Local energetic frustration affects the dependence of green fluorescent protein folding on the chaperonin GroEL

Received for publication, July 25, 2017, and in revised form, October 21, 2017. Published, Papers in Press, October 24, 2017, DOI: 10.1074/jbc.M117.808576

Boudhayan Bandyopadhyay*, Adi Goldenzweig*, Tamar Unger*, Orit Adato†, Sarel J. Fleishman‡,1 Ron Unger‡,‡, and Amnon Horovitz*‡,3

From the Departments of Structural Biology and Biomolecular Sciences and the Israel Structural Proteomics Centre, Weizmann Institute of Science, Rehovot 7610001, Israel and the Mina & Everard Goodman Faculty of Life Sciences, Bar-Ilan University, Ramat-Gan 52900, Israel

Edited by Norma Allewell

The GroE chaperonin system in Escherichia coli comprises GroEL and GroES and facilitates ATP-dependent protein folding in vivo and in vitro. Proteins with very similar sequences and structures can differ in their dependence on GroEL for efficient folding. One potential but unverified source for GroEL dependence is frustration, wherein not all interactions in the native state are optimized energetically, thereby potentiating slow folding and misfolding. Here, we chose enhanced green fluorescent protein as a model system and subjected it to random mutagenesis, followed by screening for variants whose in vivo folding displays increased or decreased GroEL dependence. We confirmed the altered GroEL dependence of these variants with in vitro folding assays. Strikingly, mutations at positions predicted to be highly frustrated were found to correlate with decreased GroEL dependence. Conversely, mutations at positions with low frustration were found to correlate with increased GroEL dependence. Further support for this finding was obtained by showing that folding of an enhanced green fluorescent protein variant designed computationally to have reduced frustration is indeed less GroEL-dependent. Our results indicate that changes in local frustration also affect partitioning in vivo between spontaneous and chaperonin-mediated folding. Hence, the design of minimally frustrated sequences can reduce chaperonin dependence and improve protein expression levels.

This work was supported by the Minerva Foundation with funding from the Federal German Ministry for Education and Research (to A. H.) and Grant 772/13 from the Israel Science Foundation (to R. U.). Research in the Fleishman laboratory was supported by a European Research Council’s Starter’s Grant and by a donation from Sam Switzer and family. The authors declare that they have no conflicts of interest with the contents of this article. This article contains supplemental Tables S1 and S2 and Figs. S1 and S2.

1 Marta S. Sagon Career Development Chair.
2 To whom correspondence may be addressed. E-mail: ron@biomodel.os.biu.ac.il.
3 Carl and Dorothy Bennett Professorial Chair in Biochemistry. To whom correspondence may be addressed. E-mail: Amnon.Horovitz@weizmann.ac.il.

The Escherichia coli GroE chaperonin system, which comprises GroEL and GroES, facilitates protein folding in vivo and in vitro in an ATP-dependent manner (1–4). GroEL consists of 14 identical subunits that form two back-to-back stacked heptameric rings with a cavity at each end (5). Capping of the cavity by GroES, which is a homoheptameric single ring, forms a compartment in which protein folding can take place in isolation from bulk solution. The volume of this compartment is ~175,000 Å³ (6) and, thus, large enough to accommodate substrate proteins that are up to 60 kDa in size. Despite more than two decades of intensive research, it is still not clear how GroEL assists folding and what distinguishes GroE substrates from all other E. coli proteins (7). Early in vitro experiments suggested that GroEL is extremely promiscuous because it binds a very wide range of E. coli proteins (8) and also artificial polypeptides with random sequences (9). Theoretical considerations indicated, however, that under normal conditions, the amount of GroEL and GroES in an E. coli cell is sufficient to assist in the folding of only ~5% of its proteins (10). In agreement with this prediction, later experimental work (11, 12) showed that the GroE system is indeed essential for the folding in vivo of only a relatively small subset (~100) of E. coli proteins, which are therefore referred to as obligatory substrates.

It has remained unclear, however, which features, other than the upper size limit, distinguish obligatory GroE substrates from all other E. coli proteins. Such features may include (i) sequence motifs and structural properties of non-native states such as exposed hydrophobic residues that confer (or allow) binding to GroEL and/or (ii) properties that render a protein GroE-dependent in the first place (e.g. propensity to aggregate). A suggestion for the first type of features was based on the fact that the mobile loops of GroES compete with protein substrates for binding to the cleft between helices H and I in the apical domains of GroEL. Hence, it was proposed that GroEL substrates contain sequences similar to the GroES mobile loops, which are separated by 10–23 residues and become exposed in non-native states (13, 14). Later work showed, however, that this feature is not strongly enriched in GroE substrates (7). A suggestion related to the second type of features was that GroE substrates have lower folding propensities, i.e. fewer hydrophobic amino acids relative to charged ones, compared with GroEL-independent proteins (15–17). Although true on average, this feature was also found to be an unreliable predictor of GroE dependence (7). Another observation that is likely to be related to the second type of features is that obligate substrates are enriched by ~4-fold in proteins belonging to the TIM barrel fold relative to their frequency among all soluble E. coli proteins with an identifiable fold (11). Given that GroEL does not recognize the native state of TIM barrels (nor of other proteins), it is clear that the GroE dependence...
Frustration affects GroEL-dependent folding

of TIM barrel substrates is conferred by other properties associated with TIM barrels, such as folding rates, structures of folding intermediates, or aggregation propensities, and not by the TIM barrel fold itself.

It should be noted, however, that TIM barrels are not universally found to be GroE substrates. YagE (an aldolase involved in formation of 2-keto-3-deoxy-galactonate) and DapA (dihydrodipicolinate synthase), for example, are both TIM barrels with sequences that share 27% identity and 46% similarity, and structures that can be superimposed with a root-mean-square deviation (RMSD)\(^4\) of \(\sim 1.5\) Å, but DapA is an obligate substrate (12, 18), whereas YagE is not. Identifying universal features for distinguishing between GroE substrates and non-substrates has remained, therefore, a challenge.

Given this lack of understanding of the determinants of GroE dependence, we reasoned that insight into this problem could be gained by generating a collection of proteins that have very similar sequences (and structures) but different GroE dependences. We decided to use eGFP as a model system because (i) it folds in a GroEL-dependent manner (19), although it is not an \(E.\ coli\) protein; and (ii) fluorescence-based screens for its folding properties can be easily designed. We generated fluorescent eGFP variants with 1–3 random mutations, along with a designed variant containing 11 mutations, which have altered GroEL dependence compared with wild-type eGFP. Strikingly, we found that the changes in GroEL dependence upon mutation can be explained using the concept of folding frustration (20, 21) and a measure of it put forward by Wolynes and coworkers (22). Frustration arises when different potentially stabilizing interactions in a protein are not satisfied concurrently. We show that such conflicting interactions, which lead to a rugged folding energy landscape with kinetic traps, can shift the kinetic partitioning from folding toward chaperone binding. Our results therefore provide a direct link between frustration in protein folding and chaperonin function \textit{in vivo}.

Results and discussion

In vivo screening for eGFP variants with altered GroEL dependence of folding

Error-prone PCR was used to generate a collection of eGFP variants each with 1–3 random mutations. In addition, one variant of eGFP with 11 mutations was obtained using the recently developed PROSS algorithm for designing proteins with increased stability and solubility (23). The randomly generated and designed variants of eGFP were then screened for changes in the dependence on GroEL for folding by using the \(E.\ coli\) MGM100 strain in which the chromosomal GroEL and GroES genes are under control of the arabinose promoter (18). In these cells, arabinose induces GroE expression and glucose suppresses it. Hence, in the presence of glucose, clones expressing eGFP variants that are less GroEL-dependent were expected to have higher folding yields of eGFP and, thus, become more fluorescent than clones expressing the wild-type eGFP. Conversely, clones expressing eGFP variants that are more GroEL-dependent were expected to contain less folded eGFP and, thus, be less fluorescent than clones expressing the wild-type eGFP, when arabinose is replaced by glucose. Using this screen for GroEL dependence \textit{in vivo} indicated, for example, that the Q204H mutant is unchanged relative to wild-type eGFP (Fig. 1A), whereas the T97I/F114I double mutant is more GroEL-dependent (Fig. 1B), and the P58S mutant and the PROSS-designed variant, des11, are less GroEL-dependent

---

\(4\) The abbreviations used are: RMSD, root-mean-square deviation; eGFP, enhanced green fluorescent protein; PDB, Protein Data Bank.

---

Figure 1. Screening for eGFP mutants whose folding displays increased or decreased GroEL dependence. \(E.\ coli\) MGM100 cells expressing either mutant or wild-type eGFP were spotted on LB plates containing 100 \(\mu\)g/ml ampicillin, 50 \(\mu\)g/ml kanamycin, 0.5 mM isopropyl \(\beta\)-d-thiogalactopyranoside, and either 0.1% arabinose or 0.1% glucose, which induce or block GroE expression, respectively. The plates were scanned after incubation using a fluorescence imager as described under "Experimental procedures." Representative plates are shown in which two colonies expressing wild-type eGFP were spotted in the bottom halves, and two colonies expressing different mutants with similar (A), increased (B), and decreased (C and D) GroEL dependence were spotted in the upper halves. The mutants in A–C were obtained by random mutagenesis and screening, whereas des11 in D was obtained by design as described (23). The fluorescence intensities of clones that express mutant or wild-type eGFP and that were grown on the same plate were compared to screen for changes in GroEL dependence. Interplate differences in fluorescence intensity are due to adjustments of the scanner and have no biological significance.
Frustration affects GroEL-dependent folding

Table 1

| Class* | eGFP variants | \( K \) | \( T_m \) | Frustation index |
|--------|---------------|---------|---------|-----------------|
| Wild-type | wild-type | \( 0.107 \pm 0.004 \) | \( 0.60 \pm 0.52 \) | -0.56 |
| Similar | Q204H | \( 0.104 \pm 0.001 \) | \( 0.56 \pm 0.13 \) | -0.56 |
| Increased | N23K, D234H | NA | NA | -0.81, NA^c |
| | N144F, P211L, N212K | NA | NA | +1.48 |
| | T97I, F114I | 0.089 \pm 0.001 | 0.06 \pm 0.05 | +0.08, +0.98 |
| | L178R, R215C | 0.112 \pm 0.001 | 0.13 \pm 0.09 | +0.98, +1.33 |
| Decreased | K162N | 0.102 \pm 0.005 | 2.97 \pm 1.10 | -0.89 |
| | K166M | 0.187 \pm 0.007 | 3.93 \pm 1.71 | -0.61 |
| | Y39N | 0.215 \pm 0.018 | 4.37 \pm 3.65 | -0.09 |
| | P58S | 0.128 \pm 0.002 | 4.55 \pm 0.09 | -0.75 |
| | S30R | 0.109 \pm 0.005 | 3.16 \pm 1.30 | -0.48 |
| des11 | S30R, N105C, K126H, Y145F, A154P, K158N, N198Y, S202H, Q204W, K206E, T225E | 0.072 \pm 0.012 | 20.70 \pm 5.6 | -1.33 |

a The division into classes is based on the in vivo screen and is relative to the GroEL dependence of wild-type eGFP.

b The data for these mutants are not available (NA) because of very tight binding to GroEL and aggregation.

c The frustration values for Asp^224 and Lys^226 could not be determined because Asp^224 is absent, and Lys^226 is replaced by alanine in the crystal structure of eGFP.

(1) In the in vivo screen and obtain more quantitative data, the GroEL dependence was determined in vitro. The in vivo assay was based on the recognition that GroEL substrates differ from non-substrates in their kinetic partitioning between (i) binding to GroEL and (ii) spontaneous folding and/or aggregation. Substrates must not contain certain structural/sequence GroEL recognition motifs but also fold (or aggregate) relatively slowly to allow binding to GroEL. Given that the kinetic partitioning between binding to GroEL and spontaneous folding (and/or aggregation) depends on GroEL concentration, we denatured wild-type eGFP and the mutant variants and then measured their initial rates of refolding as a function of GroEL concentration. For most of the variants (including wild-type eGFP, the Q204H and P58S single mutants and the T97I/F114I double mutant; Fig. 2, A–C), excellent fits of the data were obtained using Equation 1 that describes partitioning between folding and reversible binding to GroEL. The folding of these mutants was completely inhibited in the presence of excess GroEL \( (K/K'_f = 0); \) Table 1, thereby indicating that they are unable to fold when they are GroEL-bound. For two variants, however, S30R (supplemental Fig. S1) and des11 (Fig. 2D), good fits of the data required using Equation 2, according to which folding can take place both spontaneously and in association with GroEL (24). These fits revealed that the rate constants for folding in association with GroEL of the S30R and des11 mutants are lower than those for their spontaneous folding by factors of 50 and 10, respectively (Table 1). Importantly, the S30R mutation was shown previously to enhance the folding of GFP (25), in agreement with the other findings here that faster folding reduces GroEL dependence. The fits of the data (Fig. 2 and supplemental Fig. S1) using Equations 1 and 2 also yielded estimates of the binding constants for the interaction of the eGFP variants with GroEL (Table 1). Strikingly, complete agreement was found between the in vivo screen results and the in vitro assays, i.e. eGFP variants showing decreased or increased GroEL dependence in the in vivo screens were found to have respectively lower or higher affinities for GroEL (Table 1). In the case of the N23K/D234H double mutant and N144F/P211L/N212K triple mutant, very tight binding to GroEL results in large errors in the estimates of the binding constants, which are therefore not given in Table 1. More generally, these results indicated that relatively conservative and even single mutations can have widely differing effects on protein substrate affinity for GroEL, with concomitant effects on the extent of GroEL dependence of folding, and that the PROS-designed variant displays the least GroEL dependence.

**Thermal stabilities of the eGFP variants**

Native-state stability is directly proportional to the folding rate and inversely proportional to the unfolding rate. According to the kinetic partitioning mechanism, decreased GroEL dependence is predicted to be associated with faster folding rates and, thus, also with increased stability of the native state if the mutational effects on the unfolding rates are relatively small. A correlation between GroEL dependence and protein stability was also expected given that des11, which displays the least GroEL dependence among our eGFP variants, was computationally designed to be more stable. To test for such a correlation, stability measurements for the eGFP variants were carried out by monitoring changes in the far ultraviolet CD signal at 218 nm as a function of increasing temperature (supplemental Fig. S2). In general, eGFP variants with decreased or increased GroEL dependence were indeed found to have higher or lower appa-
ent melting temperatures ($T_m$), respectively (Table 1). The only exception is the P58S mutant, which has decreased GroEL dependence but is less stable than wild-type eGFP. The decreased GroEL dependence of this mutant may be due to elimination of a slow folding phase associated with cis/trans-prolyl isomerization, as shown before for a TIM barrel protein (26). A larger increase in the rate constant for unfolding than that for folding may, therefore, accounts for the lower stability of this mutant. The $T_m$ of two other mutants, N23K/D234H and N144I/P211L/N212K, could not be determined because of aggregation, in agreement with their increased GroEL dependence seen in the in vivo assays and tighter binding to GroEL in the in vitro assays.

**GroEL dependence of eGFP folding correlates with frustration**

Frustration in protein folding is reflected in a rough folding energy landscape with local minima that result in a slower folding rate (21). Our results indicating that slower folding rates are associated with increased GroEL dependence therefore led us to hypothesize that mutational effects on GroEL dependence were due to changes in local frustration. We tested this hypothesis using the Frustratometer web server, which quantifies the extent of local frustration in an input structure by computing the contribution of a residue to the native state energy relative to its contribution in a set of decoys in which, for example, neighboring amino acids are replaced (22). In cases of eGFP variants with more than one mutation, the extent of frustration was determined by summing the frustration values for all the mutated positions. The relationship between local frustration values and the cumulative extent of frustration is not known. The additivity assumption may, however, be justified in this case because most of the mutated positions in each of the variants are distant from each and thus likely to contribute to frustration in an independent manner. We found that decreased GroEL dependence is linked to mutations of frustrated residues, i.e. with an average negative frustration value of $-0.64 \pm 0.33$. Conversely, increased GroEL dependence is associated with mutations of residues that are minimally frustrated, i.e. with an average positive frustration value of $1.31 \pm 0.79$ (Table 1 and supplemental Table S1). The 30 mutations that did not affect GroEL dependence were found to be at positions with a near-neutral average frustration value of $0.18 \pm 0.80$, i.e. with an average value close to zero as expected, but with a large spread that reflects exceptions. One such exception is the Q204H mutation (Table 1), which has little effect on GroEL dependence but occurs at a moderately frustrated position. Such exceptions arise, according to our interpretation, because frustration determines the potential for a change in GroEL dependence upon mutation but the effect of the mutation in practice is also a function of the mutation type. A conservative mutation at a frustrated position, for example, is less likely to lead to a change in GroEL dependence than a non-conservative one. Overall, our findings suggest that changes in GroEL dependence upon mutation could be predicted from the levels of frustration of the mutated positions.

eGFP contains 14 highly frustrated residues (i.e. residues with a frustration index $\approx -1$ (22)), of which 7 were identified.
Effects of alanine mutations at highly frustrated positions in eGFP on its GroEL-dependent folding

Table 2

| eGFP variant | Local frustration index | K  | k_f/k_B |
|--------------|-------------------------|----|---------|
| Wild-type    |                         | 0.60 ± 0.52 | 0.02 ± 0.01 |
| D36A         | −1.200                  | 9.31 ± 5.43 | 1.23 ± 2.29  |
| N105A        | −1.052                  | 2.66 ± 1.19 | 1.98 ± 0.49  |
| Y143A        | −2.239                  | 3.55 ± 1.45 | 3.13 ± 1.45  |
| K156A        | −0.975                  | 8.12 ± 1.84 |              |
| P187A        | −1.017                  | 1.98 ± 0.49 |              |

In summary, the work described here indicates that differences in GroEL dependence between proteins with very similar sequences and structures can be attributed to changes in local frustration. Significant differences in local frustration also appear to account for the differences in GroEL dependence between several other highly similar protein pairs, although the overall average frustration is not an indicator. It is likely that other factors that affect folding rates, such as chain length and contact order, may also determine GroEL dependence. More extensive experimental work on other proteins is needed to determine the extent of frustration’s predictive value. Recently, it

Frustration affects GroEL-dependent folding

The initial rates of refolding of the eGFP variants were measured as a function of GroEL concentration (supplemental Fig. S1), and the data were fitted using Equation 1 or, in the case of the N105A mutant, Equation 2. The fits yielded estimates for the values (± S.E.) of the dissociation constants, K, for the interaction of the eGFP variants with GroEL and the ratio between the rate constants of folding in association with GroEL and spontaneous folding, k_f/k_B. The frustration values were determined using the Frustratometer server (22) and the crystal structure of eGFP (PDB code 4EUL).

Increasing the stability of a protein can be achieved by increasing its folding rate, decreasing its unfolding rate, or both. It is usually difficult, however, to determine from theoretical considerations whether success in increasing protein stability by computational design is achieved because of changes in folding or unfolding rates (or both). Importantly, similar gains in protein stability caused by changes in folding versus unfolding rates can have very different consequences in vivo because of kinetic partitioning between folding and competing processes such as aggregation and chaperone binding, some of which may be irreversible. A gain in stability caused by faster folding will also favor partitioning in favor of the native state over other states due, for example, to aggregation or misfolding, thereby increasing folding yields in vivo. The PROSS algorithm (23) has already been shown to be powerful in increasing folding yields in vivo, in some cases generating variants with 3 orders of magnitude higher bacterial expression levels than wild type. The design reported here of an eGFP variant, des11, that is less GroEL-dependent provides additional evidence that the success of PROSS in optimizing protein stability is due, at least in part, to increasing folding rates. It also indicates that this is achieved by minimizing folding frustration. It is interesting to note in this connection that one of the steps in the PROSS algorithm involves scanning single amino acid replacements at each position and selecting those that increase the stability of the native state compared with wild type; this step is conceptually very similar to the method used by the “Frustratometer” to detect frustrated residues.

Concluding remarks

In summary, the work described here indicates that differences in GroEL dependence between proteins with very similar sequences and structures can be attributed to changes in local frustration. Significant differences in local frustration also appear to account for the differences in GroEL dependence between several other highly similar protein pairs, although the overall average frustration is not an indicator. It is likely that other factors that affect folding rates, such as chain length and contact order, may also determine GroEL dependence. More extensive experimental work on other proteins is needed to determine the extent of frustration’s predictive value. Recently, it
was shown that highly frustrated regions in the colicin immunity protein Im7 are involved in its interaction with the ATP-independent chaperone Spy (28). It remains to be established, however, whether local frustration is important also for the in vivo interaction of substrates with Spy and for substrate interactions with other ATP-dependent and -independent chaperone systems.

**Experimental procedures**

**Molecular biology**

All the mutagenesis and cloning were carried out using restriction-free cloning (29) and confirmed by DNA sequencing of the entire genes. The gene for eGFP with an N-terminal His\textsubscript{6} tag was inserted into the pMALc2 plasmid using the forward and reverse primers 5′-TCACCAACAAAGGACCATAGCATCCTCGAGATGATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATCACCATAC-.

![Figure 3](image.png)
The primers designed for the single point mutations in the eGFP gene are provided in supplemental Table S2.

**Error-prone PCR-mediated mutagenesis**

Error-prone PCR-mediated mutagenesis was performed using the GeneMorph II random mutagenesis kit (Agilent Technologies). Briefly, 500–1000 ng of plasmid DNA (pMALc2 containing the gene for eGFP) was mutagenized by PCR according to the manufacturer’s protocol for a low mutation frequency (0–4.5 mutations/kb) using the forward and reverse primers 5’-CAATTTCACAGGAAACAGGCACGTC-3’ and 5’-CGGCCAGGCAAATCTGTATTCACAGC-3’, respectively. PCR products were separated by gel electrophoresis and purified using a gel extraction kit (Qiagen). The purified PCR products were then used as megaprimer for whole plasmid amplification and integration (29) into the pMALc2 vector. Following the whole plasmid amplification, the product was digested with DpnI and transformed into DH5α cells. Plasmids from single colonies were purified using QIAprep spin miniprep kit (Qiagen). The entire eGFP genes of interest were sequenced after screening using the E. coli MGM100 strain (18).

**Design of the eGFP variant des11**

eGFP (PDB code 2WUR, chain A) was submitted to PROSS (23), an algorithm for designing stable protein variants available as a web server (http://pross.weizmann.ac.il). All residues were allowed to mutate, and default parameters were used for MSA generation. PROSS provided seven designs, each comprising several mutations. A design with 12 mutations (the one with the highest number of mutations) was selected for experimental testing to increase the probability to obtain a significant effect on folding. The designed and codon-optimized gene of eGFP containing the 12 mutations was ordered from Gen9 Bio Inc. Initial assays indicated that this designed variant has reduced fluorescence. Reversing one of the mutations, V68G, which is near the chromophore, led to des11, the designed variant with restored fluorescence, which is described in this work.

**Screening for eGFP mutants with increased or decreased GroEL dependence**

E. coli MGM100 cells in which the chromosomal GroEL and GroE genes are under control of the arabinose promoter (18) were transformed with pMALc2 containing the gene for mutagenized or wild-type eGFP and then spread on LB plates containing 100 μg/ml ampicillin, 50 μg/ml kanamycin, and 0.1% arabinose. Single colonies were then picked and grown for 2 h at 37 °C in 50 μl of LB medium containing 100 μg/ml ampicillin, 50 μg/ml kanamycin, and 0.1% arabinose. The cells were then centrifuged at 12,000 rpm for 2 min, and the pellets were washed with 50 μl of LB medium containing 100 μg/ml ampicillin and 50 μg/ml kanamycin. The cell pellets were then resuspended in 25 μl of LB medium containing 100 μg/ml ampicillin and 50 μg/ml kanamycin. LB plates (preincubated at 37 °C for 1 h) containing 100 μg/ml ampicillin, 50 μg/ml kanamycin, 0.5 mM isopropyl β-d-thiogalactopyranoside, and either 0.1% arabinose or 0.1% glucose were divided into two parts that were then spotted, respectively, with 4–7 samples of 3 μl of cell suspensions expressing mutant or wild-type eGFP. The plates were incubated at 37 °C for 6 h and then at room temperature for 12 h. The plates were then scanned using a fluorescence imager (Typhoon FLA 9500; GE Healthcare) with an excitation filter at 473 nm (blue LD laser) and an emission filter of LPB (510LP). The respective average fluorescence intensities, \(<F>_{\text{mut}}/\langle F \rangle_{\text{wild-type}}\) is expected to be higher in the presence of glucose than arabinose when the mutant is less GroEL-dependent than wild-type eGFP and lower when the mutant is more GroEL-dependent than wild-type eGFP.

**Protein purification**

Purification of wild-type eGFP and its mutants was carried out as described previously (30). GroEL was purified by growing E. coli TG1 cells harboring the plasmid pOA overnight at 37 °C in 2× YT medium containing 50 μg/ml ampicillin. The overnight culture was diluted 1:100 in 2× YT medium containing 50 μg/ml ampicillin, grown overnight at 37 °C, and harvested. The cells were then resuspended in lysis buffer (20 mM Tris-HCl buffer, pH 7.5, containing 60 mM KCl, 10 mM MgCl2, 2 mM DTT, 0.1 mM EDTA, and 0.5 mM PMSF) and disrupted using sonication. The cell lysate was centrifuged at 20,000 rpm for 30 min at 4 °C, and the supernatant was subjected to 55% ammonium sulfate precipitation. The pellet obtained after centrifugation at 18,000 rpm for 30 min was kept at 4 °C and then resuspended in buffer A (50 mM Tris-HCl buffer, pH 7.5, containing 0.1 mM EDTA, and 1 mM DTT) and loaded on a HiTrap Q FF column pre-equilibrated with buffer A. The elution was carried out using a step gradient of 0–1 M NaCl in buffer A. The GroEL-containing fractions were combined and subjected to ammonium sulfate precipitation as above. The pellet was then resuspended in G10K buffer (50 mM Tris-HCl buffer, pH 7.5, containing 10 mM KCl, 10 mM MgCl2, and 1 mM DTT) and loaded onto a Superdex 6 column (Amersham Biosciences Pharmacia) equilibrated with G10K buffer. The GroEL-containing fractions were combined and subjected to ammonium sulfate precipitation and centrifugation as described above. The pellet was then resuspended in MES buffer (50 mM MES, pH 6.0, containing 1 mM EDTA, 1 mM DTT, and 25% methanol) and loaded on a MonoQ column (Amersham Biosciences Pharmacia) equilibrated with MES buffer. GroEL was eluted using a 0–1 M NaCl gradient in MES buffer. The GroEL-containing fractions were then pooled and subjected to acetone precipitation. The pellet was resuspended in G10K buffer, and aliquots of the protein were snap-frozen in liquid nitrogen and stored at −80 °C. Protein concentrations were determined as described (31).

**Folding assays**

Wild-type or mutant eGFP was denatured in 200 μl of 30 mM HCl at a final concentration of 12.5 μM and incubated at 25 °C for 1 h. The denatured protein was then diluted 100-fold in 200 μl of refolding buffer (50 mM MOPS, pH 7.0, containing 100 mM...
Frustration affects GroEL-dependent folding

KCl, 5 mM DTT, and 0.0125% Tween 20), in the absence or presence of GroEL, and folding was monitored at 25 °C by exciting at 465 nm and measuring the emission at 509 nm using an ISS PC1 spectrofluorimeter with slit widths of 1 mm (8-nm bandwidth). The oligomer concentrations of GroEL were varied from 0 to 7 µM. All the experiments were carried out using siliconized test tubes. The slope of the initial linear change in fluorescence as a function of time, V, was then plotted as a function of the total concentration of GroEL, [EL]T, and the data were fitted using OriginPro 8 to the following equation,

\[
\frac{V}{V_{\text{max}}} = \frac{[U]_T - [EL]_T - K + \sqrt{[U]_T + [EL]_T + K^2 - 4[U]_T[EL]_T}}{2[U]_T} \tag{Eq. 1}
\]

where \(V_{\text{max}}\) is the slope in the absence of GroEL, [U]_T is the total concentration of unfolded eGFP, and \(K\) is the dissociation constant for the unfolded protein with GroEL. Equation 1 was derived for the scheme \(EL \cdot U \rightleftharpoons U \rightleftharpoons F\), where \(F\) stands for the folded state, assuming tight binding. Alternatively, folding can also occur, whereas the protein is associated with GroEL as described by the scheme: \(F \rightleftharpoons EL \cdot U \rightleftharpoons U \rightleftharpoons F\). In such a case, zero velocity is not reached in the presence of excess GroEL and the data were fitted to Equation 2,

\[
\frac{V}{V_{\text{max}}} = 1 + \left(\frac{k_f}{k'_f} - 1\right)\frac{[U]_T + [EL]_T + K - \sqrt{[U]_T + [EL]_T + K^2 - 4[U]_T[EL]_T}}{2[U]_T} \tag{Eq. 2}
\]

where \(k_f\) and \(k'_f\) are the rate constants of spontaneous folding and folding in association with GroEL, respectively. The values and standard errors of [U]_T, K, and \(k'_f/k_f\) were obtained by global fitting.

Stability measurements

Protein stability measurements were carried out by monitoring changes in the far ultraviolet CD signal at 218 nm as a function of increasing temperature from 50 to 94 °C (steps of 2 °C) with an equilibration time of 5 min at each temperature before taking a reading. All the CD measurements were performed using a Chirascan CD spectrometer (Applied Photophysics, Leatherhead, UK). The temperature was controlled using a TC125 temperature controller (Quantum Northwest, Liberty Lake, WA) and monitored with a temperature probe in the sample. All the measurements were made with a quartz suprasil cell of 1-mm path length (Hellma, Forest Hills, NY), and the protein concentration was 12 µM in 10 mM phosphate buffer (pH 7.0) containing 0.0125% Tween. Values of melting temperatures, \(T_m\), were obtained by fitting the data as described before (30).

Frustration analysis

The Frustratometer server (22) was used to quantify frustration in proteins at the single residue level that includes electrostatics. The crystal structures of eGFP (PDB code 4EUL), mouse DHFR (PDB code 1U70), E. coli DHFR (PDB code 1RH3), DapA (PDB code 1DHP), YagE (PDB code 2V8Z), E. coli NanA (PDB code 2WO5), and M. synoviae NanA (PDB code 4N4P) were used for these calculations.

References

1. Thirumalai, D., and Lorimer, G. H. (2001) Chaperonin-mediated protein folding. Annu. Rev. Biophys. Biomol. Struct. 30, 245–269
2. Saibil, H. R., Fenton, W. A., Clare, D. K., and Horwich, A. L. (2013) Structure and allostery of the chaperonin GroEL. J. Mol. Biol. 425, 1476–1487
3. Hayer-Hartl, M., Bracher, A., and Hartl, F. U. (2016) The GroEL-GroES chaperonin machine: a nano-cage for protein folding. Trends Biochem. Sci. 41, 62–76
4. Gruber, R., and Horovitz, A. (2016) Allosteric mechanisms in chaperonin machines. Chem. Rev. 116, 6588–6606
5. Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L., and Sigler, P. B. (1994) The crystal structure of the bacterial chaperonin at 2.8 Å. Nature 371, 578–586
6. Xu, Z., Horwich, A. L., and Sigler, P. B. (1997) The crystal structure of the asymmetric GroEL-GroES-(ADP),chaperonin complex. Nature 388, 741–750
7. Azia, A., Unger, R., and Horovitz, A. (2012) What distinguishes GroEL substrates from other Escherichia coli proteins? FEBS J. 279, 543–550
8. Vitanen, P. V., Gatenby, A. A., and Lorimer, G. H. (1992) Purified chaperonin 60 (groEL) interacts with the nonnative states of a multitude of Escherichia coli proteins. Protein Sci. 1, 363–369
9. Aoki, K., Motojima, F., Taguchi, H., Yomo, T., Yoshida, M. (2000) GroEL binds artificial proteins with random sequences. J. Biol. Chem. 275, 13755–13758
10. Lorimer, G. H. (1996) A quantitative assessment of the role of the chaperonin proteins in protein folding in vivo. FASEB J. 10, 5–9
11. Kerner, M. J., Naylor, D. J., Ishihama, Y., Maier, T., Chang, H. C., Stines, A. P., Georgopoulos, C., Frishman, D., Hayer-Hartl, M., Mann, M., and Hartl, F. U. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. Cell 122, 209–220
12. Fujiwara, K., Ishihama, Y., Nakahigashi, K., Soga, T., and Taguchi, H. (2010) A systematic survey of in vivo obligate chaperonin-dependent substrates. EMBO J. 29, 1552–1564
13. Chaudhuri, T. K., and Gupta, P. (2005) Factors governing the substrate recognition by GroEL chaperone: a sequence correlation approach. Cell Stress Chaperoners 10, 24–36
14. Stan, G., Brooks, B. R., Lorimer, G. H., and Thirumalai, D. (2005) Identifying natural substrates for chaperonins using a sequence-based approach. Protein Sci. 14, 193–201
15. Noivirt-Brik, O., Unger, R., and Horovitz, A. (2007) Low folding propensity and high translation efficiency distinguish in vivo substrates of GroEL from other Escherichia coli proteins. Bioinformatics 23, 3276–3279
16. Raineri, E., Ribeca, P., Serrano, L., and Maier, T. (2010) A more precise characterization of chaperonin substrates. Bioinformatics 26, 1685–1689
17. Tartaglia, G. G., Dobson, C. M., Hartl, F. U., and Vendruscolo, M. (2010) Physicochemical determinants of chaperone requirements. J. Mol. Biol. 400, 579–588
18. McLenan, N., and Masters, M. (1998) GroEL is vital for cell-wall synthesis. Nature 392, 139
19. Makino, Y., Amada, K., Taguchi, H., and Yoshida, M. (1997) Chaperonin-mediated folding of green fluorescent protein. J. Biol. Chem. 272, 12468–12474
20. Onuchic, J. N., and Wolynes, P. G. (2004) Theory of protein folding. Curr. Opin. Struct. Biol. 14, 70–75
21. Ferreiro, D. U., Komives, E. A., and Wolynes, P. G. (2014) Frustration in biomolecules. Q. Rev. Biophys. 47, 285–363
22. Parra, R. G., Schafer, N. P., Radusky, L. G., Tsai, M. Y., Guzovsky, A. B., Wolynes, P. G., and Ferreiro, D. U. (2016) Protein frustratometer 2: a tool to localize energetic frustration in protein molecules. Nucleic Acids Res. 44, W356–W360

Author contributions—B. B. performed and analyzed all the experiments; A. G. designed a mutant; T. U. provided technical assistance; O. A. analyzed data; and R. U., S. J. F., and A. H. conceived the study, analyzed the data, and wrote the paper.
23. Goldenzweig, A., Goldsmith, M., Hill, S. E., Gertman, O., Laurino, P., Ashani, Y., Dym, O., Unger, T., Albeck, S., Prilusky, J., Lieberman, R. L., Aharoni, A., Silman, I., Sussman, J. L., Tawfik, D. S., and Fleishman, S. J. (2016) Automated structure- and sequence-based design of proteins for high bacterial expression and stability. *Mol. Cell* 63, 337–346
24. Gray, T. E., and Fersht, A. R. (1993) Refolding of barnase in the presence of GroE. *J. Mol. Biol.* 232, 1197–1207
25. Pédelaq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., and Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* 24, 79–88
26. Wu, Y., and Matthews, C. R. (2003) Proline replacements and the simplification of the complex, parallel channel folding mechanism for the alpha subunit of Trp synthase, a TIM barrel protein. *J. Mol. Biol.* 330, 1131–1144
27. Clark, A. C., Hugo, E., and Frieden, C. (1996) Determination of regions in the dihydrofolate reductase structure that interact with the molecular chaperonin GroEL. *Biochemistry* 35, 5893–5901
28. He, L., Sharpe, T., Mazur, A., and Hiller, S. (2016) A molecular mechanism of chaperone-client recognition. *Sci. Adv.* 2, e1601625.
29. Unger, T., Jacobovitch, Y., Dantes, A., Bernheim, R., and Peleg, Y. (2010) Applications of the restriction free (RF) cloning procedure for molecular manipulations and protein expression. *J. Struct. Biol.* 172, 34–44.
30. Sokolovski, M., Bhattacherjee, A., Kessler, N., Levy, Y., and Horovitz, A. (2015) Thermodynamic protein destabilization by GFP tagging: a case of interdomain allostery. *Biophys. J.* 109, 1157–1162
31. Gill, S. C., and von Hippel, P. H. (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal. Biochem.* 182, 319–326