THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 289, NO. 46, pp. 32073–32080, November 14, 2014

Conversion of a Chaperonin GroEL-independent Protein into an Obligate Substrate*

Received for publication, September 8, 2014, and in revised form, September 29, 2014 Published, JBC Papers in Press, October 6, 2014, DOI 10.1074/jbc.M114.610444

Takuya Ishimoto[‡], **Kei Fujiwara**[§], **Tatsuya Niwa**[‡], **and Hideki Taguchi**[‡1]

From the [‡]*Graduate School of Bioscience and Biotechnology, Tokyo Institute of Technology, B-56, 4259 Nagatsuta, Midori-ku, Yokohama 226-8501 and the* [§]*Department of Biosciences and Informatics, Keio University, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama 223-8522, Japan*

Background: Chaperonin GroEL and GroES (GroE) assist a subset of proteins in the cell.
Results: Conversion of a GroE-independent MetK into an obligate MetK occurred; GroE dependence was correlated with the propensity to form protein aggregates.
Conclusion: Subtle differences, even at single point mutations, determine the GroE dependence.
Significance: Buffering the aggregation-prone mutations by GroE plays a role in maintaining diversity of proteins.

Chaperones assist protein folding by preventing unproductive protein aggregation in the cell. In *Escherichia coli*, chaperonin GroEL/GroES (GroE) is the only indispensable chaperone and is absolutely required for the *de novo* folding of at least ~60 proteins. We previously found that several orthologs of the obligate GroE substrates in *Ureaplasma urealyticum*, which lacks the *groE* gene in the genome, are *E. coli* GroE-independent folders, despite their significant sequence identities. Here, we investigated the key features that define the GroE dependence. Chimera or random mutagenesis analyses revealed that independent multiple point mutations, and even single mutations, were sufficient to confer GroE dependence on the *Ureaplasma* MetK. Strikingly, the GroE dependence was well correlated with the propensity to form protein aggregates during folding. The results reveal the delicate balance between GroE dependence and independence. The function of GroE to buffering the aggregation-prone mutations plays a role in maintaining higher genetic diversity of proteins.

Protein folding is basically a spontaneous process (1). However, folding often competes with a side reaction, intermolecular aggregate formation, which usually impairs the functions of proteins (2, 3). Indeed, a global aggregation analysis of thousands of *Escherichia coli* proteins, using a reconstituted cell-free translation system, revealed that ~30% of proteins are aggregation-prone (4). The cellular milieu, where proteins and other molecules exist in highly crowded conditions, further increases the risk of aggregate formation (5, 6). In addition, most of the amino acid mutations in proteins are deleterious, and eventually result in aggregate formation (7).

To counteract the inherent tendency toward protein aggregation, cells have evolved a variety of chaperones (8). The canonical chaperones assist the folding of many proteins by preventing irreversible aggregate formation (3, 9, 10). *In vitro*, most of the aggregation-prone proteins, which were revealed by the reconstituted cell-free translation system (4), are rescued by one or more conserved chaperones, such as GroEL/GroES (GroE)[2] or DnaK/DnaJ/GrpE (DnaK system) (11). Overall, the effects of GroE and the DnaK system are similar, but certain fractions of proteins are biased toward either GroE or DnaK, probably reflecting the chaperone specificities (11). *In vivo*, recent proteome-wide analyses to identify the substrate proteins of chaperones confirmed the commonality and the specificity among the chaperone substrates (12–14). These previous observations suggested that determinants for the chaperone requirements would be included in their amino acid sequences, although the mechanism by which each chaperone recognizes non-native proteins in a different manner remains to be elucidated. To decode the determinants for the chaperone requirements, uncovering the relationship between chaperones and their obligate substrates is a key approach. The chaperonin GroE is a highly conserved molecular chaperone and is the only chaperone essential for *E. coli* viability (15). In *E. coli* cells, GroE interacts with ~250 proteins (12), including ~60 obligate substrates termed Class IV substrates, which are inherently aggregation-prone (13). The *in vivo* GroE dependence of Class IV substrates is not conserved among species; Class IV homologs (orthologs) in GroE-lacking organisms, such as *Ureaplasma urealyticum,* fold to the native state in GroE-depleted *E. coli* cells (13, 16). A recent study using structural homologs of obligate GroE substrates revealed that the folding properties were different between the GroE-dependent and GroE-independent proteins (16). Further comparative analyses using such homologs with different GroE requirements will be necessary to reveal the GroE dependence at an amino acid sequence level.

MetK, one of the Class IV substrates, is a good choice to analyze the determinants of the GroE requirement among homologs. The GroE requirement of MetK in *E. coli* (*Ec*MetK)

* This work was supported by Grants-in-aid for Scientific Research 19058002 and 17049009 (to H. T.) from Japan Society for the Promotion of Science (JSPS) and Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan. This work was also supported by grants from the Mochida Memorial Foundation, Takeda Science Foundation, and Uehara Memorial Foundation (to H. T.).

[1] To whom correspondence should be addressed. Tel./Fax: 81-45-924-5785; E-mail: taguchi@bio.titech.ac.jp.

[2] The abbreviations used are: GroE, GroEL and GroES; *Ec*MetK, *Escherichia coli* MetK; *Uu*MetK, *Ureaplasma urealyticum* MetK; sfGFP, superfolder GFP; MBP, maltose-binding protein.

has been analyzed very well both *in vitro* and *in vivo* (12, 13, 17). *Ec*MetK, which is an essential protein for methionine biosynthesis, links a well characterized phenotype of GroE-depleted cells; the loss of functional MetK in GroE-depleted *E. coli* induces the overexpression of the MetE protein (17–19). The MetK ortholog in *U. urealyticum* (*Uu*MetK), which shares 45% identity with *Ec*MetK, does not require GroE in its folding process and complements the MetE overexpression in GroE-depleted cells (13, 17).

In this study, we investigated the key features that define the GroE dependence, by using the MetK proteins. Chimera or random mutagenesis analyses revealed that independent multiple point mutations, and even single mutations, were sufficient to convert the GroE-independent *Uu*MetK into the GroE-dependent MetK. Strikingly, the GroE dependence was well correlated with the propensity to form protein aggregates during folding. The results suggest that the maintenance of GroE independence is marginal, and thus provide insight into protein evolution buffered by chaperones.

EXPERIMENTAL PROCEDURES

Plasmids—The plasmids pMCS, pMCS-HA, pMetE240-ParM-FLAG, and pMetE-sfGFP were constructed previously (13, 17, 20, 21). The *Uu*MetK gene optimized for *E. coli* codon use was chemically synthesized in the previous study (13). To construct chimeric *metK* genes, each region was independently amplified from pMCS-*Ec*MetK and pMCS-*Uu*MetK and combined by PCR. Local or point mutations were introduced by whole plasmid PCR, using PrimeSTAR® MAX DNA polymerase (Takara) and DNA primers with the corresponding mutations. The resultant genes encoding chimeric or mutated MetKs were digested and then cloned into the NdeI/BamHI sites of pMCS-HA. The *tac* promoter and the HA tag at the C terminus were attached by the cloning. To construct the plasmids for protein purification, the *metK* genes were cloned into the NdeI/XhoI sites of pET15b(+) (Novagen). To construct the plasmids for the colony assay, the BamHI/XhoI sites in pACYCDuet-1 (Novagen) were replaced with the BglII/XhoI sites of pMCS-*Uu*MetKHA, by digestion and ligation.

Proteins—E. coli BL21(DE3) cells harboring each *metK* plasmid were grown at 37 °C in LB medium with 50 μg/ml ampicillin. At log phase, protein overexpression was induced by 1 mM isopropyl β-D-thiogalactopyranoside. *Ec*MetK and *Uu*MetK were overexpressed for 3 h at 37 °C. *Uu*MetK mutants (K215F and K215F/I216F) were expressed by overnight induction at 18 °C because they formed inclusion bodies at 37 °C. The cells were then harvested, suspended in lysis buffer (20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 10% glycerol), and sonicated (Branson Sonifier). Soluble fractions were obtained by centrifugation at 15,000 × *g* for 30 min. The proteins were applied to nickel-nitrilotriacetic acid-agarose resin (Qiagen), which was washed and eluted with lysis buffer containing 50 or 400 mM imidazole, respectively. The eluted fractions were dialyzed against 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, and 10% glycerol. The MetK concentration was typically >400 μM after dialysis.

In vivo GroE Requirement Assay—The *in vivo* GroE requirement of each *metK* gene was assessed by the previously reported procedure, with slight modifications (13). *E. coli* MGM100 (MG1655 *groE*::*araC*-P_{BAD}-*groE* (Kan^r) (22)) cells harboring each MetK plasmid were grown in LB medium, with 50 μg/ml ampicillin and 0.2% arabinose, at 37 °C to log phase. After washing, the cells were diluted into LB with 1 mM diaminopimelate, 50 μg/ml ampicillin, and either 0.2% arabinose or 0.2% glucose. The dilution ratios were 1:5,000 for the 0.2% arabinose condition and 1:500 for the 0.2% glucose condition. The cells were cultivated for 5.5 h after the sugar shift, and then were harvested. The total quantities of the cells were adjusted with a buffer (50 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM EDTA), and then the cells were disrupted by sonication (Branson Sonifier). Soluble fractions were obtained by centrifugation at 15,000 × *g* for 30 min. The total and soluble extracts were examined by SDS-PAGE, and the MetK levels were detected by immunoblotting, using an anti-HA monoclonal antibody (conjugated with HRP) (Sigma).

Light Scattering Assay—Protein aggregation during refolding was monitored by light scattering for 20 min at room temperature (23). The light scattering intensity was measured with an FP-6500 spectrofluorometer (JASCO), with both excitation and emission at 640 nm. The purified recombinant MetK variants (100 μM) were denatured by 6 M guanidine hydrochloride and were refolded by dilution to 5 μM in 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 1 mM EDTA, and 1 mM dithiothreitol.

Random Mutagenesis—To generate random mutations in the *Uu*MetK gene, error-prone PCR with Mn^{2+} was performed. The PCR mixtures (50 μl) contained Blend Taq® -Plus- (TOYOBO), 20 μM $MnCl_2$, 2 mM dNTPs, 1 ng of pACYCtac-UuMetK, 0.4 μM primers (5′-CATGCCTGCAGGTAAGGAGATATACATATG-3′ and 5′-GGGACGTCGTATGGGTAGGATCC-3′), and 10% v/v dimethyl sulfoxide (DMSO). The PCR conditions were 95 °C for 5 min and 30 cycles of 94 °C for 30 s, 52 °C for 30 s, and 72 °C for 90 s. The error rate under these conditions was approximately 1–3 substitutions per 1,000 bases.

Colony Assay to Screen MetK Mutations That Convert the GroE Requirement—E. coli MGM100 cells harboring both pMetEp-sfGFP and the *metK* genes, under the *tac* promoter cloned in the pACYC vector, were cultured at 37 °C on LB plates with 50 μg/ml ampicillin, 12.5 μg/ml chloramphenicol, and 0.2% arabinose. Colonies were transferred onto gauze (BEMCOT, Asahi Kasei Fibers) to make replicas on LB plates with 1 mM diaminopimelate, 50 μg/ml ampicillin, 12.5 μg/ml chloramphenicol, 1 mM isopropyl β-D-thiogalactopyranoside, and 0.2% glucose. After an incubation at 37 °C for 5 h, the GFP fluorescence of each colony was detected with a trans-illuminator (BioSpeed).

RESULTS

Systematic Chimeric Analysis between EcMetK and UuMetK—The sequence similarities between *Ec*MetK and *Uu*MetK are very high, and they share 45% amino acid identity (Fig. 1*A*). In particular, the middle (M) domain, which corresponds to residues 221–301 of *Ec*MetK, is well conserved, with 74% identical residues (Fig. 1*A*). We divided MetK into three domains, including the N-terminal and C-terminal sides of the M domain (N, M, and C domains, Fig. 1*B*). To identify which domains are critical for the GroE dependence of *Ec*MetK, all combinations

A

```
Ec    1 MA-KHLFTSESVSEGHPDKIADQISDAVLDAILEQDPKARVACETYVKTGMVLVGGEITTSAWVDIEEITRNTVREIGYVHSDMGFDANSCAVLSAI
Uu    1 MKYKKIITSESVGAGHPDKICDQISDAILDECLAQDKNSRVACEVLACNRLIVIAGEITTHAYVDVVKTAWEIIKPLGYNENDF-------TIISNV
        *  *:::*****. ******.******:**  * ** ::*****. . . ::::.***** *:**: : : : :: :** ..*:       :::* :

Ec   97 GKQSPDINQGVDRADPLEQGAGDQGLMFGYATNETDVLMPAPITYAHRLVQRQAEVRKNGTLPWLRPDAKSQVTFQYDDGKIVGIDAVVLSTQHSEE
Uu   91 NKQSVDIAQSVNKANKDLIGAGDQGIVFGYACDETPQYMPLTTVLAHELLKEIEKQRRNKEFVKIQADMKSQVSIDYSNCTPV-IETMLVSIQHDED
        .*** ** *.*::*:    ******::**** :**   ** . . **.*::.  : *:*  :  ::.* ****:::*.: . * *:::::* **.*:

Ec  194 IDQKSLQEAVMEEIIKPILPAEWLTSATKFFINPTGRFVIGGPMGDCGLTGRKIIVDTYGGMARHGGGAFSGKDPSKVDRSAAYAARYVAKNIVAAG
Uu  187 YNIEYFNKKVS-TIMDQIAIKYGLNTDFKKIINSSGRFVIGGPIGDTGLTGRKIIVDTYGGIGHHGGGAFSGKDPTKVDRSASYFARWIAKNVVAAK
         : : ::: *   *:. *     *. : * :**.:********:** ***************.:***********:******:* **::***:***

Ec  201 LADRCEIQVSYAIGVAEPTSIMVETFGTEKVPSEQLTLLVREFFDLRPYGLIQMLDLLHPIYKETAAYGHFGREHF--PWEKTDKAQLLRDAAGLK
Uu  193 LAKQCEIQLAFAIGEPQPVAMFVNTFKTNLVDEKKIFEAINKSFNFDIKTFINDLELWTTKYLPVATYGHFGRDDLNLSWEKLNKVQDLIKYSK--
        **.:****:::*** .:*.:::*:** *: * .::: :.: *::    :*: *:*  . *  .*:******::.: .*** :*.* * . :
```

B

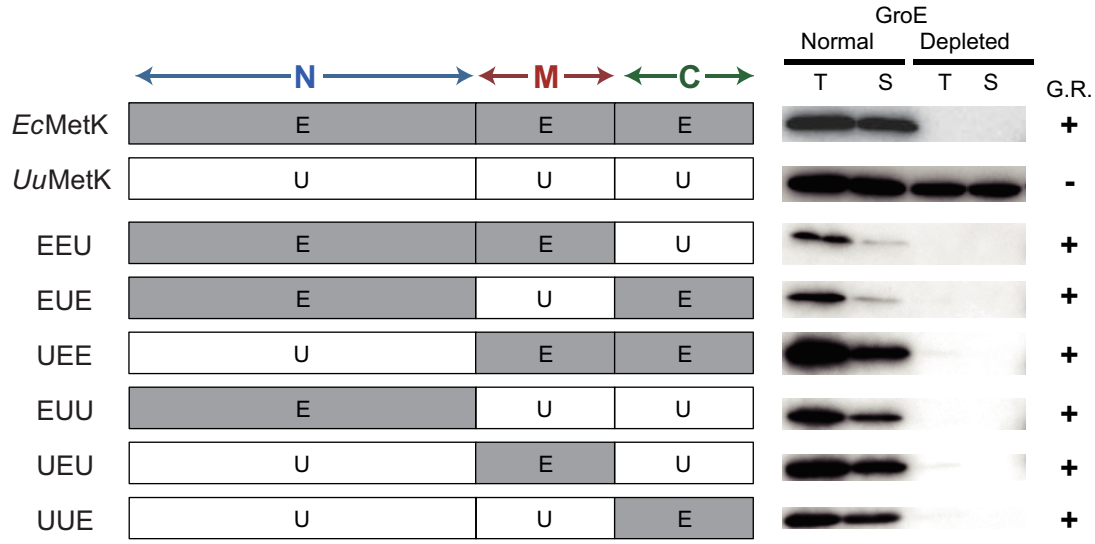

FIGURE 1. **GroE requirement of MetK chimeric mutants.** *A*, amino acid sequence alignment between *Ec*MetK (*Ec*) and *Uu*MetK (*Uu*). The N-terminal, middle (M), and C-terminal domains are colored *blue*, *red*, and *green*, respectively. *Asterisks, periods,* and *colons* indicate that the corresponding residues are identical, similar (weak), or similar (strong), respectively, between the two MetK sequences. *B*, GroE requirements of chimeric MetKs. Fractions of total (*T*) and soluble (*S*) MetKs in GroE-normal or -depleted cells were detected by Western blotting. Domains of *Ec*MetK and *Uu*MetK are indicated by *E* or *U*, and colored *gray* or *white*, respectively. *G.R.*: GroE requirement.

of chimeric MetKs (EEU, EUE, UEE, EUU, UEU, and UUE) were constructed (Fig. 1*B*).

To evaluate the *in vivo* GroE requirement of the MetK chimeras, we applied the method using the conditional GroE expression strain developed previously (13). This method utilizes the following characteristic: the obligate GroE substrates become aggregated or are degraded under the GroE-depleted conditions (13, 22). Various MetKs, including the chimeras under the *tac* promoter, were expressed in the conditional GroE strain MGM100, in which the expression of GroE is controlled by the arabinose-inducible *BAD* promoter (22). To avoid potential aggregate formation due to MetK overexpression, leaky expression by the *tac* promoter was used in this study. The MetK proteins in the total and soluble fractions were separated by centrifugation and detected by Western blotting. *Ec*MetK was degraded under the GroE-depleted conditions, whereas *Uu*MetK remained in the soluble fraction even under these conditions (Fig. 1*B*). We also observed that the *Uu*MetK expression suppressed the overexpression of MetE (data not shown) (13), which is one of the hallmarks of GroE-depleted *E. coli* (17–19), indicating that the soluble *Uu*MetK was functional. These results confirmed that *Ec*MetK is GroE-dependent but *Uu*MetK is not, as described previously (13). We investigated the MetK chimeras using the *in vivo* GroE dependence assay and found that, strikingly, all of them were degraded under the GroE-depleted conditions, indicating that all of the chimeric MetKs depend on GroE for folding.

However, we noticed that several chimeras (*e.g.* EEU, EUE) tended to be degraded even under the normal GroE conditions, suggesting that their stabilities were lower than those of their authentic counterparts, *Ec*MetK and *Uu*MetK. These results revealed that at least one replacement of the GroE-independent *Uu*MetK domain with the GroE-dependent *Ec*MetK domain confers the GroE dependence on the chimeric MetKs *in vivo*. The results also implied that all three domains have determinants that confer the GroE dependence.

Single or Double Amino Acid Substitutions Are Sufficient to Convert UuMetK into an Obligate GroE Substrate—Next, we sought to identify the critical regions that confer the GroE dependence. We studied the UEU chimera because it shares 94% amino acid sequence identity with wild-type *Uu*MetK (Figs. 1*A* and 2*A*). In the crystal structure of *Ec*MetK (Protein Data Bank (PDB) code: 1fug), the M domain is composed of three β-sheets (β8–β10) and two α-helices (α5 and α6). Among these secondary structures, β-sheet 9 and α-helix 5 are identical between *Uu*MetK and *Ec*MetK. Therefore, we substituted the amino acids in the regions of the other three secondary structures (β8, α6, and β10, Fig. 2*A*) with the corresponding *Ec*MetK amino acids. Again, all three mutants (U-$E_{\beta 8}$-U, U-$E_{\alpha 6}$-U, and U-$E_{\beta 10}$-U) showed the GroE dependence (Fig.

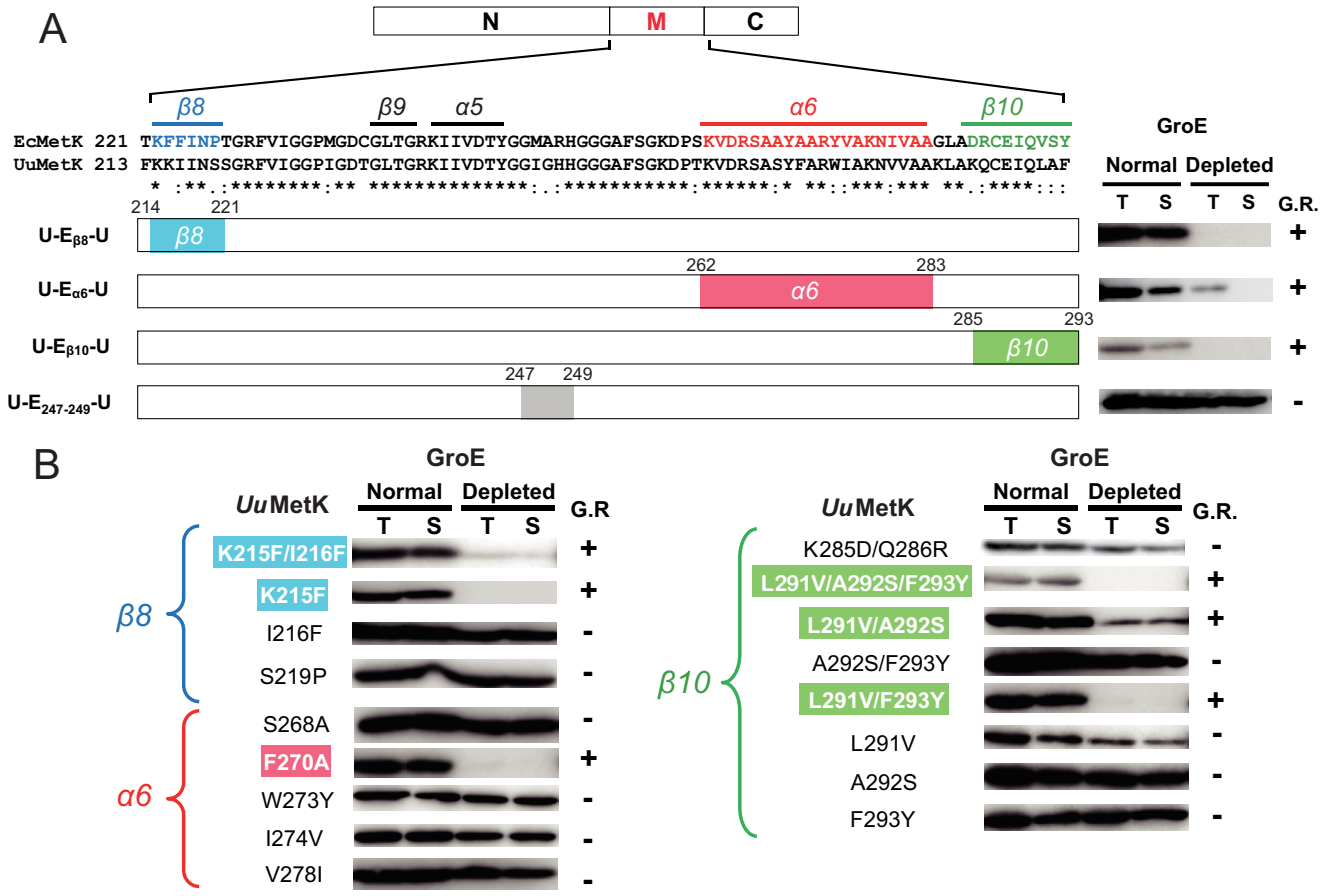

FIGURE 2. **Substitutions and mutagenesis analyses in the M domain of *Uu*MetK.** Fractions of total (*T*) and soluble (*S*) MetKs in GroE-normal or -depleted cells were detected by Western blotting. *G. R.:* GroE requirement. *A*, local replacement of *Uu*MetK with *Ec*MetK residues. *Asterisks, periods,* and *colons* indicate that the corresponding residues are identical, similar (weak), or similar (strong), respectively, between the two MetK sequences. β and α represent β-sheet and α-helix, respectively, in the *Ec*MetK crystal structure (25). The first and last amino acid numbers of *Uu*MetK in the secondary structures are described. *B*, comprehensive substitutions of *Uu*MetK with *Ec*MetK residues in the secondary structures. Mutations that converted the GroE requirement are shown by *white letters* on *colored backgrounds*.

2*A*). In contrast, the variant altered in the residues from 247 to 249 (U-$E_{247-249}$-U), which are not conserved between *Ec*MetK and *Uu*MetK and are not included in any secondary structures, did not change the GroE-independent property of *Uu*MetK (Fig. 2*A*).

To identify the minimum regions required to convert GroE-independent *Uu*MetK to GroE dependence, we constructed *Uu*MetK mutants with several amino acid substitutions in the secondary structure regions, β8, α6, and β10, which altered the GroE dependence of *Uu*MetK. Among the tested mutants, six mutants, K215F/I216F, K215F, F270A, L291V/A292S/F293Y, L291V/A292S, and L291V/F293Y, showed GroE dependence *in vivo* (Fig. 2*B*). Among these mutants, two mutants, K215F and F270A, had only a single amino acid substitution, indicating that single point mutations are sufficient to covert the GroE-independent substrate to GroE dependence.

Aggregation Propensities of GroE-dependent UuMetK Mutants—Almost all GroE obligate substrates show strong aggregation propensities, according to the global aggregation analysis under the chaperone-free conditions (4, 13). Thus, we investigated the *in vitro* aggregation properties of representative *Uu*MetK mutants, K215F and K215F/I216F, which were converted to GroE-dependent substrates, by a conventional light scattering assay. Under the conditions where *Ec*MetK formed aggregates but wild-type *Uu*MetK did not, the purified K215F and K215F/I216F *Uu*MetK mutants aggregated (Fig. 3). The double mutant, K215F/I216F, was more prone to aggregation than the single K215F mutant (Fig. 3), suggesting an additive effect of each mutation on the aggregation propensity. These results suggest that the *in vivo* GroE dependence of *Uu*MetK mutants correlates with the aggregation propensities of the proteins.

Screening System to Identify Other GroE-dependent UuMetK Mutants—The above results, which were restricted to the M domain, showed that point mutations, including single amino acid substitutions, convert the *in vivo* GroE requirement of *Uu*MetK. Besides the M domain, the replacement of the N or C domain also converted the GroE requirement (Fig. 1*B*), indicating that the determinants that confer the GroE dependence on *Uu*MetK are distributed throughout the entire *Uu*MetK sequence. To identify these determinants, we developed a screening system to select the GroE-dependent *Uu*MetK proteins from randomly mutated *Uu*MetK proteins.

To evaluate the GroE requirement of the numerous *Uu*MetK random mutants, we established a screening system that can distinguish the *Uu*MetK mutants that confer GroE dependence

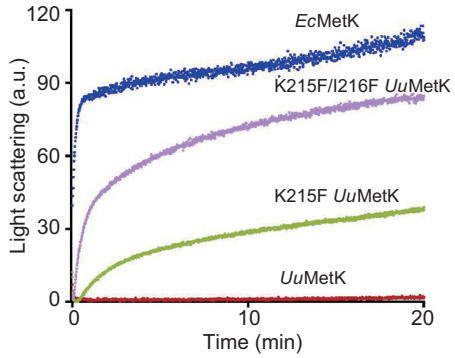

FIGURE 3. **Light scattering assay to evaluate the aggregation propensities of MetKs.** Time courses of protein aggregation, monitored by light scattering at 640 nm, for wild-type *Ec*MetK (*blue*), wild-type *Uu*MetK (*red*), the *Uu*MetK K215F mutant (*green*), and the *Uu*MetK K215F/I216F mutant (*purple*) are shown. Purified recombinant MetKs were denatured by 6 M guanidine HCl and then were diluted 20-fold (5 μM, final concentration) into the buffer. *a. u.*, arbitrary units.

from wild-type *Uu*MetK, at the colony level (Fig. 4*A*). We focused on one of the *E. coli* phenotypes associated with the GroE depletion, the massive induction of methionine synthase (MetE) (13, 17–19). The MetE overexpression is induced by the activation of the *metE* promoter (*metE*p) due to the impaired folding of *Ec*MetK under the GroE-depleted conditions (13, 17). Briefly, the impaired *Ec*MetK activity, induced by the GroE depletion, leads to the deficiency of *S*-adenosyl methionine, the product of the MetK enzyme reaction, which then activates the *metE* promoter (17). Because we previously showed that the MetE overexpression in GroE-depleted cells is suppressed by the *Uu*MetK expression (13, 17), we reasoned that the conversion of *Uu*MetK to a GroE-dependent substrate would activate the *metE* promoter. Therefore, we constructed a GFP reporter plasmid in which the expression of superfolder GFP (sfGFP (24)) was under the control of the *metE* promoter (21) (Fig. 4, *A* and *B*). The cells were also transformed with additional plasmids harboring randomly mutated *Uu*MetK genes. We expected to observe green fluorescent colonies when the cells expressed the *Uu*MetK mutants that are converted to the GroE obligate substrates, under the GroE-depleted conditions (Fig. 4, *A* and *B*).

Identification of Mutations in UuMetK That Confer GroE Dependence—The expression of wild-type *Uu*MetK in this system under the GroE-depleted conditions did not generate GFP fluorescence (Fig. 4*C*), confirming that *Uu*MetK complemented the impaired folding of *Ec*MetK due to the lack of GroE. Therefore, we expressed randomly mutated *Uu*MetK proteins, generated by error-prone PCR using Mn^{2+}, and evaluated their GroE requirements. We obtained 809 colonies in total and then selected 169 colonies that showed bright GFP fluorescence intensities. We should note that these 809 colonies are only a small fraction of the possible mutations generated by this mutation rate. Because the simple insertion of a stop codon or mutations in the *Uu*MetK ORF that result in decreased MetK expression would lead to the activation of *metE* and the bright GFP fluorescence, we cultured all of the selected colonies individually and eliminated the false positives by multiple procedures, as follows. First, the cells that expressed full-length *Uu*MetK mutants were selected by Western blotting. Second, the *Uu*MetK DNAs were analyzed by PCR and restriction enzyme digestion. Third, to eliminate the *Uu*MetK mutants that lost the *S*-adenosyl methionine synthesis activity, the *Uu*MetK mutants that remained soluble upon isopropyl β-D-thiogalactopyranoside-induced overexpression under the GroE-depleted conditions were discarded because such mutants were expected to lack the *S*-adenosyl methionine synthesis function. After eliminating the false positives through these selections, we finally obtained 35 mutants of *Uu*MetK that showed the GroE requirement for folding *in vivo* (Table 1).

Characteristics of Mutations That Convert UuMetK into Obligate GroE Substrates—Among the obtained 35 mutants, 54% of them (19 out of 35 clones) had only a single amino acid substitution, and the others had two or three substitutions (Table 1). The mutation positions were widely distributed throughout the three domains (N, M, and C), although the single mutants tended to accumulate in either the M-domain or the N-terminal region (Fig. 5*A* and Table 1).

Next, we analyzed the correlation of the mutation positions with the secondary structures in the MetK tertiary structure (Structural Classification of Proteins (SCOP) fold ID: d.130). The extremely high homology between *Uu*MetK and *Ec*MetK permitted us to generate a homology-modeled *Uu*MetK structure (Fig. 5*B*), based on the crystal structure of *Ec*MetK (PDB code: 1fug (25)). Mapping of the single mutations on the *Uu*MetK structure revealed their dispersed distribution in the *Uu*MetK structure (Fig. 5*B*). The mutations were not accumulated in secondary structure regions (Fig. 5, *B* and *C*). Instead, the mutations tended to be found in the flanking regions of the secondary structures, defined as the ±2 residues from the boundary of the secondary structures; 10 out of 19 single mutations and 29 out of 54 all mutation points were found in the flanking regions (Fig. 5*C*). Furthermore, the mutation rates in the conserved residues between *Ec*MetK and *Uu*MetK were remarkably higher than the identity of these proteins (45%); the ratio was 63% (12 out of 19) in the case of single mutations and 63% (34 out of 54) in the case of total mutations.

DISCUSSION

Recent proteome-wide approaches to identify *in vivo* GroE substrates revealed that a subset of the proteome is obligate GroE substrates, which absolutely require GroE for correct folding. *Ec*MetK, one of the obligate GroE substrates, has a highly homologous ortholog, *Uu*MetK in *Ureaplasma*, which does not require GroE for folding in *E. coli* (13). In this study, we tried to convert the GroE-independent *Uu*MetK to GroE dependence by chimera analysis, site-directed amino acid substitutions, and random mutagenesis. We found that independent multiple point mutations, and even single mutations distributed throughout the entire ORF, were sufficient to convert the GroE-independent *Uu*MetK to GroE dependence.

In addition, our results showed that the mutations that converted the GroE dependence were not rare; at least ~5% of the amino acids in the ORF were identified as GroE determinants, although this might be underestimated because our random mutagenesis experiment was not saturated. Based on the multiple determinants of the GroE dependence distributed throughout the entire sequence, one of the main messages of

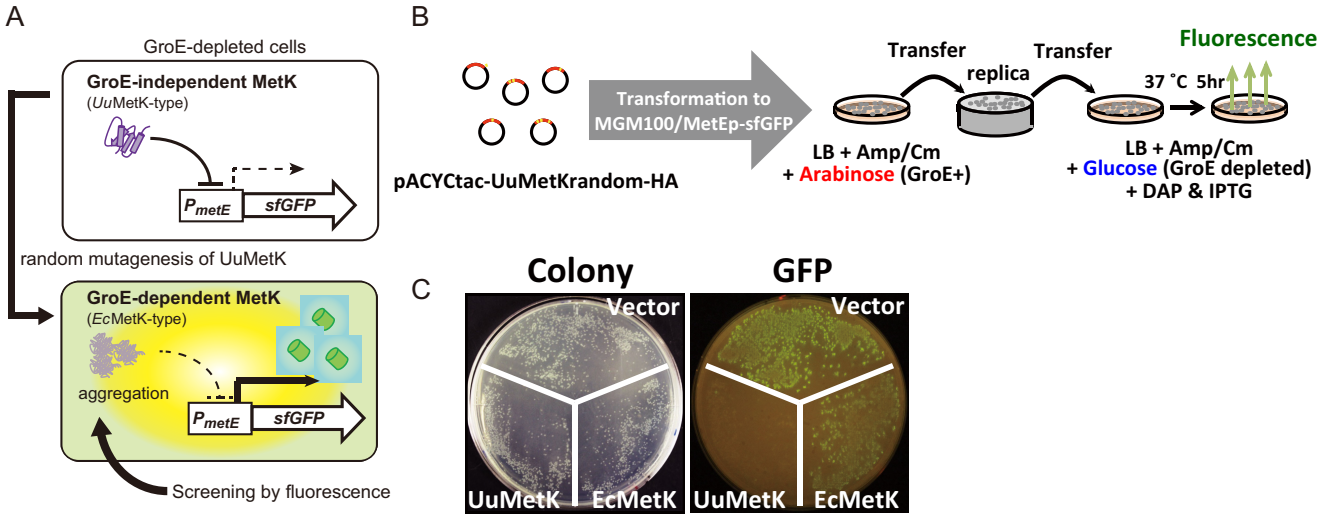

FIGURE 4. **Random mutagenesis analysis to convert *Uu*MetK into obligate GroE substrates.** *A*, the screening system to identify GroE-dependent *Uu*MetK mutants. *Ec*MetK encoded in the genome of the *E. coli* MGM100 strain is impaired under the GroE-depleted conditions, and consequently, the *metE* promoter is activated. *Uu*MetK suppresses the *metE* activation, but when mutations convert *Uu*MetK into the obligate GroE substrate, the suppression does not occur. The activation of the *metE* promoter was monitored by GFP fluorescence. *B*, a schematic illustration of the experimental procedure. *E. coli* MGM100 cells harboring the pMetEp-sfGFP plasmid were transformed with a randomly mutagenized library of *Uu*MetK. After the formation of colonies under the GroE-expressing conditions, the colonies were transferred to LB agar plates containing 0.2% glucose, which suppresses the expression of the *groESL* gene regulated by the *araBAD* promoter. After an incubation for 5 h at 37 °C, the fluorescence intensities were detected by an LAS-4000 FluorImager (Fujifilm). *Amp*, ampicillin; *Cm*, chloramphenicol; *DAP*, diaminopimelate; *IPTG*, isopropyl β-D-thiogalactopyranoside. *C*, pictures of the colonies and the GFP fluorescence under the *Ec*MetK- and *Uu*MetK-expressing conditions.

this work is that the difference between the GroE-dependent and GroE-independent properties is marginal; a delicate balance determines the GroE dependence.

Although more than half of the mutations that conferred GroE dependence are single amino acid substitutions (Table 1), we detected additive effects for the acquisition of the GroE requirement. For example, the single mutations of L291V, A292S, or F293Y in *Uu*MetK did not affect the GroE dependence, although the solubility of L291V in the GroE-normal cells was slightly lower than that of wild-type *Uu*MetK (Fig. 2*B*). However, the additional A292S or F293Y mutation converted the GroE-independent L291V mutant into a GroE-dependent protein, indicating that the accumulation of mutations had additive effects. This kind of additive effect on the GroE dependence might be related to the case of maltose-binding protein (MBP) mutants. The double mutant MBP(V8G/Y283D) has a more stringent folding property as a GroE substrate than the single mutant MBP(Y283D) (26), although MBP has not been identified as an *in vivo* GroE substrate protein.

What properties are affected by the amino acid substitutions in the GroE-dependent *Uu*MetK variants? One immediate possibility to convert the GroE dependence might be the acquisition of GroEL-binding motifs, due to the mutations. However, this is unlikely because no consensus GroEL-binding motif has been identified based on primary amino acid sequences, although several attempts have been made to identify sequence motifs for GroEL substrates (*e.g.* Refs. 27–29).

Instead, we raised the possibility that the GroE dependence is correlated with the aggregation propensity because GroEL assists in protein folding by suppressing the formation of the aggregates (*e.g.* Refs. 23 and 30). In fact, we previously found that *in vivo* obligate GroE substrates are strongly aggregation-prone (13). Strikingly, the *in vitro* aggregation analysis clearly showed that the *in vivo* GroE dependence is closely correlated with the inherent aggregation propensity (Fig. 3), supporting the idea that the acquisition of aggregation propensity could cause the conversion to GroE dependence.

The next point to be discussed is the determinants that confer the aggregation propensities. A previous global aggregation analysis under chaperone-free conditions only revealed that the aggregation property is partly correlated with the topologies of the secondary structures (*i.e.* Structural Classification of Proteins (SCOP)) (4). The bias on a tertiary structure itself is not applicable to this study because this work only considered MetK. However, we noticed that the mutation sites were enriched in the flanking regions of the secondary structures (Fig. 5*C*). Although there is no plausible explanation for the enrichment, mutations in the vicinity of the secondary structures might affect the folding rate of MetK. For example, the slower folding rates caused by the mutations would prolong the lifetimes of the folding intermediates, which tend to self-associate and form aggregates. The prolonged folding rate would be compensated by GroE. Indeed, prolonged folding rate caused by a single point mutation of MBP, MBP(Y283D), results in a much higher stability of the MBP(Y283D)-GroEL complex than that of wild-type MBP (31). In addition, this scenario, in which GroEL accelerates the folding rate, is consistent with the results from a recent detailed analysis using DapA, an *in vivo* GroE obligate substrate (22), and its homologs, in which GroE accelerates the rate of TIM barrel domain folding by catalyzing segmental structure formation (16). Alternatively, the stability of the folding intermediates may be lower in the mutants. It was recently proposed (32) that GroEL stabilizes these impaired folding intermediates, which eventually leads to the acquisition of GroE dependence.

TABLE 1
Mutation positions that convert GroE requirement of *Uu*MetK

Number	Mutations in *Uu*MetK[a] N	M	C
Single mutations			
1	G(S)13E		
2	G(G)15E		
3	K(K)19E		
4	D(D)26G		
5	C(K)48R		
6	Q(R)156P		
7		G(G)227D	
8		L(L)234P	
9		D(D)242G	
10		H(H)250R	
11		S(S)256L	
12		G(G)257S	
13		F(A)270S	
15		I(V)274T	
14		V(V)279A	
16		F(Y)293S	
17			L(L)337P
18			L(K)345P
19			W(W)362R
Two or more mutations			
20	K(K)19R A(N)98V		
21	D(D)30G N(V)80S		
22	L(L)33P A(V)47V		
23	N(D)102D Q(Q)114R		
24	E(E)10K	F(A)270S	
25	P(P)17S	T(T)243I	
26	F(F)57G S(G)100L	I(I)240T	
27	D(D)113G	F(A)270L	
28	G(G)115E	F(F)255S	
29	E(A)31G		L(F)358P
30	E(E)144G		L(L)337Q
31	D(E)186G		Y(Y)350C
32	L(L)29P	G(G)233R	K(K)367R
33		L(L)234P Y(Y)269H	
34		L(L)234P	F(F)326L P(E)346S
35		G(G)252E	T(A)349A

[a] Letters in parentheses indicate corresponding residues of *Ec*MetK.

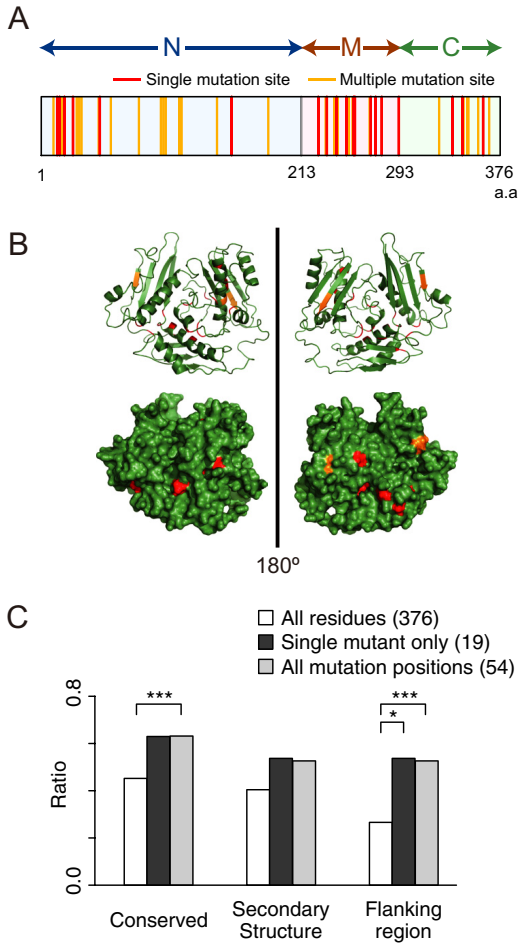

FIGURE 5. **Mutation sites that converted the GroE requirement of *Uu*MetK.** *A*, mutation sites were colored in *red* or yellow for single or multiple mutations, respectively, plotted on the ORF of *Uu*MetK. *a.a*, amino acids. *B*, the single mutations described above were plotted on a tertiary structure model of *Uu*MetK, obtained through the SWISS-MODEL Repository (SIB) and derived from the crystal structure of *Ec*MetK (PDB code: 1fug) (25). The regions colored *red* and *orange* represent the mutation points obtained by the random mutagenesis (Table 1) and by the amino acid substitutions (Fig. 2*B*), respectively. *C*, enrichment of mutation positions in the conserved residues and the flanking regions of secondary structures. Shown is comparison of the ratios of the conserved residue, the secondary structure, and the flanking region of the secondary structure between all amino acid residues and the mutation positions that cause conversion to the GroE requirement of *Uu*MetK. *Dark gray bars* show the ratios in 19 single mutations, and *light gray bars* show all 54 mutation positions, including the mutants with multiple mutations (***, $p < 0.01$, *, $p < 0.05$, binomial test).

Most protein mutations are deleterious because the small margin of protein stability prevents the acquisition of a large fraction of mutations (33). It has been proposed that chaperones have facilitated protein evolution by buffering the deleterious mutations that cause impaired folding (34). This concept, which was originally developed from studies on Hsp90 chaperones, has now been extended to the chaperonin GroEL (35–37). In particular, directed evolution studies clearly revealed that GroE overexpression can buffer the destabilizing mutations, thereby promoting the folding of compromised proteins to improve the enzyme activity (32, 36). In connection with our study, we would like to note that the evolved enzymes in GroE-overexpressing *E. coli* gained aggregation-prone properties and were converted to GroE-dependent proteins (36). The relationship between the aggregation propensity and the GroE dependence in the directed evolution experiments is strikingly consistent with that in our experiment, supporting the conclusion that the difference between the GroE-dependent and GroE-independent properties was marginal. The function of GroE to buffer the aggregation-prone mutations therefore plays a critical role in maintaining higher genetic diversity, eventually leading to the expansion of the protein world, including metabolic networks (38).

Finally, one might ask about the possibility of the opposite direction of the conversion on the GroE dependence. Is it possible to engineer the *Ec*MetK to become a GroE-independent protein? The answer would be yes in principle because our data show that a delicate balance determines the GroE dependence. However, the conversion of GroE-dependent substrates such as *Ec*MetK into a GroE-independent protein might be difficult because destabilizing mutations might be accumulated in the GroE substrates during protein evolution. Studies converting GroE-dependent substrates into independent ones, which are in progress, would provide further insight about how the GroE dependence is determined.

REFERENCES

1. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science* **181,** 223–230
2. Dobson, C. M. (2003) Protein folding and misfolding. *Nature* **426,** 884–890
3. Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2011) Molecular chaperones in protein folding and proteostasis. *Nature* **475,** 324–332
4. Niwa, T., Ying, B. W., Saito, K., Jin, W., Takada, S., Ueda, T., and Taguchi, H. (2009) Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of *Escherichia coli* proteins. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 4201–4206
5. Ellis, R. J., and Minton, A. P. (2006) Protein aggregation in crowded environments. *Biol. Chem.* **387,** 485–497
6. Gershenson, A., and Gierasch, L. M. (2011) Protein folding in the cell: challenges and progress. *Curr. Opin. Struct. Biol.* **21,** 32–41
7. Tokuriki, N., and Tawfik, D. S. (2009) Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **19,** 596–604
8. Ellis, J. (1987) Proteins as molecular chaperones. *Nature* **328,** 378–379
9. Tyedmers, J., Mogk, A., and Bukau, B. (2010) Cellular strategies for controlling protein aggregation. *Nat. Rev. Mol. Cell Biol.* **11,** 777–788
10. Richter, K., Haslbeck, M., and Buchner, J. (2010) The heat shock response: life on the verge of death. *Mol. Cell* **40,** 253–266
11. Niwa, T., Kanamori, T., Ueda, T., and Taguchi, H. (2012) Global analysis of chaperone effects using a reconstituted cell-free translation system. *Proc. Natl. Acad. Sci. U.S.A.* **109,** 8937–8942
12. 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
13. 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
14. Calloni, G., Chen, T., Schermann, S. M., Chang, H. C., Genevaux, P., Agostini, F., Tartaglia, G. G., Hayer-Hartl, M., and Hartl, F. U. (2012) DnaK functions as a central hub in the *E. coli* chaperone network. *Cell Rep.* **1,** 251–264
15. Fayet, O., Ziegelhoffer, T., and Georgopoulos, C. (1989) The groES and groEL heat shock gene products of *Escherichia coli* are essential for bacterial growth at all temperatures. *J. Bacteriol.* **171,** 1379–1385
16. Georgescauld, F., Popova, K., Gupta, A. J., Bracher, A., Engen, J. R., Hayer-Hartl, M., and Hartl, F. U. (2014) GroEL/ES chaperonin modulates the mechanism and accelerates the rate of TIM-barrel domain folding. *Cell* **157,** 922–934
17. Fujiwara, K., and Taguchi, H. (2012) Mechanism of methionine synthase overexpression in chaperonin-depleted *Escherichia coli*. *Microbiology* **158,** 917–924
18. Chapman, E., Farr, G. W., Usaite, R., Furtak, K., Fenton, W. A., Chaudhuri, T. K., Hondorp, E. R., Matthews, R. G., Wolf, S. G., Yates, J. R., Pypaert, M., and Horwich, A. L. (2006) Global aggregation of newly translated proteins in an *Escherichia coli* strain deficient of the chaperonin GroEL. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 15800–15805
19. Masters, M., Blakely, G., Coulson, A., McLennan, N., Yerko, V., and Acord, J. (2009) Protein folding in *Escherichia coli*: the chaperonin GroE and its substrates. *Res. Microbiol.* **160,** 267–277
20. Fujiwara, K., and Taguchi, H. (2007) Filamentous morphology in GroE-depleted *Escherichia coli* induced by impaired folding of FtsE. *J. Bacteriol.* **189,** 5860–5866
21. Fujiwara, K., Aoi, K. B., and Nomura, S. M. (2013) A bacterial salt sensor created by multiplying phenotypes of GroE-depleted *Escherichia coli*. *Anal. Methods* **5,** 5918–5922
22. McLennan, N., and Masters, M. (1998) GroE is vital for cell-wall synthesis. *Nature* **392,** 139
23. Taguchi, H., Makino, Y., and Yoshida, M. (1994) Monomeric chaperonin-60 and its 50-kDa fragment possess the ability to interact with nonnative proteins, to suppress aggregation, and to promote protein folding. *J. Biol. Chem.* **269,** 8529–8534
24. Pédelacq, 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
25. Takusagawa, F., Kamitori, S., Misaki, S., and Markham, G. D. (1996) Crystal structure of *S*-adenosylmethionine synthetase. *J. Biol. Chem.* **271,** 136–147
26. Tang, Y. C., Chang, H. C., Roeben, A., Wischnewski, D., Wischnewski, N., Kerner, M. J., Hartl, F. U., and Hayer-Hartl, M. (2006) Structural features of the GroEL-GroES nano-cage required for rapid folding of encapsulated protein. *Cell* **125,** 903–914
27. Stan, G., Brooks, B. R., Lorimer, G. H., and Thirumalai, D. (2006) Residues in substrate proteins that interact with GroEL in the capture process are buried in the native state. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 4433–4438
28. Tartaglia, G. G., Dobson, C. M., Hartl, F. U., and Vendruscolo, M. (2010) Physicochemical determinants of chaperone requirements. *J. Mol. Biol.* **400,** 579–588
29. Azia, A., Unger, R., and Horovitz, A. (2012) What distinguishes GroEL substrates from other *Escherichia coli* proteins? *FEBS J.* **279,** 543–550
30. Buchner, J., Schmidt, M., Fuchs, M., Jaenicke, R., Rudolph, R., Schmid, F. X., and Kiefhaber, T. (1991) GroE facilitates refolding of citrate synthase by suppressing aggregation. *Biochemistry* **30,** 1586–1591
31. Sparrer, H., Lilie, H., and Buchner, J. (1996) Dynamics of the GroEL-protein complex: effects of nucleotides and folding mutants. *J. Mol. Biol.* **258,** 74–87
32. Wyganowski, K. T., Kaltenbach, M., and Tokuriki, N. (2013) GroEL/ES buffering and compensatory mutations promote protein evolution by stabilizing folding intermediates. *J. Mol. Biol.* **425,** 3403–3414
33. Tokuriki, N., Stricher, F., Serrano, L., and Tawfik, D. S. (2008) How protein stability and new functions trade off. *PLoS Comput. Biol.* **4,** e1000002
34. Rutherford, S. L., and Lindquist, S. (1998) Hsp90 as a capacitor for morphological evolution. *Nature* **396,** 336–342
35. Fares, M. A., Ruiz-González, M. X., Moya, A., Elena, S. F., and Barrio, E. (2002) Endosymbiotic bacteria: GroEL buffers against deleterious mutations. *Nature* **417,** 398
36. Tokuriki, N., and Tawfik, D. S. (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459,** 668–673
37. Bogumil, D., and Dagan, T. (2012) Cumulative impact of chaperone-mediated folding on genome evolution. *Biochemistry* **51,** 9941–9953
38. Takemoto, K., Niwa, T., and Taguchi, H. (2011) Difference in the distribution pattern of substrate enzymes in the metabolic network of *Escherichia coli*, according to chaperonin requirement. *BMC Syst. Biol.* **5,** 98