SUPPLEMENTAL INFORMATION

Conserved and divergent chaperoning effects of Hsp60/10 chaperonins on protein folding landscapes

SUPPLEMENTAL METHODS

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Strains, Plasmids and Proteins

E. coli strain DH5α was used for cloning, WT E. coli K-12 (BW25113) strain was used for expression of arabinose inducible pBAD GFP and BL21 (DE3) was used for protein expression and purification. Protein concentrations were determined spectrophotometrically at 562 nm using BCA kit (Pierce-ThermoFisher Scientific). Deletion strains were obtained from CGSC as part of Keio collection (31).

METHOD DETAILS

E. coli GroEL/ES, Yeast and Human EL chimera Expression and Purification

GroEL/ES, GroEL chimeras were purified using E. coli BL21 (DE3) as described (7, 38, 39). GroEL/ES, cfl-EL/cfl-ES, smh-EL/smh-ES, and chimeras were expressed from pOFX series plasmids for co-expression studies (40).

Identification of inner-lining residues of Hsp60/10 cavities

To identify the inner-lining constituent amino acids of Hsp60/10 cavities, we used the method CICLOP (8). CICLOP (Characterization of Inner Cavity Lining Of Proteins) can accurately and speedily identify these regions at an atomic resolution that has been qualitatively and quantitatively benchmarked against leading methods in the field. Using Voronoi diagram, the method shows high fidelity and precision along with novel sturdiness to handle complex mega-structures.

Structural analysis using in silico methods

Modelling was performed using Modeler v9.17 (32) with the crystal structure of asymmetric chaperonin complex of GROEL/ES as the template (PDB: 1AON) (33). FASTA formatted protein sequences of the Hsp60/10 in all the bacterial species was downloaded from the Microbial Genome Database for Comparative Analysis (MBGD) (41) using in-house script. To model the structures, first the sequence of EL (Chains: ABCDEFGHIJKLMNOP) and ES (Chains: OPQRSTUVW) of the target organism was aligned to the corresponding template chain's sequence using Salign (42). The alignments of the individual chains were then assembled, and the missing template flanking region at the N-terminal was removed for modelling purposes. In-house developed scripts modelled 5 putative structures for the chain of G and H (EL) and Chain O (ES). The models were assessed based on the Discrete Optimized Protein Energy (DOPE) statistical potential (32) and ProQ2 (43) scores and the best models were chosen to build the final structure.

The quality and degree of completeness of all modelled structures was assessed using TMscore (11). TMscore performs sequence-dependent superposition of two protein structures. It assumes correspondence between residues with the same residue number. The higher the TM-score (can range from 0 to 1), the higher the degree of similarity between the considered structures. All the modelled structures were assessed with respect to 1AON (PDB ID -
https://www.rcsb.org/structure/1aon). The score of 0.68 was considered (Table S1) as the benchmark as anything above 0.5 was considered as a structurally similar protein model (44).

**Generation of Centroid Models:** The final structure is subdivided into two categories of models containing chains:

(I) -A-B-C-D-E-F-G- (EL) and -O-P-Q-R-S-T-U (ES) (CIS)

(II) -H-I-J-K-L-M-N (EL) (TRANS)

Model I contains the cis-part of the EL along with the ES cap. Model II contains only the trans-part of EL.

The centroid for these two models were then calculated using the pdb_centermass script of pdbtools (https://github.com/harmslab/pdbtools) and the coordinates were saved as C1 and C2 for the models I-II, respectively. The script calculates the center of mass of a protein and returns either the center of the model.

Electrostatic Calculation: We used the standalone version of PDB2PQR (45) to assign charge and radius parameters for CHARMM force field to our modeled Hsp60/10 structures (Model I and II, individually). The electrostatics was calculated using Adaptive Poisson-Boltzmann Solver (APBS) (35). APBS uses iterative solvers to solve the nonlinear algebraic equations resulting from the discretized Poisson-Boltzmann equation with a fixed error tolerance of $10^{-6}$. The values for the centroid coordinates were then extracted from its corresponding electrostatic potential file using themultivalue script (provided in APBS).

**Purification of Yeast HSP60 and Human HSP60**

Yeast Hsp60 and human HSP60 were purified from *E. coli* BL21 (DE3) transformed with pET Duet-1 yHSP60 or pET Duet-1 hHSP60 vector. 3L of LB medium containing Ampicillin (100 μg/ml) and 0.1% overnight grown inoculum, was grown until OD$_{600}$~0.4 at 37°C, 200 rpm and induced overnight with 0.5mM IPTG at 18°C. Cells were harvested at 4°C, 6000 rpm for 10 min and pellet was resuspended in 40 ml lysis buffer [25mM Tris-HCl pH 8.0, 1mM DTT, 5mM EDTA and complete protease inhibitor tablet (Roche)]. Cells were lysed by sonication followed by ultracentrifugation at 4°C, 18000 rpm for 60 min to separate supernatant from cell debris. Supernatant was applied to 15 ml ANX-Sepharose 4 fast flow column (GE Healthcare) equilibrated with 25mM Tris-HCl pH 8.0, 1mM DTT, 5mM EDTA. After washing with 5 column volumes of 25mM Tris-HCl pH 8.0, 1mM DTT, 5mM EDTA buffer, the protein was eluted using NaCl batch elution in 0.1 M to 0.5 M NaCl of 50 ml each. Protein specifically eluted in 0.4 M NaCl. Sample was dialyzed overnight to remove salt against 25mM Tris-HCl pH 8.0, 1mM DTT, 5mM EDTA buffer. Desalted yeast Hsp60 was concentrated to 20 ml using 30 kDa cut off amicon concentrators (Millipore) and applied to 15 ml DEAE-Sepharose column (GE Healthcare), equilibrated in 25mM Tris-HCl pH 8.0, 2mM DTT, 5mM EDTA. Proteins were eluted with NaCl batch elution from 0.1 to 0.5 M NaCl. Yeast Hsp60 specifically eluted in 0.1 M NaCl. 0.1 M fraction was collected and dialyzed against 25mM Tris-HCl pH 7.4, 150mM KCl, 2mM DTT, 10mM MgCl$_2$ and 10% Glycerol. The sample was concentrated by 30 kDa cut-off amicon concentrators (Millipore) around to 5 ml and applied to a 24 ml Superose-6 10/300 gel filtration column (GE Healthcare) equilibrated in 25mM Tris-HCl pH 7.4, 150mM KCl, 2mM DTT, 10mM MgCl$_2$ and 10% Glycerol, attached to an AKTA Explorer chromatography system. For each run maximum 500 μl sample was loaded. Yeast Hsp60 fractions were collected and concentrated and stored at -80°C.

**Purification of Human and Yeast Hsp10**

Yeast Hsp10 and human HSP10 were purified from *E. coli* BL21 (DE3) transformed with pET Duet-1 yHSP10 or pET Duet-1 hHSP10 vector. 2L of LB medium containing Ampicillin (100
μg/ml) was inoculated with 0.1% overnight grown inoculum, and was grown till OD600~0.4 at 37°C, 200 rpm. This was induced with 0.5mM IPTG at 18°C overnight. Cells were harvested at 4°C, 6000 rpm for 10 minutes and pellet was resuspended in 40 ml lysis buffer (25mM Tris-HCl pH 7.4, 150mM KCl, 2mM DTT, 10mM MgCl₂, 10mM Imidazole, 10% Glycerol and Complete protease inhibitor tablet (Roche). Cells were lysed by sonication followed by ultracentrifugation at 4°C, 18000 rpm for 60 minutes to separate supernatant from cell debris. The supernatant was bound to 2 ml Ni-NTA resin (Sigma-Aldrich) pre-equilibrated with equilibration buffer A (25mM Tris-HCl pH 7.4, 150mM KCl, 2mM DTT, 10mM MgCl₂, 5mM Imidazole, 10% Glycerol). Column was washed first with 20mM imidazole and then 30mM imidazole in buffer A (15 column volumes) and the proteins were eluted with 250mM imidazole in buffer A in 10 elutions, 1 ml each by gravity flow. Fractions containing purified proteins were concentrated to 2 ml using 3 kDa cutoff amicon concentrators (Millipore) and loaded on to 24 ml Superdex- 200 10/300 gel filtration column (GE Healthcare) equilibrated in 25mM Tris-HCl pH 7.4, 150mM KCl, 2mM DTT, 10mM MgCl₂ and 10% Glycerol, attached to an AKTA Explorer chromatography system. For each run, a maximum of 500ul sample was loaded. Protein fractions were collected, concentrated and stored at -80°C.

**Solubility of sGFP and other GroEL substrates in vivo**

WT (either WT for sGFP and DM-MBP and WT BL21 DE3 cells for MetK solubility) E. coli cells containing pBAD sGFP or pBAD DM-MBP or pET Duet-1 MetK were transformed with pOFX GroEL/ES. 0.1% inoculation was done in 10ml LB medium added with Ampicillin (100μg/ml) or Ampicillin and Chloramphenicol (35 μg/ml) from overnight grown cultures and grown till OD₆₀₀~0.5 at 37°C, 200rpm. 0.5mM IPTG was added to cells with pOFX GroEL/ES 30 minutes before induction of sGFP or DM-MBP with 0.1% arabinose for 3 hours. The cell type in which we transformed pBAD sGFP only were directly induced with 0.1% arabinose for 3 hours after reaching OD₆₀₀~0.5. Cells were harvested at 4000 rpm for 10 minutes and resuspended in 1 ml PBS pH 7.4, 2mM DTT. Lysis was done using sonication followed by high speed centrifugation to separate soluble and pellet fraction which were loaded onto 12% SDS-PAGE. Gel visualized by Coomassie staining.

**Spontaneous and chaperonin assisted in vitro refolding of sGFP**

Wt GFP and sGFP (20μM each) were denatured in buffer containing 6 M GuHCl in buffer A (20mM Tris, 20mM KCl, 5mM MgCl₂, 2mM DTT, pH 7.4) for 1 hour at 25°C and refolded upon 100 fold dilution into buffer A. Either of three refolding buffers were used for refolding, buffer A (20mM Tris, 20mM KCl, 5mM MgCl₂, 2mM DTT, pH 7.4) to mimic reducing conditions, buffer B (20mM Tris, 20mM KCl, 5mM MgCl₂, pH 7.4) to mimic non-reducing conditions and buffer C (20mM Tris, 20mM KCl, 5mM MgCl₂, 2μM CuCl₂, pH 7.4) to mimic oxidizing conditions.

GroEL/ES assisted refolding was done in the presence of (400nM) GroEL (tetradecamer), (800nM) GroES (heptamer), (substrate:GroEL:ES :: 1:2:4) and the refolding was started by addition of 2mM ATP. SR-EL/ES assisted refolding was done in presence of (800nM) of SR-EL (heptamer) (800nM) GroES (heptamer) and the refolding was started by addition of 2mM ATP. GFP fluorescence at 480 nm excitation (slit width 2 nm) and 515 nm emission (10 nm slit width) was monitored as a read out of refolding using Fluorolog 3 Spectrofluorometer (Horiba). Buffer conditions are described in the figures. All the unfolding and refolding experiments were carried out at 25°C unless specified (16).

**ATPase assay**

GroEL and its chimera y-EL and h-EL were added in a reaction mixture containing 2mM PEP (Phosphoenolpyruvate), 20/30 units/ml PK/LDH (Pyruvate Kinase/Lactic Dehydrogenase)
enzyme, 1mM NADH, 1mM ATP (and ES in EL/ES ATPase assay) (34), incubated at 25°C for 3 minutes. This removed any ADP by pyruvate kinase. The reaction was started by addition of GroEL or its chimera (600nM). Upon addition of GroEL, ATP is hydrolyzed to ADP. Hydrolysis of ATP to ADP is linked to the oxidation of NADH to NAD⁺ by the combined action of Pyruvate kinase and L-lactate dehydrogenase. The absorbance of NADH is measured at 360 nm. As soon as the kinetics start by addition of GroEL (or chimera), the ATP is hydrolyzed to ADP and NADH is oxidized to NAD⁺ which leads to decrease in absorbance at 360 nm. In the reaction where ES was present, ES was added in the reaction mixture and EL was added last to start the kinetics.

**Analysis of temperature dependent refolding using Arrhenius equation**

As described in Sadat et.al.2020 (16)

In order to obtain thermodynamic parameters that define the barrier between the refolding intermediate I₁ and the transition state (TS) of folding we used the following equation essentially as defined in (4).

\[
k_f = \rho T e^{-\frac{\Delta H^\# + \Delta C_p^\#}{R} \left( T - T_0 - T \ln \left( \frac{T}{T_0} \right) \right)}
\]

Where,

\[
\rho = \left( \frac{k_B}{R} \right) \left( k \cdot e^{\frac{\Delta S^\#}{R}} \right)
\]

and \( k_f \) is refolding rate, \( R \) is the universal gas constant, \( \Delta H^\# \), \( \Delta S^\# \) and \( \Delta C_p^\# \) are the differences in enthalpy, entropy and heat capacity at constant pressure between TS and I₁, respectively, \( T \) is the temperature of refolding reaction, \( T_0 \) is the reference temperature at which the parameters are determined (here it is 298.15). \( \kappa \) is the transmission factor, that reports the proportion of activations that lead to the formation of the native state (N), and \( k_B \) and \( h \) are Boltzmann and Planck constant respectively. Since \( \Delta S^\# \) and \( A \) linearly combine, it is not possible to obtain independent estimates of these two parameters by non-linear regression. Hence, we combine it to obtain \( \rho \). This term indicates the ease of barrier crossing, either because the diffusion is faster or because the barrier is less broad (due to lower \( \Delta S^\# \)). A high \( \Delta H^\# \) indicates a higher barrier, and a low \( \rho \) indicates an entropic (diffusion-limited) barrier. The equations were fitted using standard nonlinear regression fitting using R or MATLAB.

**Spot assay**

MGM100 (GroEL depletion strains) was transformed with different GroEL/ES and chimeras. Single colonies were picked and grown till OD₆₀₀ ~1. Five µl of logarithmically diluted culture was spotted on LB Agar plates containing 100 µg ampicillin and 0.2% glucose and visualized post 12 hours incubation at 37°C.

**Pooled GFP mutant library transformation**

Plasmid constructs of GroEL/ES, cfl-EL/ES, smh-EL/ES, yHsp60/10 and hHsp60/10 in pOFX vector were transformed to chemical competent *E. coli* Wtk cells. These strains were co-transformed with a plasmid library of low fluorescent GFP mutants in pBAD vector. For transforming the library of plasmids, 5*100 µL aliquots of chemical competent cells were taken. 20 ng of plasmid library was added to each aliquot which was then incubated on ice for 15 min. Cells were subjected to heat shock at 42°C for 60 seconds and immediately transferred on ice for 5 mins. 900 µL of recovery media was added to each and the cells were allowed to
recover at 37°C, 200 rpm for 45 mins. Then, all 5 fractions of recovered cells were pooled and inoculated to 100 μL of LB media containing Carbenicillin (100 μg/mL). 1/100th of the cells were plated on an LB agar plate containing Carbenicillin (100 μg/mL) for estimating the total number of transformants. The cells were allowed to grow overnight at 37°C, 200 rpm and then glycerol stocks were prepared from the entire pool of cells.

**Sorting of differently fluorescent population from the library**

wtk cells co-expressing low fluorescent GFP mutant library in the background of different chaperonins were grown overnight at 37°C, 200rpm. From this primary culture, secondary cultures were inoculated in 10ml LB containing Ampicillin (100μg/ml) and Chloramphenicol (35μg/ml) and grown till O.D₆₀₀ 0.4-0.6. Cultures were first induced with 1mM IPTG to overexpress chaperonins and then with 0.2% Arabinose after 45 mins to induce GFP expression. Cells were sorted on BD FACSaria III 3h post induction with Arabinose at 37°C after diluting with 1X PBS and incubating at 37°C for additional 1 hr. The library from each sample was sorted into two populations, the major peak population (PF) and the top 5% fluorescent population (HF). Sorted populations were then recovered in 10ml LB containing Ampicillin (100μg/ml) and Chloramphenicol (35μg/ml) and grown for 16hrs at 37°C ,200 rpm. From this overnight grown culture, the plasmid pool was isolated using an RBC mini prep kit.

**GFP amplification for Sequencing**

PCRs were performed for amplification of GFP using upstream and downstream primers. An upstream forward primer (5' CGGCAGAAAGTCCACATTGATTATTTCAC 3') and a downstream reverse primer (5' CATGTTATCCTCCTCGCCCTTGCTCACCAT 3') were used. The reaction mixture contained 50ng of template DNA, 1X buffer containing MgCl₂, 250 μM dNTPs, 25 pmol each of forward and reverse primers and 1U of Q5 polymerase. For amplification, the reaction comprised of initial denaturation at 95°C for 5 minutes, 40 cycles of denaturation at 98°C for 30 seconds, annealing at 58°C for 30 seconds and elongation at 72°C for 40 seconds. A final extension was done at 72°C for 10 minutes and the reaction was held at 4°C before processing further.

**High Throughput Sequencing**

The library was prepared according to Illumina (inc.) protocol using Illumina Nextera XT Library preparation kit for PCR amplified 24 samples after doing PCR-Cleanup using AmPure beads (BD). PCR amplified DNA was then processed for tagmentation and adapter ligation using Illumina Nextera XT Index kit v2 Set D. Samples were quantified using Qubit 2.0 fluorometer (Invitrogen Inc.) and pooled accordingly. Further dilution and denaturation of the library was performed as per protocols provided by the manufacturer (Illumina Inc). The final library was normalised to a concentration of 1.2pm and loaded onto the flow cell. The run was set up on Illumina Miniseq system using High Output Cartridge (2x150 cycles).

**NGS data and structural analysis of mutants**

The reads were mapped onto GFP sequence using BWA (Burrows-Wheeler Aligner) (46) after quality check by FastQC. The mapped reads were processed using SAMtools (47) which were then identified for amino acid changes and counts using in-house scripts in python. The Esc score and BI scores were calculated for each of the mutants for each condition (chaperonin overexpression conditions).

Esc scores were calculated for each mutant using the following equation

\[
\text{Esc}_{ij} = \frac{\text{Read count in } HF_{ij}}{\text{Read count in } MF_{ij}}
\]
Where, $\text{Esc}_{ij}$ is the enrichment score of mutant $i$ in condition $j$. Each case of different chaperonin overexpression is a distinct $j$.

$$\text{Esc}_{ij} = \text{Enrichment score of mutant } i \text{ with chaperonin } j$$

$\text{Read count in HF}_{ij} = \text{Read count of mutant } i \text{ in HF population with chaperonin } j$

$\text{Read count in MF}_{ij} = \text{Read count of mutant } i \text{ in MF population with chaperonin } j$

The BI of the mutants for each chaperone was calculated using the following equation

$$\text{BI}_{ij} = \log_2 \frac{\text{Esc}_{ij}}{\text{Esc}_{i0}}$$

$\text{Esc}_{ij} = \text{Enrichment score of mutant } i \text{ with chaperonin } j$

$\text{Esc}_{i0} = \text{Enrichment score of mutant } i \text{ with no overexpressed chaperonins}$

$\text{BI}_{ij} = \text{Buffering index of mutant } i \text{ in presence of overexpressed chaperonin } j$

The BI values were used for clustering the chaperonins and the mutants using the scipy cluster package. ANOVA was performed using the seaborn package. $\Delta \Delta G$ values were calculated using the software PremPS (37). Number of contact sites for a residue were calculated by identifying the number of residues within 5Å of the side chain atoms of the residue under consideration. Surface accessibility of residues were obtained from WHATIF web server (https://swift.cmbi.umcn.nl/servers/html/index.html).

**QUANTIFICATION AND STATISTICAL ANALYSIS**

Student t test and R package for non-linear regression was used for statistical analysis. Flow-cytometry data was analyzed using MATLAB using in-house scripts.
| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---------------------|--------|------------|
| **Antibodies**      |        |            |
| Rabbit anti GFP    | Abcam  | AB290      |
| **Bacterial and Virus Strains** | | |
| *E. coli* BL21 (DE3) | | |
| *E. coli* DH5α      | | |
| *E. coli* MGM100    | | |
| *E. coli* WT (K-12, BW25113) | | 💼 CGSC#: 7636 |
| F-, DE(araD-araB)567, lacZ4787(del)::rrnB-3, LAM-, rph-1, DE(rhaD-rhaB)568, hsdR514 | | (31) |
| **Chemicals, Peptides, and Recombinant Proteins** | | |
| NADH                | Sigma  | Cat#N8129  |
| ATP                 | Sigma  | Cat#A2383  |
| AMP-PNP             | Sigma  | Cat#A2647  |
| PK/LDH              | Sigma  | Cat#P0294  |
| PEP-K               | Sigma  | Cat#860077 |
| Sodium succinate hexahydrate dibasic | SRL Chemical | Cat#6106-21-4 |
| **Oligonucleotides** | | |
| Yeast Hsp60 forward primer: AAACCATGGATGGAATTGAAATTCCGCTTGAAGAAGGAAGGAA GAGCC | | This Paper |
| Yeast Hsp60 reverse primer: CCAAGCCTTTTCATCATACCTGCCATTCCTGGCATACACACCTG | | This Paper |
| Yeast Hsp10 forward primer: AAACCATGGATGTCACCCCCCTTTGAAGTCTGCTAAATC TATC | | This Paper |
| Yeast Hsp10 reverse primer: CCCCAGCTTTTAGCTCTTG GCAATCTTAGCCAGGATTT CAGCG | This Paper |
|---------------------------------------------------------------------|-----------|
| Human Hsp60 forward primer: AAACCATGGATGTATGCCA AAGATGTTAAAATTTGCTGC AGATGCC | This Paper |
| Human Hsp60 reverse primer: CCCCAGCTTTTAGAAACATG CCACCTCCATACCACCTC CCATTCC | This Paper |
| Human Hsp10 forward primer: AAACCATGGATGGCAGGA CAAGCGTTTAGAAAGTTTC TTCC | This Paper |
| Human Hsp10 reverse primer: CCCCAGCTTTTAGAACATG CCACCTCCATACCACCTC CCATTCC | This Paper |
| Human Hsp10 reverse primer: CCCCAGCTTTTAGAACATG CCACCTCCATACCACCTC CCATTCC | This Paper |
| GFP upstream forward primer: CGGCAGAAAAGTCCACATT GATTATTTGCAC | This Paper |
| GFP downstream reverse primer: CATGTTATCCTCCTCGCCC TTGCTCACCAT | This Paper |
| Recombinant DNA | |
| pET duet1 hHSP60 | This Paper |
| pET duet1 yHSP60 | This Paper |
| pET duet1 yHSP10 | This Paper |
| pET duet1 hHSP10 | This Paper |
| pET SUMO sGFP | This Paper |
| pET duet MetK | This Paper |
| pOFX y-EL/y-ES | This Paper |
| pOFX h-EL/h-ES | This Paper |
|----------------|------------|
| pOFX cfIEL/cfIES | This Paper |
| pOFX smhEL/smhES | This Paper |

**Software and Algorithms**

| Chimera | |
|---------|---|
| ABPS    | |
References (Supplemental)

38. Mapa K, Tiwari S, Kumar V, Jayaraj GG, & Maiti S (2012) Information encoded in non-native states drives substrate-chaperone pairing. *Structure* 20(9):1562-1573.

39. Tiwari S, Kumar V, Jayaraj GG, Maiti S, & Mapa K (2013) Unique structural modulation of a non-native substrate by cochaperone DnaJ. *Biochemistry* 52(6):1011-1018.

40. Castanié M-P, Bergès H, Oreglia J, Prere M-F, & Fayet O (1997) A Set of pBR322-Compatible Plasmids Allowing the Testing of Chaperone-Assisted Folding of Proteins Overexpressed in Escherichia coli. *Analytical biochemistry* 254(1):150-152.

41. Uchiyama I, Mihara M, Nishide H, & Chiba H (2015) MBGD update 2015: microbial genome database for flexible ortholog analysis utilizing a diverse set of genomic data. *Nucleic acids research* 43(Database issue):D270-276.

42. Braberg H, et al. (2012) SALIGN: a web server for alignment of multiple protein sequences and structures. *Bioinformatics* 28(15):2072-2073.

43. Leitner A, et al. (2012) The molecular architecture of the eukaryotic chaperonin TRiC/CCT. *Structure* 20(5):814-825.

44. Xu J & Zhang Y (2010) How significant is a protein structure similarity with TM-score= 0.5? *Bioinformatics* 26(7):889-895.

45. Jurrus E, et al. (2018) Improvements to the APBS biomolecular solvation software suite. *Protein science: a publication of the Protein Society* 27(1):112-128.

46. Li H & Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *bioinformatics* 25(14):1754-1760.

47. Li H, et al. (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16):2078-2079.
Supplemental Table: (Table S1 and S3 are provided as separate excel files).

Table S2: Eyring-Arrhenius Parameters for sGFP refolding under different conditions.

$\Delta H^\#$ denotes the activation enthalpy change at 25°C, $\Delta C_p^\#$ is the activation associated change in heat capacity at constant pressure, and $\rho^\#$ is the change in entropy and transition frequency associated with activation at 25°C. a These are parameters reported earlier for spontaneous refolding of sGFP under identical conditions.

| Chaperonin | $\Delta H^\#$ (kCal/mol) | $\Delta H^\#$ (error) | $\Delta C_p^\#$ (kCal/mol/K) | $\Delta C_p^\#$ (error) | $\rho^\#$ (kCal/mol/K) | $\rho^\#$ (error) |
|------------|--------------------------|------------------------|-------------------------------|-------------------------|------------------------|-------------------|
| yHsp60/10  | 3.7                      | 0.54                   | 0.011                         | 0.179                   | 5.14E-03               | 4.68E-03          |
| hHsp60/10  | 0.37                     | 1.05                   | -0.422                        | 0.345                   | 1.44E-05               | 2.53E-05          |
| y-EL/ES    | 5.33                     | 0.27                   | -0.049                        | 0.085                   | 0.098                  | 0.044             |
| h-EL/ES    | 4.56                     | 0.26                   | 0.131                         | 0.085                   | 0.029                  | 0.013             |
| Cfl-EL/ES  | 9.44                     | 0.51                   | -0.595                        | 0.14                    | 225                    | 191               |
| Smh-EL/ES  | 6.43                     | 0.32                   | 0.039                         | 0.01                    | 0.679                  | 0.363             |
| Spont a    | -0.41                    | 1.42                   | -1.70                         | 0.45                    | 5.89E-06               | 1.40E-05          |
| GroEL/ES   | 6.99                     | .6220                  | 0.146                         | 0.19                    | 3.23                   | 3.3               |
Supplemental Figure Legends

Figure S1 Amino acid abundances in the central cavity of GroEL/ES homologs and the sequence of engineered chimera

(A) Abundance of each of the amino acids (labelled in the inset) in the cis (blue) and trans (orange) cavity of GroEL/ES homologs. The p-value shown for each amino acid (in the inset) is for a paired t-test between the cis and trans cavity in all the organisms combined. The data are shown as violin plots separately for a subset of chosen bacteria (except thermophiles, criteria for choosing the species is in the text), thermophilic bacteria, representative eukaryotes. E. coli GroEL/ES parameters are shown as the fourth column line for comparison. In red are marked the residues that are significantly more enriched in the cis-cavity than in the trans-cavity.

(B & C) Multiple sequence alignment of GroEL with either hHsp60 and the chimeric h-EL (B) or with yHsp60 and the chimeric y-EL (C) to show the regions that have been grafted (effective regions marked in red boxes).

Figure S2 Differences between GroEL/ES homologs

(A) SDS PAGE gel showing the induction level of different Hsp60/10 homologs or the chimeric constructs.

(B) Gel filtration of E. coli GroEL, y-EL, h-EL were done using the Superdex 200 16/600 gel filtration column (GE Healthcare). The elution peaks were compared with the standard elution profile of the column used and represent the molecular weights of proteins eluting in that elution volume.

(C) Native PAGE showing formation of tetradecameric complexes by purified y-EL, GroEL and h-EL.

(D) ATPase activity assay of E. coli GroEL, h-EL, and y-EL chimera. ATPase activity is shown as decrease in ATPase activity upon GroES addition and compared with control E. coli GroEL.

(E) Schematic of the MGM100 strain. GroEL/ES operon is repressed in the presence of glucose and prevents growth. Strains grow only upon functional complementation with a suitable chaperonin system expressed from the plasmid.

(F) Spot assay of MGM100 strain in presence of repressor (glucose) or inducer (arabinose).

(G) Spot assay of MGM100 strain containing different overexpressing chimeras (cfl-EL/ES, E. coli GroEL/ES, yHsp60/10, smh-EL/ES, hHsp60/10, y-EL/ES, h-EL/ES). All the MGM100 strains containing overexpressing EL were serially diluted by a factor of 10.

Figure S3 Chaperonin having higher negative charge increases the solubility of its substrates

(A) Bar graph showing MetK solubility in the presence of E. coli GroEL/ES, smh-EL/ES, cfl-EL/ES, h-EL/ES, y-EL/ES (top). Representative Solubility assay SDS PAGE gel (bottom).

(B) Refolding of sGFP in the presence of cfl-EL, GroEL, yHsp60/10, smh-EL, hHsp60, y-EL, h-EL or in the absence of any chaperonin (spontaneous).

(C) Bar graph showing anisotropy of refolded sGFP alone or in presence of different GroEL and GroES-ADP-AIF₆. All the experiments were performed in biological triplicates, the error bars shown are standard deviation of the measurements. P-values shown above the bars are
obtained by comparing chaperonin assisted refolding and spontaneous refolding using Student’s unpaired t-test.

(D) Correlation between the in vitro refolding rate of sGFP with and without chaperonin complexes with in vivo activity of sGFP (GFP/mCherry ratio) with or without overexpressed chaperonin complexes.

(E) Apparent refolding rate of spontaneously refolding of sGFP or chaperonin assisted refolding of sGFP in the presence of different concentrations of succinate.

(F) Eyring-Arrhenius fits for temperature dependent refolding rate of sGFP in the presence of the different chaperonins. Black circles represent the experimentally observed rates, black line, the fitted line blue line, the 95% confidence interval and red line, the 95% prediction interval obtained from non-linear regression.

Figure S4 Chaperonin with high negative charge in cavity entropically favors folded state

(A) Arrhenius plot of sGFP refolding in the presence of monomeric GroEL, GroES and ATP. Error bars are standard deviation from three different experiments. Solid line shows the fitted curve.

(B) Mutation matrix from deep-sequencing of pooled GFP-mutant library. The heatmap shows the read-counts, as percentage of total counts, from 0.05 to 10%. X-axis is the residue number of yeGFP, and Y-axis are the different amino acids detected at each of these positions. The count of only missense mutations are shown.

(C) Histogram of GFP/mCherry fluorescence of the pooled GFP-mutant library in the presence (blue lines) and absence (pink shaded) of chimeric chaperonins. The control histogram (pink shaded) is the same in both the panels and in Figure 5A.

(D) Relative expression levels of the different chaperonins when expressed in the background of the pooled GFP-mutant library.

(E) SDS-PAGE showing expression of GroEL/ES (left gel) and hHsp60/10 (right gel) after inducing expression of the chaperonins with different concentrations of IPTG. The expression of the chaperonins are checked in cells expressing the mutant GFP library. The same cells are used to check the GFP/mCherry fluorescence in the panel below.

(F) Median GFP/mCherry fluorescence of the mutant GFP library in the absence (control) or presence of GroEL/ES or hHsp60/10 induced with different concentrations of IPTG. Orange lines show the comparison and p-value between conditions that show equivalent expression of GroEL/ES and hHsp60/10 (expression comparison in the previous panel). Cyan line (in this and the previous panel) shows a more conservative comparison under conditions where hHsp60/10 expression is higher than GroEL/ES. 3 biological experiments were performed, and the bars show the average of these measurements with their standard deviation. P-values are calculated using Student’s unpaired 2-tailed t-test.

(G) Correlation between read counts of the different GFP mutants in control cells and in cells overexpressing the different chaperonins. Blue dots mark the mutants that are differentially enriched in the chaperonin expressing cells.

(H) GFP/mCherry of GFP(V12L) in cells overexpressing chimeric chaperonins. Dashed line shows the fluorescence in control cells not overexpressing any chaperonins.

(I) GFP/mCherry of GFP(V12L) in cells overexpressing hHsp60/10, DnaK/J/E or in control cells not overexpressing any chaperonins.
(J) The median BI of mutants that has a particular endogenous amino acid (provided as AMINOACID in the inset) substituted by other residues, in the presence of different chaperonins. In blue shade is shown the 99% confidence interval. p-value shown in the inset is calculated using ANOVA. Number of mutants in each of the sets is provided as N in the inset.
**FIGURE S2**

A. Gel elution profile of GroEL, yEL, and hEL

B. Gel elution profile of GroEL, yEL, and hEL

C. 14 mer

D. % Reduction in ATPase by GroES

E. Glucose

F. LB+Glucose (0.2%)

G. Exogenous expression and Growth Defect

- MGM100
- MGM100 δGroEL/ES
- MGM100
- MGM100 δGroEL/ES
- MGM100
- MGM100 δGroEL/ES
- MGM100
- MGM100 δGroEL/ES
- MGM100
- MGM100 δGroEL/ES
- MGM100
- MGM100 δGroEL/ES
- MGM100
- MGM100 δGroEL/ES
- MGM100
- MGM100 δGroEL/ES
- MGM100
- MGM100 δGroEL/ES
- MGM100
- MGM100 δGroEL/ES
- MGM100
