FoldEco: A Model for Proteostasis in *E. coli*

Evan T. Powers, 1,* David L. Powers, 2 and Lila M. Gierasch 3,*

1 Department of Chemistry, The Scripps Research Institute, La Jolla, CA 92037, USA
2 Department of Mathematics and Computer Science, Clarkson University, Potsdam, NY 13699, USA
3 Departments of Chemistry and Biochemistry & Molecular Biology, University of Massachusetts-Amherst, Amherst, MA 01003, USA

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

DOI 10.1016/j.celrep.2012.02.011

SUMMARY

To gain insight into the interplay of processes and species that maintain a correctly folded, functional proteome, we have developed a computational model called FoldEco. FoldEco models the cellular proteostasis network of the *E. coli* cytoplasm, including protein synthesis, degradation, aggregation, chaperone systems, and the folding characteristics of protein clients. We focused on *E. coli* because much of the needed input information—including mechanisms, rate parameters, and equilibrium coefficients—is available, largely from in vitro experiments; however, FoldEco will shed light on proteostasis in other organisms. FoldEco can generate hypotheses to guide the design of new experiments. Hypothesis generation leads to system-wide questions and shows how to convert these questions to experimentally measurable quantities, such as changes in protein concentrations with chaperone or protease levels, which can then be used to improve our current understanding of proteostasis and refine the model. A web version of FoldEco is available at http://foldeco.scripps.edu.

INTRODUCTION

As they fold to their native states, most proteins sample marginally stable misfolded states that are susceptible to self-association (Chiti and Dobson, 2006), which can turn the marginally stable misfolded states that are susceptible to self-association. They fold to their native states, most proteins sample marginally stable misfolded states that are susceptible to self-association. As they fold to their native states, most proteins sample marginally stable misfolded states that are susceptible to self-association. 

As they fold to their native states, most proteins sample marginally stable misfolded states that are susceptible to self-association. As they fold to their native states, most proteins sample marginally stable misfolded states that are susceptible to self-association. As they fold to their native states, most proteins sample marginally stable misfolded states that are susceptible to self-association. As they fold to their native states, most proteins sample marginally stable misfolded states that are susceptible to self-association. As they fold to their native states, most proteins sample marginally stable misfolded states that are susceptible to self-association. As they fold to their native states, most proteins sample marginally stable misfolded states that are susceptible to self-association. As they fold to their native states, most proteins sample marginally stable misfolded states that are susceptible to self-association. As they fold to their native states, most proteins sample marginally stable misfolded states that are susceptible to self-association. As they fold to their native states, most proteins sample marginally stable misfolded states that are susceptible to self-association. As they fold to their native states, most proteins sample marginally stable misfolded states that are susceptible to self-association. As they fold to their native states, most proteins sample marginally stable misfolded states that are susceptible to self-association. As they fold to their native states, most proteins sample marginally stable misfolded states that are susceptible to self-association. As they fold to their native states, most proteins sample marginally stable misfolded states that are susceptible to self-association.

RESULTS AND DISCUSSION

The FoldEco Model

FoldEco (Figures 1, S1, S2, S3, S4, and S5) tracks what happens to soluble proteins of interest (“client proteins”) as they are produced in the *E. coli* cytosol (membrane and periplasmic proteins are currently not included in FoldEco; see Extended Experimental Procedures). FoldEco has five systems:

1. Protein Synthesis and Folding

   The synthesis of a client protein (protein “i”) in FoldEco is represented as a simplified three-step process (see Extended...
Experimental Procedures). It begins with ribosomes (R) becoming translationally active (Rₐ,i). The protein then emerges from the ribosomal exit tunnel and the ribosome:nascent protein complex is formed (Rₑ,Ui). Finally, the ribosome releases the unfolded client protein (Ui) when synthesis is complete. A trigger factor monomer, T, can bind to the ribosome.
ribosome:nascent protein complex yielding the species R:T, R$_{42}$:T, or R$_{42}$:U:T (Deuerling and Bukau, 2004). The ribosome releases the unfolded client protein free (U$_i$) or in complex with trigger factor (T:U$_i$). The unfolded client protein reversibly folds to the native state (N$_i$) or reversibly misfolds to form an off-pathway intermediate (M$_i$). M$_i$ kinetically partitions between self-association by nucleated polymerization (Oosawa and Asakura, 1975) to form aggregates ($A_{ij}$, where there are $j$ monomers in the aggregate) or chaperone binding. We have not incorporated the possibility of cotranslational folding or misfolding (see Extended Experimental Procedures).

2. The DnaK/DnaJ/GrpE, or KJE, System

The role of the KJE system in proteostasis is to bind misfolded proteins and unfold them, thereby giving them another chance to fold (Mayer and Bukau, 2005). In FoldEco, the mechanism of the KJE system begins with unfolded or misfolded protein binding to dimeric DnaJ (J$_2$) or ATP-bound DnaK (K$_T$), forming binary complexes (J$_2$:U$_i$/M$_i$ or K$_T$:U$_i$/M$_i$, where the “/” indicates “either-or”). J$_2$:U$_i$/M$_i$ or K$_T$:U$_i$/M$_i$ can then add K$_T$ or J$_2$, respectively, to form the ternary complex K$_T$:U$_i$/J$_2$. However, K$_T$ binds weakly to substrates (Gisler et al., 1998; Mayer et al., 2000), so binding to J$_2$ is preferred. J$_2$ binding accelerates ATP hydrolysis by DnaK in the K$_T$:U$_i$/M$_i$:J$_2$ complex, which causes a conformational change within DnaK. This conformational change is propagated to the bound client protein, causing it to unfold. Thus, both K$_T$:U$_i$:J$_2$ and K$_T$:M$_i$:J$_2$ produce K$_D$:U$_i$:J$_2$ upon ATP hydrolysis (K$_D$ = ADP-bound DnaK). K$_D$:U$_i$:J$_2$ then releases J$_2$ to give K$_D$:U$_i$(K$_D$:U$_i$:J$_2$ can also be produced by slow ATP hydrolysis in K$_T$:U$_i$/M$_i$). Nucleotide exchange in DnaK is induced by binding to dimeric GrpE (E$_2$), producing K$_D$:U$_i$:E$_2$. Dissociation of E$_2$ yields K$_T$:U$_i$, which can dissociate to yield free unfolded protein, U$_i$. Note that because either U or M can enter the KJE cycle, and U$_i$ is released, this chaperone team provides a conduit from M to U. Note also that K$_D$:U$_i$ can also unproductively reenter the KJE cycle by binding to J$_2$, depending on the relative rates of substrate dissociation and J$_2$ binding. The total number of KJE cycles that a particular client protein experiences before it folds is determined by the relative rates of entry into the KJE cycle and folding. Our model for the KJE cycle is in part based on that of Hu et al. (2006).

3. The GroEL/GroES, or GroELS, System

The role of the GroELS system in proteostasis is to encapsulate client proteins, thereby providing them with an isolated environment in which to fold (Horwich et al., 2000). In FoldEco, the GroELS cycle begins when an unfolded or misfolded client protein binds to the cis ring of GroEL, which is either ATP-bound (GrL$_c$) or not (GrL), thus yielding GrL:U$_i$/M$_i$, or GrL$_c$:U$_i$/M$_i$. Therefore, client proteins can bind to either the ATP-free or ATP-bound state of GroEL, but there should be very little of the former at typical in vivo ATP concentrations (Tyagi et al., 2009). GrL:U$_i$ and GrL:M$_i$ are both converted to GrL$_c$:U$_i$/M$_i$ by ATP-binding-induced forced unfolding of the bound client protein (Lin et al., 2008). GrL$_c$:U$_i$/M$_i$ is then capped by GroES (GrS), which releases the client protein into the cis cavity (GrL$_c$:U$_i$/M$_i$:GrS), where it can fold or, in principle, misfold, to give GrL$_c$:U$_i$/M$_i$:N$_i$:GrS. In FoldEco, the rate of folding of a client protein in the GroEL cavity can be the same as it is in solution or can be different, as desired by the user; both circumstances have been reported (Apetri and Horwich, 2008; Chakraborty et al., 2010), and both are therefore allowed. ATP hydrolysis in the cis ring yields GrL$_c$:U$_i$/M$_i$:N$_i$:GrS, which can follow one of two pathways. In the first, ATP binding in the trans ring results in GrL$_c$:U$_i$/M$_i$:N$_i$:GrS:GrL$_t$ and release of ADP, GrS, and the client protein from the cis ring. In the second pathway, binding of unfolded or misfolded client protein (U$_i$ or M$_i$; the subscript k indicates that it could be a different type of client protein from the one in the cis cavity) in the trans ring results in GrL$_c$:U$_i$/M$_i$:N$_i$:GrS:GrL$_t$:U$_i$/M$_i$, which then binds ATP in the trans ring, stimulating release of ADP, GrS, and client protein from the cis ring. The trans ring is left in the GrL$_t$:U$_i$/M$_i$ state, ready to reenter the cycle. The partitioning between the first and second pathways depends on the rates of binding of ATP versus substrate. As with the KJE cycle, the total number of GroELS cycles that a particular client protein experiences before it folds is determined by the relative rates of entry into the GroELS cycle and folding. Our model of the GroELS cycle is based in part on those of Tehver and Thirumalai (2008) and Jewett and Shea (2008).

4. The ClpB/DnaK/DnaJ/GrpE, or B+KJE, Disaggregation System

Disaggregation is critical for the recovery of E. coli from heat shock (Mognon et al., 2003). This function can be performed for small aggregates by the KJE cycle (Diamant et al., 2000), but larger aggregates require the AAA+ ATPase ClpB, “B” for short (Mognon et al., 2003; Weibezahn et al., 2004). We based the mechanism of the B+KJE cycle in FoldEco on several studies that suggest that the KJE system prepares aggregates for B (Acebron et al., 2008; Acebron et al., 2009; Doyle et al., 2007), which then removes monomers from aggregates. Thus, client protein aggregates of size $j$ ($A_{ij}$) bind first to J$_2$ or K$_T$ (we have assumed a 1:1 stoichiometry of chaperone:aggregate in these complexes; see Extended Experimental Procedures). The resulting complexes (J$_2$:A$_{ij}$ or K$_T$:A$_{ij}$) then bind to K$_T$ or J$_2$, respectively, to form the ternary complex K$_T$:A$_{ij}$:J$_2$. ATP hydrolysis produces K$_D$:A$_{ij}$:J$_2$, where “i” indicates that the aggregate is now prepared for ClpB binding. Release of J$_2$ yields K$_D$:A$_{ij}$, which has two possible fates. First, it can bind to E$_2$, which induces nucleotide exchange to give K$_T$:A$_{ij}$:E$_2$. For small oligomers ($j \leq 4$ in FoldEco), a monomer is lost and E$_2$ dissociates to give K$_T$:A$_{i-1}$; and a free monomer, U$_i$ (Diamant et al., 2000). For larger aggregates, E$_2$ dissociates and K$_T$:A$_{ij}$ is produced without monomer loss, consistent with the observation that the KJE system on its own cannot disperse large aggregates (Diamant et al., 2000). K$_T$:A$_{i-1}$ and K$_T$:A$_{ij}$ can then either reenter the cycle or dissociate to give a free aggregate and K$_T$. Second, K$_D$:A$_{ij}$ can bind to B to form K$_D$:A$_{ij}$:B: $E_2$ binding followed by nucleotide exchange gives $E_2$:K$_D$:A$_{ij}$:B, from which E$_2$ and K$_T$ dissociate to give A$_{ij}$:B. Finally, a monomer is translocated through the central pore of B (Weibezahn et al., 2004) and the complex dissociates, giving A$_{ij}$:U$_i$, and B.

5. Protein Degradation

The most important proteases for proteostasis in E. coli are the energy-dependent proteases (Gottesman, 2003). FoldEco incorporates two degradation pathways. The first features Lon, a protease that degrades unfolded and misfolded client proteins (Gottesman, 2003). This pathway begins with Lon and U$_i$ or M$_i$, associating to form the reversible complex Lon:U$_i$/M$_i$. The bound substrate is then transferred to the proteolytic chamber (with concurrent forced unfolding, if the substrate is misfolded), to
form Lon:U\(_i^*\), where * indicates that the substrate is now committed to degradation. Degradation is processive, with the ATPase domains of Lon feeding the substrate into the proteolytic chamber. In the second pathway, ClpAP-type proteases (Dn) degrade native proteins (N) that have been tagged with a degradation signal, or degron (Gottesman, 2003; Varshavsky, 2011). This pathway begins with Dn binding to N, perhaps after delivery by an adaptor protein such as ClpS (Erse et al., 2006), to form the reversible complex Dn:N. The bound client protein is then forcibly unfolded and transferred to the proteolytic chamber, yielding Dn:U\(_i^*\). The substrate is degraded processively as described above. Note that degron installation, adaptor protein binding, and delivery of the substrate to the protease are subsumed into the binding step in FoldEco (see Extended Experimental Procedures).

Implementing and Parametrizing FoldEco
FoldEco is implemented by writing the differential equations that describe the time-dependent concentrations of each species in the model. Solving this system of equations requires the model parameters (the rate constants and initial concentrations) to have numerical values. Fortunately, the rate constants for virtually every step in FoldEco can be estimated from the literature (Figures S1, S2, S3, S4, and S5). Note that parameters such as chaperone-substrate binding affinity and on/off rates are protein-specific. The default parameters that we have used here and on the FoldEco website were measured for model proteins and can be changed by users on the input page of the FoldEco website. Literature estimates of the initial concentrations of the proteostasis network components are also available (Table S1). Parameters for the folding of a particular client protein can be estimated from literature data, experiments, or models; for example, that of Ghosh and Dill for protein stability (Ghosh and Dill, 2010; Ghosh and Dill, 2009) or those of Plaxco et al. (1998) or Ouyang and Liang (2008) for folding rates. The differential equations are then solved numerically, and the time-dependent concentrations of all species are the output.

Any model of biological networks must by necessity contain some approximations. Thus, FoldEco in its present form does not account for the effect of bacterial growth on proteostasis, stress responses, or possible changes in ATP levels. These and other approximations and their potential effects on the model are discussed in the Extended Experimental Procedures.

Models such as FoldEco are powerful tools for rationalizing experimental observations and generating hypotheses. We describe in the following sections some illustrative examples of findings and predictions derived from our own exploration of FoldEco. We have deployed a web version of FoldEco (described in the final section) to allow others to use it for their own purposes.

The Importance of Synthesis and Degradation Rates to Protein Aggregation
An important use for any model of proteostasis is to characterize the circumstances under which a client protein aggregates. Aggregation is possible when the client protein’s aggregation-prone intermediate (in FoldEco, the M\(_i\) state) has a concentration higher than its critical concentration for aggregation, or \(C_{\text{crit}}\) (Oosawa and Asakura, 1975). The concentration of M\(_i\), in turn, is determined by the balance between the rates of protein synthesis and of degradation by Lon, the folding and misfolding energetics, and of course chaperones (see Extended Experimental Procedures).

To illustrate the effect of synthesis rate and Lon concentration on protein aggregation, we used FoldEco to examine the behavior of a model client protein in the nonphysiological, but nevertheless instructive, situation of being in the absence of any chaperones. The model protein was assigned a moderately aggregation-prone “biophysical profile,” which includes the following (see Figures 1 and S1): the folding rate and equilibrium constants (\(k_b\) and \(K_b\)), the misfolding rate and equilibrium constants (\(k_m\) and \(K_m\)), the monomer-aggregate association constant (\(k_a\)), and the critical concentration for aggregation (\(C_{\text{crit}}\)). Thus, the model protein was given moderately favorable misfolding parameters (\(k_m\) and \(K_m\) = 1 s\(^{-1}\) and 10, respectively), more favorable folding parameters (\(k_b\) and \(K_b\) = 1 s\(^{-1}\) and 1000, respectively), moderately fast aggregation kinetics (\(k_a\) = 0.1 \(\mu\)M\(^{-1}\) s\(^{-1}\)), and a strong aggregation propensity (\(C_{\text{crit}}\) = 0.1 \(\mu\)M). The chosen value of \(C_{\text{crit}}\) is comparable to the in vitro critical concentration of A\(_{\beta_{22}}\), the 42-residue-long and most aggregation-prone of the major isoforms of the A\(_{β}\) peptide, the aggregation of which is implicated in Alzheimer disease (Usui et al., 2009). The other parameters used in these simulations are listed in Table S2. FoldEco was solved for this model protein at synthesis rates varying from 0.01 \(\mu\)M s\(^{-1}\) to 1 \(\mu\)M s\(^{-1}\). The upper end of this range is 70% of the maximal protein synthesis rate given the parameters used in this simulation (see Extended Experimental Procedures). At each synthesis rate, we determined the minimum concentration of Lon hexamer that would suppress aggregation to below 5% of total protein after 10,000 s of simulation time (about 2.75 hr; Figure 2A, black curve). This length of simulation time was arbitrarily chosen, but qualitatively similar results are obtained at different times and at steady state (Figure S6A). The time required to reach steady state appears to depend primarily on the folding energetics (\(k_b\) and \(K_b\)) and degradation rates (see Extended Experimental Procedures).

These simulations show the importance of synthesis and degradation rates to protein aggregation. The typical concentration of Lon hexamers in E. coli is about 0.3 \(\mu\)M (Table S1), which is the minimum concentration needed to suppress aggregation at the relatively slow synthesis rate of 0.019 \(\mu\)M s\(^{-1}\) (Figure 2A). Higher synthesis rates for this aggregation-prone model protein require higher concentrations of Lon (or the introduction of chaperones; see below). Heat shock can increase the concentration of Lon by up to four-fold (Zhao et al., 2005); a four-fold increase in Lon concentration would increase the synthesis rate that can be tolerated to 0.074 \(\mu\)M s\(^{-1}\) (Figure 2A). Expression from a plasmid can give even higher concentrations of Lon, which would enable higher synthesis rates.

Higher synthesis and degradation rates may be desirable for protein production because they lead to higher concentrations of native protein (Figure 2A, blue numbers). For example, at a synthesis rate of 0.01 \(\mu\)M s\(^{-1}\) and Lon hexamer concentration of 0.12 \(\mu\)M (the minimum required to suppress aggregation), the native state concentration of our model protein at 10,000 s...
As in (A), but with a full complement of chaperones. The parameters used in this simulation are listed in Table S2. Fractions of the total synthesized protein that have been degraded at 10,000 s at the same points are indicated in red font. The concentrations (μM) of native protein present at 10,000 s for points along the black curve are indicated in blue font. The fractions of the total synthesized protein that have been degraded at 10,000 s at the same points are indicated in red font.

See also Figure S6.

Figure 2. Aggregation and the Balance between Synthesis and Degradation

(A) The black curve represents the minimum Lon concentration, or [Lon], required to suppress aggregation to <5% of total extant protein after 10,000 s of simulation time for an aggregation-prone model protein as a function of synthesis rate in the absence of chaperones. Below the curve (gray), aggregation is largely suppressed. The synthesis rates that can be tolerated at the normal and heat-shock Lon concentrations are indicated by dashed lines. The concentrations (μM) of native protein present at 10,000 s for points along the black curve are indicated in blue font. The fractions of the total synthesized protein that have been degraded at 10,000 s at the same points are indicated in red font.

(B) As in (A), but with a full complement of chaperones. The parameters used in this simulation are listed in Table S2. See also Figure S6.

is 41 μM. Increasing the synthesis rate 10-fold to 0.1 μM s⁻¹ and the Lon hexamer concentration to 1.6 μM gives a native protein concentration of 112 μM at 10,000 s. Analyzing the kinetics of protein accumulation (Figure S6B) and a simplified model for protein synthesis and degradation (Figure S6C) reveals that this increase in the native protein concentration arises not just from the higher synthesis rate, but rather from the suppression of misfolding by degradation, which increases the concentration of the unfolded state during protein synthesis, and thereby increases the rate of production of the native state (see Extended Experimental Procedures).

The strategy of jointly increasing the synthesis rate and Lon concentration to increase native protein concentration comes at a cost to the cell. Our simulations suggest that although the native protein concentration increases as the synthesis and degradation rates increase, the fraction of protein that is degraded increases as well (Figure 2A, red numbers). Thus, while our model protein can be kept soluble when the synthesis rate is as high as 1 μM s⁻¹, very high Lon concentrations (13 μM; Figure 2A) are required to do so. In these circumstances, 94% of the total synthesized protein becomes degraded. The enormous metabolic cost of so much wasted protein synthesis would be unacceptable to an experimenter expressing a protein in E. coli, but would be a hindrance to an organism in its natural setting (Stoebel et al., 2008).

Proteostasis can be efficiently maintained at considerably lower metabolic cost by using chaperones to minimize aggregation. To demonstrate this notion, we repeated the simulations described above with a full complement of chaperones (20 μM trigger factor, 30 μM DnaK, 1 μM DnaJ, 15 μM GrpE, 3 μM GroEL tetradecamers, 5 μM GroES heptamers, 0.3 μM ClpB hexamers; Table S1). At physiological concentrations of Lon (0.3 μM), chaperones permit synthesis rates up to almost 0.3 μM s⁻¹, about 15-fold higher than in the absence of chaperones, and yield much higher concentrations of native protein (Figure 2B). Higher synthesis rates, however, require sharply increasing concentrations of Lon to suppress aggregation. Thus, chaperones do not entirely obviate the need for degradation during the expression of aggregation-prone proteins.

Misfolding-Prone Proteins Benefit the Most from the KJE System

In addition to enhancing our understanding of in vivo protein aggregation, models for proteostasis should also be useful for revealing how client proteins with particular biophysical profiles benefit from particular chaperoning mechanisms. To find client proteins that benefit strongly from the KJE system, we generated 4,000 random biophysical profiles, input each of them into FoldEco, and solved the model with a modest synthesis rate (0.02 μM s⁻¹), typical concentrations of Lon (0.3 μM hexamer) and trigger factor (20 μM; Table S1), and with or without the KJE system (30 μM DnaK, 1 μM DnaJ, and 15 μM GrpE; Table S1). The GroELS system and ClpB were excluded from the simulations in order to yield a focused view of profiles for which the KJE system alone influenced client outcome. Although such conditions cannot exist in reality, because GroEL is an essential protein in E. coli, isolating the KJE system like this is the best way to understand its client preferences and overall capacity to maintain proteostasis.

To generate the random biophysical profiles, $k_f$ and $k_m$ were varied between 10⁻³ and 10² s⁻¹; $k_f$ was varied between 10⁻³ and 10² μM⁻¹ s⁻¹; $K_f$ was varied between 10² and 10⁶; $K_m$ varied between 10⁻³ and 10³; and $C_m$ was varied between 1 mM and 1 nM (Table S3). Where possible, the bounds on these variables were based on available protein-folding data (see Extended Experimental Procedures). The parameters used for the
association between client proteins and DnaK and DnaJ, were those measured for the heat shock transcription factors Hsf1 and Hsp105 (Gamer et al., 1996; Mayer et al., 2000). The other parameters used in these simulations are listed in Table S4.

We examined the biophysical profiles of the client proteins with the largest (top 5%) increases in the concentration of the native state at an arbitrarily chosen time point (10,000 s) upon adding the KJE system (Table S3). The greatest increases varied from about +50 to +139 μM, which are substantial enhancements given that 200 μM of protein was synthesized. The outstanding feature of the biophysical profiles of the top KJE substrates is their uniformly high misfolding propensities (Figure 3A). This feature shows that client proteins that benefit most from the KJE system have a strong tendency to misfold. By pumping client proteins from the M state, which cannot fold, to the U state, which can, the KJE system promotes folding and rescues client proteins from aggregation or degradation. Which of these two fates a client protein is rescued from depends on its aggregation propensity, as measured by \( C_{\text{crit}} \).

At a synthesis rate of 0.02 μM s\(^{-1}\), client proteins need an exceptionally strong aggregation propensity (low \( C_{\text{crit}} \) value) to aggregate because so little protein is produced: the median \( C_{\text{crit}} \) value for proteins that aggregated more than 5% (total aggregate concentration > 10 μM) in the absence of the KJE system was 7.4 nM (Figure S7A). Thus, only 25% of the top KJE substrates are rescued primarily from aggregation (Table S3). The rest are rescued from degradation.

Figure 3. Biophysical Profiles for Client Proteins that Benefit the Most from the KJE and GroELS Systems at a Low Synthesis Rate

(A) Medians (red points) and interquartile ranges (black lines) for the misfolding, folding, and aggregation parameters of the best substrates for the KJE system. These substrates were identified from a pool of 4,000 client proteins with randomly generated parameters as having the largest (top 5%) KJE-system-mediated increases in native state concentration after 10,000 s at a synthesis rate of 0.02 μM s\(^{-1}\). The median values of the parameters for the underlying pool of 4,000 randomly generated client proteins are shown as horizontal dashed lines. The parameters represented are as follows: \( k_m \), misfolding rate constant; \( K_m \), misfolding equilibrium constant; \( k_f \), folding rate constant; \( K_f \), folding equilibrium constant; \( k_a \), monomer-aggregate association rate constant; \( C_{\text{crit}} \), critical concentration of misfolded protein required for aggregation. Note the inverted scale for \( C_{\text{crit}} \).

(B) As in (A), but for the top GroELS system substrates.

(C) Increase in native state concentration, [N], due to the simultaneous presence of the KJE and GroELS systems (the “combined effect” ; y axis) plotted against the sum of the increases due to the individual presence of either the KJE or the GroELS system (the “summed effects” ; x axis). Region 1 contains the 415 cases for which the combined effects are less than the summed effects, region 2 contains the 161 cases for which the combined and summed effects are roughly equal, and region 3 contains the 81 cases for which the combined effects are greater than the summed effects. Red open circles indicate cases in which the KJE or GroELS systems by themselves almost completely rescue the client protein. Blue open circles indicate cases in which neither system by itself completely rescues the client protein.

(D) As in (A), but for the substrates in region 3 of (C), for which the combined effects of the KJE and GroELS systems are greater than the summed effects. Table S3 lists all of the 4,000 randomly generated biophysical profiles and the effects of adding the KJE system, GroELS system, or both systems at a synthesis rate of 0.02 μM s\(^{-1}\). Table S4 lists all of the other parameters used.

See also Figure S7.
Slow Folding Proteins Benefit the Most from the GroELS System

To find the biophysical profiles that benefit the most from the GroELS system, we repeated the simulation experiment described above with or without the GroELS system (3 μM GroEL tetradecamers, 5 μM GroES heptamers; Table S1). The KJE system and ClpB were excluded in order to focus on the impact of the GroELS system, and the folding and misfolding parameters were assumed to be the same inside and outside the GroEL cavity. Binding parameters for typical GroELS system substrates were used (Figure S3, Table S4). At a synthesis rate of 0.02 μM s⁻¹, the top 5% of increases in the concentration of the native state upon addition of the GroELS system varied from +102 to +153 μM (Table S3). In addition, the native state concentration increased by more than 10 μM for 2,223 of the 4,000 randomly generated client proteins when the GroELS system was added, but for only 737 client proteins when the KJE system was added (Table S3). This result demonstrates that the GroELS system is beneficial to a broader range of our randomly generated client proteins than the KJE system at this synthesis rate. The efficiency of the GroELS system is especially remarkable considering that the folding and misfolding rates in the GroEL cavity were the same as they are in solution. If faster folding rates inside the cavity had been used, the effect of the GroELS system on native state concentration would presumably have been even larger.

A striking feature of the top GroELS substrates is their low and narrowly distributed folding rate constants (Figure 3B). This rate selectivity arises from two factors: fast-folding client proteins (kf > 0.1 s⁻¹) tend not to need chaperone assistance, while very slow-folding client proteins (kf < 0.01 s⁻¹) fold too slowly for the effect of the GroELS system to take hold on the time scale of the simulation. The optimal folding rate for GroELS substrates is determined by the rate constant for ATP hydrolysis by GroEL, which determines how long substrates remain encapsulated. In FoldEco, the rate constant for this process is set to 0.1 s⁻¹ (corresponding to a half-life of about 7 s), as generally found experimentally (Ranson et al., 1997; Rye et al., 1999). In principle, the optimal folding rate constant for GroELS substrates could be changed by changing the ATP hydrolysis rate constant. However, if this rate constant is too low, the GroELS system retains substrates for so long that it quickly becomes saturated; if this rate constant is too high, the GroELS system does not retain slow-folding client proteins long enough for them to fold appreciably and, as noted above, fast-folding substrates have no need of the GroEL system. The ATP hydrolysis rate constant of GroEL appears to have evolved to an optimum.

The other noteworthy feature of the biophysical profiles of the top GroELS substrates is their low misfolding propensities (Figure 3B). In fact, only 12% of the top GroELS substrates are rescued from aggregation, compared with 25% of the top KJE substrates (Table S3). The GroELS system is apparently not optimized to correct misfolding. It is important to note, however, that while misfolding-prone proteins may not be among the top GroELS substrates, they nevertheless benefit substantially from the GroELS system. For example, of the 737 proteins whose native state concentration increases by > 10 μM upon addition of the KJE system, 657 (89%) benefit similarly upon addition of the GroELS system. This observation suggests, and experiments have shown (Vorderwülbecke et al., 2004), that the GroELS system can complement the loss of the KJE system, at least partially.

Cooperation between the KJE and GroELS Systems

To determine whether the biophysical profiles that benefit the most from the KJE or GroELS systems are different when one system is introduced in the presence of the other than when they are introduced individually, we repeated the simulation experiment described above with both systems present. The results from these simulations were similar to those from the previous simulations. The top 5% of the increases in the concentration of the native state upon addition of the KJE system in the presence of the GroELS system ranged from +36 to +122 μM, and the native state concentration increased by more than 10 μM for 572 of the 4,000 client proteins. The top 5% of the increases in the concentration of the native state upon addition of the GroELS system in the presence of the KJE system ranged from +101 to +156 μM, and the native state concentration increased by more than 10 μM for 2,013 of the 4,000 client proteins. The biophysical profiles of the top KJE and GroELS substrates were nearly identical in the presence (Figures S7B and S7C, respectively) and absence (Figures 3A and 3B, respectively) of the other system.

The observations above suggest that the activities of the KJE and GroELS systems are independent and complementary. The extent to which this idea is true is demonstrated by Figure 3C. As noted above, 657 of the 4,000 biophysical profiles studied had increases in native state concentration upon addition of the KJE system or the GroELS system that were >10 μM. For this subset, we have plotted in Figure 3C the increase in native state concentration due to the simultaneous presence of the KJE and GroELS systems (the “combined effect”; y axis) against the sum of the increases due to the individual presence of either the KJE or the GroELS system (the “summed effects”; x axis). This plot can be split into three regions: region 1 contains the 415 cases for which the combined effects are less than the summed effects, region 2 contains the 161 cases for which the combined and summed effects are roughly equal, and region 3 contains the 81 cases for which the combined effects are greater than the summed effects.

The cases in region 2 are perhaps the easiest to understand: the effects of the KJE and GroELS systems are simply independent and additive. In contrast, the KJE and GroELS systems appear to interfere with each other to some extent for the cases in region 1. This is partly due to an artifact, as 247 of the cases in region 1 (60%) are there because the KJE or GroELS systems by themselves completely rescue the client protein (red open circles). In these cases, a second chaperone system will always be superfluous. For the remaining 168 cases in region 1, the KJE and GroELS systems are partially redundant (blue open circles).

For the cases in region 3, the KJE and GroELS systems are synergistic. The biophysical profiles of these cases (Figure 3D) show that they combine the features of the top KJE and GroELS substrates: they have high propensities to misfold (high Km) and
slow folding rates (low $k_f$). For such substrates, the ability of the KJE system to recover protein from the misfolded state apparently increases the input into the GroELS system, which then increases the flux of protein into the native state. The cases in regions 1 (Figure S7D) and 2 (Figure S7E) have lower misfolding propensities and faster folding rates, so that the KJE and GroELS systems are less able to amplify each other’s effects.

The Effect of Increasing the Synthesis Rate on the Biophysical Profiles of the Top KJE and GroELS Substrates

We probed the effect of a higher synthesis rate on the biophysical profiles of the top KJE and GroELS substrates by repeating the simulation experiments described above with a 10-fold higher synthesis rate ($0.2 \, \mu M \, s^{-1}$) used in the simulation. The top 5% of the increases in the concentration of the native state at 10,000 s upon addition of the KJE system by itself or in the presence of the GroELS system are proportionately higher than they were at the lower synthesis rate, ranging from +643 to +1,526 $\mu M$ without the GroELS system and from +512 to +1,507 $\mu M$ with the GroELS system (2,000 $\mu M$ of protein was synthesized; Table S5). The biophysical profiles of the top KJE substrates at this synthesis rate (Figure 4A) are nearly identical with and without GroELS (Figure 4A and Figure S8A) and are both similar to those of the top KJE substrates at the lower synthesis rate, but with a new feature. The misfolding propensities of the top KJE substrates are still high, but now their aggregation propensities are high as well (Figure 4A). Thus, aggregation appears to be more important at this synthesis rate than at the lower synthesis rate. Indeed, all of the top KJE substrates are rescued from aggregation rather than degradation at this synthesis rate. Rescue from aggregation is not guaranteed by the presence of the KJE system, however. The aggregation rate constants for the top KJE substrates tend to be low (Figure 4A), suggesting that fast aggregation interferes with the KJE system.

At the higher synthesis rate, the top 5% of the increases in the concentration of the free native state upon addition of the GroELS system ranged from +279 to +1,119 $\mu M$ without the KJE system and from +135 to +1,119 with the KJE system (Table S5). Unlike the situation with the KJE system, the biophysical profiles of the top GroELS substrates in the absence (Figure 4B) or presence (Figure S8B) of the KJE system are different at this synthesis rate than at the lower synthesis rate. The folding rate constants are not as low as they were before for the top GroELS substrates. Instead, all of the parameters associated with misfolding and aggregation ($k_m$, $K_m$, $k_a$, and $C_{crit}$) are high, indicating that the top beneficiaries of the GroELS system at higher synthesis rates are misfolding and aggregation prone rather...
than slow folding. Consistent with this observation, all of the top GroELS substrates are rescued primarily from aggregation.

These observations suggest that aggregation rather than degradation becomes the dominant problem in proteostasis as the synthesis rate increases, and the KJE and GroELS systems are comparably able of rescuing proteins from misfolding and aggregation. This assertion is demonstrated by the concentration of the native state increasing by more than 100 fold when expressed in bacteria at a slow or fast translation rate. Specific activity is normalized to the total mass of protein.

The experimental data for (A) and (B) were taken from Siller et al., 2010; see Extended Experimental Procedures for details. See also Figure S9.

Figure 5. Tests of FoldEco Predictions
(A) Comparison of the experimental (gray bars) and predicted (blue bars) fraction of Luc that is soluble when expressed in bacteria at a slow (left) or fast (right) translation rate.
(B) Comparison between experimentally determined (gray bar) and FoldEco-predicted ratios of the specific activities (SA) of Luc when expressed in bacteria at a slow or fast translation rate. Specific activity is normalized to the total mass of protein.

The results of our simulations at these chaperone concentrations match the experimental results well. They reproduce the increase in aggregation as well as the decrease in specific activity (Figures 5A and 5B, respectively; see also Table S6 and Extended Experimental Procedures) observed experimentally upon increasing the translation rate (Siller et al., 2010). FoldEco also reproduces the trends in the data at lower chaperone concentrations, but the extent of aggregation is generally overestimated (Figures S9C and S9D; Table S7). Thus, FoldEco currently encompasses most of the aspects of in vivo proteostasis needed to understand the effect of protein synthesis rate on the partitioning of protein between aggregates and native, functional protein.

Despite this success, one observation made by Siller et al. was not reproduced in these simulations: that the absolute activity (not just the specific activity) of Luc is higher in their expression system when translation is slower, despite less protein being produced per cell. Siller et al. attribute this effect to cotranslational folding at lower translation rates (Siller et al., 2010). The current version of FoldEco does not explicitly account for cotranslational folding, but we can mimic it by increasing the folding rate constant (kf) by a factor of about 2 in the simulations at the slower translation rate, but not at the faster translation rate (Figure S9E). This adjustment enables us to reproduce their observation (see Extended Experimental Procedures) and suggests that cotranslational folding doubles the effective folding rate of Luc.

This exercise demonstrates one of the most important uses of models for complicated systems: they enable concrete tests of whether a given set of processes is sufficient to explain a given...
set of experimental results. When they are not, it is clear that at least one important process have been left out of the model. In this case, cotranslational folding is not necessary to explain the decrease in aggregation of Luc when it is synthetized more slowly, but the importance of cotranslational folding to the production of active Luc is nevertheless unambiguous. The current version of FoldEco can be adapted to account for cotranslational folding by adjusting $k_i$, but future versions will include it explicitly.

The other hypotheses that are suggested by our results with FoldEco have to do with the characteristics of the substrates for the KJE and GroEL systems. For example, on the basis of Figures 3A and 4A, we expect that DnaK substrates in E. coli should be aggregation prone. This prediction is consistent with the results from a recent study of DnaK interactors. Calloni et al. found that proteins that are highly enriched on DnaK (Calloni et al., 2012) tend to be less soluble than the average proteins of soluble cell lysate (where solubility measurements were taken from a proteome-wide measurement of E. coli protein solubilities) (Niwa et al., 2009). On the basis of Figures 3B and 4B, we expect that GroEL substrates in E. coli should be slow folding, aggregation prone, or both. The latter prediction is consistent with experimental results (Chapman et al., 2006; Fujiwara et al., 2010; Kerner et al., 2005). The former prediction must await proteome-wide determination of folding rate constants to be evaluated, but it is consistent with the finding that obligate GroEL substrates tend to have topologically complicated structures (Fujiwara et al., 2010; Kerner et al., 2005), because proteins with such structures generally fold slowly (Plaxco et al., 1998).

**FoldEco as a Web Resource**

To enable those interested to perform their own experiments with FoldEco, we have created a web version of FoldEco, the home page of which is http://foldeco.scripps.edu. This website has an overview of the model, descriptions of its components and subsystems, and basic and advanced interactive pages into which parameters can be input and then sent to a server. The server solves FoldEco with the user-defined parameters and returns the output. There are several options for the output: fully graphical depictions of the simulation results (including static images, movies, and interactive animation), concentration versus time plots for a selection of the species in the model, and downloadable tables with concentration versus time data for some or all of the species in the model. Complete guides to the input parameters and the output data are provided on the website. We expect this publicly available resource to prove useful as a means to both generate hypotheses and rationalize existing data about proteostasis.

**EXPERIMENTAL PROCEDURES**

**General**

The differential equations that make up FoldEco were solved numerically with the use of Mathematica 8.0.1 (Wolfram Research) on a Dell Precision T7500 Workstation running Windows 7 Professional (64 bit) with a 3.33 GHz Intel Xeon W5590 Quadcore CPU and 24 GB of RAM. FoldEco is implemented on the web with the use of Wolfram webMathematica 3.0 running on a VMware virtual server running CentOS installed on an Oracle Sun Blade server. Details about setting up and solving FoldEco can be found in the Extended Experimental Procedures. A Mathematica application file with the code for FoldEco and a Mathematica notebook file demonstrating how to use FoldEco are provided in Folder S1.

**Setting Up and Solving FoldEco**

The differential equations that make up FoldEco can be derived from Figure 1 (or Figures S1, S2, S3, S4, and S5). The net rate of change of each species’ concentration is equal to the sum of the rates of the processes that produce it, less the sum of the rates of the processes that consume it (where the rate of a process is its rate constant multiplied by the concentrations of the species that the process consumes). For example, free inactive ribosomes (R in Figure 1) are consumed when they become translationally active (forming Ra,i) or when they bind trigger factor (T). R is produced by dissociation of R:T, and by dissociation of Ra:i, and Ra:i:T after translation is complete. The rate of change of the concentration of R is therefore

$$\frac{d[R]}{dt} = -s_{2,1}[R:T] - s_{3,1}[R] + s_{2,2}[R:T] + s_{1,11}[R:i:U] + s_{1,12}[R:i:U:T]$$  (1)

where the square brackets indicate concentrations, $d[R]/dt$ is the time derivative of $[R]$, and $s_{2,1}, s_{3,1}, s_{2,2}, s_{1,11}, and s_{1,12}$ are rate constants (the first subscript of these rate constants is 0 if the process does not involve a client protein and if it does; the second subscript indicates the reaction number in the synthesis system of FoldEco; see Figure S1). FoldEco is solved by using the numerical differential equation solver in Mathematica, which adaptively chooses its integration methods and step sizes depending on the properties of the system of differential equations (e.g., smoothness and stiffness of the solutions).

**SUPPLEMENTAL INFORMATION**

Supplemental information includes nine figures, seven tables, a Mathematica application file with the code for FoldEco, a Mathematica notebook file demonstrating how to use FoldEco, and Extended Experimental Procedures. This material can be found with this article online at doi:10.1016/j.celrep.2012.02.011.

**LICENSING INFORMATION**

This is an open-access article distributed under the terms of the Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported License (CC-BY-NC-ND; http://creativecommons.org/licenses/by-nc-nd/3.0/legalcode).

**ACKNOWLEDGMENTS**

We acknowledge with gratitude the contribution of Daniel W. Farrell (Stony Brook University) to the implementation of the FoldEco program. We thank Jeffery W. Kelly, R. Luke Wiseman, Amber N. Murray (The Scripps Research Institute), Ken A. Dill (Stony Brook University) and members of the Gierasch group (University of Massachusetts-Amherst) for helpful discussions. We thank F.U. Hartl and M. Hayer-Hartl for giving us early access to their DnaK interactome results. This work was funded by National Institutes of Health grants OD009845 and GM027616.

Received: October 24, 2011
Revised: February 21, 2012
Accepted: February 27, 2012
Published online: March 15, 2012

**REFERENCES**


