Supporting Information

Niwa et al. 10.1073/pnas.1201380109

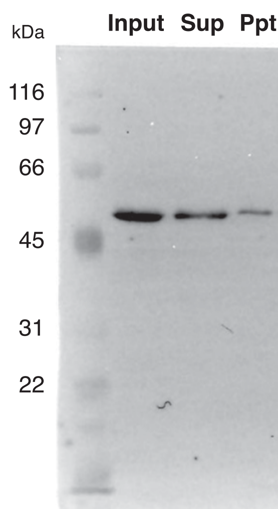

Fig. S1. Association of trigger factor (TF) with the ribosome. Sucrose cushion experiment to detect chaperones in the ribosome fraction was conducted as previously described (1) with a slight modification. C-terminal 6xHis-tagged TF (5 μM) and the ribosome (1 μM) were incubated in the PURE system reaction buffer (2) without translation factors at 37 °C for 10 min. Then, the mixture was overlaid onto 1.6 mL sucrose buffer [20 mM Hepes-KOH, 10 mM $Mg(OAc)_2$, 30 mM NH_4Cl, 20% (wt/vol) sucrose, pH 7.6] and centrifuged at 50,000 rpm in a TLA100.3 rotor (Beckman Coulter) for 4 h. The resulting supernatant was concentrated by a trichloroacetic acid precipitation. The resulting pellet (ribosome fraction) was dissolved with Hepes buffer [20 mM Hepes-KOH, 10 mM Mg $(OAc)_2$, 30 mM NH_4Cl, pH 7.6]. Uncentrifuged solution (Input), the supernatant (Sup), and the pellet (Ppt) fractions were analyzed by Western blotting with anti-6xHis monoclonal antibody (Wako) and Cy3-conjugated anti-mouse antibody (Invitrogen). Quantification using FLA7000 fluoroimager and multigauge software (Fujifilm) revealed that TF in the pellet fraction was 0.59 μM, indicating that TF is associated with more than half of the ribosome (1 μM).

1. Ying BW, Taguchi H, Ueda T (2006) Co-translational binding of GroEL to nascent polypeptides is followed by post-translational encapsulation by GroES to mediate protein folding. *J Biol Chem* 281:21813–21819.
2. Niwa T, et al. (2009) Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of *Escherichia coli* proteins. *Proc Natl Acad Sci USA* 106: 4201–4206.

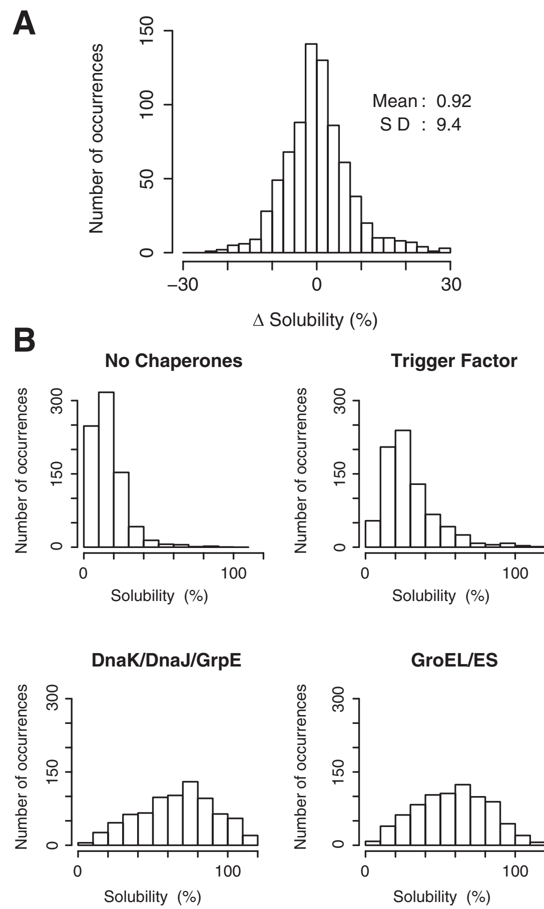

Fig. S2. Reproducibility and raw data of global aggregation analysis. (*A*) Reproducibility of the experimental data under the chaperone-free conditions. Comparison of the solubilities under the chaperone-free conditions obtained in this study with those from the previous study (1). The difference (Δ) solubility percentage was calculated by subtracting the solubility evaluated in the previous study from the solubility in the absence of chaperone evaluated in this study. The mean Δ value and the SD of this distribution were 0.92% and 9.4%, respectively. (*B*) Histograms of raw solubility data obtained under various chaperone conditions.

1. Niwa T, et al. (2009) Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of *Escherichia coli* proteins. *Proc Natl Acad Sci USA* 106: 4201–4206.

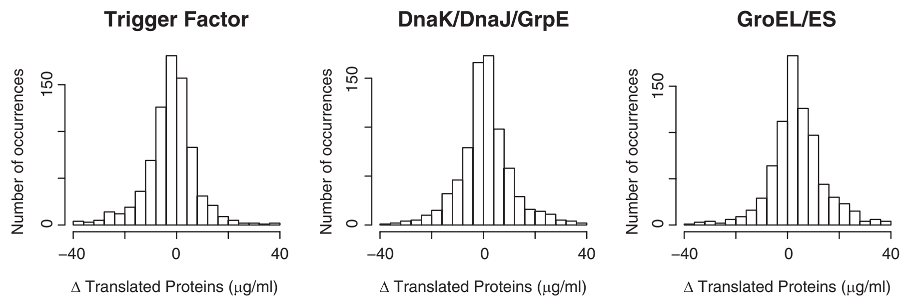

Fig. S3. Histograms of difference (Δ) synthesis yields. The difference in the synthesis yield (Δtranslated protein yield) was calculated by subtracting the yield in the absence of chaperone from that in the presence of chaperone.

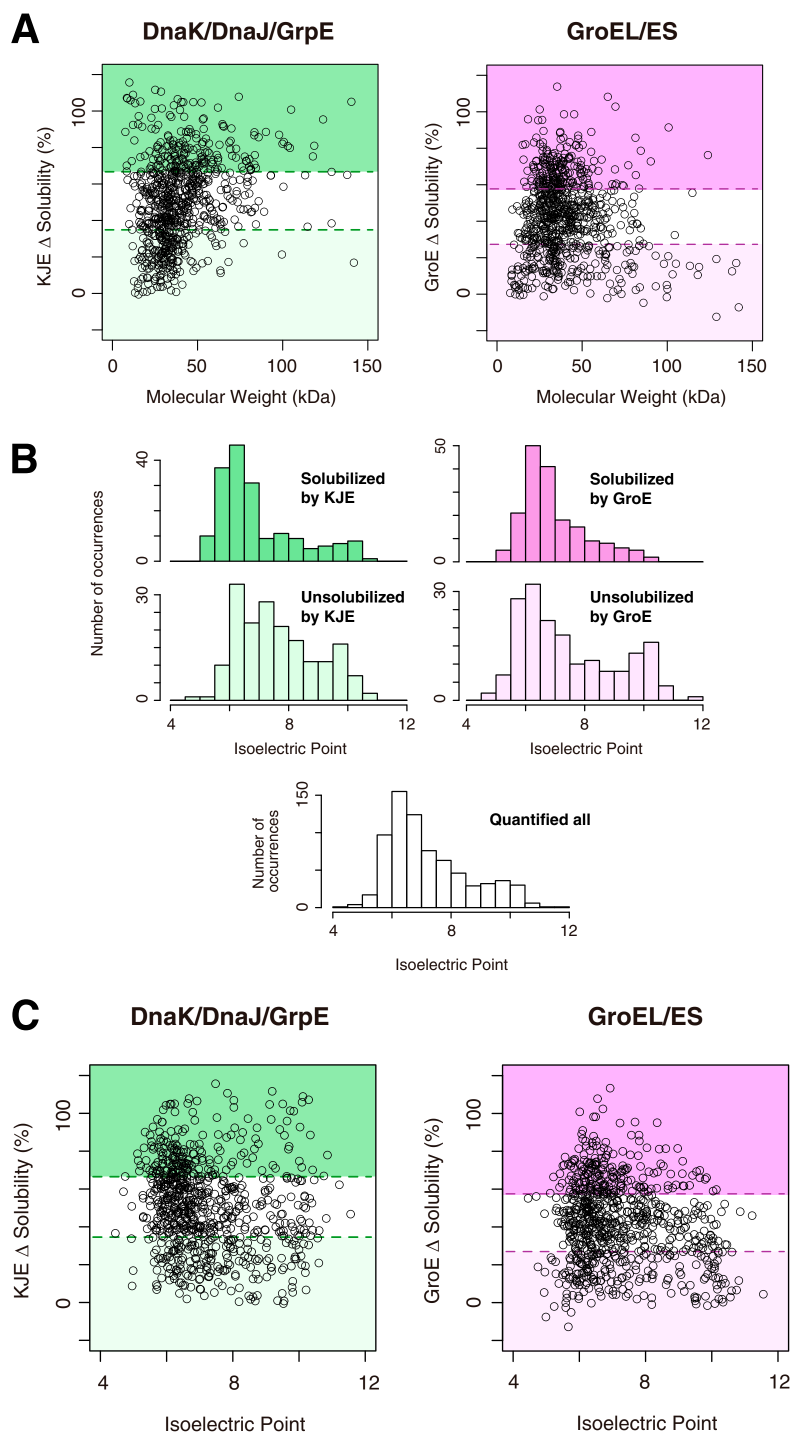

Fig. S4. Correlation between the solubility and the molecular weights or isoelectric point (pI). (*A*) Scatter plots of Δsolubility for DnaKJE (*Left*) and GroE (*Right*) versus the molecular weights. Shades indicate the upper (≥75th percentile, dark) and lower (≤25th percentile, light) quartiles for the Δsolubility in each distribution. (*B*) Histograms of pI. Only the upper and lower quartiles in *C* are aligned in the histograms. The histogram of pI for all quantified aggregation-prone proteins is presented at the bottom. (*C*) Scatter plots of Δsolubility for DnaKJE (*Left*) and GroE (*Right*) versus pI. pI were calculated from the deduced amino acid sequences. Shades indicate the upper (dark) and lower (light) quartiles of the Δsolubility in each distribution. The details of the calculation process were previously described (1). The comparison for TF was omitted because of the narrow distribution of Δsolubility.

1. Niwa T, et al. (2009) Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of *Escherichia coli* proteins. *Proc Natl Acad Sci USA* 106: 4201–4206.

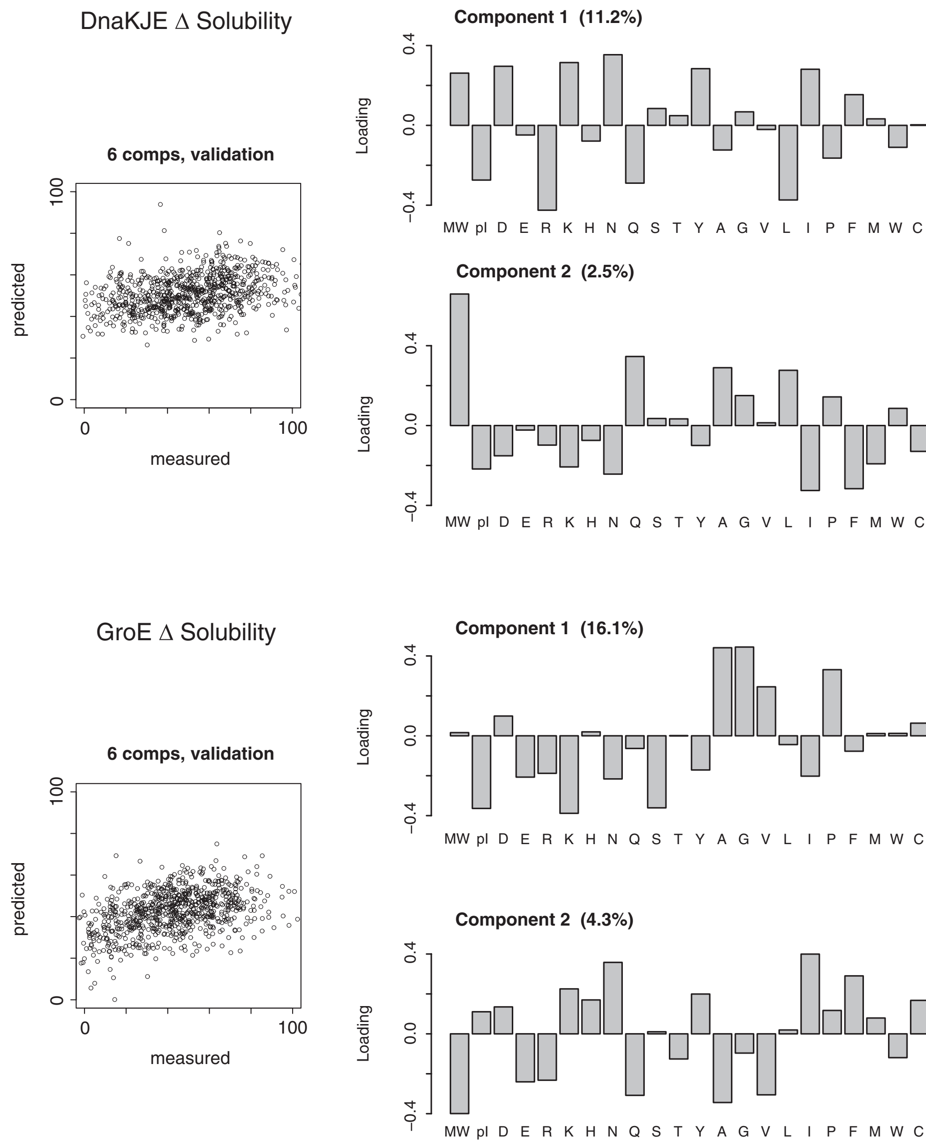

Fig. S5. Contribution of molecular weight, pI, and the amino acid contents to Δsolubilities estimated by the partial least-squares (PLS) regression analysis. (*Left*) The predicted-measured plot derived from 10-fold cross validation calculated with six latent variables. The RMSEP values were 20.0 for GroE Δsolubility and 23.1 for DnaKJE Δsolubility, and the percentages of variance explained up until the sixth components were ∼15% and ∼21% for the DnaKJE and GroE, respectively. (*Right*) The loading patterns from the PLS regression analysis. MW, molecular weight; and the other capital letters in the *x* axis are one-letter abbreviations of amino acids. Only two first components are shown. The number in parentheses represented the percentages of variance explained of each component. In the first component of GroE, alanine (A) and glycine (G), were significant, which is consistent with previous observation that the composition of the A/G residues was enriched in the obligate GroE substrates (1).

1. Fujiwara K, Ishihama Y, Nakahigashi K, Soga T, Taguchi H (2010) A systematic survey of in vivo obligate chaperonin-dependent substrates. *EMBO J* 29:1552–1564.

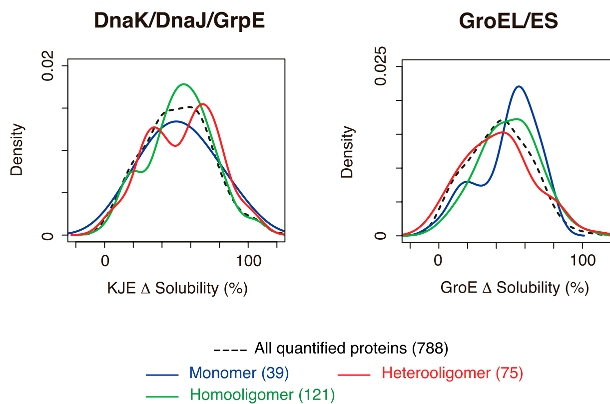

Fig. S6. Correlation between the solubility and the quaternary structure. A keyword search (monomer, heterooligomer, homooligomer) in the SUBUNIT entry in the Uniprot database was executed to select the proteins with quaternary structure annotations. The distribution of Δsolubility was drawn by the Kernel density estimation method. The number of proteins within each category is indicated in parentheses.

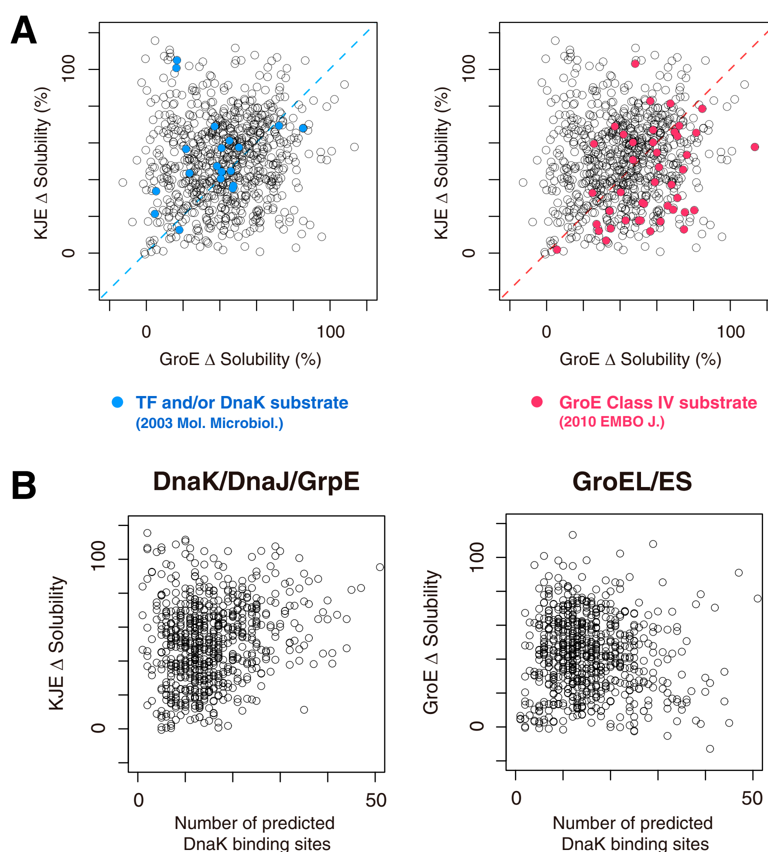

Fig. S7. Comparison between the solubility and known chaperone substrates. (*A*) (*Left*) The distribution of in vivo TF and DnaK substrates, denoted as blue circles, previously identified by Deuerling et al. (1). Among the 94 identified substrates, 19 substrates were examined in this work. (*Right*) The distribution of in vivo obligate GroE substrates, denoted as red circles, previously identified by Fujiwara et al. (2). Among the 57 substrates, 43 substrates were examined in this work. (*B*) Scatter plots of Δsolubility versus the number of predicted DnaK binding sites, calculated by the LIMBO algorithm (3). The calculation script was developed in-house (the parameters for the 20 amino acids were obtained from ref. 3). The threshold value for the peak determination was set at 8.26, which is the value for high sensitivity conditions. The comparison for TF was omitted because of the narrow distribution of Δsolubility.

1. Deuerling E, et al. (2003) Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. *Mol Microbiol* 47:1317–1328.
2. Fujiwara K, Ishihama Y, Nakahigashi K, Soga T, Taguchi H (2010) A systematic survey of in vivo obligate chaperonin-dependent substrates. *EMBO J* 29:1552–1564.
3. Van Durme J, et al. (2009) Accurate prediction of DnaK-peptide binding via homology modelling and experimental data. *PLOS Comput Biol* 5:e1000475.

Table S1. Number of solubilized/unsolubilized proteins by each chaperone

Table S1

Dataset S1. Raw solubilities and synthesis yields in the presence of chaperones

Dataset S1

Dataset S2. Solubilities and synthesis yields for "recalcitrant" proteins in the presence of chaperone combinations (Fig. 5)

Dataset S2

PNAS PNAS PNAS PNAS