Thermosensitivity of growth is determined by chaperone-mediated proteome reallocation

Ke Chen[a], Ye Gao[b], Nathan Mih[a,c], Edward J. O'Brien[a], Laurence Yang[a], and Bernhard O. Palsson[a,d,e,1]

[a]Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093; [b]Division of Biological Sciences, University of California, San Diego, La Jolla, CA 92093; [c]Bioinformatics and Systems Biology, University of California, San Diego, La Jolla, CA 92093; [d]Department of Pediatrics, University of California, San Diego, La Jolla, CA 92093; and [e]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kongens Lyngby, Denmark

Edited by Eugene I. Shakhnovich, Harvard University, Cambridge, MA, and accepted by Editorial Board Member Daniel L. Hartl September 8, 2017 (received for review April 4, 2017)

Maintenance of a properly folded proteome is critical for bacterial survival at notably different growth temperatures. Understanding the molecular basis of thermoadaptation has progressed in two main directions, the sequence and structural basis of protein thermostability and the mechanistic principles of protein quality control assisted by chaperones. Yet we do not fully understand how structural integrity of the entire proteome is maintained under stress and how it affects cellular fitness. To address this challenge, we reconstruct a genome-scale protein-folding network for *Escherichia coli* and formulate a computational model, FoldME, that provides statistical descriptions of multiscale cellular response consistent with many datasets. FoldME simulations show (*i*) that the chaperones act as a system when they respond to unfolding stress rather than achieving efficient folding of any single component of the proteome, (*ii*) how the proteome is globally balanced between chaperones for folding and the complex machinery synthesizing the proteins in response to perturbation, (*iii*) how this balancing determines growth rate dependence on temperature and is achieved through nonspecific regulation, and (*iv*) how thermal instability of the individual protein affects the overall functional state of the proteome. Overall, these results expand our view of cellular regulation, from targeted specific control mechanisms to global regulation through a web of nonspecific competing interactions that modulate the optimal reallocation of cellular resources. The methodology developed in this study enables genome-scale integration of environment-dependent protein properties and a proteome-wide study of cellular stress responses.

thermoadaptation | proteome allocation | bacterial growth law | genome-scale model | molecular chaperones

Temperature is one of the most important environmental parameters that dictate the evolution of bacterial species. Our current understanding of thermoadaptation is based on deep investigations from a few different standpoints. First, sequence and structural determinants of thermosensitivity are identified through comparison of homologous enzymes between psychrophilic, mesophilic, and thermophilic organisms (1, 2) or fitness-increasing mutations that arise during laboratory evolution at high temperatures (3–5). Second, efforts have been made to comprehend the detailed mechanisms by which molecular chaperones promote efficient folding, minimize toxic aggregation, and maintain a properly folded proteome under stressful perturbations (6–8). In particular, two major chaperone families that are well conserved across bacteria, the Hsp70 (9) and Hsp60 (10) systems, share the majority of the folding load in a cell. Therefore, physicochemical principles (11–15) and chaperone–substrate interactions (16–18) that regulate efficient folding for a single protein are extensively studied in in vitro experiments and theoretical models.

Empirical and population genetics models of bacterial growth try to explain the general principles for various species to adapt to diverse thermal niches. For example, temperature responses fit nicely using the activation enthalpy of a single rate-limiting reaction and thermodynamic parameters of its catalyzing enzyme (19). In another approach, focusing on marginally stable proteomes that may experience sharp and cooperative denaturation, thermo-response can be simply described by two parameters, a dominant metabolic activation barrier and the number of proteins controlling replication process in an organism (20, 21).

Apparently, evolution of protein thermodynamics and the function of a rate-limiting response reaction (presumably chaperone-assisted folding) are both critical to the temperature dependence of bacterial growth. To date, most studies provide in-depth investigation on the effect of only one aspect. However, the cytoplasm of a living cell is an active complex medium, where a large number of protein molecules with varied thermal qualities compete to achieve diverse cellular functions. How do proteins with different evolutionarily developed thermal features respond to instant temperature perturbations? How do molecular chaperones catch these changes and distribute their folding service to optimize growth? How do these two determinants interact at the systems level to modulate temperature response of a cell under different environmental conditions?

To answer these questions, we use the genome-scale network reconstructions and computational models of metabolism and protein expression [ME Models (22–24)] for *Escherichia coli*.

Significance

How do bacteria adapt to the diverse thermal niches on earth? Evidence accumulates in the protein sequence and structural determinants of thermosensitivity and mechanisms by which molecular chaperones aid protein folding. However, a comprehensive understanding of how thermoadaptation is achieved at the systems level is still missing. Here we reconstruct an integrated genome-scale protein-folding network for *Escherichia coli*, termed FoldME, that couples both contributing factors to the metabolic state of a cell. FoldME simulations reproduce the asymmetrical bacterial temperature response and delineate the multiscale strategies cells use to resist unfolding stresses induced by high temperature and destabilizing mutations in a single gene. The results highlight how global proteome allocation regulates thermoadaptation through balance between chaperones for folding and translational machinery for biosynthesis.

Author contributions: K.C. and B.O.P. designed research; K.C. and Y.G. performed research; N.M., E.J.O., and L.Y. contributed new analytic tools; K.C. analyzed data; and K.C., Y.G., and B.O.P. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. E.I.S. is a guest editor invited by the Editorial Board.

[1]To whom correspondence should be addressed. Email: palsson@ucsd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1705524114/-/DCSupplemental.

The *E. coli* ME Model is capable of generating fine-grained descriptions of proteome composition that optimizes cellular growth in a given environment. Furthermore, the decoupling of protein expression from metabolic requirement enables incorporation of multiple protein states, providing the framework to characterize the changing properties of proteins and lay out how chaperones are distributed to maintain protein quality control in vivo.

Herein, we present the reconstruction of the genome-scale protein-folding and chaperone network in *E. coli* K-12 MG1655. This reconstruction is then integrated into the ME Model to form FoldME. FoldME describes the in vivo protein folding as a competition between de novo spontaneous folding and assisted folding, using the HSP70 (DnaK/DnaJ/GrpE) and the HSP60 chaperonin system (GroEL/GroES). With the chaperones being allowed to respond dynamically to changes in the proteomic folding state, FoldME delineates how organismal fitness is affected by a variety of perturbations, such as temperature fluctuations, nutrient availability shifts, and genetic mutations. Importantly, we demonstrate that cellular response to unfolding stresses is more complicated than a simple decision about which folding pathway to choose for each unfolded peptide. It involves a systems-level proteome reallocation in accordance with empirical bacterial growth laws (25), to balance availability of chaperones for folding and the biosynthesis machinery to synthesize the proteome, including the chaperones.

Results

Model Reconstruction. Environmental and genetic perturbations modulate cell growth by changing the properties of the cell's protein components, followed by a subsequent reallocation of cellular resources in response to that alteration. To assess this effect, we first associate all biochemical reactions in the *E. coli* genome-scale ME Model [*i*OL1650 (24)] with the sequences and structures of their catalytic enzymes, using the protocols developed in our group (26, 27). Next, we compute the temperature-dependent protein kinetic folding rate $[k_f(T)]$, thermostability [free energy of unfolding $\Delta G(T)$], equilibrium constant of unfolding $K_{eq}(T) \triangleq \exp\{-\Delta G(T)/RT\}$, and aggregation propensity (*agg*) from first principles (*Materials and Methods*) for every protein in FoldME. The calculation provides us with a condition-specific characterization of the folding state of the proteome, which is then coupled to cell growth through flux balance formulation of the folding reactions described below.

Three pathways that actively fold nonspecific protein targets in vivo are incorporated to form the folding network (Fig. 1*A*): (*i*) the spontaneous folding pathway, (*ii*) the DnaK-assisted folding pathway, and (*iii*) the GroEL/ES-mediated folding pathway. Spontaneous folding occurs in the presence of trigger factor once a nascent peptide chain exits the ribosome. We describe the temperature-dependent unfolded fraction of an individual protein with the coupling constraint $V_{dilution} \geq (\frac{\mu}{k_f(T)} + K_{eq}(T)) \cdot V_{folding}$, where μ is the growth rate.

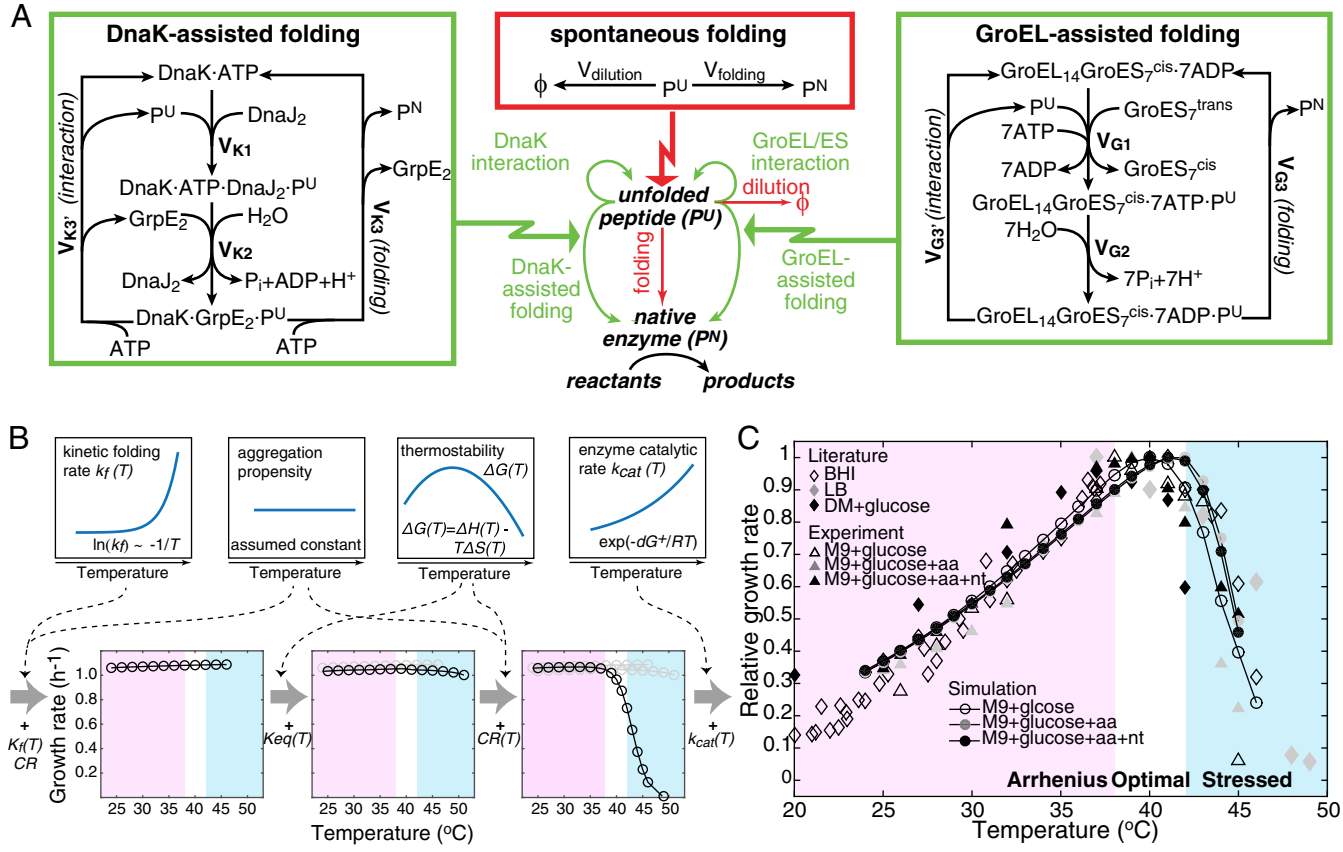

Fig. 1. FoldME reconstruction and validation. (*A*) Elementary model reactions for the three folding pathways. The flux going through each reaction is denoted $V_{reaction\_label}$, and the coupling constraints are explained in the text. (*B*) Illustration of how temperature dependence of each biophysical property is combined to compute their collective effect on cell growth. CR stands for chaperone requirement calculated from *agg* alone; CR(T) takes into account both *agg* and $\Delta G(T)$. (*C*) FoldME predictions (circles connected with a solid line) of relative growth rates of *E. coli* over temperatures, compared with data obtained from the literature (16, 50, 51) (diamonds) and in-house experiments (triangles).

DnaK- and GroEL/ES-assisted folding pathways have been studied extensively due to their central role in maintaining cellular proteostasis (12, 18, 28). Without loss of generality, we describe each chaperone-assisted folding pathway with three kinetically controlled elementary steps (Fig. 1*A*) that are constrained with the corresponding enzymatic turnover rates measured experimentally (29, 30). To reflect the fact that multiple repeated cycles of complete release and rebinding of the peptide are required for the substrate to reach its native state (31, 32), we design a pair of duplicated reactions, one for the successful folding event (V_{K3} and V_{G3}) and another unfruitful chaperone-interaction cycle that releases the unfolded peptide ($V_{K3'}$ and $V_{G3'}$). The ratio between the flux of these two reactions is set to an effective temperature-dependent chaperone requirement, $CR(T) = agg \cdot (1 + K_{eq}(T))$, to estimate the number of repeated cycles required for a particular peptide. Derivations and additional details for modeling the three folding pathways are described in *SI Materials and Methods*.

The *E. coli* protein-folding network has been developed in more detail based on the concerted action of the molecular chaperones and proteases to predict the folding outcome of a single protein (14, 15, 33). However, such a kinetic model gives no clue on how the chaperones should be partitioned simultaneously among the folding request from the whole proteome and how proteome folding is coupled to the metabolic state of a growing cell. Here, we allow the three folding pathways to compete for folding of any protein instead of being designated a priori to any particular clients. Through integration into *i*OL1650, this unique computational model, termed FoldME, is capable of dynamically adjusting the in vivo folding pathway of each protein based on its folding characteristics, as well as the proteome composition and metabolic state in a given environment.

Asymmetrical Temperature Dependence of Cell Growth. We simulated FoldME at different temperatures by computing the enzyme catalytic rate k_{cat} and proteomic biophysical profile according to statistical mechanical laws depicted in Fig. 1*B*. FoldME computes the proteins' folding properties separately from their metabolic activities and can thus assess how each property contributes to the nonlinear nature of the cell's temperature response over a wide range of temperatures (Fig. 1*C*). Remarkably, without any further assumptions or parameter fitting, the predicted relative growth rate agrees quantitatively with the independent experimental data from 24 °C to 46 °C, in both minimal glucose and defined rich media.

Over the Arrhenius growth temperatures (24∼37 °C, region in pink in Fig. 1*C*), change in growth rate is governed by the temperature dependence of enzyme catalytic rates. We estimated the equivalent Arrhenius activation energy for cell growth to be 55.9 ± 1.1 kJ/mol, consistent with the experimental value 56.5 kJ/mol previously measured for *E. coli* in rich media (34). Between 38 °C and 42 °C, growth rate varies only in a small range, and the optimal growth temperature is dependent on the medium type. Consistent with our experimental measurements, FoldME predicted that the optimal growth temperature in rich medium was slightly higher (∼1 °C) than that in the minimal glucose medium. Relatively constant growth in this temperature range is maintained by an intricate competition between all contributing factors. At higher temperatures ($T \geq 42$ °C, region in blue in Fig. 1*C*), neither the increased kinetic folding rate nor the elevated enzyme catalytic rate is enough to compensate for the cost of maintaining stability of the unfolding proteins; hence, the growth rate decreases sharply.

In addition to predictions of growth behavior, FoldME simulations correctly capture the intracellular abundance and temperature response of the molecular chaperones. At 37 °C, FoldME estimates DnaK to contribute 0.72% to the total mass in defined rich medium, consistent with the estimation of ∼1% total proteome mass during exponential growth (35). At 42 °C, DnaK is calculated to increase by 2.3-fold, which is very close to our experimental measurements (2.1 ± 0.1). Although abundances of GroEL vary in different experiments, FoldME predictions capture the general trend that GroEL is slightly lower in mass fraction than DnaK at physiological temperatures (Fig. S1*A*).

Quantitative consistency in the up-regulation of chaperones at higher temperatures is obscured due to the difference between FoldME simulations that reflects an evolved global optimum and experiments usually performed for the WT cells. Nevertheless, in a study that evolved *E. coli* to an extreme temperature of 48.5 °C, GroEL is determined to increase 16-fold over its WT level (36), partially supporting our FoldME estimations. The increased level of chaperones is used to keep the total unfolded protein fraction below 1% of the proteome mass under all temperatures (Fig. S1*B*). Therefore, FoldME simulations faithfully recapitulate the critical function of the chaperone network in buffering the temperature-induced unfolding stresses and maintaining robust cell growth.

Chaperone-Mediated Proteome Reallocation Details the Empirical Bacterial Growth Law. The phenotypic change adapted to different temperatures is reflected in proteome allocation strategies. We compared computed gene expression profiles at 28 °C and 45 °C, where the growth rates are similar (Fig. 2 and Fig. S2). In the Arrhenius temperature range, the shift in gene expression is minor and homogeneously distributed to all metabolic enzymes to compensate for the overall decrease in enzymatic efficiency. In contrast, under severe unfolding stress at 45 °C, the up-regulation of chaperones significantly drains cellular resources away from ribosome synthesis, limiting the synthesis of all other cellular components. Thus, chaperones not only respond to unfolding needs, but also mediate global proteome reallocation by setting a constraint on the use of cellular resources for biomass synthesis.

We further detailed the constraints associated with proteome allocation, using large-scale simulations of 21 nutrients that were

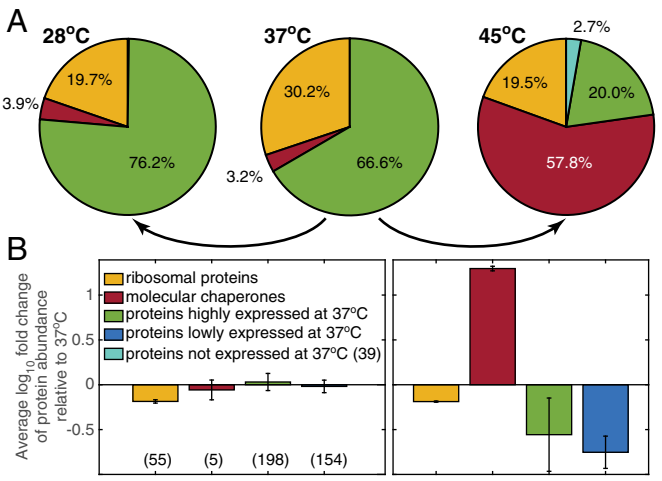

Fig. 2. Proteome reallocation with change in temperature. (*A*) Pie charts of the computed mass fraction for ribosomal proteins (yellow); molecular chaperones (red); and proteins that are highly expressed (green), lowly expressed (blue), and not expressed (cyan) at 37 °C. Percentage is calculated with respect to all expressed proteins in FoldME (454, 446, and 448 at 28 °C, 37 °C, and 45 °C, respectively). (*B*) The average fold change in protein abundance at 28 °C (*Left*) and 45 °C (*Right*) with respect to 37 °C. The number of expressed proteins used for fold change calculation is shown in parentheses for each colored category. The error bars indicate variation of fold change in the particular protein fraction shown.

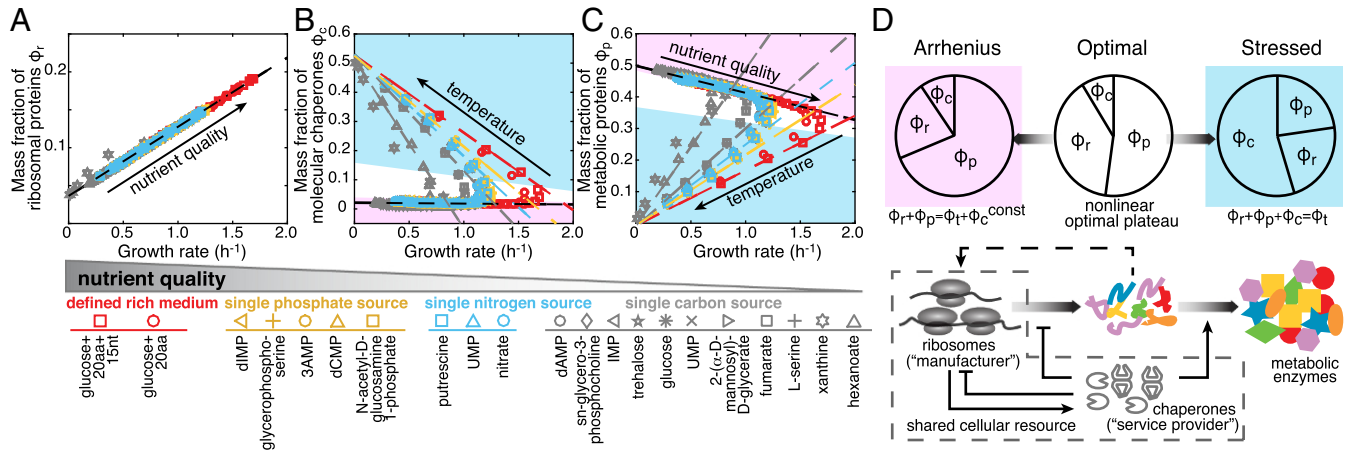

Fig. 3. Growth law for the chaperone-regulated proteome. (*A–C*) The linear relationship between growth rate and mass fraction of the ribosomal proteins (Φ_r), molecular chaperones (Φ_c), and metabolic proteins (Φ_p). Arrows indicate the direction of nutrient quality or temperature increase. Nutrient quality is ordered within each nutrient group by predicted growth rate at 37 °C (Fig. S3). (*D*) Schematic of the bacterial growth law and chaperone's regulatory role in proteome allocation. Φ_t denotes the total expressed proteome.

previously shown to best represent the diversity of the proteome under different conditions (37). With implementation of the chaperone network, computed growth rates showed consistent temperature response from 24 °C to 46 °C for all nutrients considered (Fig. S3).

The mechanism underlying these detailed predictions is consistent with reported coarse-grained bacterial growth laws (25, 38, 39). Three overall fractions of the expressed proteome, ribosomal proteins (Φ_r), molecular chaperones (Φ_c), and metabolic proteins (Φ_p), each show a different type of growth-rate dependency (Fig. 3 *A–C*). The number of ribosomes in a cell directly determines how much biomass a cell can produce; therefore Φ_r increases linearly with growth rate under all simulated conditions. Consistent with the empirical bacterial growth law, growth rate increases with better nutrient quality, representing a higher translational efficiency. In the Arrhenius temperature range, Φ_c remains constant so that the allocation tradeoff between ribosomal and metabolic proteins gives rise to the conjugate growth-related change between Φ_r and Φ_p. In the stressed temperature range, growth is modulated by the partition among all three fractions. Φ_c becomes linearly dependent on the growth rate with a negative slope, similar to the relationship between Φ_r and growth rate under translational inhibition by antibiotics (25). The slope changes according to the nutrient quality, which likely modulates the overall folding efficiency of the chaperone. Between the two temperature ranges, nonlinearity arises within an "optimal plateau" (38), due to the suboptimal level of both folding and translational efficiency.

The chaperones' regulatory role in growth-coupled proteome partition originates from the need to maintain a low cellular level of unfolded peptides at the lowest biosynthetic cost. Under stress, increasing translational efficiency leads to increased levels of unfolded peptides and native enzymes simultaneously, resulting in an inefficient regulation and significant waste of cellular resources (Fig. 3*D*). The evolutionary invention of chaperones resolves this dilemma by producing a balanced cellular process that increases the flux from a pool of unfolded peptides to native proteins while suppressing the production of both unfolded peptides and ribosomes.

Multiscale Predictions for the Cellular Adaptation Mechanisms. FoldME is a multiscale model that describes not only global regulation of proteome composition but also the statistical effects on in vivo folding at the level of metabolic pathways or upon perturbation of a single gene. At the pathway level, FoldME predicts that at high temperatures, DeoA, an enzyme involved in the pyrimidine degradation pathway, becomes unstable and extremely expensive to produce. The high cost of maintaining DeoA leads to a shift in sugar uptake from pyrimidine degradation to the glycolysis and pentose phosphate pathway (Fig. S4*A*). This prediction is confirmed by our experiment showing that *E. coli* cells grow on glucose, but not on thymidine at high temperatures (Fig. S4*B*). Additional support comes from a long-term evolution experiment of *E. coli* subjected to high temperature in LB medium, where the steady-state levels of enzymes involved in pyrimidine degradation including DeoA, DeoB, and DeoC are significantly down-regulated upon adaptation (36).

To evaluate the effect of perturbation of a single gene, we used FoldME to compute the consequences of point mutations in the core metabolic enzyme dihydrofolate reductase (DHFR), which were shown to affect the cellular abundance of a large number of *E. coli* proteins (40). We reproduced the sharp decrease in growth rate within small variations of DHFR stability (Fig. 4*A*) and correctly predicted that DHFR mutants used GroEL/ES for folding (41). Upon destabilization of the DHFR protein, FoldME predicted the differential expression of a large number of proteins. Consistent with discoveries reported in Bershtein et al. (40), the overall SD of protein expression level increases as DHFR stability and organism fitness decrease (Fig. 4*B*).

For quantitative comparisons between experiments and FoldME predictions, we calculated z scores for genes expressed in both experiment and FoldME: $z = \frac{Y_i - \langle Y \rangle}{\sigma^Y}$, where Y_i is the logarithm of relative protein abundance (LRPA) with respect to the WT level for gene i. Average variations in expression for individual proteins correlated quantitatively between transcriptomic data and FoldME predictions for the majority of the clusters of orthologous group (COG) categories (Fig. 4*C*). Importantly, the biosynthetic resource is distributed to the three proteome partitions with the same growth-coupling relationships as shown in Fig. 3 (Fig. S5). Similar down-regulation of coenzyme biosynthetic pathways was observed for temperature elevation and destabilizing mutations (Fig. S6), indicating a consistent energy allocation strategy in chaperone-mediated adaptation to environmental and genetic perturbations.

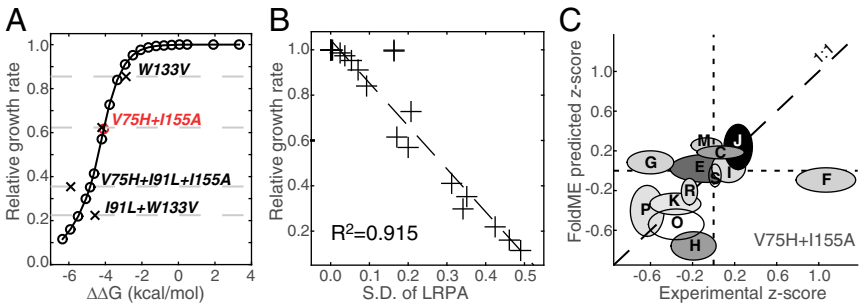

Fig. 4. Systems-level responses to DHFR mutations. (*A*) FoldME predicts sharp decrease in growth rate (circles) as the stability of DHFR decreases. Experimental values for four DHFR mutants (crosses) are taken from the literature (40, 52). (*B*) SD of the LRPA is anticorrelated with growth rate. (*C*) The average *z* scores for transcriptomic data (40) and model prediction are highly correlated (Pearson's $r = 0.84$) for the shown COG categories (53). The width and height of each oval are proportional to the SD of *z* scores within the group. Gray scales with the number of proteins included in each category (darker color represents larger number).

Discussion

To obtain a comprehensive understanding of bacterial thermoadaptation that bridges the knowledge from the molecular to the systems level, we have reconstructed the *E. coli* protein-folding network. We used this reconstruction to formulate a computational model, FoldME, through integration with the genome-scale model of metabolism and protein expression. Even with uncertainties in model parameters for single proteins (*SI Materials and Methods* and Figs. S7–S9), FoldME simulations are capable of reproducing robust cell growth in conditions that vary in nutrient, temperature, and gene content. The results illustrate the complexity of protein folding in vivo, such that thermostability of a single protein has only very limited influence on cellular fitness (Fig. S9*E*). Instead, interwoven interactions between protein thermodynamics and chaperone regulation lead to multilevel strategies, ranging from gene to pathway to network that a cell uses to deal with unfolding stresses. Moreover, we highlight the systems-level regulatory role of the chaperone network that has been overlooked in previous studies. During bacterial thermoadaptation, the molecular chaperone takes a "service" function that mediates proteome allocation and cellular fitness in two interconnected ways. First, the chaperone pool is shared by the whole proteome; thus occupancy of a chaperone by one unfolded protein sets a constraint on the structural integrity of all other proteins. Second, the increased expression of chaperones under stress drains available resources from protein synthesis, setting a stringent translational constraint on the entire proteome.

Three critical factors contribute to FoldME's ability to achieve a deep multiscale understanding of bacterial thermoadaptation. First, we incorporate a metabolically inactive unfolded state of protein to facilitate assessment of the proteome's biophysical profile. This profile serves as an internal "sensor" to reflect the environmental and genetic perturbations. Second, we design a mathematical formulation for the chaperones to respond to the folding request of the proteome independent of its metabolic state. Third, instead of imposing the chaperone-assisted folding reactions only on the few experimentally validated substrates, we enable competition among the spontaneous, DnaK-assisted, and GroEL-mediated folding pathways, for all modeled proteins. As such, changes in the proteostatic state of the cell induced by environmental and genetic perturbations can be calculated based on first principles, evaluated by the protein quality-control machinery, and coupled to the whole cell's economics. The approach adopted in this study opens up fundamental unique possibilities for genome-scale integration of environment-dependent protein properties, which enables proteome-wide study of cellular stress responses to environmental and genetic perturbations.

Materials and Methods

Kinetic Folding-Rate Calculation. The kinetic folding rate k_f is calculated using the Gromiha method, which is reported with a correlation of 0.97 between predicted and experimentally measured folding rates for a sample of 32 proteins (42). To calculate k_f for each modeled protein, we first compute its secondary structures using the DSSP tool (43) in ProDy (44) and then submit the protein sequence along with the assigned secondary structure class to the FOLD-RATE web server (https://www.iitm.ac.in/bioinfo/fold-rate/). Finally, the predicted values are set as the reference folding rate at 37 °C and scaled according to the relationship $\ln k_f \propto 1/T$ to cover the temperature range between 24 °C and 46 °C (45).

Thermostability Calculation. Gibbs free energy of unfolding ($\Delta G = G_{unfolded} - G_{folded}$) is predicted using a combination of the Dill expression (21) and the Oobatake method (46). The Dill expression is formulated based on the empirical correlation between protein length and thermostability. It generates a homogeneously stable proteome with melting temperatures varying in a small range, between 53.9 °C and 58.7 °C. The Oobatake method generates a more diverse thermostability profile, using information from protein sequence and structure. To maintain both heterogeneity and a low level (5%) of *E. coli* proteome being intrinsically disordered (47), we assign the unfolding free energy in two steps: (*i*) Calculate $\Delta G(T)$ for $T \in [24, 46]$ °C using the Oobatake method; and (*ii*) if $\Delta G(T) < 0$ for all calculated temperatures, recalculate $\Delta G(T)$ using the Dill expression; otherwise, assign $\Delta G(T)$ using the values calculated in step *i*. Folding constraints and chaperone requirements in FoldME are expressed in equilibrium constant ($K_{eq} = [U]_{eq}/[N]_{eq}$) calculated using $\Delta G(T) = -RT\ln(K_{eq}(T))$. More details for the calculation and sensitivity analysis of $k_f(T)$ and $\Delta G(T)$ are discussed in *SI Materials and Methods* and Figs. S8 and S9.

***agg* Calculation.** *agg* is defined as the number of "aggregation-prone" segments on an unfolded protein sequence. These aggregation-prone regions have been extensively studied and shown to be highly correlated with chaperone selectivity (48). We choose a consensus method (49), which incorporates 11 popular algorithms that use different aspects of the sequence property to predict aggregation propensity of the *E. coli* proteome. Sequences of the modeled proteins are submitted to the web server (aias.biol.uoa.gr/AMYLPRED2/) for evaluation. To obtain the best balance between sensitivity and specificity, we follow the author's guidelines to consider every five consecutive residues agreed among at least five methods contributing 1 to the *agg*.

Bacteria Strains and Growth Media. *E. coli* strains, culture conditions, and characterizations of the media used in this study are described in *SI Materials and Methods*.

ACKNOWLEDGMENTS. We thank Daniel Zielinski, Nathan E. Lewis, Adam Feist, Zak King, and Jonathan Monk (all at University of California, San Diego) for helpful discussions. This work was funded by National Institutes of Health grants (Awards GM102098 and GM057089) and the Novo Nordisk Foundation (Award NNF10CC1016517). This research used resources of the National Energy Research Scientific Computing Center, supported by the US Department of Energy under Contract DE-AC02-05CH11231.

1. Vieille C, Zeikus GJ (2001) Hyperthermophilic enzymes: Sources, uses, and molecular mechanisms for thermostability. *Microbiol Mol Biol Rev* 65:1–43.
2. Nguyen V, et al. (2017) Evolutionary drivers of thermoadaptation in enzyme catalysis. *Science* 355:289–294.
3. Bennett AF, Lenski RE (2007) An experimental test of evolutionary trade-offs during temperature adaptation. *Proc Natl Acad Sci USA* 104:8649–8654.
4. Tenaillon O, et al. (2012) The molecular diversity of adaptive convergence. *Science* 335:457–461.
5. Sandberg TE, et al. (2014) Evolution of Escherichia coli to 42 °C and subsequent genetic engineering reveals adaptive mechanisms and novel mutations. *Mol Biol Evol* 31:2647–2662.
6. Bukau B, Weissman J, Horwich A (2006) Molecular chaperones and protein quality control. *Cell* 125:443–451.
7. Hartl FU, Bracher A, Hayer-Hartl M (2011) Molecular chaperones in protein folding and proteostasis. *Nature* 475:324–332.
8. Saibil H (2013) Chaperone machines for protein folding, unfolding and disaggregation. *Nat Rev Mol Cell Biol* 14:630–642.
9. Genevaux P, Georgopoulos C, Kelley WL (2007) The Hsp70 chaperone machines of Escherichia coli: A paradigm for the repartition of chaperone functions. *Mol Microbiol* 66:840–857.
10. Horwich AL, Farr GW, Fenton WA (2006) GroEL-GroES-mediated protein folding. *Chem Rev* 106:1917–1930.
11. Hu B, Mayer MP, Tomita M (2006) Modeling Hsp70-mediated protein folding. *Biophys J* 91:496–507.
12. Tehver R, Thirumalai D (2008) Kinetic model for the coupling between allosteric transitions in GroEL and substrate protein folding and aggregation. *J Mol Biol* 377: 1279–1295.
13. Hingorani KS, Gierasch LM (2014) Comparing protein folding in vitro and in vivo: Foldability meets the fitness challenge. *Curr Opin Struct Biol* 24:81–90.
14. Powers ET, Powers DL, Gierasch LM (2012) FoldEco: A model for proteostasis in E. coli. *Cell Rep* 1:265–276.
15. Cho Y, et al. (2015) Individual and collective contributions of chaperoning and degradation to protein homeostasis in E. coli. *Cell Rep* 11:321–333.
16. Deuerling E, et al. (2003) Trigger factor and DnaK possess overlapping substrate pools and binding specificities. *Mol Microbiol* 47:1317–1328.
17. Kerner MJ, et al. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. *Cell* 122:209–220.
18. Calloni G, et al. (2012) Dnak functions as a central hub in the E. coli chaperone network. *Cell Rep* 1:251–264.
19. Corkrey R, et al. (2014) Protein thermodynamics can be predicted directly from biological growth rates. *PLoS One* 9:e96100.
20. Chen P, Shakhnovich EI (2010) Thermal adaptation of viruses and bacteria. *Biophys J* 98:1109–1118.
21. Dill KA, Ghosh K, Schmit JD (2011) Physical limits of cells and proteomes. *Proc Natl Acad Sci USA* 108:17876–17882.
22. Lerman JA, et al. (2012) In silico method for modelling metabolism and gene product expression at genome scale. *Nat Commun* 3:929.
23. Thiele I, et al. (2012) Multiscale modeling of metabolism and macromolecular synthesis in E. coli and its application to the evolution of codon usage. *PLoS One* 7:e45635.
24. O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BØ (2013) Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol Syst Biol* 9:693.
25. Scott M, Gunderson CW, Mateescu EM, Zhang Z, Hwa T (2010) Interdependence of cell growth and gene expression: Origins and consequences. *Science* 330:1099–1102.
26. Chang RL, et al. (2013) Structural systems biology evaluation of metabolic thermotolerance in Escherichia coli. *Science* 340:1220–1223.
27. Brunk E, et al. (2016) Systems biology of the structural proteome. *BMC Syst Biol* 10:26.
28. Mayer M, Bukau B (2005) Hsp70 chaperones: Cellular functions and molecular mechanism. *Cell Mol Life Sci* 62:670–684.
29. Pierpaoli EV, et al. (1997) The power stroke of the DnaK/DnaJ/GrpE molecular chaperone system. *J Mol Biol* 269:757–768.
30. Jewett A, Shea JE (2008) Do chaperonins boost protein yields by accelerating folding or preventing aggregation? *Biophys J* 94:2987–2993.
31. Schröder H, Langer T, Hartl F, Bukau B (1993) DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage. *EMBO J* 12:4137–4144.
32. Weissman JS, Kashi Y, Fenton WA, Horwich AL (1994) GroEL-mediated protein folding proceeds by multiple rounds of binding and release of nonnative forms. *Cell* 78: 693–702.
33. Santra M, Farrell DW, Dill KA (2017) Bacterial proteostasis balances energy and chaperone utilization efficiently. *Proc Natl Acad Sci USA* 114:E2654–E2661.
34. Herendeen SL, Vanbogelen RA, Neidhardt FC (1979) Levels of major proteins of Escherichia coli during growth at different temperatures. *J Bacteriol* 139:185–194.
35. Seyer K, Lessard M, Piette G, Lacroix M, Saucier L (2003) Escherichia coli heat shock protein DnaK: Production and consequences in terms of monitoring cooking. *Appl Environ Microbiol* 69:3231–3237.
36. Rudolph B, Gebendorfer KM, Buchner J, Winter J (2010) Evolution of Escherichia coli for growth at high temperatures. *J Biol Chem* 285:19029–19034.
37. Yang L, et al. (2015) Systems biology definition of the core proteome of metabolism and expression is consistent with high-throughput data. *Proc Natl Acad Sci USA* 112:10810–10815.
38. Scott M, Klumpp S, Mateescu EM, Hwa T (2014) Emergence of robust growth laws from optimal regulation of ribosome synthesis. *Mol Syst Biol* 10:747.
39. Maitra A, Dill KA (2015) Bacterial growth laws reflect the evolutionary importance of energy efficiency. *Proc Natl Acad Sci USA* 112:406–411.
40. Bershtein S, Choi JM, Bhattacharyya S, Budnik B, Shakhnovich E (2015) Systems-level response to point mutations in a core metabolic enzyme modulates genotype-phenotype relationship. *Cell Rep* 11:645–656.
41. Bershtein S, Mu W, Serohijos AW, Zhou J, Shakhnovich EI (2013) Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness. *Mol Cell* 49:133–144.
42. Gromiha MM (2005) A statistical model for predicting protein folding rates from amino acid sequence with structural class information. *J Chem Inf Model* 45:494–501.
43. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637.
44. Bakan A, Meireles LM, Bahar I (2011) Prody: Protein dynamics inferred from theory and experiments. *Bioinformatics* 27:1575–1577.
45. Scalley ML, Baker D (1997) Protein folding kinetics exhibit an Arrhenius temperature dependence when corrected for the temperature dependence of protein stability. *Proc Natl Acad Sci USA* 94:10636–10640.
46. Oobatake M, Ooi T (1993) Hydration and heat stability effects on protein unfolding. *Prog Biophys Mol Biol* 59:237–284.
47. Oldfield CJ, et al. (2005) Comparing and combining predictors of mostly disordered proteins. *Biochemistry* 44:1989–2000.
48. Rousseau F, Serrano L, Schymkowitz JW (2006) How evolutionary pressure against protein aggregation shaped chaperone specificity. *J Mol Biol* 355:1037–1047.
49. Tsolis AC, Papandreou NC, Iconomidou VA, Hamodrakas SJ (2013) A consensus method for the prediction of 'aggregation-prone' peptides in globular proteins. *PLoS One* 8:e54175.
50. Blaby IK, et al. (2012) Experimental evolution of a facultative thermophile from a mesophilic ancestor. *Appl Environ Microbiol* 78:144–155.
51. Cooper VS, Bennett AF, Lenski RE (2001) Evolution of thermal dependence of growth rate of Escherichia coli populations during 20,000 generations in a constant environment. *Evolution* 55:889–896.
52. Bershtein S, Mu W, Shakhnovich EI (2012) Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations. *Proc Natl Acad Sci USA* 109:4857–4862.
53. Galperin MY, Makarova KS, Wolf YI, Koonin EV (2014) Expanded microbial genome coverage and improved protein family annotation in the cog database. *Nucleic Acids Res* 43:D261–D269.
54. Bar-Even A, et al. (2011) The moderately efficient enzyme: Evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* 50:4402–4410.
55. Davidi D, et al. (2016) Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro k_{cat} measurements. *Proc Natl Acad Sci USA* 113: 3401–3406.
56. Ebrahim A, et al. (2016) Multi-omic data integration enables discovery of hidden biological regularities. *Nat Commun* 7:13091.
57. Leuenberger P, et al. (2017) Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science* 355:eaai7825.
58. Ouyang Z, Liang J (2008) Predicting protein folding rates from geometric contact and amino acid sequence. *Protein Sci* 17:1256–1263.
59. LaCroix RA, et al. (2014) Discovery of key mutations enabling rapid growth of Escherichia coli K-12 MG1655 on glucose minimal media using adaptive laboratory evolution. *Appl Environ Microbiol* 83:02246-14.
60. Bremer H, Dennis P (1996) *Modulation of Chemical Composition and Other Parameters of the Cell at Different Exponential Growth Rates in Escherichia Coli and Salmonella*, ed Neidhardt FC (ASM, Washington, DC).
61. Forchhammer J, Lindahl L (1971) Growth rate of polypeptide chains as a function of the cell growth rate in a mutant of Escherichia coli 15. *J Mol Biol* 55:563–568.