Soluble oligomerization provides a beneficial fitness effect on destabilizing mutations.

Shimon Bershtein and Eugene I. Shakhnovich

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford St, Cambridge, MA 02138

Abstract

Protein stability is widely recognized as a major evolutionary constraint. However, the relation between mutation-induced perturbations of protein stability and biological fitness has remained elusive. Here we explore this relation by introducing a selected set of mostly destabilizing mutations into an essential chromosomal gene of E.*coli* encoding dihydrofolate reductase (DHFR) to determine how changes in protein stability, activity and abundance affect fitness. Several mutant strains showed no growth while many exhibited fitness *higher* than wild type. Overexpression of chaperonins (GroEL/ES) buffered the effect of mutations by rescuing the lethal phenotypes and *worsening* better-fit strains. Changes in stability affect fitness by mediating the abundance of active and soluble proteins; DHFR of lethal strains aggregates, while destabilized DHFR of high fitness strains remains monomeric and soluble at 30^{o}C and forms soluble oligomers at 42^{o}C. These results suggest an evolutionary path where mutational destabilization is counterbalanced by specific oligomerization protecting proteins from aggregation.

In order to evolve, an organism must acquire genetic mutation(s), yet these very mutations can cause structural destabilization of the proteins they encode[1,2,3,4], potentially affecting the ability of an organism to survive and reproduce (fitness). It is critical to the organism's survival, therefore, that mutated, potentially unstable, proteins that perform important physiological roles are folded in an active form[3]. This dichotomy of how cells can accommodate evolutionarily beneficial, but structurally destabilizing mutations is central to Biology, yet poorly understood. It sculpts the fitness effects of mutations, which determine evolutionary trajectories of populations. However, prediction or even rationalization of fitness effects of mutations remains a challenge because the bridge between the molecular effect of mutations on proteins and the resulting effect on fitness of the whole organism is not well understood.

Many past efforts have attempted to bridge this gap ''in one shot'' – from sequences to phenotype[5]. However, the relationship between genomic sequences and phenotypic traits is likely to be indirect with many sequence variations leading to the same phenotypic effect, making it difficult to pinpoint phenotypic changes to specific mutations. A promising approach, introduced in recent multi-scale theoretical models [6,7,8,9] is to bridge the scales gap mid-way by coarse graining molecular traits. These models postulate certain relationships between fitness and ''macroscopic'' properties of proteins, such as their stability, abundance in soluble and aggregated forms and interactions with other proteins.

Of all the ''macroscopic'' molecular traits, protein stability has been widely recognized as one of the most evolutionarily important[3,7,10,11,12]. Indeed, to be functional, almost all proteins must be either stably folded or, in the case of many intrinsically-disordered proteins, assume a specific structure upon binding to a partner[13]. However, the relationship between stability and fitness, while often postulated[10,11 12], has not been sufficiently explored experimentally.

Here, we aim to partially close this gap by exploring the relationship between the stability of an essential enzyme, dihydrofolate reductase (DHFR) encoded by the *folA* gene in *Escherichia coli*, and bacterial fitness. DHFR is present in relatively low abundance in *E.coli* (~40 copies/cell[14]), so toxicity from its aggregation should be negligible. We therefore might expect that the main effect on fitness would be through the modulation of active enzyme copy number.

We have explored the fitness effects of mutations through an extensive set of carefully designed mutational perturbations. Unlike earlier approaches, which considered how overexpression of non-endogenous destabilized proteins from a plasmid affects fitness[15,16], we aim to establish the link between the molecular properties of an *endogenous* essential protein and organism fitness at maximally realistic conditions through site-specific mutations directly on the chromosome and without perturbation of the regulatory region controlling the endogenous expression level (Fig. 1). Our main objective is to determine how protein stability affects organism fitness and which particular molecular properties of DHFR, besides stability, have the most pronounced impact on the fitness of *E. coli.*

Selection of mutants and their in vitro analysis.

The rationale behind choosing DHFR mutations was to cover a broad range of stabilities with minimum effect on activity. To that end, an extensive search of the literature identified 8 loci, all far from the active site (most buried in the protein's hydrophobic core; Table I). Multiple sequence alignments at the selected loci were examined and substitutions with both low and high conservation propensity were identified (Supplementary Figure 1).

The stabilities of mutants were measured against thermal denaturation using Differential Scanning Calorimetry (DSC) and urea denaturation monitored by circular dichroism. This provided three closely related measures of stability for each mutant; (1) 'unfolding transition temperature', T_m; (2) 'urea mid-transition concentration', C_m; and, (3) 'free energy of denaturation at 25°C', ΔG - based on a two-state model fit (see Table 1 and Online Methods). The three quantities are linearly correlated in the studied range of temperatures, as expected from protein thermodynamics[17,18,19] (Supplementary Figure 2). We chose to use T_m to characterize the stability of mutants, as it is directly measured experimentally. In addition, we measured catalytic activity for each single mutant (k_{cat} and K_M) using a standard assay (see Table I, and Online Methods). There was no statistically significant correlation between protein activity and stability (Supplementary Figure 3).

Fitness of mutant strains

We measured laboratory fitness of all 27 mutant strains in a range of temperatures by competing them with a wtDHFR strain. (Fig. 1 and Online Methods). As a complementary measure of fitness we determined individual growth rates for all mutants at the same temperature range (Online Methods and Supplementary Figure 4). In what follows, we report fitness measurements obtained in competition experiments, unless indicated otherwise (growth rate data are qualitatively consistent with competition data). Below we focus on 3 temperatures; 30°C as the lower limit of Arrenius-like dependence of growth rate on temperature (Supplementary Figure 5), 42°C as its highest limit, and 37°C as '' physiological'' temperature for *E. coli* in humans.

Protein abundance determines fitness at 30°C

No significant correlation was found between catalytic activity and fitness (Fig. 2A and Supplementary Figure 6), suggesting a catalytic saturation regime for DHFR (as found by Hartl for another enzyme[20]). Next, we measured intracellular abundances of DHFR for all mutants in cell lysates using Western Blot (see Methods Online) and found a significant correlation between abundance of DHFR and organism fitness (Fig. 2B). It appears, therefore, that fitness depends more strongly on product abundance than catalytic activity (Fig. 2C). Statistical mechanics of two-state protein folding [21,22] would predict the relationship between protein stability and intracellular abundance of folded proteins to follow the Boltzmann law:

$$C_f = \frac{C_{tot} e^{-\frac{\Delta G}{k_B T}}}{\left(1 + e^{-\frac{\Delta G}{k_B T}}\right)} \qquad (1)$$

Where T is temperature, k_B is Boltzmann constant, C_{tot} and C_f are total and folded abundances. The Boltzmann relationship in Eq (1) predicts a very weak dependence between stability and folded abundance for even moderately stable (few kT) proteins, and a plateau for the observed range of free energies of mutants at $\Delta G \leq -2kcal/m$ (see Table 1), provided that total abundance is fixed. In contrast, we observed a stronger, linear, dependence between abundance and stability (Fig.2D), suggesting that cytoplasm is an active medium, where stability critically affects total abundance through homeostatic balance between protein production and degradation.

Destabilization of a protein apparently shifts this balance towards degradation - as has been observed *in vitro*[23,24].

Anti-correlation between fitness and T_m at 42^{o}C

At 30 and 37°C there is no statistically significant correlation between T_m and fitness, while at 42°C, there appears to be a clear *anti-correlation*. Furthermore, figure 3A-C shows that the Distribution of Fitness Effects (DFE) is bimodal at all temperatures, and especially at 42°C, where only a few mutants totally lose the competition with wt (fitness near 0, essentially meaning the organism cannot grow) but many more are considerably *more* fit than wt. These findings at 42°C imply an *inversion* in the correlation between fitness and stability - *i.e.* mutant strains encoding *less* stable DHFR appear to be *more* fit at the higher temperature!(Fig. 3D-F).

Soluble oligomerization at 42^{o}C

The results of our fitness experiments presented a paradoxical situation where the main determinants of fitness reverse themselves for strains growing at higher temperature. While fitness dependence on protein abundance and activity at 30°C is intuitive and consistent with theoretical views based on flux balance theory[25,26], it is totally unexpected (and counterintuitive) at 42°C where strains carrying less stable mutants appear to be more fit! A critical hint comes from hyperthermophile *Thermotoga Maritima* whose DHFR exists as a stable homodimer[27]. This observation suggests that, at higher temperatures, destabilized mutants can form soluble oligomers preventing their further aggregation and preserving activity. Such behavior at elevated temperature close to unfolding transition has been predicted for another protein, SH3 domain [28]. To test this hypothesis we carried out *in vitro* cross-linking experiments at room temperature (25°C) and at 42°C on purified DHFR from wt and 19 mutants (see Fig. 4, Online Methods, and Supplementary Fig. 5). Strikingly, we found that mutants that exhibit higher fitness at 42°C (e.g. I155A) have a strong tendency to form oligomers at 42°C but not at room temperature (Figure 5A). Furthermore, oligomer-forming mutants did not exhibit high molecular weight bands typical of aggregated proteins, while mutants of low fitness strains showed a pronounced aggregation and a much weaker homodimer band (see Fig. 4 and Supplementary Figure 7).
Quantitatively, we found a significant correlation between the propensity to oligomerize (assessed by density of DHFR oligomerized species bands in the gels) with both fitness at 42°C (Figure 5B) and unfolding transition temperature T_m (Fig. 4C). A cautionary note is that cross-linking as any other method to determine protein association *in vitro* is not free from artifacts, as cross-links might induce an unnatural association. However, three key observations suggest that the effect observed here is real: *i*) it is only at 42°C where we observe anti-correlation between fitness and stability, and not at RT; *ii*) 42°C fitness significantly correlates with the density of oligomerization band (Fig. 4B); *iii*) cross-linking of lethal mutation strains shows high molecular weight bands (non-specific cross-linking leading to aggregation), while the most fit mutants show no such bands – instead we see pronounced bands corresponding to dimers and, to a lesser extent, trimer and tetramer bands.

GroEL/ES controls oligomerization and aggregation in vivo.

Our results so far point out to two types of protein association in cytoplasm: ‘’bad’’ aggregation, resulting in high molecular weight associates and low fitness, and ‘’good’’ aggregation – soluble oligomerization that maintains proteins in stable and soluble yet homodimeric form and helps to maintain high fitness. We reasoned that overexpression of

GroEL/ES can serve as an *in vivo* probe for these two types of protein aggregation and an *in vivo* test of this view.

While the mechanism of GroEL action is still under debate [29,30], recent work from the Horwich lab has provided strong evidence that the main role of GroEL/ES is to prevent aggregation [31]. Therefore we expect that overexpression of GroEL/ES should interfere with both ''good'' and ''bad'' aggregation, and, as a result, shrink the DFE by rescuing lethal mutants from ''bad'' aggregation and worsening better fit ones whose enhanced fitness came from soluble oligomerization.

To test this hypothesis we transformed all mutant strains with the pGro7 plasmid, conditionally overexpressing GroEL/ES, and measured their laboratory fitness. The marker we used to distinguish between the competing strains appeared no longer neutral with pGro7 induction, therefore here we report fitness in terms of growth rates measured for individually grown strains (competition fitness assays give qualitatively similar results). The effect on lethal mutations is striking; while GroEL/ES overexpression does not affect growth of the wt strain, it dramatically increases growth of low fitness strains at 42^{o}C (see Fig. 5A-C). On the other hand, overexpression of GroEL/ES had, as predicted, an *inverse* effect on the fittest mutants by actually reducing their fitness. (See Fig. 4D and growth curves data for all 27 mutant strains in Supplementary Figure 4). While worsening growth rates of the fittest mutants is not as dramatic as rescue of lethal ones, it remains a consistent and statistically significant observation.

Evolutionary implications

The metaphor of a 'rugged fitness landscape' is often invoked to reflect the notion that fitness effects cannot be predicted from sequence variation. This study shows that fitness effects of mutations, while hardly rationalizable at the level of sequence variation can be, to a significant extent, predicted from changes in coarse-grained properties of proteins, such as stability. A common expectation is that destabilization would be detrimental to fitness, first, because active proteins are lost to unfolding (see Eq.1) and, second, because destabilized proteins tend to aggregate causing both irreversible loss of function and possible toxicity [8,32]. While such expectations are qualitatively consistent with behavior observed at 30^{o}C (with significant caveats as explained above), the opposite is observed at higher temperature because an unexpected physical factor interferes. Less stable DHFR proteins tend to escape aggregation by forming soluble oligomers, and this phenomenon is responsible for a surprising anti-correlation between stability and fitness at 42^{o}C.

Soluble dimerization has been predicted elsewhere, in *simulations* of protein aggregation for an SH3 domain protein [28]. These simulations [28] found that domain swapping [33] could lead to two types of protein homodimers,– soluble closed-form dimers and aggregation-prone open-form dimers. Importantly, at low temperature SH3 domains formed monomers, while soluble closed form dimers were observed in simulations only at elevated temperatures close to folding transition temperature for the SH3 domain[28]– similar to the trend observed *experimentally* here.
It is likely therefore, that a similar domain swapping mechanism is responsible for the formation of soluble dimers in DHFR at 42^{o}C, although this remains to be established in structural studies. It is also likely that two types of oligomers found in [28]- open, prone to further aggregation and closed, soluble homodimers - are responsible for bimodal DFE (Fig. 3) which is especially apparent at 42^{o}C where soluble oligomerization is most pronounced.

This study reveals the complexity of the concept of fitness, which is central to population genetics. Mutations that provide higher fitness under one set of conditions can be detrimental

under another. In particular, a destabilizing mutation that has an immediate fitness advantage can bring a protein closer to its stability threshold whereby a subsequent destabilizing mutation will likely result in a loss of structural integrity and lethal phenotype (see Supplementary Figure 8). The new phenomenon reported here can be described as ''stability by association''. Apparently, it provides a route to escape the detrimental effects of destabilizing mutations by opening sequence space for a broader exploration at higher temperatures. It is tempting to suggest that such a mechanism can serve as a universal route to ''sequence-based'' [34] thermal adaptation for many proteins of originally mesophilic organisms which colonize warmer environments.

A complementary mechanism to promote evolvability *in vivo* is through GroEL/ES overexpression [35]. As this study clearly shows, chaperonins rescue highly deleterious aggregated DHFR mutants, thus providing new evolutionary paths to survival under conditions that otherwise would have lead to extinction (Supplementary Figure 8). But this study also reveals two important caveats in the evolutionary role of GroEL/ES. First, chaperonins are efficient only at the stoichiometric excess over the aggregated protein species. This immediately brings the notion that they are mostly efficient for lowly expressed proteins, while highly expressed ones might simply not have enough GroEL/ES to the rescue. This factor might contribute to the differences observed in the evolutionary rates of highly and lowly expressed proteins [36]. Second, chaperonins are likely to interfere with the oligomerization induced by destabilizing mutations, thus counteracting the evolutionary potential of the 'stability by association". This finding adds another level of complexity to our understanding of chaperonins-proteome co-evolution, and is especially interesting since both mechanisms - ''stability by association'' and overexpression of GroEL/ES- are likely to enhance evolvability of proteins in any circumstances at which protein destabilization might occur, such as high mutational load [35,37] or drop in population size [12,38].

Methods Summary

DHFR mutants fused to His-tag were generated by site-directed mutagenesis, cloned in a pET24 vector under a T7 inducible promoter, and purified on Ni-NTA columns. Thermodynamic properties (ΔG, T_m) of the purified recombinant DHFR mutants were analyzed by urea induced unfolding and followed by changes in circular dichroism signals and Differential Scanning Calorimentry. Catalytic parameters (k_{cat}, K_M) were derived from full-progress DHFR kinetic curves, monitored by coenzyme fluorescence using Global Kinetic Explorer. Chromosomal incorporation of the *in vitro* analyzed mutations into E.*coli*'s MG1655 *folA* gene was performed using a modified gene knock-out protocol based on lambda Red induced homologous recombination. Laboratory fitness was determined by competition assays. Strains carrying mutations in the *folA* gene were competed with a wtDHFR strain using *lacZ* gene knock-out as a neutral marker. The outcome of the competition performed over 18 hours in liquid culture was monitored by plating onto LB agar plates supplemented with X-gal and IPTG**,** and counting the ratio between blue and white colonies. Each competition was performed with the neutral marker swap. Chaperonin overexpression experiments were performed with a pGro7 plasmid (Takara), which conditionally overexpresses GroEL/ES. Growth rates of the DHFR mutant strains were derived from the individual growth measurements (OD at 600 nm) and performed in 96-well microtiter plates(see Supplementary Table 1 and Figure 4 for summary of all growth rates) Intracellular protein abundance was determined for the soluble fraction of cell lyzates by Western blot.

1 DePristo, M. A., Weinreich, D. M. & Hartl, D. L. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat Rev Genet* **6**, 678-687 (2005).

2 Bloom, J. D., Labthavikul, S. T., Otey, C. R. & Arnold, F. H. Protein stability promotes evolvability. *Proc Natl Acad Sci U S A* **103**, 5869-5874 (2006).

3 Zeldovich, K. B., Chen, P. & Shakhnovich, E. I. Protein stability imposes limits on organism complexity and speed of molecular evolution. *Proc Natl Acad Sci U S A* **104**, 16152-16157 (2007).

4 Soskine, M. & Tawfik, D. S. Mutational effects and the evolution of new protein functions. *Nature Reviews Genetics* **11**, 572-582, doi:Doi 10.1038/Nrg2808 (2010).

5 Weinreich, D. M., Delaney, N. F., DePristo, M. A. & Hartl, D. L. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111-114, doi:Doi 10.1126/Science.1123539 (2006).

6 Zeldovich, K. B., Chen, P., Shakhnovich, B. E. & Shakhnovich, E. I. A First-Principles Model of Early Evolution: Emergence of Gene Families, Species, and Preferred Protein Folds. *PLoS Comput Biol* **3**, e139 (2007).

7 Taverna, D. M. & Goldstein, R. A. Why are proteins marginally stable? *Proteins* **46**, 105-109 (2002).

8 Drummond, D. A. & Wilke, C. O. Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**, 341-352 (2008).

9 Heo, M., Maslov, S. & Shakhnovich, E. Topology of protein interaction network shapes protein abundances and strengths of their functional and nonspecific interactions. *Proc Natl Acad Sci U S A* **108**, 4258-4263 (2011).

10 Bloom, J. D., Raval, A. & Wilke, C. O. Thermodynamics of neutral protein evolution. *Genetics* **175**, 255-266 (2007).

11 Pena, M. I., Davlieva, M., Bennett, M. R., Olson, J. S. & Shamoo, Y. Evolutionary fates within a microbial population highlight an essential role for protein folding during natural selection. *Molecular systems biology* **6**, 387 (2010).

12 Wylie, C. S. & Shakhnovich, E. I. A biophysical protein folding model accounts for most mutational fitness effects in viruses. *Proc Natl Acad Sci U S A* **108**, 9916-9921 (2011).

13 Dyson, H. J. & Wright, P. E. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6**, 197-208 (2005).

14 Taniguchi, Y. *et al.* Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533-538 (2010).

15 Bershtein, S., Segal, M., Bekerman, R., Tokuriki, N. & Tawfik, D. S. Robustness-epistasis link shapes the fitness landscape of a randomly drifting protein. *Nature* **444**, 929-932 (2006).

16 Geiler-Samerotte, K. A. *et al.* Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 680-685 (2011).

17 Privalov, P. Stability of proteins: small globular proteins. *Advances in Protein Chemistry* **33**, 167 (1979).

18 Robertson, A. D. & Murphy, K. P. Protein Structure and the Energetics of Protein Stability. *Chemical reviews* **97**, 1251-1268 (1997).

19 Chen, P. & Shakhnovich, E. I. Thermal adaptation of viruses and bacteria. *Biophysical journal* **98**, 1109-1118 (2010).

[20] Dykhuizen, D. E., Dean, A. M. & Hartl, D. L. Metabolic flux and fitness. *Genetics* **115**, 25-31 (1987).
[21] Privalov, P. L. & Khechinashvili, N. N. A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. *J Mol Biol* **86**, 665-684 (1974).
[22] Shakhnovich, E. Protein folding thermodynamics and dynamics: where physics, chemistry, and biology meet. *Chem Rev* **106**, 1559-1588 (2006).
[23] Parsell, D. A. & Sauer, R. T. The structural stability of a protein is an important determinant of its proteolytic susceptibility in Escherichia coli. *The Journal of biological chemistry* **264**, 7590-7595 (1989).
[24] Park, C. & Marqusee, S. Pulse proteolysis: a simple method for quantitative determination of protein stability and ligand binding. *Nature methods* **2**, 207-212 (2005).
[25] Heo, M., Kang, L. & Shakhnovich, E. I. Emergence of species in evolutionary "simulated annealing". *Proc Natl Acad Sci U S A* **106**, 1869-1874 (2009).
[26] Wang, Z. & Zhang, J. Impact of gene expression noise on organismal fitness and the efficacy of natural selection. *Proc Natl Acad Sci U S A* **108**, E67-76 (2011).
[27] Dams, T. *et al.* The crystal structure of dihydrofolate reductase from Thermotoga maritima: molecular features of thermostability. *J Mol Biol* **297**, 659-672 (2000).
[28] Ding, F., Dokholyan, N. V., Buldyrev, S. V., Stanley, H. E. & Shakhnovich, E. I. Molecular dynamics simulation of the SH3 domain aggregation suggests a generic amyloidogenesis mechanism. *J Mol Biol* **324**, 851-857 (2002).
[29] Thirumalai, D. & Lorimer, G. H. Chaperonin-mediated protein folding. *Annu Rev Biophys Biomol Struct* **30**, 245-269 (2001).
[30] Horwich, A. L. & Fenton, W. A. Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. *Q Rev Biophys* **42**, 83-116 (2009).
[31] Tyagi, N. K., Fenton, W. A., Deniz, A. A. & Horwich, A. L. Double mutant MBP refolds at same rate in free solution as inside the GroEL/GroES chaperonin chamber when aggregation in free solution is prevented. *FEBS Lett* **585**, 1969-1972 (2011).
[32] Dobson, C. M. Protein aggregation and its consequences for human disease. *Protein Pept Lett* **13**, 219-227 (2006).
[33] Schlunegger, M. P., Bennett, M. J. & Eisenberg, D. Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Adv Protein Chem* **50**, 61-122 (1997).
[34] Berezovsky, I. N. & Shakhnovich, E. I. Physics and evolution of thermophilic adaptation. *Proc Natl Acad Sci U S A* **102**, 12742-12747 (2005).
[35] Tokuriki, N. & Tawfik, D. S. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459**, 668-673 (2009).
[36] Drummond, D. A., Bloom, J. D., Adami, C., Wilke, C. O. & Arnold, F. H. Why highly expressed proteins evolve slowly. *Proc Natl Acad Sci U S A* **102**, 14338-14343 (2005).
[37] Fares, M. A., Ruiz-Gonzalez, M. X., Moya, A., Elena, S. F. & Barrio, E. Endosymbiotic bacteria: groEL buffers against deleterious mutations. *Nature* **417**, 398 (2002).
[38] Fernandez, A. & Lynch, M. Non-adaptive origins of interactome complexity. *Nature* **474**, 502-505 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements. We are grateful to Roy Kishony for help at the early stages of this project, to Art Horwich for discussions and to Dan Tawfik, Bill Eaton, Maxim Frank-Kamenetskii, and Stephen M. Gould for comments on the manuscript. We thank Adrian Serohijos, Yakov Pechersky, Abhishek Chintapalli, Mu Wanmeng and Phil Snyder for invaluable technical assistance. This work was supported by NIH Grant No GM 068670 and long-term postdoctoral fellowship from the Human Frontier Science Program (SB).

Authors Contributions: EIS and SB designed research, SB performed research, SB and EIS analyzed data, EIS and SB wrote the paper.

Author Information:
Reprints and permissions information is available at www.nature.com/reprints
Authors declare no competing financial interests
Correspondence and requests for materials should be addressed to EIS shakhn@fas.harvard.edu

Table 1 | Thermodynamic, structural, and catalytic properties of the *in vitro* purified DHFR mutants

Mutant	T_m[a] (°C)	$\Delta G_{(H2O)}$[a] (kcal/mole) 25°C	k_{cat}[a] (s-1)	k_{cat}/K_m[a] (s-1μM-1)	ASA[b]	Conservation[c] (%)	
						native	substitution
wt	50.3	-4.5	11.65	3.6			
V40A	42.6	-3.32	24.05	3.22	0	43.6	0.34
I61V	53.2	-4.72	8.99	3.7	0	19.9	79
V75H	39.9	-1.92	15.5	4.57	0.05	32	0
V75I	38.5	-2.43	18.25	5.51	0.05	32	10.3
I91V	48.2	-3.82	17.87	5.13	0.07	33.3	79
I91L	40.4	-2.71	ND[d]	ND[d]	0.07	33.3	20
L112V	46.7	-2.73	11.45	2.34	0	34	27.5
W133F	43.7	-5.12	13.45	6.79	0.05	53.6	20.6
W133V	ND[d]	-1.75	ND[d]	ND[d]	0.05	53.6	0.34
I155T	42.6	-3.68	11.27	2.72	0.12	15.8	21
I155L	45.2	-2.92	12.9	5.12	0.12	15.8	1
I155A	36.8	-2.43	11.95	2.71	0.12	15.8	1
I115V	50.3	-3.43	10.26	2.08	0.02	55.3	34.4
I115A	45	-3.41	7.59	0.5	0.02	55.3	0
V88I	43.7	-4.22	13.8	3.12	0.34	12.4	1.4
A145T	50.3	-4.14	10	3.6	0.89	12.4	1.4

[a]For detailed description of stability and activity measurements see Online Methods

[b]ASA (accessible surface area) was calculated by Vadar package (http://vadar.wishartlab.com) ; cutoff for buried residues is around 0.25

[c]Conservation (%) of a native E. coli's DHFR residue (left column) or a substituted residue (right column) in a given position of 290 aligned mesophylic prokaryotic DHFR sequences retrieved from the Optimal Growth Temperature database (http://pgtdb.csie.ncu.edu.tw)

[d]Not determined

Figure Legends

Figure 1|The experimental approach. (1) 9 DHFR residues, predominantly from the hydrophobic core of the enzyme and distant from the active site, were chosen for mutagenesis based on structural and phylogenetic predictions and published biophysical and biochemical data (see Methods). (2) 16 single mutants (Table 1) were generated, cloned into pET vector, expressed and purified. (3) Gibbs free energy difference between folded and unfolded state (ΔG), mid-transition temperature of unfolding (T_m), and catalytic parameters (k_{cat}, K_m) were measured (Table 1). (4a) A site-directed chromosomal mutagenesis method was developed to introduce *in vitro* characterized mutations into the chromosomal *folA* gene of E. *coli*'s MG1655 strain without perturbing the gene's regulatory region. In addition to 16 single mutant strains, 11 multiple mutant strains were generated by combining the 4 most destabilizing mutations (V75H+I91L, V75H+W133F, I91L+W133F, V75H+I91L+W133F, V75H+I155A, I91L+I155A, V75H+I91L+I155A, V75H+W133F+I155A, I91L+W133F+I155A, W133F+I155A, V75H+I91L+W133F+I155A). (4b) All DHFR mutant strains were transformed with pGro7 plasmid that conditionally over-expresses chaperoninsGroEL/ES. (5) Fitness effects of the introduced mutations were measured by *i*) growth competition of the mutant strains with a wtDHFR strain; *ii*) monitoring growth curves of the individually grown strains.

Figure 2 | Correlating fitness at 30^{o}C with molecular properties. **A**, Fitness of single DHFR mutant strains, relative to wt, measured at 30^{o}C by competition assay is plotted against catalytic parameters. **B**, Fitness, relative to wt, *vs* intracellular soluble fraction of the DHFR proteins determined in lysates of mutant strains by Western blot. **C**, Fitness of single DHFR mutant strains, relative to wt, *vs* a product of catalytic parameters and soluble intracellular abundance. **D**, Intracellular protein abundance, relative to wt, of single DHFR mutant strains *vs* T_m.S.d. for fitness values is given for 2 independent measurements obtained from blue-white swaps (see Methods). ANOVA testing was used to calculate R, and *P*-values.

Figure 3 | Bi-modal distribution of fitness effects. A, B, C, Histograms of distributions of fitness effects at 30^{o}C, 37^{o}C, and 42^{o}C. Fitness, relative to wt, was determined for all mutant strains (16 single and 11 multiple mutants) by competition assays with wt strain. **D, E, F**, Correlation of competition fitness values for single mutant strains at 30^{o}C, 37^{o}C, and 42^{o}C with their *in vitro* measured T_m. S.d. for fitness values is given for 2 independent measurements obtained from blue-white swaps (see Methods). ANOVA testing was used to calculate R, and *P-value*.

Figure 4 | *In vitro* oligomerization assay. **A**, A representative cross-linking experiment. 11 μM of purified wt DHFR, W133V, or I155A in 25 mM potassium-phosphate buffer (pH 7.8) and 100 μM NADPH were incubated for 45 min at room temperature (RT) or 42^{o}C with or without 2 mM of cross-linking (c-l) agent (glutaraldehyde). Shown is the SDS-PAGE analysis of 10 μl of protein samples after Coomassie staining. Red arrows indicate molecular weight equivalents of monomeric and oligomeric protein species. Smears at high molecular weight seen for W133V at 37^{o}C and 42^{o}C in the presence of c-l are due to extensive aggregation. **B**, The *in vitro* propensity to oligomerize correlated with fitness values measured by competition at 42^{o}C. Oligomerization propensity was determined by measuring the relative density of the SDS-PAGE bands corresponding to DHFR oligomerized species for each of the purified mutants (Supplementary Figure 1). **C**, Correlation of oligomerization propensity (as measured in **B**) to mutant's T_m. S.d. for fitness values is given for 2 independent measurements obtained from blue-white swaps (see Methods). ANOVA testing was used to calculate R, and *P-value*.

Figure 5 | Fitness effects of GroEL/ES overexpression. A, Growth rate values of all DHFR mutant strains grown independently at 42^{o}C are arranged from smallest to largest (black). Effect of GroEL/ES overexpression on growth rate is shown in red. **B**, full growth curves of the wtDHFR strain grown at 42^{o}C without (black) and with (red) GroEL/ES overexpression. **C**, **D**, as above for mutant strains V75H+I155A, and I91L+W133F.S.d.are given for 16 independent measurements for each strain.

Methods

Protein expression and purification

folA genes carrying the mutations and a C-terminal His-tag were cloned in pET24 expression vector under inducible T7 promoter, transformed into BL21(DE3) cells and expressed by a standard protocol. The recombinant proteins were purified on Ni-NTA columns (Qiagen).

Enzyme kinetics

DHFR kinetic parameters were measured by progress-curve kinetics, essentially as described[39]. Purified enzymes (10 nM) were preincubated with 120 μM NADPH in MTEN buffer (50 mM 2-(N-morpholino)ethanesulfonic acid, 25 mM tris (hydroxymethyl) aminomethane, 25mM ethanolamine, and 100 mM sodium chloride, pH7). The reaction was initiated by addition of dihydropholate (final concentrations of 20, 15, 10 μM) and monitored to completion by coenzyme fluorescence (ex. 290 nm, em. 450 nm). The kinetic parameters (k_{cat} and K_M) were derived from progress-curves analysis using Global Kinetic explorer[40].

Stability measurements

Thermodynamic parameters (T_m and ΔG) were measured for 25 μM protein in 10 mM potassium phosphate buffer pH8.2, 1mM beta-meracpoethanol, and 0.2 mM EDTA by differential scanning calorimetry (DSC) and urea unfolding monitored by circular dichroism (CD). Briefly, DSC measurements were performed under 1^{o}C/min temperature increase regime. T_m was derived from a two-state model fit [41] provided by the instrument software (TA instruments). Urea unfolding experiments were carried out at 25^{o}C by following a CD signal at 221 nm (J-7.1 CD spectrometer, Jasco)[42]. Data were fit to a two-state model using Kaleidagraph (Synergy Software).

Site-directed chromosomal mutagenesis

The method is a modification of a chromosomal gene knock-out protocol[43]. Briefly, *folA* gene carrying the desired mutation(s) with an entire endogenous regulatory region (191 bp separating stop codon of the upstream *kefC* gene and start codon of *folA* gene) was placed in a pKD13 plasmid flanked by two different antibiotic markers (genes encoding for kanamycin (kanR) and chloramphenicol (cmR) resistances). The entire cassette was then amplified with two primers tailed with 50 nucleotides homologous to the region of a chromosome intended for recombination (*kefC* gene upstream and *apaH* gene downstream to *folA)*. The amplified product was transformed into E.*coli* BW25113 with induced Red helper plasmid, and the recombinants were selected on plates carrying both antibiotics. Strains carrying the desired mutation in the chromosome were verified by sequencing. Identified chromosomal mutations were then moved to E.*coli* MG1655 by P1 transduction and double antibiotic selection (kan and cm) and again verified by sequencing.

Competition assay

Laboratory fitness was determined by competing each of the strains expressing mutant DHFR protein with wt DHFR strain in M9 minimal media supplemented with 0.2% glucose, 1mM $MgSO_4$, 0.1% casamino acids, and 0.5 μg/ml thiamine. To this end, E. *coli* strain carrying wt *folA* gene flanked by cmR and kanR genes was mixed with one of the MG1655 DHFR mutant strains in a 1:1 ratio (≈ 10^4 cells each) in 50 ml of

medium. Prior to mixing, cells were grown separately overnight from a single colony, diluted 1/100 and re-grown to early exponential phase (OD_{600}≈0.2) at 30°C. The competition was performed for 18 hours at 30°C, 37°C, and 42°C. To distinguish between the competing strains, a knock-out mutation was introduced in the *lacZ* gene. Two identical competition experiments were performed with *lacZ* knockout (neutral marker under the condition of the competition experiment) always present in one of the competing strains. The ratio before and after competition was determined by plating the culture on LB agar plates supplemented with X-gal and IPTG (*lacZ*$^-$ strain generates white colonies, whereas *lacZ*$^+$ strain generated blue colonies, hence "blue-white" swap). Around 3,000-5,000 colonies were counted for each competition experiment.

GroEl/ES overexpression

Cells were transformed with pGro7 plasmid (Takara) carrying GroEL and GroES genes under arabinose inducible promoter. pGro7 was modified by introducing an ampicillin resistance gene. Transformed cells were grown overnight at 30°C from a single colony in M9 (supplemented with 100 μg/ml ampicillin), diluted 1/100 and chaperonin overexpression was induced by arabinose (1 mg/ml).

Growth curves analysis

Cells were grown overnight at 30°C from a single colony in supplemented M9, diluted 1/100 and transferred into 96-well microtiter plates (16 wells per each strain). OD data were collected at 600nm at 35 min intervals. The resulting curves were fitted to a bacterial growth model[44].

Intracellular protein abundance

Cells were grown in supplemented M9 medium for 6 hours at 30°C, chilled on iced for 30 min and lysed with BugBuster (Novagen). DHFR amounts in the soluble cell lysate fractions were determined by SDS-PAGE followed by Western Blot using Rabbit-anti E.*coli*'s DHFR polyclonal antibodies (custom raised by Pacific Immunology).

[39] Fierke, C. A., Johnson, K. A. & Benkovic, S. J. Construction and evaluation of the kinetic scheme associated with dihydrofolate reductase from Escherichia coli. *Biochemistry* **26**, 4085-4092 (1987).

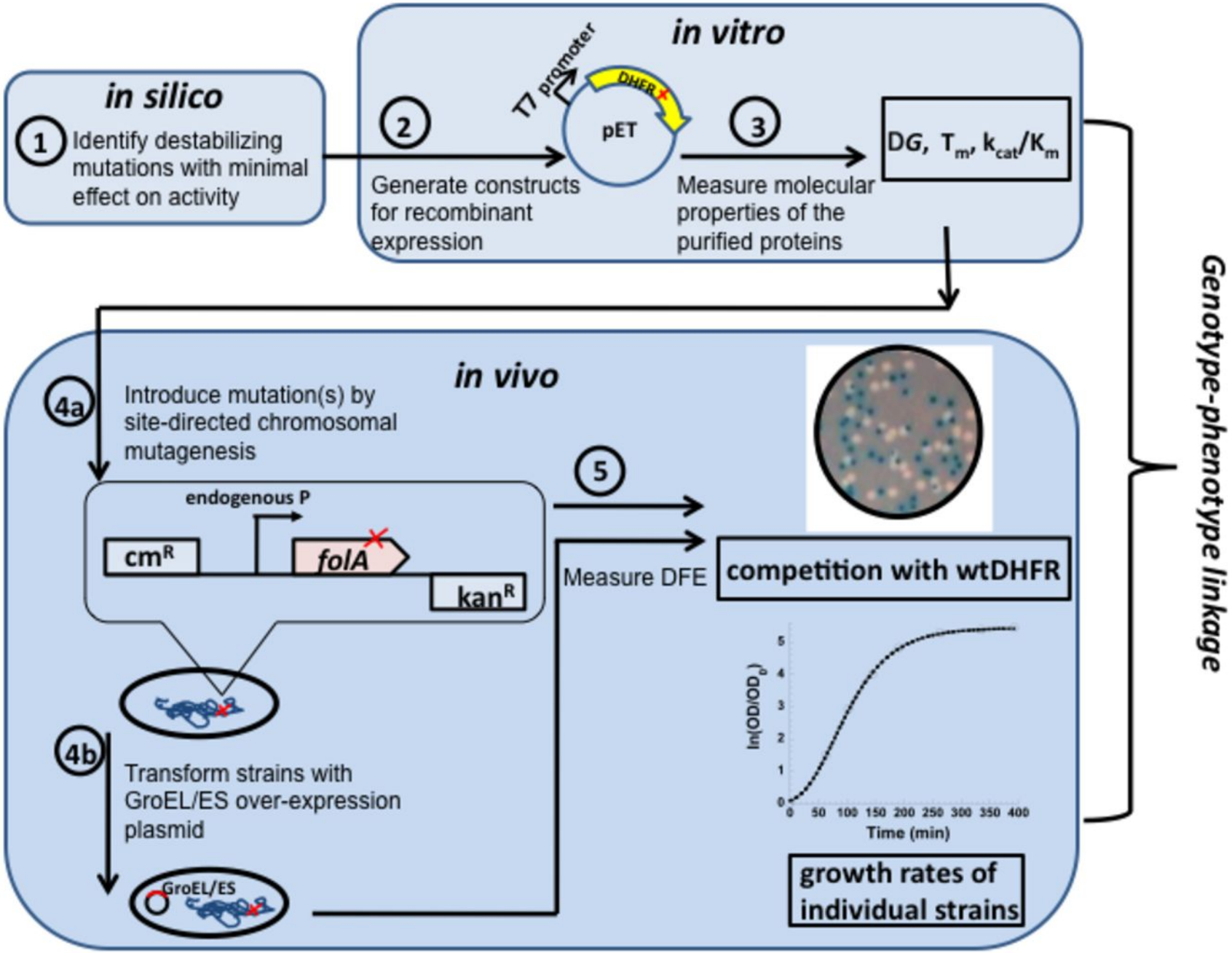
[40] Johnson, K. A., Simpson, Z. B. & Blom, T. Global kinetic explorer: a new computer program for dynamic simulation and fitting of kinetic data. *Anal Biochem* **387**, 20-22 (2009).

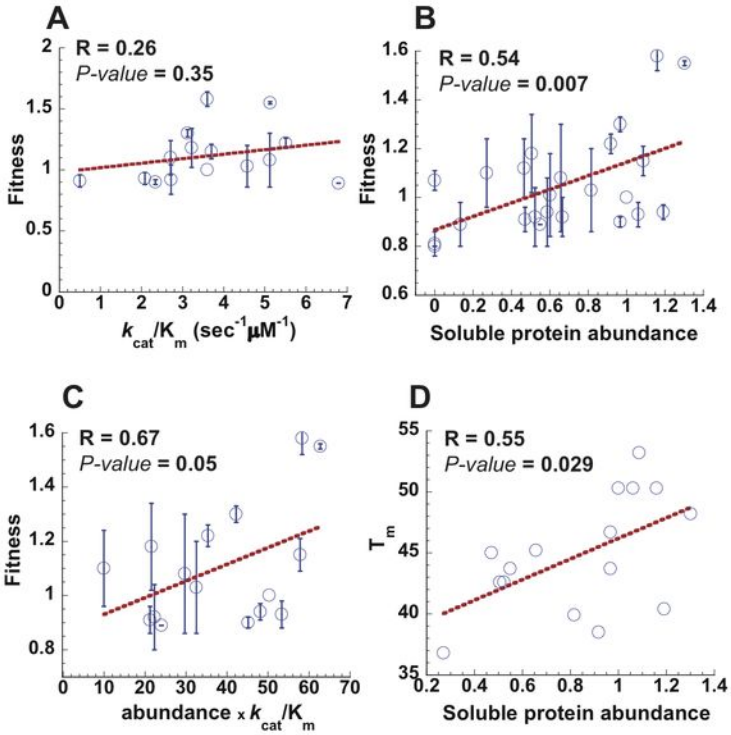
[41] Lopez, M. M. & Makhatadze, G. I. Differential scanning calorimetry. *Methods Mol Biol* **173**, 113-119 (2002).

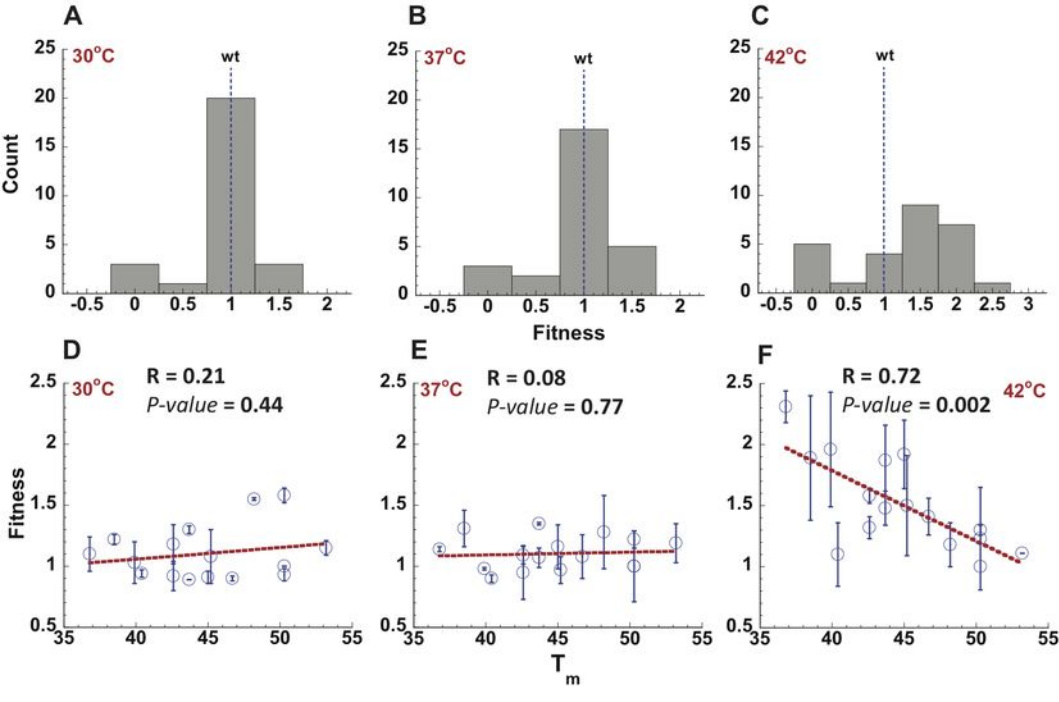
[42] Pace, C. N. Determination and analysis of urea and guanidine hydrochloride denaturation curves. *Methods Enzymol* **131**, 266-280 (1986).

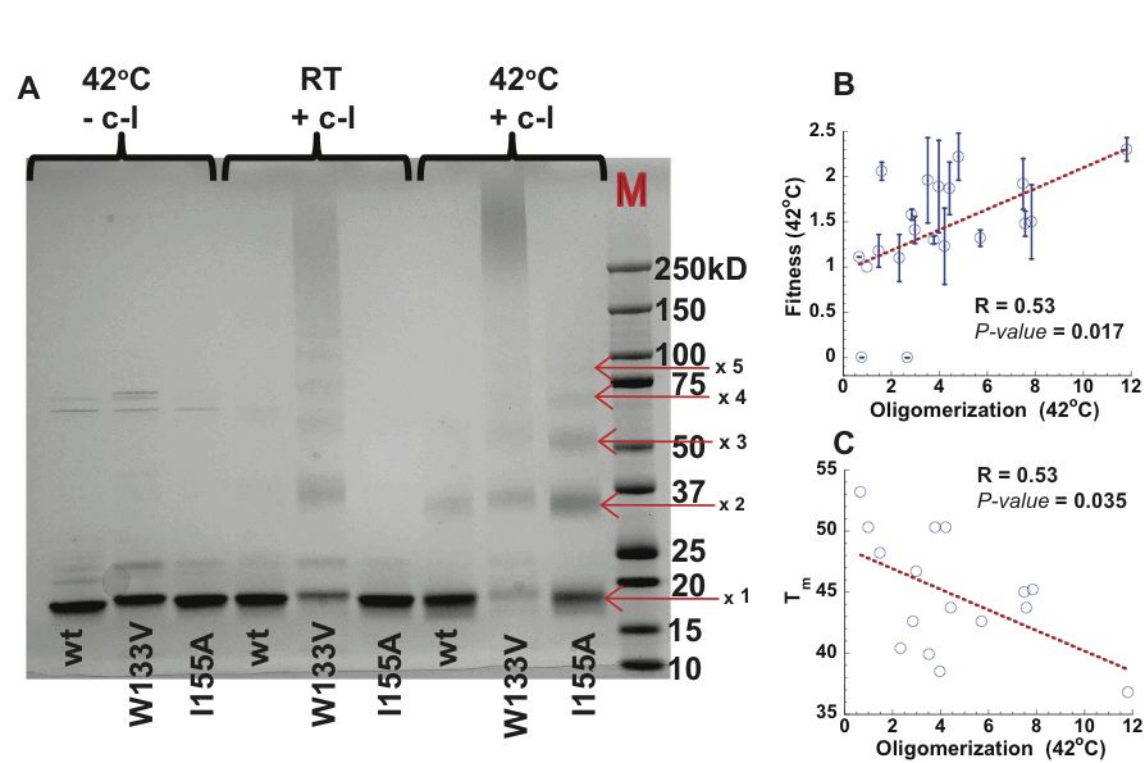
[43] Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97**, 6640-6645 (2000).

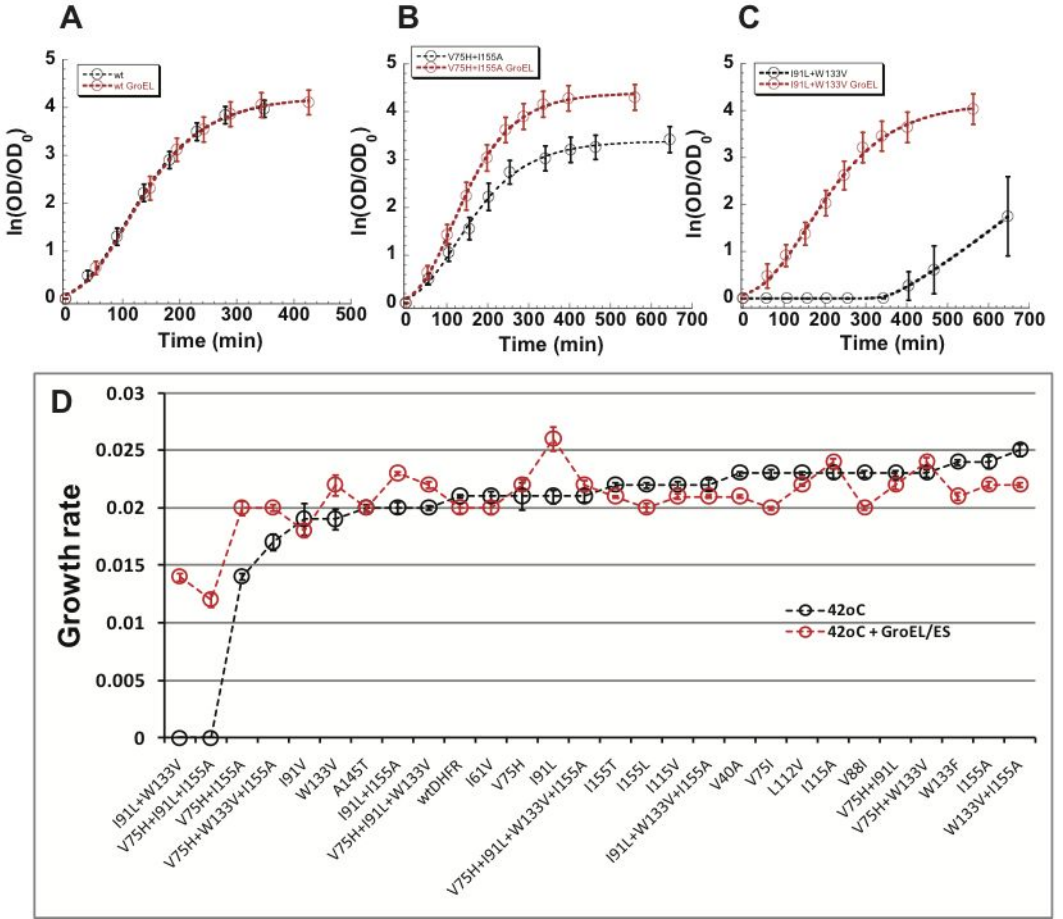
[44] Zwietering, M. H., Jongenburger, I., Rombouts, F. M. & van 't Riet, K. Modeling of the bacterial growth curve. *Appl Environ Microbiol* **56**, 1875-1881 (1990).


in silico
1
Identify destabilizing mutations with minimal effect on activity
in vitro
T7 promoter
DHFR
pET
2
Generate constructs for recombinant expression
3
Measure molecular properties of the purified proteins
DG, T_m, k_{cat}/K_m
in vivo
4a
Introduce mutation(s) by site-directed chromosomal mutagenesis
endogenous P
cm^R
folA
kan^R
4b
Transform strains with GroEL/ES over-expression plasmid
GroEL/ES
5
Measure DFE
competition with wtDHFR
ln(OD/OD0)
Time (min)
growth rates of individual strains
Genotype-phenotype linkage


A
R = 0.26
P-value = 0.35
Fitness
kcat/Km (sec-1µM-1)
B
R = 0.54
P-value = 0.007
Fitness
Soluble protein abundance
C
R = 0.67
P-value = 0.05
Fitness
abundance x kcat/Km
D
R = 0.55
P-value = 0.029
Tm
Soluble protein abundance


A
30°C
wt
B
37°C
wt
C
42°C
wt
Count
Fitness
D
30°C
R = 0.21
P-value = 0.44
E
37°C
R = 0.08
P-value = 0.77
F
42°C
R = 0.72
P-value = 0.002
Fitness
T m


A
42°C
- c-I
RT
+ c-I
42°C
+ c-I
M
250kD
150
100
75
50
37
25
20
15
10
x 5
x 4
x 3
x 2
x 1
wt
W133V
I155A
wt
W133V
I155A
wt
W133V
I155A
B
Fitness (42°C)
R = 0.53
P-value = 0.017
Oligomerization (42°C)
C
R = 0.53
P-value = 0.035
Tm
Oligomerization (42°C)


A
wt
wt GroEL
ln(OD/OD0)
Time (min)
B
V75H+I155A
V75H+I155A GroEL
ln(OD/OD0)
Time (min)
C
I91L+W133V
I91L+W133V GroEL
ln(OD/OD0)
Time (min)
D
Growth rate
42oC
42oC + GroEL/ES
I91L+W133V
V75H+I91L+I155A
V75H+I155A
V75H+W133V+I155A
I91V
W133V
A145T
I91L+I155A
V75H+I91L+W133V
wtDHFR
I61V
V75H
I91L
V75H+I91L+W133V+I155A
I155T
I155L
I115V
I91L+W133V+I155A
V40A
V75I
L112V
I115A
V88I
V75H+I91L
V75H+W133V
W133F
I155A
W133V+I155A