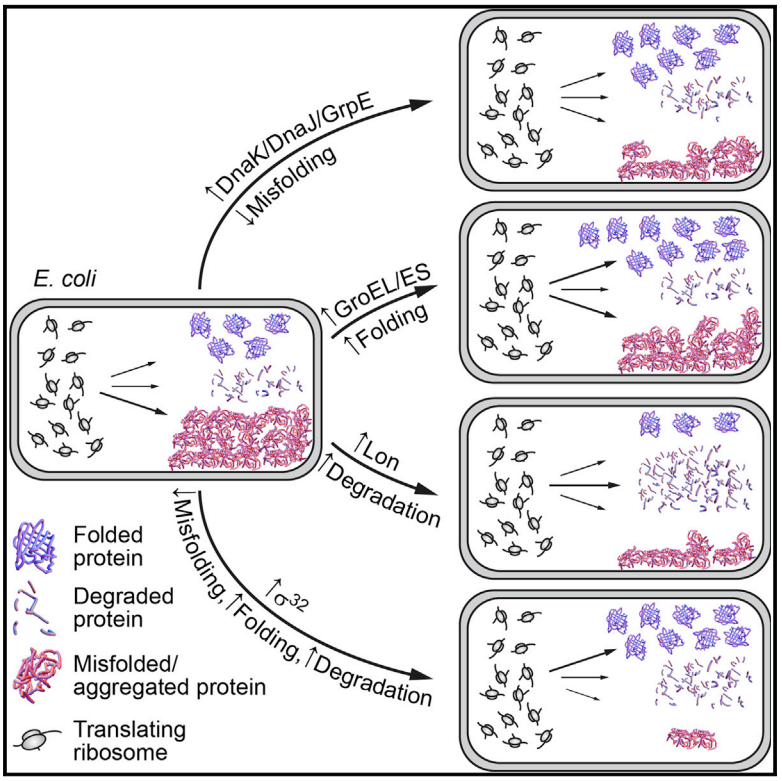
Article

Cell Reports

Individual and Collective Contributions of Chaperoning and Degradation to Protein Homeostasis in *E. coli*

Graphical Abstract



Highlights

- DnaK/DnaJ/GrpE, GroEL/GroES, and Lon dominate the *E. coli* proteostasis network.
- The effects of the proteostasis network components are largely additive.
- Modeling reveals relationships between chaperone efficacy and folding energetics.

Authors

Younhee Cho, Xin Zhang, ...,
Lila M. Gierasch, Evan T. Powers

Correspondence

gierasch@biochem.umass.edu (L.M.G.),
epowers@scripps.edu (E.T.P.)

In Brief

Cho et al. interrogate the folding fate of three test proteins in *E. coli* as a function of the composition of the proteostasis network. Quantitative data analysis using both phenomenological and mechanistic models reveals relationships between folding energy landscapes and the efficacy of the proteostasis network as a system.

Cho et al., 2015, Cell Reports *11*, 321–333
April 14, 2015
http://dx.doi.org/10.1016/j.celrep.2015.03.018

CellPress

Cell Reports
Article

Individual and Collective Contributions of Chaperoning and Degradation to Protein Homeostasis in *E. coli*

Younhee Cho,[1,2,6] Xin Zhang,[1,2,6] Kristine Faye R. Pobre,[3] Yu Liu,[1,2] David L. Powers,[4] Jeffery W. Kelly,[1,2,5] Lila M. Gierasch,[3,*] and Evan T. Powers[2,*]
[1]Department of Molecular and Experimental Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA
[2]Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037, USA
[3]Departments of Chemistry and Biochemistry and Molecular Biology, University of Massachusetts-Amherst, Amherst, MA 01003, USA
[4]Department of Mathematics and Computer Science, Clarkson University, Potsdam, NY 13699, USA
[5]The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
[6]Co-first author
*Correspondence: gierasch@biochem.umass.edu (L.M.G.), epowers@scripps.edu (E.T.P.)
http://dx.doi.org/10.1016/j.celrep.2015.03.018

SUMMARY

The folding fate of a protein in vivo is determined by the interplay between a protein's folding energy landscape and the actions of the proteostasis network, including molecular chaperones and degradation enzymes. The mechanisms of individual components of the *E. coli* proteostasis network have been studied extensively, but much less is known about how they function as a system. We used an integrated experimental and computational approach to quantitatively analyze the folding outcomes (native folding versus aggregation versus degradation) of three test proteins biosynthesized in *E. coli* under a variety of conditions. Overexpression of the entire proteostasis network benefited all three test proteins, but the effect of upregulating individual chaperones or the major degradation enzyme, Lon, varied for proteins with different biophysical properties. In sum, the impact of the *E. coli* proteostasis network is a consequence of concerted action by the Hsp70 system (DnaK/DnaJ/GrpE), the Hsp60 system (GroEL/GroES), and Lon.

INTRODUCTION

Protein homeostasis, or proteostasis, is achieved when an organism has enough natively folded proteins to carry out its essential functions but not enough misfolded and aggregated proteins to interfere with organismal fitness (Balch et al., 2008; Powers et al., 2009). In a simplified view of proteostasis, new proteins can have three fates: they can fold to their native states, they can misfold and/or aggregate, or they can be degraded (Figure 1). Proteins that experience the latter two fates are not functional. All organisms regulate the health of their proteomes via a collection of chaperones, folding enzymes, proteases, and other components that together make up the proteostasis network (PN) (Kim et al., 2013; Powers and Balch, 2013).

Since protein folding, misfolding, and aggregation equilibria are linked, PN components that modulate any of these processes indirectly affect the others. However, each PN component seems to have an "assigned responsibility": a process that it affects most directly. Using the *E. coli* PN as an example, native folding is promoted most directly by GroEL and GroES, the *E. coli* chaperonin/co-chaperonin pair (Chapman et al., 2006; Horwich and Fenton, 2009). Misfolding and aggregation are opposed most directly by DnaK, DnaJ, and GrpE, the *E. coli* Hsp70/Hsp40/nucleotide exchange factor trio (Calloni et al., 2012; Mayer and Bukau, 2005; Sharma et al., 2010), and by the collaboration of this trio with the disaggregating chaperone ClpB (Doyle et al., 2013). Finally, many proteases degrade proteins in *E. coli*, but Lon appears to be the most important for degrading misfolded protein (Gottesman, 1996; Gur and Sauer, 2008).

While the main functions of individual PN components are fairly well understood, their contributions to the integrated, system-level function of the whole PN are not as clear (Bershtein et al., 2013; Dickson and Brooks, 2013; Hingorani and Gierasch, 2014; Kim et al., 2013; Powers et al., 2012; Wiseman et al., 2007). How do PN components complement each other? Do they perform multiple or redundant functions? To what extent does a protein's folding energy landscape determine its route through the PN and its fate? We have used a combination of experiments and computational modeling to address these and related questions.

Here, we focus on how proteins with low stabilities behave when overexpressed in *E. coli* because expression of such proteins challenges proteostasis (Gidalevitz et al., 2006; Olzscha et al., 2011). We chose *E. coli* as a model organism because of the availability of FoldEco (Powers et al., 2012), a computational model of *E. coli*'s PN that is essential to answer the mechanistic questions posed above. In addition, *E. coli* is widely used as a microbial factory for producing heterologous

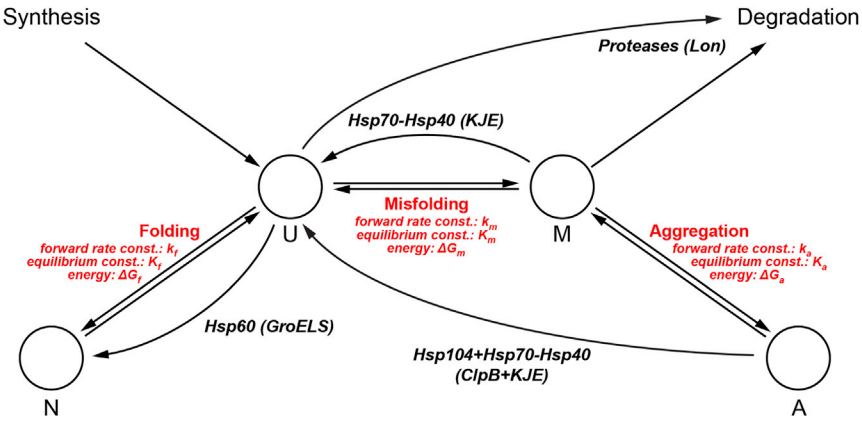


Figure 1. Schematic of Kinetic Partitioning during Protein Folding In Vivo

Protein folding in vivo begins with ribosomal synthesis ("Synthesis"), which yields unfolded protein molecules ("U"; note that we neglect the possibility of co-translational folding). Unfolded protein can fold to the native state ("N"), misfold to the misfolded state ("M"), or be degraded ("Degradation"). Misfolded protein can be degraded or self-associate to form aggregates ("A"; note that aggregation is reversible in principle, as shown here, but in many cases is irreversible in practice unless assisted by the proteostasis network). A protein's folding energy landscape dictates its partitioning among the unfolded, native, misfolded, and aggregated states in vitro (red text). However, each folding process is modulated by components of the proteostasis network in vivo. In *E. coli*, DnaK, DnaJ, and GrpE (KJE; the Hsp70/Hsp40/nucelotide exchange factor system) oppose misfolding by binding to misfolded protein molecules and forcing them to resume the unfolded state. GroEL and GroES (GroELS; the Hsp60//Hsp10 chaperonin system) promote folding by encapsulating unfolded protein molecules and enabling them to fold in an isolated cavity. Degradation is carried out by proteases, in particular Lon. Finally, ClpB (Hsp104) collaborates with KJE to solubilize aggregates.

proteins. Failures in proteostasis were quantified by measuring total expression levels and the amount of aggregated versus soluble protein. These quantities report on the extent of degradation, aggregation, and native folding experienced by our test proteins and therefore cover each of a protein's potential fates. We interrogated the roles of the DnaK/DnaJ/GrpE pathway (KJE), the GroEL/GroES pathway (GroELS), and Lon by overexpressing these PN components individually or in combinations. The point at which, and the extent to which, proteostasis failed for the test proteins then informed us as to the limits of the *E. coli* PN.

The test proteins in this work are unstable variants of *E. coli* dihydrofolate reductase (EcDHFR), murine cellular retinoic acid-binding protein 1 (MmCRABP1), and a de novo designed retroaldolase enzyme (RA114.3). These proteins span a range of origins (endogenous *E. coli* versus mammalian versus de novo designed, respectively) and folds ($\alpha\beta\alpha$ sandwich, β barrel, and α/β barrel, respectively; Figure 2) (Bjelic et al., 2014; Kleywegt et al., 1994; Liu et al., 2014; Sawaya and Kraut, 1997) and have no significant sequence similarity. By examining how each member of this diverse group partitions between being soluble, aggregating, and being degraded as a function of the composition of the *E. coli* PN, we hoped to extract general lessons about the attributes of the PN as a system in its interactions with as broad as possible a selection of proteins, as well as lessons about the dominant contributors to the PN's various functions. It is important to note here that this undertaking requires the assumption that the PN of *E. coli* handles heterologous proteins and its own endogenous proteins similarly. While a PN component from one organism generally cannot complement the loss of the orthologous component in another organism, chaperones from one organism are generally capable of assisting the folding of proteins from another. For example, upregulation of *E. coli* chaperones improved the expression yields for most of a set of 64 heterologous proteins (de Marco et al., 2007). Thus, KJE, GroELS, and Lon from *E. coli* appear to be quite general in their selection and handling of substrates.

RESULTS

The Test Proteins and Their Overexpression in the *E. coli* Proteostasis Network in the Absence of Other Perturbations

The test proteins studied here are the M42T/H114R mutant of EcDHFR (m-EcDHFR), the R131Q/Y133S mutant of MmCRABP1 (m-MmCRABP1), and the E10K/D120V/N124S/L225P mutant of RA114.3 with a C-terminal His tag (m-RA114). These proteins are small to medium sized (DHFR: 159 amino acids; CRABP1: 137 amino acids; RA114: 258 amino acids) and monomeric (Figure 2). That each of these mutants is less stable than the corresponding wild-type (WT) protein is demonstrated by their susceptibilities to urea denaturation for m-EcDHFR and m-RA114 (Figures S1A and S1B) and was reported previously for m-MmCRABP1 (Budyak et al., 2013).

Each of these test proteins was expressed by isopropyl β-d-1-thiogalactopyranoside (IPTG) induction of a plasmid under control of the lac promoter in *E. coli* K12 HMS174 (DE3) cells growing in Luria-Bertani media at 30°C. After 2-hr induction, the cells were lysed, and aggregated and soluble proteins were separated by centrifugation. Total, soluble, and aggregated (i.e., in the pellet) protein fractions were analyzed by SDS-PAGE and the absolute amount of test protein in each was determined by comparison to a calibration line constructed using purified recombinant protein (Figure S1C; Supplemental Experimental Procedures). The total protein concentrations present at the end of the 2-hr expression period were 498 ± 58, 106 ± 5, and 385 ± 40 μM (mean ± SEM) for m-EcDHFR, m-MmCRABP1, and m-RA114, respectively. Substantial aggregation was observed for each test protein, with the aggregated fractions amounting to 46% ± 3%, 76% ± 1%, and 86% ± 1% of total protein for m-EcDHFR, m-MmCRABP1, and m-RA114, respectively (Figures 2B and 3). Thus, m-EcDHFR appears to be the best behaved of our test proteins, while m-MmCRABP1 and m-RA114 are more aggregation prone. Expression of the WT versions of these proteins under the same conditions resulted in negligible (WT-EcDHFR

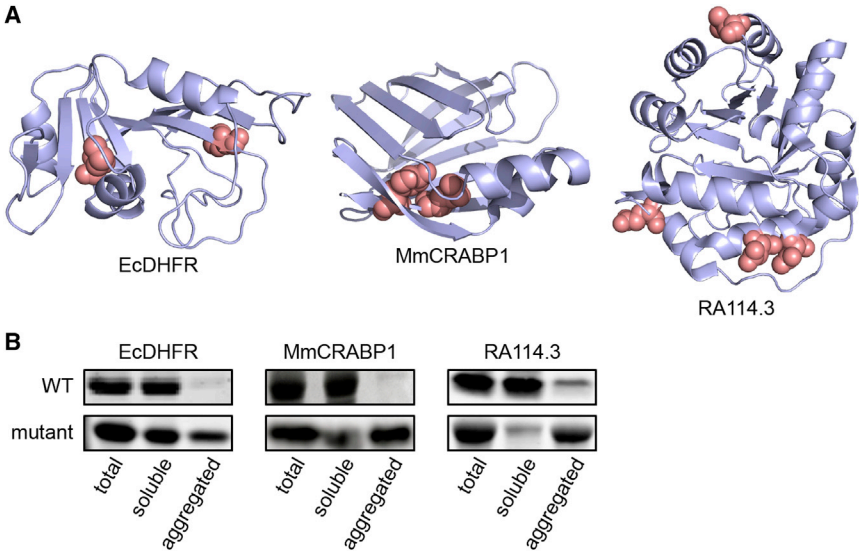


Figure 2. Test Proteins Used in This Work

(A) Structures of wild-type (WT) test proteins: EcDHFR (PDB 1RX2) (Sawaya and Kraut, 1997), MmCRABP1 (PDB 2CBR) (Kleywegt et al., 1994), and RA114.3 (PDB 4OU1) (Liu et al., 2014). The sites of the mutations in the destabilized test proteins are highlighted as pink spheres.

(B) Bands corresponding to the test proteins in SDS-PAGE gels run on samples derived from *E. coli* overexpressing the WT and mutant forms of the test proteins for 2 hr at 30°C. The lane labeled "total" is from pre-centrifugation cell lysates, and the lanes labeled "soluble" and "aggregated" are from the supernatants and pellets, respectively, after cell lysates were centrifuged for 10 min at 13,500 × *g*. The WT variants of the test proteins form little or no aggregates when overexpressed in *E. coli*, but the mutants aggregate substantially. See also Figure S1.

and WT-MmCRABP1) to low (WT-RA114) levels of aggregates (Figure 2B).

Heterologous protein expression can cause stress and lead to the upregulation of PN components (Gasser et al., 2008; Hoffmann and Rinas, 2004). We therefore measured the levels of GroEL, DnaK, and Lon before and 2 hr after induction of m-EcDHFR and m-RA114. Compared to an empty vector control, DnaK and GroEL increased ~20% to 40% and Lon ~80% to 160% (Figure S2A). Similar increases in PN component levels were observed when WT-RA114, which is more stable than m-RA114 but still aggregates, was overexpressed. In contrast, overexpression of WT-EcDHFR, which is stable and well behaved, resulted in much smaller changes in PN component levels (Figure S2A). These results suggest that it is the overexpression of aggregation-prone proteins, and not the overexpression of proteins per se, that causes PN component levels to increase. The PN as it exists after being perturbed by expression of the test proteins will be referred to as the "adapted-basal" PN.

Individual Upregulation of KJE, GroELS, or Lon Moderately Decreases Test-Protein Aggregation

To assess the effects of KJE, GroELS, and Lon on the test proteins, we introduced them into pBAD expression vectors (Figure S2B) so that their expression could be titratably induced by arabinose. Co-transformation with separate plasmids carrying the test protein and the PN components enabled us to express each independently. The expression of KJE, GroELS, or Lon was induced by adding arabinose 1 hr before induction of the destabilized test protein (Figure 3A). The level of upregulation of the PN components was controlled by the concentration of arabinose added. The "high" and "low" upregulation levels resulted in respective concentration increases of ~4- to 8-fold and ~2- to 3-fold for the major PN components (DnaK, GroEL, and Lon; Figures S2C and S2D). The "medium" upregulation level, which was used only for Lon because the high-Lon conditions caused a drastic decrease in total protein levels (see below), increased Lon ~4-fold (Figures S2C and S2D).

Upregulation of KJE decreased, but did not eliminate, aggregation for each of the test proteins. The aggregated fraction decreased significantly from 46% ± 3% to 25% ± 4% of the total protein for m-EcDHFR ($p = 0.001$, one-tailed t test), from 76% ± 1% to 46% ± 5% for m-MmCRABP1 ($p < 0.001$, one-tailed t test), and from 86% ± 1% to 37% ± 3% for m-RA114 ($p < 0.001$, one-tailed t test) under the high-KJE upregulation conditions (Figures 3B–3D and S2E; Table S1). In addition, the total test-protein concentrations decreased by 30%–40% under the high-KJE upregulation conditions (Figures 3B–3D; Table S1). Low-KJE upregulation conditions also decreased total concentrations of the three proteins, but to a lesser extent. We cannot exclude the possibility that this result is due to direct delivery of substrates to proteases by DnaK or DnaJ (Sherman and Goldberg, 1992). However, based on modeling results with FoldEco that are presented in a later section, it appears more likely that the observed decreases in protein concentration are due to proteins that are rescued from misfolding and/or aggregation by KJE being degraded before they can fold or re-aggregate.

Upregulation of GroELS similarly decreased aggregation of m-EcDHFR and m-RA114. The aggregated fractions decreased from 46% ± 3% to 22% ± 9% for m-EcDHFR ($p = 0.004$, one-tailed t test) and from 86% ± 1% to 39% ± 3% for m-RA114 ($p < 0.001$, one-tailed t test) under the high-GroELS upregulation conditions (Figures 3B and 3D; Table S1). In contrast, GroELS upregulation did not significantly decrease the aggregation of m-MmCRABP1 (from 76% ± 1% to 66% ± 11%; $p = 0.26$, one-tailed t test) (Figure 3C; Table S1). This observation suggests that m-MmCRABP1 is not a good substrate for GroELS; we examine this possibility in a later section.

Upregulation of Lon decreased the levels of both the soluble and aggregated forms of the test proteins, but, especially at the low and medium upregulation levels, the decrease was larger for the aggregates than for the soluble protein for two of the

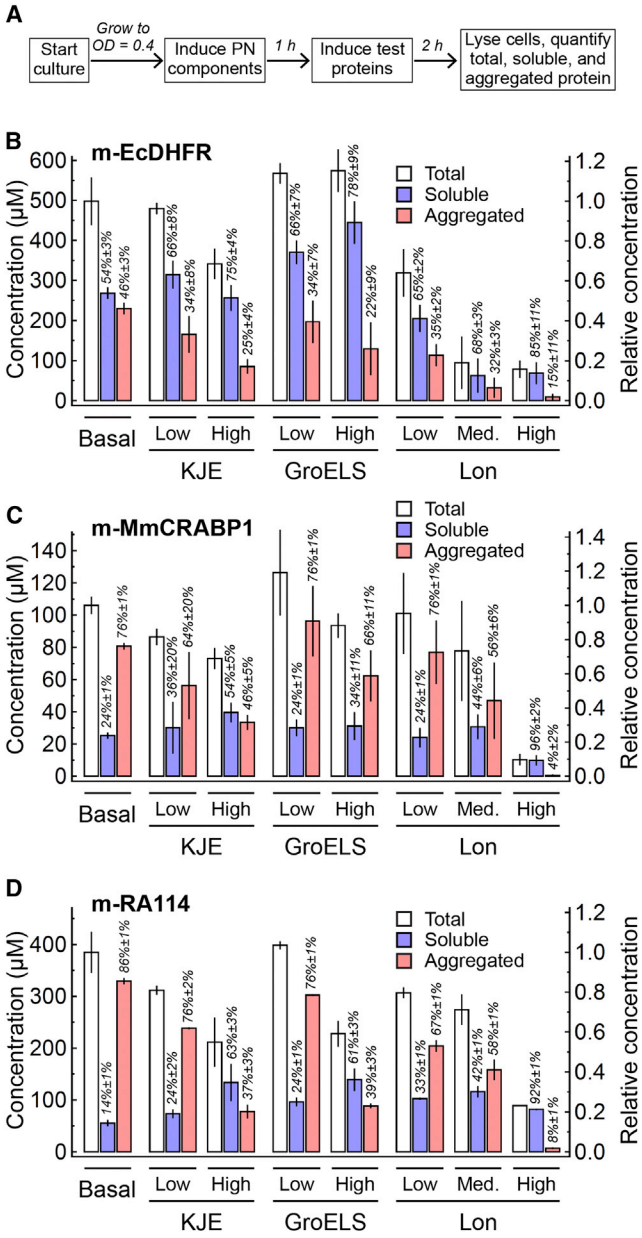


Figure 3. Folding Fates of Test Proteins upon Overexpression in *E. coli* under Adapted-Basal Conditions and with Individual PN Components Upregulated

(A) Experimental scheme for overexpression of test proteins. All experiments were performed at 30°C.

(B) Bar graph showing the results of overexpressing m-EcDHFR in *E. coli* under various conditions. Left axis: cytoplasmic concentration of m-EcDHFR as determined by quantitative analysis of gels like those shown in Figures 2B and S2E. Right axis: concentration of m-EcDHFR relative to the total concentration of m-EcDHFR (soluble + aggregated) produced under adapted-basal conditions. These concentrations are referred to in the text as $[Sol]_{rel,X}$ and $[Agg]_{rel,X}$ for soluble and aggregated forms of a given test protein "X." White bars represent total protein concentration under each condition. Blue bars represent the concentration of soluble protein under each condition (i.e., the concentration in the supernatant after lysis and centrifugation). Red bars represent the concentration of aggregated protein under each condition (i.e., the concentration in the pellet after lysis, centrifugation, and re-suspension). Error bars represent SEM. Numbers above bars are percentage of the total protein that is soluble or aggregated under each condition ± SEM.

(C) As in (B), but for m-MmCRABP1.

(D) As in (B), but for m-RA114. The "high" and "low" upregulation levels resulted in respective concentration increases of ~4- to 8-fold and ~2- to 3-fold for the major PN components (DnaK, GroEL, and Lon; Figures S2C and S2D). The "medium" upregulation level (for Lon only) resulted in a ~4-fold increase in the concentration of Lon (Figures S2C and S2D).

See also Figure S2 and Table S1.

three (m-MmCRABP1 and m-RA114; compare Figure 3B with Figures 3C and 3D).

Upregulation of KJE, GroELS, or Lon in Pairs Further Decreases Test-Protein Aggregation

To determine the effects of joint upregulation of KJE, GroELS, and Lon on our test proteins, we introduced pairs of these systems into the same expression vectors with one system under an arabinose promoter and the other under a tetracycline promoter (Figure S2B). We then repeated the test-protein overexpression experiments described above (Figure 3A), except that the PN pathways were upregulated in pairs using inducer concentrations at the high upregulation level. In all cases, the levels of the PN components did not increase as much as when they were overexpressed on their own (~2- to 4-fold instead of 4- to 8-fold; Figure S3A).

GroELS+Lon and KJE+Lon were the most effective pairs of PN components for suppressing test-protein aggregation (Figure 4; Table S1). The GroELS+KJE combination was less effective at suppressing aggregation (i.e., the fraction aggregated was higher under this condition) than either GroELS+Lon or KJE+Lon for m-MmCRABP1 and m-RA114 (for m-MmCRABP1: $p = 0.0025$ and 0.0016 for GroELS+KJE versus KJE+Lon and GroELS+Lon; for m-RA114: $p = 0.001$ and 0.023 for GroELS+KJE versus KJE+Lon and GroELS+Lon, one-tailed t test). The same trends are apparent for m-EcDHFR, although the null hypothesis cannot be rejected with the same level of confidence ($p = 0.089$ and 0.176 for GroELS+KJE versus KJE+Lon and GroELS+Lon, one-tailed t test).

Simultaneous Upregulation of KJE, GroELS, or Lon via Expression of σ^{32} Virtually Eliminates Aggregation

As noted above, pairwise upregulation of KJE, GroELS, and Lon resulted in lower levels of these PN components than upregulating them individually. Extrapolating this result suggests that it could be difficult to attain sufficiently high levels of upregulation if all three PN pathways were upregulated using arabinose- and tetracycline-induced expression systems. We therefore sought another way to simultaneously upregulate KJE, GroELS, and Lon.

KJE, GroELS, and Lon, as well as many other PN components, are in the regulon of the heat shock transcription factor, σ^{32} (Guisbert et al., 2008; Zhang et al., 2014). We therefore overexpressed the I54N mutant of σ^{32}, which evades the post-translational regulation of σ^{32} (Guisbert et al., 2008; Yura et al., 2007; Zhang et al., 2014), to simultaneously increase the levels of KJE, GroELS, and Lon. Induction of I54N σ^{32} for 1 hr prior to inducing the test proteins yielded 3- to 4-fold increases in the

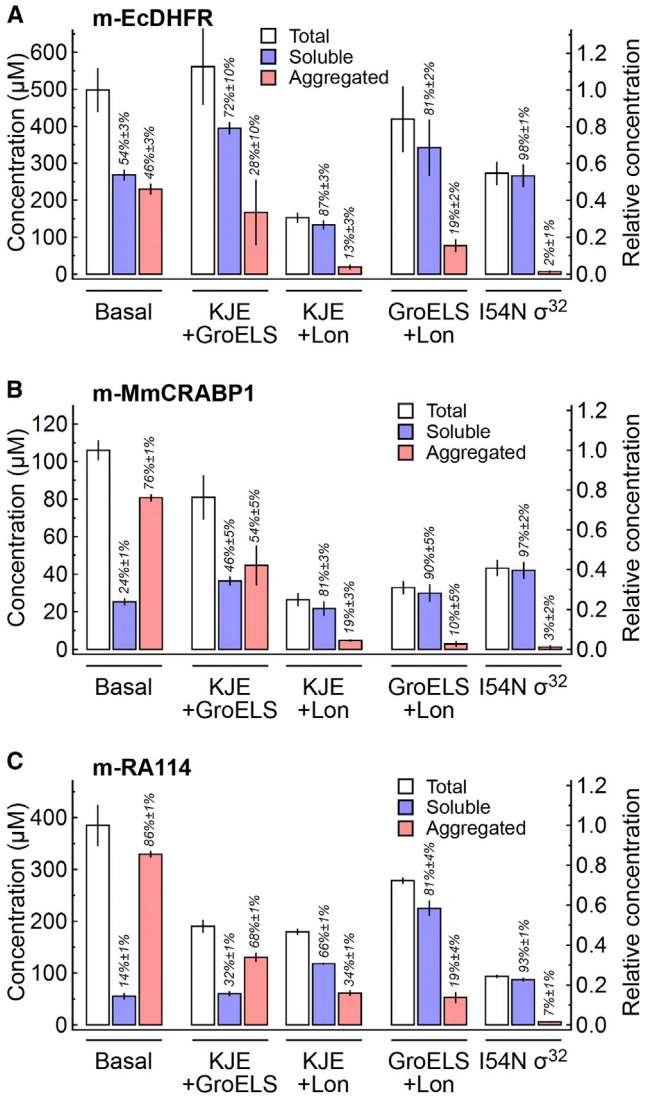


Figure 4. Folding Fates of Test Proteins upon Overexpression in *E. coli* under Adapted-Basal Conditions and with PN Components Upregulated in Pairs or after Overexpression of I54N σ^{32}

(A) As in Figure 3B, but with multiple PN components upregulated.

(B) As in Figure 3C, but with multiple PN components upregulated.

(C) As in Figure 3D, but with multiple PN components upregulated. The levels of DnaK, GroEL, and Lon increased by ~2- to 4-fold when they were upregulated in pairs or via overexpression of I54N σ^{32} (Figure S3A).

Error bars represent SEM. See also Figure S3 and Table S1.

levels of DnaK, GroEL, and Lon (Figure S3A), as reported previously (Zhang et al., 2014).

Using I54N σ^{32} expression to upregulate the σ^{32} regulon nearly eliminated aggregation of the test proteins (Figures 4A–4C). To determine the extent to which this result was due to KJE, GroELS, and Lon upregulation and not to other PN pathways that are part of the σ^{32} regulon, we examined the effect of upregulating ClpB (Doyle et al., 2013), or HtpG, the *E. coli* Hsp90 (Pearl and Prodromou, 2006), together with KJE, since both of these chaperones can cooperate with KJE (Doyle et al., 2013; Genest et al., 2011; Nakamoto et al., 2014). Upregulating ClpB+KJE or HtpG+KJE by expressing them simultaneously did not decrease the aggregated fractions of the test proteins beyond what was observed by upregulating KJE alone, even at the low level of upregulation (compare Figure S3B with the low-KJE upregulation results in Figures 3B–3D). Simultaneous upregulation of IbpA and IbpB, the small heat shock proteins of *E. coli* (Kuczyńska-Wiśnik et al., 2002; Thomas and Baneyx, 1998), tended to increase the extent of aggregation of m-EcDHFR and m-MmCRABP1 relative to adapted-basal conditions, although this increase was only significant for m-EcDHFR ($p = 0.003$, one-tailed t test). Upregulation of IbpA and IbpB had no effect on m-RA114 (Figure S3B).

Taken together, these results show that upregulating the PN using σ^{32}, which evolved to counter the protein folding stress caused by heat shock, enables *E. coli* to suppress aggregation even for highly destabilized proteins at high expression levels, consistent with previous results (Zhang et al., 2014). Moreover, our results argue that KJE, GroELS, and Lon are primarily responsible for the effects observed with σ^{32} overexpression.

Analysis of Test-Protein Folding Fates

Understanding how the *E. coli* PN manages proteostasis for our test proteins requires a quantitative analysis of our data. Thus, we first extract protein-specific trends, so far as they exist, by a phenomenological method. We then model how a test protein's response to the PN reports on its energy landscape by using FoldEco, a mechanistic model for proteostasis in *E. coli*.

Phenomenological Models

To quantify how KJE, GroELS, and Lon affect our test proteins, we fit the overexpression data for each test protein to the phenomenological models below:

$$[\mathrm{Agg}]_{\mathrm{rel,X}} = c_{\mathrm{Agg,X}} + a_{\mathrm{K,X}}[\mathrm{DnaK}]_{\mathrm{rel}} + a_{\mathrm{G,X}}[\mathrm{GroEL}]_{\mathrm{rel}} + a_{\mathrm{L,X}}[\mathrm{Lon}]_{\mathrm{rel}} \quad \text{(Equation 1)}$$

$$[\mathrm{Sol}]_{\mathrm{rel,X}} = c_{\mathrm{Sol,X}} + s_{\mathrm{K,X}}[\mathrm{DnaK}]_{\mathrm{rel}} + s_{\mathrm{G,X}}[\mathrm{GroEL}]_{\mathrm{rel}} + s_{\mathrm{L,X}}[\mathrm{Lon}]_{\mathrm{rel}}, \quad \text{(Equation 2)}$$

where $[\mathrm{Agg}]_{\mathrm{rel,X}}$ and $[\mathrm{Sol}]_{\mathrm{rel,X}}$ are the concentrations of the aggregated and soluble forms of test protein "X" normalized to the total concentration under adapted-basal conditions; $[\mathrm{DnaK}]_{\mathrm{rel}}$, $[\mathrm{GroEL}]_{\mathrm{rel}}$, and $[\mathrm{Lon}]_{\mathrm{rel}}$ are the concentrations of the PN components relative to their adapted-basal concentrations; $a_{\mathrm{K,X}}$, $a_{\mathrm{G,X}}$, and $a_{\mathrm{L,X}}$ are the gradients of $[\mathrm{Agg}]_{\mathrm{rel,X}}$ with respect to $[\mathrm{DnaK}]_{\mathrm{rel}}$, $[\mathrm{GroEL}]_{\mathrm{rel}}$, and $[\mathrm{Lon}]_{\mathrm{rel}}$; $s_{\mathrm{K,X}}$, $s_{\mathrm{G,X}}$, and $s_{\mathrm{L,X}}$ are the gradients of $[\mathrm{Sol}]_{\mathrm{rel,X}}$ with respect to $[\mathrm{DnaK}]_{\mathrm{rel}}$, $[\mathrm{GroEL}]_{\mathrm{rel}}$, and $[\mathrm{Lon}]_{\mathrm{rel}}$; and $c_{\mathrm{Agg,X}}$ and $c_{\mathrm{Sol,X}}$ are the model intercepts. Values for $[\mathrm{Agg}]_{\mathrm{rel,X}}$ and $[\mathrm{Sol}]_{\mathrm{rel,X}}$ under the various PN conditions can be read off the right axes in Figures 3B–3D and 4A–4C. Values for $[\mathrm{DnaK}]_{\mathrm{rel}}$, $[\mathrm{GroEL}]_{\mathrm{rel}}$, and $[\mathrm{Lon}]_{\mathrm{rel}}$ under all conditions are shown in Figures S2D and S3A. The gradient parameters quantify the efficacies of the PN components for a given test protein. For example, a protein that benefits greatly from GroELS would have a large, positive value of $s_{\mathrm{G,X}}$ (indicating that $[\mathrm{Sol}]_{\mathrm{rel,X}}$ increases sharply

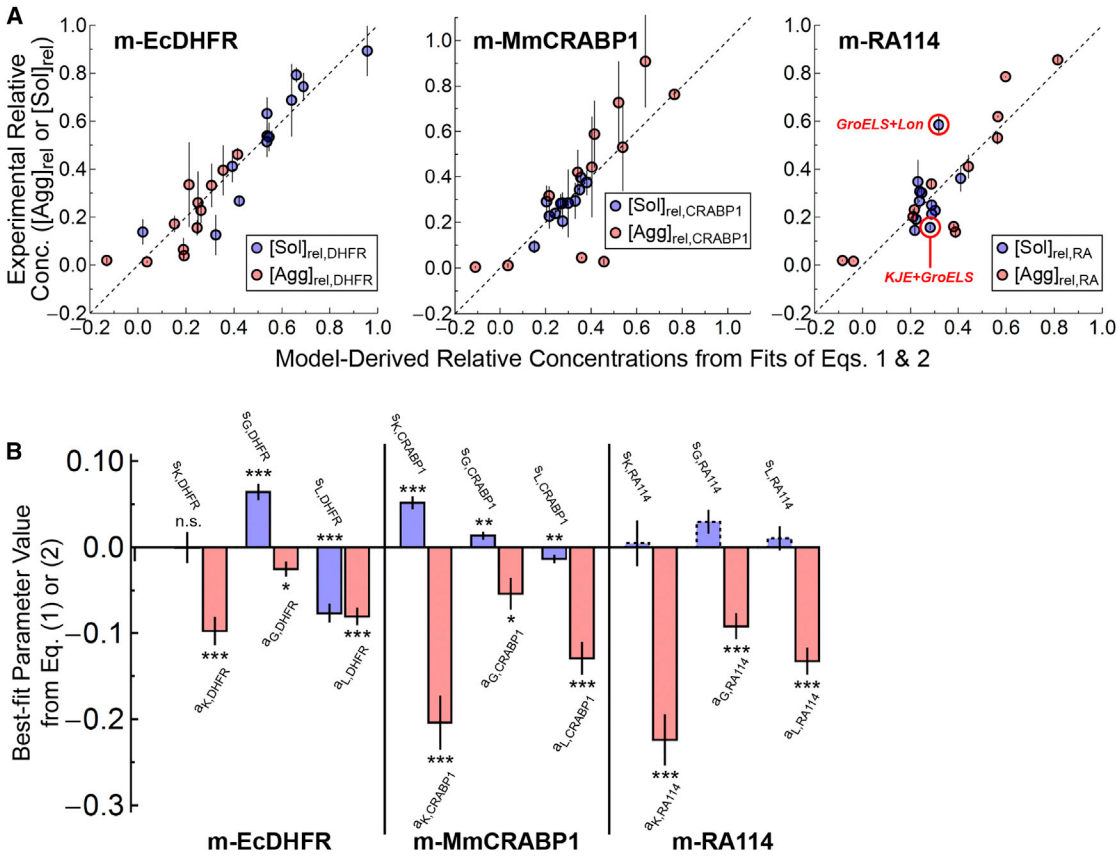


Figure 5. Dependences of Test-Protein Folding Fates on Different PN Components Based on Phenomenological Fits of Overexpression Data

(A) Plots of the experimental values of $[Agg]_{rel,X}$ and $[Sol]_{rel,X}$ of m-EcDHFR (left), m-MmCRABP1 (middle), and m-RA114 (right) from Figures 3 and 4 versus the corresponding model-derived values from the fits of Equations 1 or 2. Red data points are data points for $[Agg]_{rel,X}$, fit with Equation 1. Blue data points are data points for $[Sol]_{rel,X}$, fit with Equation 2. Dashed line: the line through the origin with a slope of 1. The extent to which the data points fall on the dashed line indicates the goodness of fit of the model. The circled data points have the largest residuals in the fit of Equation 2 to the $[Sol]_{rel,RA114}$ data.

(B) Bar graph showing the gradient parameters and their SEs from the fits of Equations 1 (red bars) and 2 (blue bars). Negative values indicate that increasing the concentration of a PN component decreases the concentration of the aggregated (red bars) or soluble (blue bars) form of a test protein. Positive values indicate the opposite. The blue bars for m-RA114 are bordered by dashed lines because these parameter values were obtained from a poor-quality fit. The p values for the gradient parameters are indicated as follows: ***$p < 0.0001$; **$p < 0.001$; *$p < 0.01$; and n.s., $p > 0.05$. Tests were not performed for $s_{K,RA114}$, $s_{G,RA114}$, or $s_{L,RA114}$ since the p value for the fit as a whole was >0.01.

See also Tables S1 and S2.

as GroELS is upregulated) and/or a large negative value of $a_{G,X}$ (indicating that $[Agg]_{rel,X}$ decreases sharply as GroELS is upregulated). Because of the differences in the expression levels of the three test proteins and the inherent non-linearity of aggregation kinetics with respect to protein concentration, one must be cautious when comparing the magnitudes of the gradient parameters of different proteins. However, differences in the signs of the gradient parameters, and whether or not they differ significantly from 0, are not subject to such concerns. Also, we have chosen to use relative concentrations of the PN components in our model rather than their absolute concentrations for ease of presentation. However, we also report the gradient parameters scaled to the absolute test protein and PN component concentrations in Table S2.

The qualities of the fits of Equation 1 to the normalized concentrations of aggregated protein are moderate to good for all three test proteins (adjusted R^2 = 0.73 for m-EcDHFR, 0.83 for m-RA114, and 0.68 for m-MmCRABP1; Figure 5A, red data points; see Table S1 for fit residuals). The parameters a_K, a_G, and a_L are negative for each test protein, indicating that all of the PN components decrease aggregation (Figure 5B, red bars). KJE decreases the extent of aggregation the most for a given fold change in its concentration, followed by Lon and GroELS ($a_{K,X} < a_{L,X} < a_{G,X}$ for all test proteins X). This result is consistent with KJE specifically antagonizing misfolding and aggregation, whereas GroELS and Lon affect these processes less directly.

The fits of Equation 2 to the normalized concentrations of soluble protein are good for m-EcDHFR and m-MmCRABP1 (adjusted R^2 = 0.82 for m-EcDHFR and 0.71 for m-MmCRABP1) but poor for m-RA114 (adjusted R^2 = 0.10) (Figure 5A, blue data points; see Table S1 for fit residuals). The poor fit for m-RA114 is

likely due to the small range of $[Sol]_{rel,RA114}$ combined with its surprisingly low value when KJE+GroELS are jointly upregulated and its surprisingly high value when GroELS+Lon are jointly upregulated (highlighted data points in Figure 5A, right). This observation suggests that the KJE and GroELS pathways may interfere with each other whereas the GroELS and Lon pathways cooperate to handle m-RA114.

The parameter $s_{L,X}$ is negative for both m-EcDHFR and m-MmCRABP1, but $s_{L,DHFR}$ is roughly the same as $a_{L,DHFR}$ whereas $s_{L,CRABP1}$ is much smaller than $a_{L,CRABP1}$ (Figure 5B). This observation indicates that upregulating Lon preferentially depletes aggregates for m-MmCRABP1, but not for m-EcDHFR. In addition, $s_{K,DHFR}$ is close to 0, indicating that while KJE is very effective for diminishing aggregation for m-EcDHFR, it does not increase the concentration of soluble protein (Figure 5B). In contrast, $s_{K,CRABP1}$ is substantial and positive for m-MmCRABP1 (Figure 5B). The situation is reversed for the GroELS system: $s_{G,DHFR}$ is substantial and positive but $s_{G,CRABP1}$ is much smaller (Figure 5B). These results suggest that m-MmCRABP1 may be a poor substrate for GroELS. While one should be cautious when interpreting this observation because of the expression level differences of m-EcDHFR and m-MmCRABP1, an analysis of the data using FoldEco (see below), which explicitly accounts for these expression level differences, corroborates this notion.

Although fitting Equations 1 and 2 to our data has enabled us to quantify the effects of KJE, GroELS, and Lon on our test proteins, these equations are phenomenological and cannot inform us about the causes of a given protein's behavior. Based on previous results with FoldEco, the extent to which a protein benefits from different chaperoning mechanisms should be a function of that protein's folding energetics (Dickson and Brooks, 2013; Powers et al., 2012). Thus, the values of the best-fit parameters for Equations 1 and 2 should reflect the folding energetics of our test proteins. To explore this possibility, we used FoldEco to fit our data by using the folding energetics as adjustable parameters.

General Analysis of Test-Protein Folding Fates using FoldEco

FoldEco comprises a system of ordinary differential equations that represent the kinetics of the processes undergone by proteins in vivo: synthesis, folding, misfolding, and aggregation; interaction with KJE and GroELS; and degradation by Lon (Figure 1) (Powers et al., 2012). To fit FoldEco to our data, we have to vary the parameters in FoldEco until the model optimally matches the experimental data. FoldEco has many dozens of parameters, but many of these are likely to be independent of, or weakly dependent on, the nature of the protein (Powers et al., 2012). For example, chaperone/co-chaperone interaction parameters should not be differentially affected by bound substrates (that is, the effect of one bound substrate on chaperone/co-chaperone interaction parameters should be similar to the effect of another). Furthermore, chaperones bind to substrates promiscuously (Aoki et al., 2000; Landry and Gierasch, 1991; Rüdiger et al., 1997; Wang et al., 1999)—GroEL can even chaperone a substrate composed entirely of d-amino acids (Weinstock et al., 2014)—so chaperone-substrate interaction parameters are likely to be similar for most substrates. Values for such parameters can be obtained from the literature, and the available data are broadly (though not perfectly) consistent with the assertions in the preceding sentences (see Powers et al., 2012 and Figures S1–S5 therein). Thus, to a first approximation, the only adjustable parameters needed to fit FoldEco to our data are those for the folding energetics and synthesis rates of the test proteins. These parameters include the folding rate and equilibrium constants (k_f and K_f), the misfolding rate and equilibrium constants (k_m and K_m), the aggregation rate and equilibrium constants (k_a and K_a, where K_a is the equilibrium constant for adding a misfolded monomer to an aggregate), and the steady-state protein synthesis rate (σ) (Figure 1). Importantly, FoldEco can account for how the different expression levels of the test proteins affect their overall behavior and in particular their tendency to aggregate via its built-in nucleated polymerization model for protein aggregation (Powers et al., 2012).

It is important to note that the parameters derived from FoldEco fits of our data are "effective parameters," since we are applying FoldEco's generic folding mechanism to the test proteins. For example, EcDHFR folds through a multi-step mechanism (Frieden, 1990). Thus, the single folding rate constant derived from FoldEco fits to the m-EcDHFR expression data (k_f) subsumes the rate constants for the individual folding steps. In addition, the equilibrium denaturation of m-RA114 reveals at least one intermediate (Figure S1B). Thus, the best-fit folding equilibrium constant for m-RA114 encompasses the energetics for all of the states in the native conformational ensemble. Nevertheless, we expect that the effective parameters derived from fitting FoldEco will faithfully capture the essences of the true folding and misfolding processes.

The qualities of the FoldEco fits range from moderate to good, with the best-fit values of $[Agg]_{rel,X}$ and $[Sol]_{rel,X}$ deviating from their experimental values on average by 0.05 for m-EcDHFR, 0.12 for m-MmCRABP1, and 0.11 for m-RA114 (Figure 6; see Table S1 for fit residuals). Unfortunately, only the fit for m-EcDHFR permitted parameters to be estimated with acceptable precision (Table S3). These are discussed in the next section. The FoldEco fits for m-MmCRABP1 and m-RA114, despite not providing definite parameter estimates, nevertheless define some relationships among the parameters and thereby enable us to discern some general features of the in vivo folding energy landscapes of these two proteins.

Analysis of the Folding Fate of m-EcDHFR using FoldEco

The effective folding parameters for m-EcDHFR are: k_m = 0.06 s^{-1}; K_m = 0.3 (corresponding to ΔG_m = +0.7 kcal mol^{-1}); k_f = 0.3 s^{-1}; K_f = 130 (corresponding to ΔG_f = −2.9 kcal mol^{-1}); and $k_a = 4 \times 10^9$ M^{-1} s^{-1} (see Table S3 for the errors in these parameters). Only a lower limit could be determined for K_a, which was 1.4×10^9 M^{-1} (corresponding to ΔG_a = −12.7 kcal mol^{-1}). These parameters indicate that m-EcDHFR folds about five times faster than it misfolds, but its native state is not very stable, and once it misfolds it aggregates rapidly to form stable aggregates (Figure 7A). In fact, the aggregation rate constant is around the diffusion-controlled limit, possibly

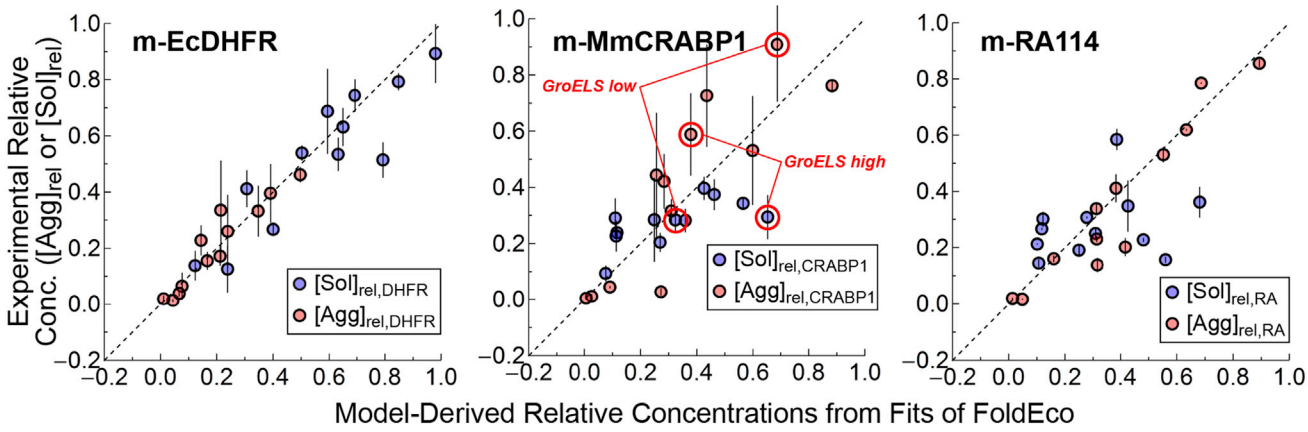


Figure 6. FoldEco-Derived Fits of Experimental Data from the Overexpression of Test Proteins

As in Figure 5A, except that the model-derived relative concentrations are determined from the FoldEco fits to the data. The circled points are those for m-MmCRABP1 under the low- and high-upregulation GroELS conditions, which have particularly large residuals. Error bars represent SEM. See also Figures S4 and S5 and Tables S1 and S3.

indicating that m-EcDHFR aggregates directly from the unfolded state rather than through a monomeric misfolded intermediate as assumed in FoldEco.

It is of interest to compare the parameters obtained from our FoldEco fits to the analogous parameters measured in vitro. However, because parameters for processes that are not on the folding pathway, like the rate and equilibrium constants for misfolding and aggregation, can be difficult to determine, we limit our comparisons to the folding equilibrium and rate constants, which can be determined using well-established methods. The in vitro free energy and rate constant for folding of m-EcDHFR were found to be $\Delta G_f = -2.6$ kcal mol^{-1} and $k_f = 1.1$ s^{-1} by equilibrium denaturation (Figure S1A) and fluorescence-monitored refolding kinetics (Figure S4; k_f here refers to the rate of tertiary structure acquisition, not subsequent steps required for NADP binding that were identified in past mechanistic studies of WT-EcDHFR folding; Frieden, 1990). These values were determined at [m-EcDHFR] = 3.5 μM and 25°C; aggregation was not detected under these conditions. The values determined for these parameters from our FoldEco fit were $\Delta G_f = -2.9$ kcal mol^{-1} and $k_f = 0.3$ s^{-1} (Table S3; Figure 7A). The good correspondence between these parameter estimates and their experimental values supports our use of FoldEco to understand the in vivo folding of our test proteins.

Some aspects of the sensitivities of m-EcDHFR proteostasis to the PN components (quantified in Figure 5B) can be explained by the FoldEco fit to the m-EcDHFR expression data. For example, Lon affects the levels of aggregated and soluble m-EcDHFR about equally ($a_{L,DHFR} = -0.080$ versus $s_{L,DHFR} = -0.076$; Figure 5B). Since the relatively slow misfolding of m-EcDHFR is followed by fast formation of stable aggregates (Figure 7A), the misfolded state of m-EcDHFR has a short lifetime and a low concentration, and Lon cannot intercept and degrade it before it aggregates. In contrast, Lon appears to preferentially diminish the aggregated state for m-MmCRABP1 ($a_{L,CRABP1} = -0.129$ versus $s_{L,CRABP1} = -0.014$; Figure 5B). The same is qualitatively true for m-RA114 (Figure 3D), although a quantitative comparison cannot be made for m-RA114 because of the poor fit of Equation 2.

This point can be further illustrated using the FoldEco simulation of m-EcDHFR under adapted-basal conditions. In this simulation, only 0.1% of the protein molecules that sample the misfolded state are degraded by Lon (Figure 7A, red numbers), and furthermore, of the almost 188 μM of m-EcDHFR that is degraded over the 2-hr time course of the expression simulation, only 0.2% is taken from the misfolded state (Figure 7A, black numbers). The remaining 99.8% of the m-EcDHFR that is degraded comes from the unfolded state. Since the unfolded state is the progenitor of both natively folded and aggregated protein, Lon affects the levels of each similarly for m-EcDHFR.

The short lifetime of the misfolded state also explains why KJE decreases levels of aggregated m-EcDHFR without increasing soluble m-EcDHFR. KJE affects the partitioning between aggregated and soluble protein only when it directly reverses the process of misfolding by converting misfolded protein to unfolded protein. Because misfolded m-EcDHFR exists so briefly and at such a low concentration, DnaJ and DnaK, like Lon, cannot bind to it before it aggregates. In our FoldEco simulation of m-EcDHFR under adapted-basal conditions, only 1% of the m-EcDHFR molecules that sample the misfolded state are engaged by KJE (Figure 7A, red numbers). Although KJE, especially with the assistance of ClpB, can recover protein from the aggregated state and return it to the unfolded state, this process happens after the "product determining step" of the protein folding pathway—that is, after unfolded protein is partitioned between folding and misfolding/aggregation pathways and therefore cannot influence the folding/misfolding decision. The protein that the ClpB+KJE pathway recovers from aggregates is, however, susceptible to degradation, which at least partly explains why upregulating KJE leads to lower total protein levels for m-EcDHFR as well as m-MmCRABP1 and m-RA114.

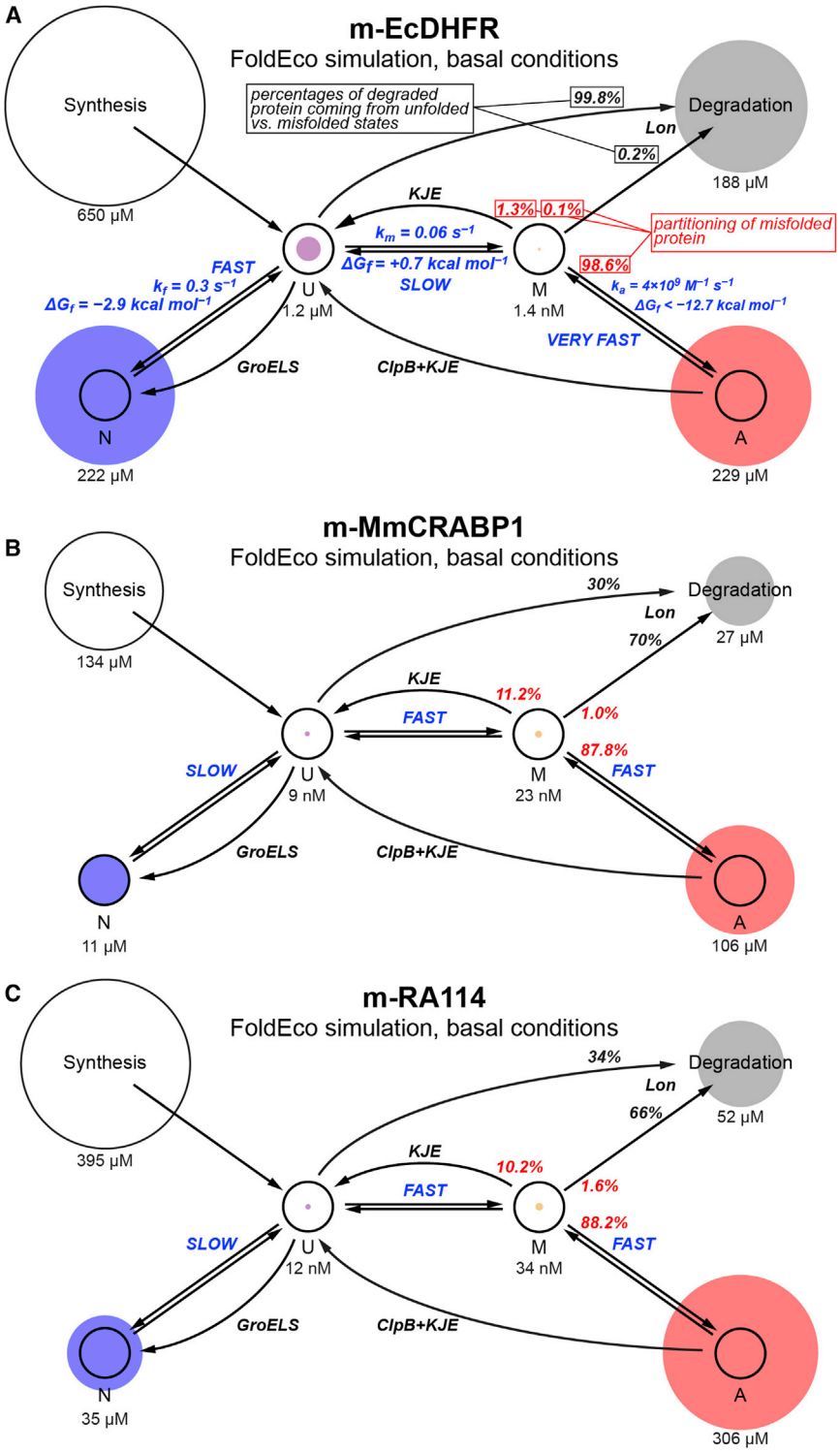


Figure 7. Partitioning of the Three Test Proteins under Adapted-Basal Conditions at t = 2 hr Based on FoldEco Simulations

(A) Summary diagram for m-EcDHFR, laid out as in Figure 1. The sizes of the colored circles indicate the concentrations of the unfolded (U), native (N), misfolded (M), and aggregated (A) states. The radii of the circles are proportional to the cube roots of the concentrations. Cube roots are used to enable the lowest and highest concentrations to be shown on the same diagram. The circles for "Synthesis" and "Degradation" represent the total concentration of protein synthesized and degraded over the 2-hr time course of the simulation. The numerical values of each concentration are written below the circles. Blue text shows the best-fit biophysical parameters from the fit of FoldEco to the data. Red numbers are percentages of misfolded protein molecules that aggregate, engage the KJE recovery pathway, or engage the Lon degradation pathway. Black italic numbers are percentages of degraded protein taken from the unfolded or misfolded states.

(B) As in (A), but for m-MmCRABP1.

(C) As in (A) but for m-RA114.

Note that qualitative descriptors are used for the biophysical processes in (B) and (C). See also Table S3.

Analysis of Folding Fates of m-MmCRABP1 and m-RA114 using FoldEco

Although the fits of FoldEco to the data for m-MmCRABP1 and m-RA114 did not permit precise estimation of the biophysical parameters for these two proteins, the fits did define some relationships between the parameters that illuminate some general features of the folding energy landscapes of m-MmCRABP1 and m-RA114. For example, the folding rate constant (k_f) of m-MmCRABP1 can be increased from its optimal value with very little decrease in the quality of the FoldEco fit provided that, first, the misfolding equilibrium constant (K_m) also increases, and second, the misfolding rate constant (k_m) is much greater than k_f so that misfolding is nearly at equilibrium with respect to folding. Then, increases in k_f and K_m have offsetting effects on the unfolded-to-native-state flux. Given that this flux is equal to the product $k_f \times$ [U] (where [U] is the concentration of unfolded protein), increasing k_f directly increases this product, whereas increasing K_m decreases this product by shifting the misfolding equilibrium away from the unfolded state, thereby decreasing [U]. The overall effect of these combined changes on the behavior of m-MmCRABP1 is therefore minimal.

Despite the uncertainties in the parameters from the FoldEco fits for m-MmCRABP1 and m-RA114, certain aspects of their folding energy landscapes are nevertheless clear. Both m-MmCRABP1 and m-RA114 misfold faster than they fold ($k_m > k_f$) but do not aggregate as fast as m-EcDHFR, resulting

in appreciable accumulation of misfolded protein under adapted-basal PN conditions (Table S3; Figures 7B and 7C). This behavior can explain the response of m-MmCRABP1 and m-RA114 to KJE and Lon overexpression. The higher concentration of misfolded protein increases the utilization of the KJE recovery and Lon degradation pathways by almost 10-fold compared to m-EcDHFR (compare Figure 7A with Figures 7B and 7C, red numbers). Thus, Lon preferentially reduces the concentration of aggregates of m-MmCRABP1 and m-RA114 by degrading mainly misfolded protein. Similarly, KJE can directly convert misfolded m-MmCRABP1 and m-RA114 to the unfolded state, simultaneously increasing the levels of soluble protein and decreasing levels of aggregated protein.

However, the FoldEco fits are unable to account for the poor chaperoning of m-MmCRABP1 by GroELS. According to the FoldEco fits, all of the test proteins should benefit similarly from GroELS upregulation. In reality, m-MmCRABP1 benefits much less from GroELS than do m-EcDHFR or m-RA114 (Figures 3B–3D), as can be seen in the residuals of the FoldEco fit for $[Agg]_{rel,CRABP1}$ and $[Sol]_{rel,CRABP1}$ under the GroELS-low and -high upregulation conditions (Figure 6, middle, highlighted data points; Table S1). These residuals average 0.21 for m-MmCRABP1, compared to 0.05 for m-EcDHFR and 0.12 for m-RA114, suggesting that m-MmCRABP1 may interact differently with GroEL than the other test proteins do.

To explore this notion, we fit FoldEco to the expression data for m-MmCRABP1 while varying the equilibrium association constant between m-MmCRABP1 and GroEL ($K_{Gro–CRABP1}$). The best fit was obtained when $K_{Gro\text{-}CRABP1}$ was 1000-fold lower than its default value (10^3 M^{-1} versus 10^6 M^{-1}; Figure S5A), in which case the overall FoldEco fit improved by about 20% (mean residuals = 0.10 versus 0.12), and the residuals for the GroELS upregulation conditions decreased from 0.21 to 0.15 (Figure S5B). The improved fit still did not yield precise estimates of the biophysical parameters of m-MmCRABP1, but it is clear that the fit is improved by the large decrease in $K_{Gro–CRABP1}$, suggesting that m-MmCRABP1 binds weakly to GroEL. This observation based solely on fitting the folding fate of this protein under different PN conditions is consistent with experimental findings in vitro that GroELS has no effect on MmCRABP1 folding and that there is no detectable binding between GroEL and incompletely folded MmCRABP1 variants (I. Budyak, H.-P. Feng, and L.M.G., unpublished data).

We note that adding more adjustable parameters does not dramatically improve the FoldEco fits for m-EcDHFR or m-RA114, nor does varying parameters beyond $K_{Gro–CRABP1}$ yield further substantive improvements to the FoldEco fit for m-MmCRABP1. Thus, the default, literature-derived parameters used in FoldEco appear to be sufficiently accurate in most cases, and when they are not, a lack of fit can indicate an unusual interaction between a PN component and a substrate.

DISCUSSION

Our results show that increasing the levels of PN components enables cells to better handle situations that involve protein misfolding, consistent with previous studies of the effect of chaperone upregulation on, for example, the yields of overexpressed heterologous proteins (de Marco, 2007; de Marco et al., 2007; Makino et al., 2011; Zhang et al., 2014), protein evolution (Bershtein et al., 2013; Bogumil and Dagan, 2012; Queitsch et al., 2002; Tokuriki and Tawfik, 2009), and stress tolerance (Feder et al., 1996; Welte et al., 1993). The "profolding" (GroELS), "anti-misfolding" (KJE), and "concentration control" (Lon) arms of the *E. coli* PN form an efficacious triad for maintaining proteostasis for a broad range of proteins, in that the concentration of soluble protein is generally maintained or increased by chaperone upregulation even when the total amount of protein decreases (see Figures 3 and 4). These arms are especially effective when upregulated in their native ratios as part of the σ^{32} transcriptional program (Figure 4). The two quantitative models of *E. coli* proteostasis, the phenomenological Equations 1 and 2 and FoldEco, provide insight into how these systems contribute to proteostasis, and how their contributions depend on their substrates' folding energy landscapes.

The two models, which generally explain 60%–80% of the variation in our data (Figures 5A and 6), share two important features: first, the only components of the PN that they include are KJE, GroELS, and Lon; and second, they do not account for any direct mechanisms by which these components can collaborate. The success of the models in fitting the data in light of the first feature suggests that KJE, GroELS, and Lon are the primary contributors to the proteostasis of our test proteins. This does not mean that other PN components, like ClpB, HtpG, and IbpA/IbpB, are not important but rather that their importance lies outside generic proteostasis in the absence of environmental stress. ClpB is vital for recovery from heat shock (Doyle et al., 2013; Squires et al., 1991; Weibezahn et al., 2004), but it is not necessary for maintaining normal growth rates under non-stress conditions (Squires et al., 1991). Overexpressing ClpB in our system has little effect on the folding fates of our test proteins, probably because protein disaggregation by ClpB+KJE is simply not as fast as protein synthesis. Similarly, HtpG may be important for the folding of specific substrates (Yosef et al., 2011), but it is not necessary for either normal growth or thermotolerance (Bardwell and Craig, 1988).

The ability of the models to fit our data without invoking collaboration among KJE, GroELS, and Lon suggests that these PN components act independently under our experimental conditions and that their effects are therefore largely, though not perfectly, additive. This assertion is corroborated by the residuals of the fits of our models to the test protein data, ~70% of which are less than ±0.1 and ~90% of which are less than ±0.2 (Table S1). However, we emphasize that many examples of collaboration between chaperone systems have been reported. In eukaryotes, the Hsp70 and Hsp90 systems work together in the proteostasis of, for example, steroid hormone receptors (Pratt and Toft, 2003; Wegele et al., 2006). PN components may also "hand off" substrates from one to another—for example, DnaK to GroEL (Langer et al., 1992)—and this sequential activity is important for in vivo protein folding. However, the large amounts of unfolded protein that are being produced in our experiments

may lead to unoccupied chaperones binding to newly synthesized protein before they can accept the transfer of a substrate from another chaperone, thereby suppressing chaperone collaboration.

Although the major PN systems in *E. coli* appear to act independently under high protein-folding loads, there may still be occasional exceptions to this rule. Such an exception may have caused the poor fit of Equation 2 to the $[Sol]_{rel,RA114}$ data. Although we can only speculate that direct interactions between PN components is the reason that Equation 2 does not fit the $[Sol]_{rel,RA114}$ data well, it remains an intriguing possibility that some client proteins can access chaperone collaboration pathways that others cannot.

FoldEco simulations allow us to go beyond the phenomenological treatment of Equations 1 and 2 because they enable correlations between folding energetics and chaperone mechanisms. For example, the response of m-EcDHFR to KJE and Lon was due to a combination of slow misfolding and fast aggregation. Even the cases where a good fit of the model to the data was elusive could be a clue to underlying biophysics. The poor fit of FoldEco to the data for m-MmCRABP1 under GroELS upregulation suggested weak binding of m-MmCRABP1 to GroEL. It should be noted that some aspects of proteostasis that could contribute to protein misfolding and aggregation are not yet modeled in FoldEco, like the effect of translation rates on proteostasis. Given that the m-MmCRABP1 and m-RA114 genes were not codon optimized for *E. coli*, such effects could contribute to the high aggregation propensities observed for these proteins. However, previous FoldEco simulations of luciferase expression found that ignoring the effect of cotranslational folding did not greatly diminish the performance of FoldEco (Powers et al., 2012).

FoldEco has enabled us, in a sense, to invert the usual reductionist process for studying in vivo protein folding. Rather than analyzing our test proteins' folding in vitro and then using this information to rationalize their behavior in vivo, we have used FoldEco to translate the behavior of our test proteins in vivo into information about their folding energy landscapes. In addition, given that FoldEco embodies our best understanding of chaperone mechanisms based on the wealth of biochemical literature that exists on this subject, using FoldEco to fit our data constitutes a test of this understanding. We believe that the results of these fits were good enough to indicate that this understanding is fundamentally sound, albeit incomplete. We expect that as FoldEco is refined and expanded, its ability to explain the nuances of how proteostasis is managed for proteins with different folding energy landscapes will improve, providing a foundation for understanding processes like organismal stress responses and protein evolution that are intimately linked to protein-folding energetics.

EXPERIMENTAL PROCEDURES

E. coli Strains and Plasmids

E. coli K12 strain HMS174 (DE3) was used for the overexpression experiments. Vectors under pBAD and/or pTet promoters on low-copy-number plasmids were used to express chaperones, Lon, and I54N σ^{32}. Test proteins were expressed using a low-copy-number pET29b vector. Additional details can be found in the Supplemental Experimental Procedures.

Measurement of the Levels of Aggregated and Soluble Protein

Overexpression of desired PN components was induced by adding arabinose to *E. coli* harboring a plasmid containing the genes for the desired PN component(s) and a plasmid containing the gene for the desired test protein. After 1 hr of overexpressing the PN component(s) at 30°C, test-protein overexpression was induced by adding IPTG. After 2 hr at 30°C, cells were lysed by sonication at 4°C. Portions of the lysates were centrifuged for 10 min at 13,500 × *g* at 4°C. Test protein in the supernatant and pellet was defined as the soluble and aggregated fractions, respectively. The total (i.e., not centrifuged), soluble, and aggregated fractions were analyzed by SDS-PAGE (12 or 15% Bio-Rad gel) and then stained with Coomassie blue (m-EcDHFR) or subjected to western blot analysis (m-MmCRABP1 and m-RA114) to quantify protein levels. Adapted-basal PN controls were included in all gels so that the total test-protein expression levels under perturbed PN conditions and adapted-basal PN conditions could be directly compared. Further details can be found in the Supplemental Experimental Procedures.

Fits of Equations 1 and 2 and FoldEco to Test-Protein Overexpression Data

Equations 1 and 2 were fit to the data by linear regression. FoldEco was fit to the data using a least-squares approach. Details can be found in the Supplemental Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.03.018.

AUTHOR CONTRIBUTIONS

Y.C., X.Z., K.F.R.P., J.W.K., L.M.G., and E.T.P. conceived the project, designed the experiments, and wrote the manuscript. Y.C., X.Z., K.F.R.P., and Y.L. performed the experiments. D.L.P. and E.T.P. fit the models to the data. All authors were involved in interpreting and discussing the results.

ACKNOWLEDGMENTS

We thank Gareth Morgan, Luke Wiseman, and Bernd Bukau for valuable discussions about this project and Colleen Fearns for critical reading of the manuscript. The anti-Lon and anti-GrpE antibodies used herein were kindly provided by the Sauer and Bukau laboratories, respectively. X.Z. was a Howard Hughes Medical Institute Fellow of the Helen Hay Whitney Foundation and is currently supported by a Burroughs Wellcome Fund Career Award at the Scientific Interface. This work was supported by the Skaggs Institute for Chemical Biology and the Lita Annenberg Hazen Foundation and by NIH grants GM101644 (to L.M.G. and E.T.P.) and DK075295 (to J.W.K.).

Received: October 31, 2014
Revised: January 29, 2015
Accepted: March 7, 2015
Published: April 2, 2015

REFERENCES

Aoki, K., Motojima, F., Taguchi, H., Yomo, T., and Yoshida, M. (2000). GroEL binds artificial proteins with random sequences. J. Biol. Chem. *275*, 13755–13758.

Balch, W.E., Morimoto, R.I., Dillin, A., and Kelly, J.W. (2008). Adapting proteostasis for disease intervention. Science *319*, 916–919.

Bardwell, J.C., and Craig, E.A. (1988). Ancient heat shock gene is dispensable. J. Bacteriol. *170*, 2977–2983.

Bershtein, S., Mu, W., Serohijos, A.W., Zhou, J., and Shakhnovich, E.I. (2013). Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness. Mol. Cell *49*, 133–144.

Bjelic, S., Kipnis, Y., Wang, L., Pianowski, Z., Vorobiev, S., Su, M., Seetharaman, J., Xiao, R., Kornhaber, G., Hunt, J.F., et al. (2014). Exploration of alternate catalytic mechanisms and optimization strategies for retroaldolase design. J. Mol. Biol. *426*, 256–271.

Bogumil, D., and Dagan, T. (2012). Cumulative impact of chaperone-mediated folding on genome evolution. Biochemistry *51*, 9941–9953.

Budyak, I.L., Krishnan, B., Marcelino-Cruz, A.M., Ferrolino, M.C., Zhuravleva, A., and Gierasch, L.M. (2013). Early folding events protect aggregation-prone regions of a β-rich protein. Structure *21*, 476–485.

Calloni, G., Chen, T., Schermann, S.M., Chang, H.C., Genevaux, P., Agostini, F., Tartaglia, G.G., Hayer-Hartl, M., and Hartl, F.U. (2012). DnaK functions as a central hub in the *E. coli* chaperone network. Cell Rep. *1*, 251–264.

Chapman, E., Farr, G.W., Usaite, R., Furtak, K., Fenton, W.A., Chaudhuri, T.K., Hondorp, E.R., Matthews, R.G., Wolf, S.G., Yates, J.R., et al. (2006). Global aggregation of newly translated proteins in an *Escherichia coli* strain deficient of the chaperonin GroEL. Proc. Natl. Acad. Sci. USA *103*, 15800–15805.

de Marco, A. (2007). Protocol for preparing proteins with improved solubility by co-expressing with molecular chaperones in *Escherichia coli*. Nat. Protoc. *2*, 2632–2639.

de Marco, A., Deuerling, E., Mogk, A., Tomoyasu, T., and Bukau, B. (2007). Chaperone-based procedure to increase yields of soluble recombinant proteins produced in *E. coli*. BMC Biotechnol. *7*, 32.

Dickson, A., and Brooks, C.L., 3rd. (2013). Quantifying chaperone-mediated transitions in the proteostasis network of *E. coli*. PLoS Comput. Biol. *9*, e1003324.

Doyle, S.M., Genest, O., and Wickner, S. (2013). Protein rescue from aggregates by powerful molecular chaperone machines. Nat. Rev. Mol. Cell Biol. *14*, 617–629.

Feder, M.E., Cartaño, N.V., Milos, L., Krebs, R.A., and Lindquist, S.L. (1996). Effect of engineering Hsp70 copy number on Hsp70 expression and tolerance of ecologically relevant heat shock in larvae and pupae of *Drosophila melanogaster*. J. Exp. Biol. *199*, 1837–1844.

Frieden, C. (1990). Refolding of *Escherichia coli* dihydrofolate reductase: sequential formation of substrate binding sites. Proc. Natl. Acad. Sci. USA *87*, 4413–4416.

Gasser, B., Saloheimo, M., Rinas, U., Dragosits, M., Rodríguez-Carmona, E., Baumann, K., Giuliani, M., Parrilli, E., Branduardi, P., Lang, C., et al. (2008). Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. Microb. Cell Fact. *7*, 11.

Genest, O., Hoskins, J.R., Camberg, J.L., Doyle, S.M., and Wickner, S. (2011). Heat shock protein 90 from *Escherichia coli* collaborates with the DnaK chaperone system in client protein remodeling. Proc. Natl. Acad. Sci. USA *108*, 8206–8211.

Gidalevitz, T., Ben-Zvi, A., Ho, K.H., Brignull, H.R., and Morimoto, R.I. (2006). Progressive disruption of cellular protein folding in models of polyglutamine diseases. Science *311*, 1471–1474.

Gottesman, S. (1996). Proteases and their targets in *Escherichia coli*. Annu. Rev. Genet. *30*, 465–506.

Guisbert, E., Yura, T., Rhodius, V.A., and Gross, C.A. (2008). Convergence of molecular, modeling, and systems approaches for an understanding of the *Escherichia coli* heat shock response. Microbiol. Mol. Biol. Rev. *72*, 545–554.

Gur, E., and Sauer, R.T. (2008). Recognition of misfolded proteins by Lon, a AAA(+) protease. Genes Dev. *22*, 2267–2277.

Hingorani, K.S., and Gierasch, L.M. (2014). Comparing protein folding in vitro and *in vivo*: foldability meets the fitness challenge. Curr. Opin. Struct. Biol. *24*, 81–90.

Hoffmann, F., and Rinas, U. (2004). Stress induced by recombinant protein production in *Escherichia coli*. Adv. Biochem. Eng. Biotechnol. *89*, 73–92.

Horwich, A.L., and Fenton, W.A. (2009). Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. Q. Rev. Biophys. *42*, 83–116.

Kim, Y.E., Hipp, M.S., Bracher, A., Hayer-Hartl, M., and Hartl, F.U. (2013). Molecular chaperone functions in protein folding and proteostasis. Annu. Rev. Biochem. *82*, 323–355.

Kleywegt, G.J., Bergfors, T., Senn, H., Le Motte, P., Gsell, B., Shudo, K., and Jones, T.A. (1994). Crystal structures of cellular retinoic acid binding proteins I and II in complex with all-*trans*-retinoic acid and a synthetic retinoid. Structure *2*, 1241–1258.

Kuczyńska-Wiśnik, D., Kedzierska, S., Matuszewska, E., Lund, P., Taylor, A., Lipińska, B., and Laskowska, E. (2002). The *Escherichia coli* small heat-shock proteins IbpA and IbpB prevent the aggregation of endogenous proteins denatured *in vivo* during extreme heat shock. Microbiology *148*, 1757–1765.

Landry, S.J., and Gierasch, L.M. (1991). The chaperonin GroEL binds a polypeptide in an α-helical conformation. Biochemistry *30*, 7359–7362.

Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M.K., and Hartl, F.U. (1992). Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. Nature *356*, 683–689.

Liu, Y., Tan, Y.L., Zhang, X., Bhabha, G., Ekiert, D.C., Genereux, J.C., Cho, Y., Kipnis, Y., Bjelic, S., Baker, D., and Kelly, J.W. (2014). Small molecule probes to quantify the functional fraction of a specific protein in a cell with minimal folding equilibrium shifts. Proc. Natl. Acad. Sci. USA *111*, 4449–4454.

Makino, T., Skretas, G., and Georgiou, G. (2011). Strain engineering for improved expression of recombinant proteins in bacteria. Microb. Cell Fact. *10*, 32.

Mayer, M.P., and Bukau, B. (2005). Hsp70 chaperones: cellular functions and molecular mechanism. Cell. Mol. Life Sci. *62*, 670–684.

Nakamoto, H., Fujita, K., Ohtaki, A., Watanabe, S., Narumi, S., Maruyama, T., Suenaga, E., Misono, T.S., Kumar, P.K., Goloubinoff, P., and Yoshikawa, H. (2014). Physical interaction between bacterial heat shock protein (Hsp) 90 and Hsp70 chaperones mediates their cooperative action to refold denatured proteins. J. Biol. Chem. *289*, 6110–6119.

Olzscha, H., Schermann, S.M., Woerner, A.C., Pinkert, S., Hecht, M.H., Tartaglia, G.G., Vendruscolo, M., Hayer-Hartl, M., Hartl, F.U., and Vabulas, R.M. (2011). Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. Cell *144*, 67–78.

Pearl, L.H., and Prodromou, C. (2006). Structure and mechanism of the Hsp90 molecular chaperone machinery. Annu. Rev. Biochem. *75*, 271–294.

Powers, E.T., and Balch, W.E. (2013). Diversity in the origins of proteostasis networks–a driver for protein function in evolution. Nat. Rev. Mol. Cell Biol. *14*, 237–248.

Powers, E.T., Morimoto, R.I., Dillin, A., Kelly, J.W., and Balch, W.E. (2009). Biological and chemical approaches to diseases of proteostasis deficiency. Annu. Rev. Biochem. *78*, 959–991.

Powers, E.T., Powers, D.L., and Gierasch, L.M. (2012). FoldEco: a model for proteostasis in *E. coli*. Cell Rep. *1*, 265–276.

Pratt, W.B., and Toft, D.O. (2003). Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery. Exp. Biol. Med. (Maywood) *228*, 111–133.

Queitsch, C., Sangster, T.A., and Lindquist, S. (2002). Hsp90 as a capacitor of phenotypic variation. Nature *417*, 618–624.

Rüdiger, 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*, 1501–1507.

Sawaya, M.R., and Kraut, J. (1997). Loop and subdomain movements in the mechanism of *Escherichia coli* dihydrofolate reductase: crystallographic evidence. Biochemistry *36*, 586–603.

Sharma, S.K., 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. Biol. *6*, 914–920.

Sherman, MYu, and Goldberg, A.L. (1992). Involvement of the chaperonin dnaK in the rapid degradation of a mutant protein in *Escherichia coli*. EMBO J. *11*, 71–77.

Squires, C.L., Pedersen, S., Ross, B.M., and Squires, C. (1991). ClpB is the *Escherichia coli* heat shock protein F84.1. J. Bacteriol. *173*, 4254–4262.

Thomas, J.G., and Baneyx, F. (1998). Roles of the *Escherichia coli* small heat shock proteins IbpA and IbpB in thermal stress management: comparison with ClpA, ClpB, and HtpG *In vivo*. J. Bacteriol. *180*, 5165–5172.

Tokuriki, N., and Tawfik, D.S. (2009). Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature *459*, 668–673.

Wang, Z., Feng, Hp., Landry, S.J., Maxwell, J., and Gierasch, L.M. (1999). Basis of substrate binding by the chaperonin GroEL. Biochemistry *38*, 12537–12546.

Wegele, H., Wandinger, S.K., Schmid, A.B., Reinstein, J., and Buchner, J. (2006). Substrate transfer from the chaperone Hsp70 to Hsp90. J. Mol. Biol. *356*, 802–811.

Weibezahn, J., Tessarz, P., Schlieker, C., Zahn, R., Maglica, Z., Lee, S., Zentgraf, H., Weber-Ban, E.U., Dougan, D.A., Tsai, F.T., et al. (2004). Thermotolerance requires refolding of aggregated proteins by substrate translocation through the central pore of ClpB. Cell *119*, 653–665.

Weinstock, M.T., Jacobsen, M.T., and Kay, M.S. (2014). Synthesis and folding of a mirror-image enzyme reveals ambidextrous chaperone activity. Proc. Natl. Acad. Sci. USA *111*, 11679–11684.

Welte, M.A., Tetrault, J.M., Dellavalle, R.P., and Lindquist, S.L. (1993). A new method for manipulating transgenes: engineering heat tolerance in a complex, multicellular organism. Curr. Biol. *3*, 842–853.

Wiseman, R.L., Powers, E.T., Buxbaum, J.N., Kelly, J.W., and Balch, W.E. (2007). An adaptable standard for protein export from the endoplasmic reticulum. Cell *131*, 809–821.

Yosef, I., Goren, M.G., Kiro, R., Edgar, R., and Qimron, U. (2011). High-temperature protein G is essential for activity of the *Escherichia coli* clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system. Proc. Natl. Acad. Sci. USA *108*, 20136–20141.

Yura, T., Guisbert, E., Poritz, M., Lu, C.Z., Campbell, E., and Gross, C.A. (2007). Analysis of σ^{32} mutants defective in chaperone-mediated feedback control reveals unexpected complexity of the heat shock response. Proc. Natl. Acad. Sci. USA *104*, 17638–17643.

Zhang, X., Liu, Y., Genereux, J.C., Nolan, C., Singh, M., and Kelly, J.W. (2014). Heat-shock response transcriptional program enables high-yield and high-quality recombinant protein production in Escherichia coli. ACS Chem. Biol. *9*, 1945–1949.