ min⁻¹ (Table I and Figure 7). The ligated product shows a slightly reduced hydrolysis rate, but this rate is within the variability observed for different DnaK isolates, which we attribute to both inherent errors in concentration and possible contamination with very small amounts of other ATPases. Peptide activation of ATPase rate is an essential allosteric function of Hsp70 chaperones. Importantly, the ATPase activity of all three proteins tested can be substantially stimulated by the addition of the substrate peptide p5, arguing for proper allosteric function.^{49,50}

Striking NMR spectral simplification is achieved using segmentally labeled DnaK

Our previous solution NMR study showed the two domains of nucleotide-free or ADP-bound DnaK to behave independently with no stable interdomain interaction.²⁴ Compelling evidence for this interpretation was that the two-dimensional NMR spectra of the individually expressed domains of DnaK showed the same chemical shifts as in the context of the two-domain construct and that the resonance line widths were only slightly broader than those of the individual domains. Conversely, upon ATP binding, global chemical shift differences are observed in the spectrum of the two-domain protein relative to the isolated SBD and ATPase domains as a consequence of major conformational changes throughout

the protein when domains come together. Moreover, as the domains dock on one another, the protein tumbles as a 60 kDa entity causing the NMR signals to broaden. Additionally, several well-dispersed SBD peaks disappear after ATP binding as a result of conformational exchange. All these phenomena complicate the analysis of the docked state of DnaK and cause major challenges for structural characterization of all the nucleotide states necessary to understand the functioning of the allosteric mechanism.

In this work, we simplified the NMR spectra of DnaK by making visible only the ATPase domain within the context of the intact protein. Figure 8 shows the NMR ^1H , ^{15}N TROSY spectrum of the ATPase domain-labeled ligated DnaK(1–552)_{ye} T199A T383C protein overlaid on the spectrum of the uniformly labeled version. It is clear in the figure that significant spectral simplification is obtained using the segmental labeling strategy, but there is still considerable overlap. For a protein of the size of DnaK, or even the ATPase domain itself, 3D spectra are essential for NMR analysis. In the 3D HNCO spectrum of the uniformly labeled ATP-bound DnaK (not shown), a total of 481 backbone peaks can be resolved (90.8% of those expected for the total of non-proline residues). For the segmentally labeled protein, 142 of these disappear because the SBD is unlabeled, making the process of identifying spin systems substantially more feasible. Even a preliminary effort using the segmentally labeled sample allowed 211 ATPase domain backbone peaks to be assigned in the context of the ATP-bound two-domain DnaK construct by transferring assignments from the ATPase domain alone. Here, the major advantage of segmental labeling is reduction of the number of spin systems that must be assigned, as there remain many overlapping signals from the ATPase domain in the segmentally labeled product (Fig. 8). Nonetheless, the removal of 142 spin systems that are attributable to the SBD reduced overlap of peaks that have yet to be assigned in the 2D TROSY spectrum from 113 to 33. Thus, our results show that segmental labeling of DnaK indeed dramatically simplified spectra, decreased the number of overlapping peaks, and reduced spectral degeneracy. The resulting spectral simplification positions us to do a complete analysis on the structural and dynamic changes in the full-length DnaK molecular chaperone.

Conclusions

The goal of this research was to design a strategy for EPL that would enable us to introduce biophysical labels into the Hsp70 molecular chaperones as a means to elucidate their mechanism of allosteric interdomain communication. Of specific use in our NMR studies of the *E. coli* Hsp70 DnaK, this work has more general significance: It illustrates the necessary steps in the use of EPL with a large (>30 kDa) multidomain protein that has functionally important conformational rearrangements, a type of system for which ligation strategies are relatively unproven.

The first step in the use of EPL is choice of ligation site. Here, this step proved to be quite difficult, as the interdomain linker region, which seemed to be a logical choice for a native state ligation, participates directly in allostery and is therefore intolerant of mutation. By performing a Cys-scanning mutagenesis through the linker and immediate flanking sequences, we were able to identify a site that tolerated a Cys substitution with retention of function. This finding put us in a favorable position to ligate the individual domains of DnaK and also as a bonus to use this Cys mutant for the study of DnaK by other techniques requiring labeling of a reactive side chain.

Optimization of ligation required testing a very large array of parameters, including temperature, pH, duration of reaction, concentration and ratios of reactant domains, and presence of a small concentration of chemical denaturant. We believe that these parameters are case-specific, which is one of the contributing reasons to the large energy barrier to use

of EPL. On the other hand, intimate knowledge of the target system will help the investigator to choose initial parameters to test. For example, we found that addition of a small amount of urea favored the ligation efficiency, when concentrations of reacting domains were relatively low. Higher concentrations of the reaction domains, however, led to higher yields of ligated product and loss of the favorable effect of urea. In addition to challenges of optimizing ligation yields, development of strategies for purification of ligated product away from unreacted domains is an essential stage in the use of EPL and should be considered when the EPL strategy is initially designed.

Ultimately, we successfully used EPL to generate a DnaK molecule wherein only the ATPase domain was ^2H , ^{13}C and ^{15}N labeled, making possible NMR analysis of an allosterically functional two-domain 60 kDa chaperone in all of its ligand-bound states. Although the optimization of the ligation conditions and purification of the ligated product were challenging, ligated DnaK could be prepared in amounts sufficient for NMR studies. Additionally, the implementation of this strategy allowed us to have in hand a method that can facilitate study of DnaK by different biophysical techniques, such as EPR and fluorescence spectroscopy, which can help provide answers to essential questions about allosteric functioning in Hsp70s.

Acknowledgments

We thank Beena Krishnan for helpful discussion throughout the course of this work. This work was supported by NIH Grant GM027616.

References

1. Smock RG, Gierasch LM. *Science*. 2009; 324:198–203. [PubMed: 19359576]
2. Roy R, Hohng S, Ha T. *Nat Methods*. 2008; 5:507–516. [PubMed: 18511918]
3. Jager M, Michalet X, Weiss S. *Prot Sci*. 2005; 14:2059–2068.
4. Ohta A, Yamagishi Y, Suga H. *Curr Opin Chem Biol*. 2008; 12:159–167. [PubMed: 18249198]
5. Wang L, Xie J, Schultz PG. *Annu Rev Biophys Biomol Struct*. 2006; 35:225–249. [PubMed: 16689635]
6. Muir TW, Sondhi D, Cole PA. *Proc Natl Acad Sci U S A*. 1998; 95:6705–6710. [PubMed: 9618476]
7. Muralidharan V, Muir TW. *Nat Methods*. 2006; 3:429–438. [PubMed: 16721376]
8. Severinov K, Muir TW. *J Biol Chem*. 1998; 273:16205–16209. [PubMed: 9632677]
9. Camarero JA, Shekhtman A, Campbell EA, Chlenov M, Gruber TM, Bryant DA, Darst SA, Cowburn D, Muir TW. *Proc Natl Acad Sci U S A*. 2002; 99:8536–8541. [PubMed: 12084914]
10. Skrisovska L, Allain FHT. *J Mol Biol*. 2008; 375:151–164. [PubMed: 17936301]
11. Xu R, Ayers B, Cowburn D, Muir TW. *Proc Natl Acad Sci U S A*. 1999; 96:388–393. [PubMed: 9892643]
12. Yagi H, Tsujimoto T, Yamazaki T, Yoshida M, Akutsu H. *J Am Chem Soc*. 2004; 126:16632–16638. [PubMed: 15600369]
13. Mayer MP, Bukau B. *Cell Mol Life Sci*. 2005; 62:670–684. [PubMed: 15770419]
14. Albanese V, Yam AY, Baughman J, Parnot C, Frydman J. *Cell*. 2006; 124:75–88. [PubMed: 16413483]
15. Frydman J, Nimmesgern E, Ohtsuka K, Hartl FU. *Nature*. 1994; 370:111–117. [PubMed: 8022479]
16. Hartl FU, Hayer-Hartl M. *Science*. 2002; 295:1852–1858. [PubMed: 11884745]
17. Melville MW, McClellan AJ, Meyer AS, Darveau A, Frydman J. *Mol Cell Biol*. 2003; 23:3141–3151. [PubMed: 12697815]
18. Yam AY, Albanese V, Lin HT, Frydman J. *J Biol Chem*. 2005; 280:41252–41261. [PubMed: 16219770]

19. Young JC, Agashe VR, Siegers K, Hartl FU. *Nat Rev Mol Cell Biol.* 2004; 5:781–791. [PubMed: 15459659]
20. Han W, Christen P. *FEBS Lett.* 2001; 497:55–58. [PubMed: 11376662]
21. Laufen T, Mayer MP, Beisel C, Klostermeier D, Mogk A, Reinstein J, Bukau B. *Proc Natl Acad Sci USA.* 1999; 96:5452–5457. [PubMed: 10318904]
22. Mayer MP, Laufen T, Paal K, McCarty JS, Bukau B. *J Mol Biol.* 1999; 289:1131–1144. [PubMed: 10369787]
23. Vogel M, Mayer MP, Bukau B. *J Biol Chem.* 2006; 281:38705–38711. [PubMed: 17052976]
24. Swain JF, Dinler G, Sivendran R, Montgomery DL, Stotz M, Gierasch LM. *Mol Cell.* 2007; 26:27–39. [PubMed: 17434124]
25. Bertelsen EB, Chang L, Gestwicki JE, Zuiderweg ER. *Proc Natl Acad Sci U S A.* 2009; 106:8471–8476. [PubMed: 19439666]
26. Revington M, Zhang Y, Yip GN, Kurochkin AV, Zuiderweg ER. *J Mol Biol.* 2005; 349:163–183. [PubMed: 15876376]
27. Buchberger A, Theyssen H, Schroder H, McCarty JS, Virgallita G, Milkereit P, Reinstein J, Bukau B. *J Biol Chem.* 1995; 270:16903–16910. [PubMed: 7622507]
28. Rist W, Graf C, Bukau B, Mayer MP. *J Biol Chem.* 2006; 281:16493–16501. [PubMed: 16613854]
29. Moro F, Fernandez V, Muga A. *FEBS Lett.* 2003; 533:119–123. [PubMed: 12505170]
30. Palleros DR, Reid KL, McCarty JS, Walker GC, Fink AL. *J Biol Chem.* 1992; 267:5279–5285. [PubMed: 1544910]
31. Wilbanks SM, Chen L, Tsuruta H, Hodgson KO, McKay DB. *Biochemistry.* 1995; 34:12095–12106. [PubMed: 7547949]
32. Liu Q, Hendrickson WA. *Cell.* 2007; 131:106–120. [PubMed: 17923091]
33. Swain JF, Schulz EG, Gierasch LM. *J Biol Chem.* 2006; 281:1605–1611. [PubMed: 16275641]
34. Suh WC, Lu CZ, Gross CA. *J Biol Chem.* 1999; 274:30534–30539. [PubMed: 10521435]
35. Willard L, Ranjan A, Zhang H, Monzavi H, Boyko RF, Sykes BD, Wishart DS. *Nuc Acids Res.* 2003; 31:3316–3319.
36. Montgomery DL, Morimoto RI, Gierasch LM. *J Mol Biol.* 1999; 286:915–932. [PubMed: 10024459]
37. Tugarinov V, Kanelis V, Kay LE. *Nat Protocols.* 2006; 1:749–754.
38. Bukau B, Walker GC. *EMBO J.* 1990; 9:4027–4036. [PubMed: 2249663]
39. Norby JG. *Methods Enzymol.* 1988; 156:116–119. [PubMed: 2835597]
40. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. *J Biomol NMR.* 1995; 6:277–293. [PubMed: 8520220]
41. Keller R, Grace CR, Riek R. *Magn Reson Chem.* 2006; 44:196–205.
42. Dawson PE, Muir TW, Clark-Lewis I, Kent SB. *Science.* 1994; 266:776–779. [PubMed: 7973629]
43. Cegielska A, Georgopoulos C. 1989; 264:21122–21130.
44. Liberek K, Osipiuk J, Zylicz M, Ang D, Skorko J, Georgopoulos C. *J Biol Chem.* 1990; 265:3022–3029. [PubMed: 2154468]
45. Buchberger A, Valencia A, Schröder H, Büttner M, Bukau B. *Nat Struct Biol.* 1994; 1:95–101. [PubMed: 7656024]
46. Buchberger A, Valencia A, McMacken R, Sander C, Bukau B. *EMBO J.* 1994; 13:1687–1695. [PubMed: 7908876]
47. Theyssen H, Schuster HP, Packschies L, Bukau B, Reinstein J. *J Mol Biol.* 1996; 263:657–670. [PubMed: 8947566]
48. Banecki B, Zylicz M, Bertoli E, Tanfani F. *J Biol Chem.* 1992; 267:25051–25058. [PubMed: 1460007]
49. Jordan R, McMacken R. *J Biol Chem.* 1995; 270:4563–4569. [PubMed: 7876226]
50. Sadis S, Hightower LE. *Biochemistry.* 1992; 31:9406–9412. [PubMed: 1356434]

Figure 1. Strategy for expressed protein ligation of the *E. coli* Hsp70 DnaK

The ATPase domain of DnaK was cloned in pTwin1 vector (NEB) upstream of the *Mxe* intein and the CBD. The SBD was cloned downstream of the CBD and the *Ssp* intein in the same vector. Both domains were over-expressed in *E. coli* and bound to a chitin-binding affinity column (CBC). Cleavage from the intein fusion was induced by treatment with MESNA at basic pH for the ATPase domain cleavage and by low pH and elevated temperature for the SBD cleavage (see Methods for details). The two cleaved and purified domains (the ATPase domain with a C-terminal α -thioester and SBD with a free N-terminal Cys) were mixed together in the presence of MESNA under optimized conditions (see text) to obtain the ligated DnaK protein DnaK(1-552)_{ye} T199A T383C.

Figure 2. *In vivo* functionality of various DnaK Cys-scanning mutants. A. λ phage assay The number of λ phage plaques produced in *E. coli* BB1553 cells expressing each DnaK are compared to the number of plaques formed in cells expressing DnaK wild type. **B. Heat shock assay.** Liquid cultures of *E. coli* BB1553 expressing DnaK wild type, DnaK T383C and DnaK T199A were grown at 43 °C and serial dilutions spotted on a plate.

Figure 3. Expression and cleavage of the ATPase domain and SBD intein fusions
 Coomassie-blue stained SDS-PAGE showing the ATPase domain and SBD intein fusion proteins over-expressed in *E. coli*. Uninduced and IPTG-induced cultures are indicated. Purified, cleaved ATPase domain with a C-terminal α -thioester and SBD with an N-terminal Cys are shown.

Figure 4. Model of ADP-bound DnaK showing linkage region

A. Two-domain ADP-bound DnaK (from KHO PDB entry 25 truncated at residue 552) is depicted (Image prepared using PyMol (DeLano Scientific LLC, <http://www.pymol.org>)). The ATPase domain is blue, the SBD is green, and the interdomain linker is yellow. T383, T382 and W102 are shown as spheres in magenta, red, and orange, respectively. **B.** The area enclosed by a box in A has been enlarged to show the region around the linkage site. All the atoms within 5 Å of L382 and T383 are shown as blue spheres.

Figure 5. Optimization of the EPL reaction

A. DnaK protein band in a Coomassie-stained SDS-PAGE obtained after ligation of the individual domains at the indicated urea concentrations. This is an example of one of the experiments quantified in **B**. **B.** Quantification of the DnaK protein band in a Coomassie-stained SDS-PAGE obtained after ligation of the individual domains at different urea concentrations under dilute (micromolar) conditions as described in Materials and Methods (plotted using Kaleidagraph (Synergy Software, Reading, PA)). Bands were quantified using GeneTools program (Syngene, Frederick, MD), corrected by protein amount using a contaminating protein band for which the concentration was constant in all the reactions, and normalized to 1 for the ligated protein obtained in the absence of urea. Each bar is the average of at least four experiments. **C.** Coomassie-stained SDS-PAGE showing the ligated DnaK obtained from the indicated relative ratios of the individual domains in mM concentrations as described (in the presence of 2 M urea, pH 8.0). The last SDS-PAGE lane shows the ligated DnaK obtained after purification.

Figure 6. Fluorescence assay for ATP-induced conformational change

A. Trp fluorescence spectra in the absence of ATP (filled symbols) or in the presence of 5 mM ATP (empty symbols) for DnaK wild type (circles), DnaK T383C (diamonds) and ligated DnaK (squares). Relative fluorescence plotted is normalized between 0 and 1 for the spectra in the absence of ATP. **B.** W102 fluorescence blue shift quantified as the shift in the emission maximum upon ATP binding, as previously described.³⁶ The bars represent the average of at least two experiments.

Figure 7. Test for allostery using peptide activation of steady-state ATP hydrolysis rate
Steady-state ATP hydrolysis rate as mol ATP hydrolyzed mol⁻¹ DnaK per min for DnaK wild type (white bars), DnaK T383C (light gray bars) and ligated DnaK (dark gray bars) in the absence and presence of substrate peptide p5 as indicated.

Figure 8. ^1H , ^{15}N TROSY NMR spectra

NMR spectrum of the ligated segmentally labeled (ATPase domain only) DnaK(1–552)_{ye} T199A T383C (red) overlaid on that of uniformly labeled DnaK(1–552)_{ye} T199A T383C (black), both in the presence of 5 mM ATP.

Table I

ATP hydrolysis rate.

DnaK	ATPase rate^a	SD^b	Stimulation^c
Wild Type	0.05	0.027	
Wild Type +p5	0.27	0.015	4.8
T383C	0.07	0.002	
T383C +p5	0.26	0.007	3.7
Ligated	0.02	0.002	
Ligated +p5	0.12	0.003	8.3

^a ATPase rate as $\text{mol ATP} \times \text{DnaK}^{-1} \times \text{min}^{-1}$.

^b The reported rates are an average of at least four independent experiments.

^c Stimulation refers to the ratio of the ATPase activity in the presence of p5 peptide substrate over the basal ATPase activity of the same protein.