PNAS

Supplementary Information for

A Proteome-Wide Map of Chaperone-Assisted Protein Refolding in a Cytosol-like Milieu

Philip To,[1] Yingzi Xia,[1] Sea On Lee[1], Taylor Devlin,[2] Karen G. Fleming[2], Stephen D. Fried[1,2]*

[1] Department of Chemistry, Johns Hopkins University, Baltimore, MD, USA

[2] Thomas C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, MD, USA

Stephen D. Fried

Email: sdfried@jhu.edu

This PDF file includes:

Supplementary Methods

1. Preparation and Characterization of Cyto-Serum

1a. Culturing of *E.* coli (Strain K12) for Making Cyto-Serum. 2 × 1 L of TB Broth (Cold Spring Harbor Recipe) were inoculated with *E. coli* (strain K12) cells in 2 L baffled flask from overnight saturated cultures and grown at 37 ˚C with agitation (220 rpm) to a final OD_{600} of 2.0. Cells were harvested by centrifugation at 2000 *g* for 15 min at 4 ˚C. Supernatants were removed and cells were washed with 200 mL MPW (Millipore Water) before centrifuging at 4000 *g* for 15 min at 4 ˚C. Washed pellets were stored at -20˚C until further use.

1b. Preparation of Cyto-Serum. Washed *E. coli* pellets were resuspended in MPW to a final volume of 15 mL and lysed by sonication (QSonica Q700R) using a ¼ inch probe for 30 mins on time (pulse: 5 sec on / 5 sec off) at 55% amplitude in a water-ice bath. Lysed cells were transferred to a JL20 centrifugation tube and clarified at 16000 *g* for 15 min at 4˚C. To deplete large macromolecules, cellular lysates were transferred to 5 mL ultra-clear ultracentrifugation tubes and ultracentrifuged at 40000 rpm for 20 h at 4˚C without sucrose cushion in a SW-55 Ti rotor. Supernatants were then carefully pipetted from the ultracentrifugation tubes to not disturb the pellet and transferred to 2K MWCO Vivaspin 15R centrifugal filters (Sartorius). The filters were spun in a swing bucket rotor (Eppendorf 5910R) centrifuge at 3000 *g* for 3 h at 4˚C to remove remaining macromolecules. Standard Tris-Glycine SDS-PAGE (6% stacking, 12% resolving) was used to confirm the removal of all protein molecules that would be stained by Coomassie. The resulting filtrate was concentrated using a Vacufuge Plus (Eppendorf) to a final volume of 1.92 mL, the total volume enclosed within the collective cytoplasms of the original *E. coli* population, creating cyto-serum (2 L of cells at OD_{600} 2.0 comprises of 3.2e12 cells, each cell having 0.6 fL of cytoplasm resulting in 1.92 mL of collected cyto-serum). The cyto-serum is aliquoted, flash frozen in liquid nitrogen and stored at -80˚C until further use. The cyto-serum as prepared above is used effectively as a 1.195× stock.

1c. UV-Vis. Ultraviolet–visible absorption spectra of cyto-serum were obtained with a NanoDrop One (Thermo Scientific) spectrophotometer. The cyto-serum was diluted 60× with MPW prior to analysis and MPW was used as the background.

1d. pH and Viscosity. The pH of cyto-serum was measured for three independent preparations of cyto-serum using a pH meter (Mettler Toledo), using a three-point calibration at pH 4.0, 7.0, and 10.0. Additional pH measurements were obtained for clarified lysates extracted into cyto-serum and cyto-serum supplemented with 4 μM GroEL and 8 μM GroES (final concentrations, from 100 μM stocks). The viscosity of cyto-serum was measured via a MRC92 rheometer (Anton-

Parr) using 25 mm parallel plates. 25 data points were collected with a point duration of 20 s with a Shear Rate ranging from .01 to 100 1/s logarithmic.

2. Limited Proteolysis Mass Spectrometry (LiP-MS) Refolding Studies

2a. Culturing of K12 for Limited Proteolysis Mass Spectrometry (LiP-MS) Refolding Studies. K12 cells were grown in 2 sets of 3 × 50 mL (biological triplicates) of in-house prepared MOPS EZ rich media(-Arginine/-Lysine) from saturated overnight cultures with a starting OD_{600} of 0.05. One set was supplemented with 0.5 mM [$^{13}C_6$]L-Arginine and 0.4 mM [$^{13}C_6$]L-Lysine and the other with 0.5 mM L-Arginine and 0.4 mM L-Lysine. Cells were cultured at 37˚C with agitation (220 rpm) to a final OD_{600} of 0.8. Each heavy/light pair was pooled together and then transferred to 2 × 50 mL falcon tubes and collected by centrifugation at 4000 *g* for 15 mins at 4˚C. The supernatants were removed, and cell pellets were stored at -20˚C until further use. Further preparation of cyto-serum, aggregation study samples, and refolding experiment samples are described in method details below.

2b. Preparation of Normalized Lysates. For proteome-wide refolding studies, frozen cell pellets were resuspended in 900 µL of cyto-serum lysis buffer (1× cyto-serum, supplemented with DNase I to a f.c of 0.1 mg mL^{-1}). Resuspended cells were flash frozen by slow drip over liquid nitrogen and cryogenically pulverized with a freezer mill (SPEX Sample Prep) over 8 cycles consisting of 1 min of grinding (9 Hz), and 1 min of cooling. Pulverized lysates were transferred to 50 mL centrifuge tubes and thawed at room temperature for 20 min. Lysates were then transferred to fresh 1.5 mL microfuge tubes and clarified at 16000 *g* for 15 min at 4 °C to remove insoluble cell debris. To deplete ribosome particles, clarified lysates were transferred to 3 mL *konical* tubes and ultracentrifuged at 33,300 rpm at 4°C for 90 min without sucrose cushions using a SW55 Ti rotor. Protein concentrations of clarified lysates were determined using the bicinchoninic acid assay (Rapid Gold BCA Assay, Pierce) in a microtiter format with a plate reader (Molecular Devices iD3) using BSA as a calibration standard. Due to the reducing nature of cyto-serum, the BCA assay is incompatible with it. Hence, to determine protein concentrations in lysates prepared in cyto-serum, cell pellets would be generated from the same original liquid culture but split into two equally sized aliquots. The aliquots were resuspended in equal volumes of lysis buffer, with one of the aliquots lysed in Tris native buffer. The two parallel lysates are simultaneously clarified and ultracentrifuged together. Hence, under these conditions the protein concentration in the Tris-lysed lysate can be used as a surrogate to ascertain protein concentrations in cyto-serum. Generally, the raw the concentrations would be between 3.5 – 4.0 mg mL^{-1} for various preparations. Protein concentrations were diluted to a standard concentration

of 3.3 mg mL^{-1} using their respective lysis buffers. This generates the normalized lysates for all downstream workflows.

2c. Preparation of Native and Refolded Samples with and without Molecular Chaperones for Limited Proteolysis Mass Spectrometry. To prepare half-isotopically-labeled native samples for experiments without molecular chaperones, 3.5 µL of normalized lysates derived from pellets in which half of the cells were grown with [$^{13}C_6$]L-Arginine and [$^{13}C_6$]L-Lysine during cell culture and half of the cells were grown with natural abundance L-Arginine and L-Lysine during cell culture (and lysed in cyto-serum), were diluted with 96.5 µL of cyto-serum native dilution buffer (1x cyto-serum, 0.1036 mM DTT, 62.17 mM GdmCl) to a final protein concentration of 0.115 mg mL^{-1}. Following dilution, the final concentrations are 1x cyto-serum, 0.1 mM DTT and 60 mM GdmCl. To prepare native samples with the addition of molecular chaperones, 3.5 µL of normalized lysates prepared in cyto-serum were diluted with 96.5 µL of cyto-serum native dilution buffer (1x cyto-serum, 0.1036 mM DTT, 62.17 mM GdmCl, 621.6 µM ATP, supplemented with either 5.19 µM DnaK, 1.04 µM DnaJ and 1.04 µM GrpE; or 4.15 µM GroEL and 8.3 µM GroES) to a protein concentration of 0.115 mg mL^{-1}. Following dilution, the final concentrations are 1x cyto-serum, 0.1 mM DTT, 60 mM GdmCl, 600 µM ATP and either 5 µM DnaK, 1 µM DnaJ, 1 µM GrpE; or 4 µM GroEL, 8 µM GroES. While preparing both native and refolding dilution buffers, molecular chaperones were added as the final component and used immediately to prevent them from prematurely utilizing all available ATP. Native samples were then equilibrated by incubating for 90 min at room temperature prior to limited proteolysis. We note here as an important detail that because cyto-serum dilution buffers containing chaperones must be used immediately, and because it is important for reproducibility's sake that the same buffer preparation is used for all samples, these experiments require three experimentalists working simultaneously to process the three biological replicate samples at the same time.

The refolding samples were prepared as follows: 600 µL of normalized lysates, 100 mg of solid GdmCl, and 2.4 µL of a freshly prepared 700 mM DTT stock solution were added to a fresh 1.5 mL microfuge tube, and solvent was removed using a Vacufuge Plus (Eppendorf) to a final volume of 170 µL, such that the final concentrations of all components were 11.6 mg mL^{-1}, 6 M GdmCl, 3.5x cyto-serum, and 10 mM DTT. These unfolded lysates were incubated overnight at room temperature to complete unfolding prior to refolding.

As above, refolding samples were prepared with or without the addition of molecular chaperones. To prepare refolding samples without molecular chaperones, 99 µL of refolding dilution buffer (0.975x cyto-serum) were added to a fresh 1.5 mL microfuge tube. 1 µL of unfolded extract was then added to the tube containing the refolding dilution buffer and quickly mixed by rapid vortexing, diluting the sample by 100x, followed by flash centrifugation to collect liquids to the

bottom of the tube. The final concentrations were 1x cyto-serum, 0.1 mM DTT and 60 mM GdmCl. To prepare refolding samples with the addition of molecular chaperones, 99 µL of refolding dilution buffer (0.975x cyto-serum supplemented with 606 µM ATP and with either 5.05 µM DnaK, 1.01 µM DnaJ and 1.01 µM GrpE; or 4.04 µM GroEL and 8.08 µM GroES) were added to a fresh 1.5 mL microfuge tube. 1 µL of unfolded lysate was then added to this refolding dilution buffer and quickly mixed by rapid vortexing, diluting the sample by 100x, followed by flash centrifugation to collect liquids to the bottom of the tube. The final concentrations were 1x cyto-serum, 0.1 mM DTT, 60 mM GdmCl, 600 µM ATP and either 5 µM DnaK, 1 µM DnaJ, 1 µM GrpE; or 4 µM GroEL, 8 µM GroES. Refolded samples were then incubated at room temperature for 1 min, 5 min, or 2 h to allow for proteins to refold prior to limited proteolysis.

2d. Limited Proteolysis Mass Spectrometry Sample Preparation. To perform limited proteolysis, 2 µL of a PK stock (prepared as a 0.067 mg mL^{-1} PK in a 1:1 mixture of Tris lysis buffer and 20% glycerol, stored at -20˚C and thawed at most only once) were added to a fresh 1.5 mL microfuge tube. After refolded proteins were allowed to refold for the specified amount of time (1 min, 5 min, or 2 h), or native proteins were allowed their 90 min equilibration, 100 µL of the native/refolded lysates were added to the PK-containing microfuge tube and quickly mixed by rapid vortexing (enzyme:substrate ratio is a 1:100 w/w ratio (1)), followed by flash centrifugation to collect liquids to the bottom of the tube. Samples were incubated for exactly 1 min at room temperature before transferring them to a mineral oil bath preequilibrated at 110°C for 5 min to quench PK activity. Boiled samples were then flash centrifuged (to collect condensation on the sides of the tube), and transferred to fresh 1.5 mL microfuge tube containing 76 mg urea such that the final urea concentration was 8 M and the final volume was 158 µL. They are then vortexed to dissolve the urea to unfold all proteins and quench any further enzyme activity indefinitely, and flash centrifuged to collect liquids to the bottom of the tubes. Addition to urea is the only allowed pause point; all samples operate on a strict timetable from the moment they are refolded until this point. Moreover, once chaperones are added to cyto-serum, they must be used immediately: in the case of native samples, cyto-serum native dilution buffers are added to proteins immediately after preparation, and then 90 min incubation begins. In the case of refolded samples, cyto-serum refolding buffers are added to unfolded proteins immediately after preparation, and then refolding times (1 min, 5 min, 120 min) begins. This method generates all limited proteolysis samples for this study. For the final studies used for the primary datasets, 51 separate samples were prepared for this experiment, they include: native and refolded in cyto-serum with and without molecular chaperones (DnaK/DnaJ/GrpE or GroEL/ES), and the appropriate biological triplicates for each category. In addition, native samples in cyto-serum prepared with and without GroEL/ES were each prepared on two separate occasions, creating a set of technical duplicates. Refolded samples for each of the three refolding timepoints were prepared in biological triplicates. The 1 min refolding timepoint in cyto-serum with and without

GroEL/ES were each prepared on two separate occasions, creating a set of technical duplicates. An additional set was prepared for the 5 min refolding in cyto-serum with the addition of GroEL/ES. A representation of all samples prepared for this study is presented in Figure S1. We note here that LiP-MS studies typically prepare a series of parallel 'control' samples in which PK is withheld; these samples are then used for standard quantitative proteomics experiments to measure protein abundance differences across conditions (1, 2). We opted to not perform this for the current study for the following reasons: 1) there is no practical way native and refolded samples that are compared to each other can have different protein abundances given that they are derived from the same lysates; indeed, the samples compared to each other for these studies are compositionally identical and differ only in history; 2) our previous study (2) confirmed that refolded/native protein abundance ratios were equal to unity at a frequency higher than the false discovery rate.

All protein samples were prepared for mass spectrometry as follows: 2.25 µL of a freshly prepared 700 mM stock of DTT were added to each sample-containing microfuge tube to a final concentration of 10 mM. Samples were incubated at 37°C for 30 minutes at 700 rpm on a thermomixer to reduce cysteine residues. 9 µL of a freshly prepared 700 mM stock of iodoacetamide (IAA) were then added to a final concentration of 40 mM, and samples were incubated at room temperature in the dark for 45 minutes to alkylate reduced cysteine residues. To assist trypsin in the digestion of samples with the addition of molecular chaperones, 1 µL of 0.4 µg μL^{-1} Lys-C (NEB) stock was added (enzyme:substrate ratio of 1:100 w/w) and digestion proceeded for 2 h at 37˚C. After digestion with Lys-C, 471 µL of 100 mM ammonium bicarbonate (pH 8) were added to the samples to dilute the urea to a final concentration of 2 M. 2 µL of a 0.4 µg μL^{-1} stock of Trypsin (NEB) were added to the samples (to a final enzyme:substrate ratio of 1:50 w/w) and incubated overnight (15-16 h) at 25°C at 700 rpm (not 37˚C, so as to minimize decomposition of urea and carbamylation of lysines).

2e. Desalting of Mass Spectrometry Samples. Peptides were desalted with Sep-Pak C18 1 cc Vac Cartridges (Waters) over a vacuum manifold. Tryptic digests were first acidified by addition of 16.6 µL trifluoroacetic acid (TFA, Acros) to a final concentration of 1% (vol/vol). Cartridges were first conditioned (1 mL 80% ACN, 0.5% TFA) and equilibrated (4 x 1 mL 0.5% TFA) before loading the sample slowly under a diminished vacuum (ca. 1 mL/min). The columns were then washed (4 x 1 mL 0.5% TFA), and peptides were eluted by addition of 1 mL elution buffer (80% ACN, 0.5% TFA). During elution, vacuum cartridges were suspended above 15 mL conical tubes, placed in a swing-bucket rotor (Eppendorf 5910R), and spun for 3 min at 350 g. Eluted peptides were transferred from Falcon tubes back into microfuge tubes and dried using a vacuum centrifuge (Eppendorf Vacufuge). Dried peptides were stored at -80°C until analysis. For analysis,

samples were vigorously resuspended in 0.1% FA in Optima water (ThermoFisher) to a final concentration of 0.5 mg mL^{-1}.

2f. LC-MS/MS Acquisition. Chromatographic separation of digests were carried out on a Thermo UltiMate3000 UHPLC system with an Acclaim Pepmap RSLC, C18, 75 µm × 25 cm, 2 µm, 100 Å column. Approximately, 1 µg of protein was injected onto the column. The column temperature was maintained at 40 °C, and the flow rate was set to 0.300 µL min^{-1} for the duration of the run. Solvent A (0.1% FA) and Solvent B (0.1% FA in ACN) were used as the chromatography solvents. The samples were run through the UHPLC System as follows: peptides were allowed to accumulate onto the trap column (Acclaim PepMap 100, C18, 75 µm x 2 cm, 3 µm, 100 Å column) for 10 min (during which the column was held at 2% Solvent B). The peptides were resolved by switching the trap column to be in-line with the separating column, quickly increasing the gradient to 5% B over 5 min and then applying a 95 min linear gradient from 5% B to 25% B. Subsequently, the gradient was increased from 35% B to 40% B over 25 min and then increased again from 40% B to 90% B over 5 min. The column was then cleaned with a sawtooth gradient to purge residual peptides between runs in a sequence.

A Thermo Q-Exactive HF-X Orbitrap mass spectrometer was used to analyze protein digests. A full MS scan in positive ion mode was followed by 20 data-dependent MS scans. The full MS scan was collected using a resolution of 120000 (@ m/z 200), an AGC target of 3E6, a maximum injection time of 64 ms, and a scan range from 350 to 1500 m/z. The data-dependent scans were collected with a resolution of 15000 (@ m/z 200), an AGC target of 1E5, a minimum AGC target of 8E3, a maximum injection time of 55 ms, and an isolation window of 1.4 m/z units. To dissociate precursors prior to their reanalysis by MS2, peptides were subjected to an HCD of 28% normalized collision energies. Fragments with charges of 1, 6, 7, or higher and unassigned were excluded from analysis, and a dynamic exclusion window of 30.0 s was used for the data-dependent scans. For pseudo-SILAC samples, mass tags were enabled with Δm of 2.00671 Th, 3.01007 Th, 4.01342 Th, and 6.02013 Th (to account for the fixed 6 or 12 Da mass shifts in different charge states) to promote selection of non-chaperone-derived peptides for isolation and data-dependent MS2 scans.

3. Methods to Study Aggregation

3a. Culturing of K12. K12 cells were grown in 3 × 50 mL (biological triplicates) of MOPS EZ rich media from saturated overnight cultures with a starting OD_{600} of 0.05. Cells were cultured at 37˚C with agitation (220 rpm) to a final OD_{600} of 0.8 before being transferred to 3 × 50 mL falcon tubes and collected by centrifugation at 4000 *g* for 15 mins at 4˚C. The supernatants were removed, and cell pellets were stored at -20˚C until further use.

3b. Preparation of Cell Lysates and Refolded Samples. Frozen cell pellets were resuspended in a lysis buffer consisting of either 900 µL of Tris pH 8.2 lysis buffer (20 mM Tris pH 8.2, 100 mM NaCl, 2 mM $MgCl_2$ and supplemented with DNase I to a final concentration (f.c.) of 0.1 mg mL^{-1}) or 900 µL of cyto-serum lysis buffer (1× cyto-serum, supplemented with DNase I to a f.c of 0.1 mg mL^{-1}). In samples prepared for analytical ultra-centrifugation, 9 µL of a 100x protease inhibitor cocktail (100 mM AEBSF.HCL, 80 µM Aprotonin, 5 mM Bestatin,1.5 mM E-64, 2 mM Leupeptin, and 1 mM Pepstatin A) was added to prevent protein degradation during ultracentrifugation. Normalized lysates were prepared as above, by lysing by cryogenic pulverization, removing ribosomes by ultracentrifugation, and normalizing to 3.3 mg mL^{-1} with the BCA assay.

Native samples were prepared as followed: normalized lysates were diluted with their respective native dilution buffers (20 mM Tris pH 8.2, 100 mM NaCl, 2 mM $MgCl_2$, 1.04 mM dithiothreitol (DTT), 62 mM GdmCl; or 1x cyto-serum, 0.1036 mM DTT, 62 mM GdmCl) to a protein concentration of 0.115 mg mL^{-1}. Following dilution, the final concentrations are: 20 mM Tris pH 8.2, 100 mM NaCl, 2 mM $MgCl_2$, 1 mM DTT and 60 mM GdmCl; or 1x cyto-serum, 0.1 mM DTT and 60 mM GdmCl. Native samples are then incubated overnight at room temperature. The refolding samples were prepared as follows: 600 µL of normalized lysates, 100 mg GdmCl as a solid, and 2.4 µL of a freshly prepared 700 mM DTT stock solution were combined into a fresh 1.5 mL microfuge tube, and solvent was removed using a Vacufuge Plus to a final volume of 170 µL, such that the final concentrations of all components were 11.6 mg mL^{-1} protein, 6 M GdmCl, 70 mM Tris pH 8.2, 350 mM NaCl, 7 mM $MgCl_2$, and 10 mM DTT; or 11.6 mg mL^{-1} protein, 6M GdmCl, 3.5x cyto-serum, and 10 mM DTT. These unfolded lysates were incubated overnight in a sealed container at room temperature to complete unfolding prior to refolding.

3c. Sedimentation Velocity Analytical Ultracentrifugation. To study the presence of smaller soluble aggregates in refolded extracts using analytical ultracentrifugation, native and unfolded lysates in Tris pH 8.2 were prepared as described above. For analytical ultracentrifugation, all studies were carried out using Tris pH 8.2 refolding buffers as cyto-serum has too many components that absorb at similar wavelengths to proteins (Fig. S2D). To prepare refolded samples, unfolded lysates were diluted 100× with refolding dilution buffer (19.5 mM Tris pH 8.2, 97.5 mM NaCl,1.95 mM MgCl2, and 0.91 mM DTT) and incubated for 2 h at room temperature before being loaded into AUC cells assembled with 1.2 mm double-sector epoxy centerpieces and sapphire windows. Prior to starting each sedimentation velocity (SV) experiment, samples were equilibrated at 20 ˚C for 1 hour in the centrifuge. Each sample was spun at 20 ˚C using a 4-hole, An-Ti60 rotor and speed of 50000 rpm. Absorbance was monitored at 280 nm, and radial scans were acquired with 0.003 cm radial steps in continuous mode and with zero time interval

between scans. All SV experiments were performed using a Beckman XL-A ultracentrifuge (Beckman Coulter). SV data were analyzed using the time derivative method in dcdt+ (3) to obtain normalized g(s*) distributions. Refolding buffer density (ρ = 1.00464 g/mL) and viscosity (η = 1.0200 cP) were calculated in *SEDNTERP* (4) and an average protein partial specific volume ($\bar{v}$) of 0.73 ml g^{-1} was used to describe the heterogenous cell lysates.

3d. Mass Photometry (MP). To monitor smaller soluble aggregates in cyto-serum, Mass Photometry (MP) experiments were conducted on a OneMP instrument (Refeyn) at room temperature (5, 6). Native samples were prepared in their respective native dilution buffers (either Tris or cyto-serum) as described above. Unfolded samples (either Tris or cyto-serum) were prepared as described above. To prepare refolded samples, 2 µL of unfolded extracts were diluted 100× with 198 µL of refolding dilution buffer (19.5 mM Tris pH 8.2, 97.5 mM NaCl,1.95 mM $MgCl_2$, and 0.91 mM DTT; or 0.95× cyto-serum) and incubated at room temperature for 2 h to allow proteins to refold. To prepare samples for MP, 10 µL of native samples or refolding reactions (both at 0.115 mg mL^{-1}) were rapidly diluted an additional 100× by addition to 990 µL of Tris lysis buffer and immediately transferred to silicone gaskets on microscope coverslips. Acquisition (which takes 2 min) was initiated within 1 min of the additional 100× dilution. To prepare the set up for sample analysis, microscope cover slips were first cleaned by washing with ethanol, isopropanol, and MPW and then dried with N_2 gas. Cleaned microscope cover slips were then fitted with a silicone gasket. 10 µL of Tris lysis buffer was loaded onto the silicone gasket to focus and sharpen the instrument. 10 µL of sample was gently pipetted into the droplet seated in the gasket without disturbing focus. Recordings were acquired using the AcquireMP (Refeyn) software and mass distributions were calculated utilizing the DiscoverMP (Refeyn) software.

3e. Quantification of Pelleting Aggregates Upon Refolding. To study the amount of insoluble aggregates that form upon global refolding, native samples were prepared in their respective native dilution buffers (either Tris or cyto-serum) as described above. Unfolded samples (either Tris or cyto-serum) were prepared as described above. To prepare refolded samples, 5 µL of unfolded extracts were diluted 100× with 495 µL of refolding dilution buffer (19.5 mM Tris pH 8.2, 97.5 mM NaCl,1.95 mM $MgCl_2$, and 0.91 mM DTT; or 0.95× cyto-serum) and incubated at room temperature for 2 h to allow proteins to refold (or precipitate). 500 µL of native and refolded samples (both at 0.115 mg mL^{-1}, final protein concentration) were centrifuged at 16000 *g* for 15 mins at 4˚C to collect aggregated proteins. The supernatant was carefully removed by pipetting to not disturb the protein pellet. The pellets in all samples were washed with 500 µL of Tris lysis buffer to reduce the interference from reducing agents in Tris or cyto-serum refolding buffers with the BCA assay. The washed pellets were then resuspended in 50 µL of 8M urea in MPW and the protein concentrations were quantified with the BCA Assay as described above. The amount of

protein in the pellet was determined using the protein concentration and the resuspension volume (50 µL) and converted to fractional precipitation by dividing by the initial amount of protein in the refolding reaction (57.5 µg). The data are reported as a mean ± standard deviations from biological triplicates, which were differentiated at the inoculation stage. Statistical significance between samples refolded in either Tris or cyto-serum were assessed using t-tests with Welch's correction for unequal population variances as implemented in Prism 9 (Graphpad). The "precipitation" measured for the native samples were treated as the background level of the measurement because they should not possess any precipitated protein.

4. Additional Characterization of Refolding Reactions

4a. GroEL/ES ATP Hydrolysis Assay. To determine the rate of ATP hydrolysis in cyto-serum by GroEL/ES in refolding reactions, unfolded lysates in cyto-serum were prepared as described above except that *E. coli* cells were cultured in MOPS EZ rich media without isotopically labeled L-Arginine or L-Lysine. To prepare refolding samples with the addition of GroEL/ES, 99 µL of refolding dilution buffer (0.975x cyto-serum supplemented with 606 µM ATP, 4.04 µM GroEL and 8.08 µM GroES) were added to a fresh 1.5 mL microfuge tube. 1 µL of unfolded lysate was then added to this refolding dilution buffer and quickly mixed by rapid vortexing, diluting the sample by 100x, followed by flash centrifugation to collect liquids to the bottom of the tube. The final concentrations were 1x cyto-serum, 0.1 mM DTT, 60 mM GdmCl, 600 µM ATP, 4 µM GroEL and 8 µM GroES. Refolding reactions were incubated at room temperature for 1 min, 5 min, 10 min, 20 min, 30 min, 60 min, and 120 min, at which point 10 µL of the refolding reaction were taken and diluted 100x with 8 M urea to quench GroEL/ES activity. ATP concentrations in these aliquots were determined using the ATP Determination Assay (Invitrogen), based on luminescence generated by luciferase. A standard curve was generated by combining 10 µL of ATP in various concentrations from 100 nM to 10 mM with 990 µL of 8 M urea. 10 µL of each sample (either standard or unknown) were loaded into an opaque white polystyrene 96-well plate (Thermo Scientific) and then to each 190 µL of working reagent was rapidly added using a repeater. Luminescence was measured immediately after on a Tecan Spark plate reader. A calibration curve corresponding luminescence to ATP concentration was constructed and data are reported as means ± standard deviations from three independent refolding reactions. The ATP concentration time series (Figure S9) were fit to an exponential decay in GraphPad Prism 9.

4b. Enzyme Reactivation Assays.

4b-i. Purified Enzymes, phosphoglucose isomerase (Pgi) and Mannose-6-phosphate isomerase (ManA). Assays were based on the Phosphoglucose isomerase colorimetric assay kit (Sigma). A concentrated Pgi stock was prepared by reconstituting the PGI positive control

with 20 µL of assay buffer, according to the manufacturer protocol. To measure its native activity, a 350-fold dilution was prepared in Tris buffer (20 mM Tris pH 8.2, 100 mM NaCl, 2 mM $MgCl_2$). 5 µL of Pgi was transferred to a clear 96-well microtiter plate, and to it was added 95 µL of working solution, itself composed of 89 µL assay buffer, 2 µL enzyme mix, 2 µL developer solution, and 2 µL substrate mix (and prepared immediately before use). Activity of Pgi was measured by its conversion of fructose-6-phosphate (in the substrate mix) to glucose-6-phosphate, which is then used to convert NAD^+ to NADH via glucose-6-phosphate dehydrogenase (both in enzyme mix). NADH's signal is amplified by binding to a chromophore in the developer kit and monitored by measuring absorbance at 450 nm. Absorbance measurements were made every minute for 10 min following initial addition of the working solution. A background was conducted in which the 2 µL of substrate mix was withheld and replaced with 2 additional µl of assay buffer. Relative activity was calculated by performing linear regression on the A450 time course in Excel and extracting the slope from the first 5 min. Measurement was conducted in technical triplicate, and the average of the background samples' slopes was subtracted. A 7-fold diluted denatured form of Pgi was prepared by combining 10 mg solid GdmCl, 0.25 µL of 700 mm DTT, 7.25 µL Tris buffer, and 2.5 µL of concentrated Pgi stock (total volume is 17.5 µL); Pgi was allowed to unfold overnight. To initiate refolding, the solution was diluted 50-fold with one of three refolding buffers. To refold intrinsically, the refolding dilution buffer was 19.53 mM Tris pH 8.2, 97.6 mM NaCl, 10.03 mM $MgCl_2$, 10.1 mM KCl, 0.91 mM DTT. To refold with GroEL, the refolding dilution buffer was 19.53 mM Tris pH 8.2, 97.6 mM NaCl, 10.03 mM $MgCl_2$, 10.1 mM KCl, 0.91 mM DTT, 2.02 mM ATP, 4.04 µM GroEL, 8.08 µM GroES. To refold with DnaKJE, the refolding dilution buffer was 19.53 mM Tris pH 8.2, 97.6 mM NaCl, 10.03 mM $MgCl_2$, 10.1 mM KCl, 0.91 mM DTT, 2.02 mM ATP, 5.05 µM DnaK, 1.01 µM DnaJ, 1.01 GrpE. At each time point, 5 µL of the refolding reaction was transferred to a clear 96-well microtiter plate, and 95 µL of working solution (described above) was added. Activity was measured, and A450 slopes calculated as above. Fractional reactivation was calculated as: $(slope_{refold(t)} - slope_{bkg})/(slope_{native} - slope_{bkg})$.

To measure activity of ManA, purified ManA from *E. coli* (Sigma) as a suspension in ammonium sulfate was resuspended in Tris buffer to produce a concentrated 100 µM stock. To measure its native activity, a 200-fold dilution of concentrated ManA was prepared in Tris buffer. 5 µL of ManA was transferred to a clear 96-well microtiter plate, and to it was added 95 µL of working solution, itself composed of 51 µL assay buffer, 10 µL of 50-fold diluted PGI positive control, 2 µL enzyme mix, 2 µL developer solution, and 30 µL of 60 mM mannose-6-phosphate in assay buffer (and prepared immediately before use). Activity of ManA was measured by its conversion of mannose-6-phosphate to fructose-6-phosphate to glucose-6-phosphate, which is then used to convert NAD^+ to NADH via glucose-6-phosphate dehydrogenase. Absorbance measurements

were made every minute for 10 min following initial addition of the working solution. A background was conducted in which the 30 µL of mannose-6-phosphate was withheld and replaced with 30 additional µl of assay buffer. Slopes were calculated from the 3 min to 8 min timepoint, and relative activity was calculated as above. Measurement was conducted in technical triplicate, and the average of the background samples' slopes was subtracted. A 2-fold diluted denatured form of ManA was prepared by combining 20 mg solid GdmCl, 0.5 µL of 700 mm DTT, and 19.5 µL of concentrated ManA stock (total volume is 39 µL); ManA was allowed to unfold overnight. To initiate refolding, the solution was diluted 50-fold with one of three refolding buffers, as described above. At each time point, 5 µL of the refolding reaction was transferred to a clear 96-well microtiter plate, and 95 µL of ManA working solution (described above) was added. Activity was measured, A450 slopes calculated, and fractional reactivation was calculated as above.

4b-ii. **Pgi and ManA in *E. coli* Lysate**. To measure Pgi activity, native *E. coli* lysates were prepared as described identically as in section 3a–3b. 5 µL of lysate (containing proteins at 0.115 mg mL^{-1} overall concentration), from three separate biological replicates, were transferred to a clear 96-well microtiter plate, and to them were added 95 µL of working solution, itself composed of 89 µL assay buffer, 2 µL enzyme mix, 2 µL developer solution, and 2 µL substrate mix (and prepared immediately before use). Relative activity was measured as above. Lysates were globally unfolded following the protocol in section 3b. To initiate refolding, lysates were diluted 100-fold with 19.53 mM Tris pH 8.2, 97.6 mM NaCl, 10.03 mM $MgCl_2$, 10.1 mM KCl, 0.91 mM DTT. Specifically, 99 µL of refolding buffer was transferred into a fresh tube and 1 µL of unfolded lysate was pipetted in, followed immediately by vortexing and flash centrifugation. At distinct refolding times, 5 µL of refolding lysate, from the three replicates, were transferred to a clear 96-well microtiter plate, and to them were added 95 µL of working solution (described above). Relative activity was obtained from the slopes, and fractional reactivation calculated from their ratio, as described in section 4b-i.

To measure ManA activity, native *E. coli* lysates were prepared as described in section 3a-3b, except instead of adding 5 µL to wells of lysates diluted to 0.115 mg mL^{-1}, lysates were diluted to 1.15 mg mL^{-1}. The ManA working solution was used, and relative activity measured as described in section 4b-i. Lysates were globally unfolded and refolded by 100-fold dilution, except at a 2-fold larger scale. To match the loading of the refolded samples to that of the native samples, 50 µL of refolding lysate was used instead of 5 µL. At distinct refolding times, 50 µL of refolding lysate, from the three replicates, were transferred to a clear 96-well microtiter plate and to them were added 50 µL of working solution, itself composed of: 6 µL assay buffer, 10 µL of 50-fold diluted PGI positive control, 2 µL enzyme mix, 2 µL developer solution, and 30 µL of 60 mM

mannose-6-phosphate in assay buffer (and prepared immediately before use). Relative activity was obtained from the slopes, and fractional reactivation calculated from their ratio, as described in section 4b-i.

5. MS Data Analysis and Computational Workflows

5a. LC-MS/MS Data Analysis. Proteome Discoverer (PD) Software Suite (v2.4, Thermo Fisher) and the Minora Algorithm were used to analyze mass spectra and perform Label Free Quantification (LFQ) of detected peptides. Default settings for all analysis nodes were used except where specified. The data were searched against Escherichia coli (UP000000625, Uniprot) reference proteome database. For peptide identification, either the PD Sequest HT node (for non-pseudo-SILAC samples) or PD MSFragger node (pseudo-SILAC) were used, each using a semi-tryptic search allowing up to 2 missed cleavages. A precursor mass tolerance of 10 ppm was used for the MS1 level, and a fragment ion tolerance was set to 0.02 Da at the MS2 level for both search algorithms. For Sequest HT, a peptide length between 6 and 144 amino acid residues was allowed. For MSFragger, a peptide length between 7 and 50 amino acid residues was allowed with a peptide mass between 500 and 5000 Da. Additionally, a maximum charge state for theoretical fragments was set at 2 for MSFragger. Oxidation of methionine and acetylation of the N-terminus were allowed as dynamic modifications while carbamidomethylation on cysteines was set as a static modification. For pseudo-SILAC samples, heavy isotope labeling ($^{13}C_6$) of Arginine and Lysine were allowed as dynamic modifications. All parameters for Sequest HT and MSFragger search algorithms are provided in the table below. The Percolator PD node was used for FDR validation for peptides identified with the Sequest HT search algorithm. For peptides identified with the MSFragger search algorithm, the Philosopher PD node was used for FDR validation. Raw normalized extracted ion intensity data for the identified peptides were exported from the .pdResult file using a three-level hierarchy (protein > peptide group > consensus feature). These data were further processed utilizing custom Python analyzer scripts (available on GitHub, and described in depth previously in (2)). Briefly, normalized ion counts were collected across the refolded replicates and the native replicates for each successfully identified peptide group. Effect sizes are the ratio of averages (reported in $\log_2$) and P-values (reported as $-\log_{10}$) were assessed using *t* tests with Welch's correction for unequal population variances. Missing data are treated in a special manner. If a feature is not detected in all three native (or refolded) injections and is detected in all three refolded (or native) injections, we use those data, and fill the missing values with 1000 (the ion limit of detection for this mass analyzer); this peptide becomes classified as an all-or-nothing peptide. If a feature is not detected in one out of six injections, the missing value is dropped. Any other permutation of missing data (e.g., missing in two injections) results in the quantification getting discarded. In many situations, our data provide multiple independent sets of quantifications for the same peptide group. This

happens most frequently because the peptide is detected in multiple charge states or as a heavy isotopomer. In this case, we calculate effect size and P-value for all features that map to the same peptide group. If the features all agree with each other in sign, they are combined: the quantification associated with the median amongst available features is used and the P-values are combined with Fisher's method. If the features disagree with each other in sign, the P-value is set to 1. Coefficients of variation (CV) for the peptide abundance in the three replicate refolded samples are also calculated. Analyzer returns a file listing all the peptides that can be confidently quantified, and provides their effect-size, P-value, refolded CV, proteinase K site (if half-tryptic), and associated protein metadata.

5b. Refoldability Analysis. Results from analyzer are digested in the following way. Proteins with only one peptide confidently quantified are discounted; proteins with more than two are kept. Peptides are considered to have significantly different abundance in the refolded sample if the effect size is 2 or greater (more than double or less than half the abundance of native), and the P-value is less than 0.01 by Welch's *t* test. All-or-nothing peptides must have abundance differences greater than 64-fold, and use a relaxed P-value cut-off of 0.0158. The number of significant and all-or-nothing peptides is counted for each protein (or, in the case of Fig. 4I, for each domain, whose residue ranges are provided and where peptides are only assigned to a given domain if the PK cut site or the full tryptic range falls within the domain boundaries). Proteins (or domains) are deemed nonrefoldable if two or more peptides with significantly different abundances in the refolded sample are identified.

Protein-level refoldability analyses proceed by counting the number of refoldable and nonrefoldable proteins within a set of categories (e.g., $5 < \text{pI} < 6$) associated with a feature (e.g., pI) and calculating the fraction refolding within the category. To determine if there is a significant enrichment for (non-)refolders within certain categories, we calculate the expected number of (non-)refolders for each category by taking the total number of proteins that are assigned a value under the feature in question, times the fraction (non-)refolding, times the fraction of proteins in that category. The chi-square test is used to determine if the observed counts and expected counts significantly differ, for all cases in which the feature has three or more categories. If it only has two, Fisher's exact test is used instead.

Peptide-level refoldability analyses are performed in a similar way. The total number of significant and nonsignificant peptides mapped to proteins within a set of categories associated with a feature are counted and the percentage significant calculated. To determine if there is a significant enrichment for (non-)significant peptides associated with certain categories, we calculate the expected number of (non-)significant peptides for each category by taking the total

number of peptides associated with proteins that are assigned a value under the feature in question, times the fraction of peptides that are (non-)significant, times the fraction of peptides associated with that category. The chi-square test is used to determine if the observed counts and expected counts significantly differ, for all cases in which the feature has three or more categories. If it only has two, Fisher's exact test is used instead.

For condition comparisons (i.e., comparing Tris to cyto-serum, or refolding with GroEL/ES vs. DnaK/J/E), we performed 12-way LFQs, and created a slightly modified analyzer script that assesses peptide quantifications separately for the six samples associated with condition 1 and the six samples associated with sample 2. The analyzer returns a file listing all the peptides that can be confidently quantified, and provides their effect-size, P-value and refolded CV for condition 1 and 2, proteinase K site (if half-tryptic), and associated protein metadata. Similar to before the number of significant and all-or-nothing peptides are counted for each protein in condition 1 and 2. Proteins are only admitted into the comparison if 2 or more peptides are identified in both conditions, and are classified as refolding in both, refolding in condition 1, refolding in condition 2, or nonrefolding in both. Proteins are discarded if they are on the border; e.g., one significant peptide assigned in condition 1 and two significant peptides assigned in condition 2.

For these analyses, we count the number of proteins associated with a given category (e.g., 5 < pI < 6) that refold in both, refold in condition 1, refold in condition 2, or do not refold in either. For each category, expected counts are calculated by taking the total number of proteins in that category times the overall fraction of proteins that refold in both, refold in condition 1, refold in condition 2, or do not refold in either. The chi-square test is used to determine if the observed counts and expected counts significantly differ. Note that these tests are conducted on individual categories (e.g., the 5 < pI < 6 category is enriched for proteins that refold with GroEL/ES but not without it), whereas previously, the test is conducted on the feature overall (e.g., pI groups do not all refold with the same frequency).

For kinetic comparisons (i.e., comparing proteins that have refolded in cyto-serum for 1 min or 5 min), we combined results from the separate timepoints by collecting the subset of proteins that were identified in both experiments and compiling together the number of significant and all-or-nothing peptides that are counted for each protein at timepoint 1 and 2. Proteins are only admitted into the comparison if 2 or more peptides are identified at both timepoints, and are classified as refolding in both (fast refolder), refolding at the later timepoint (slow refolder), or refolding at the earlier timepoint (fold loser). Nonrefolders are not used for kinetic comparisons. Proteins are discarded if they are on the border; e.g., one significant peptide assigned at

timepoint 1 and two significant peptides assigned in timepoint 2. The analyses and chi-square tests are done analogously as above, for the condition comparisons.

5c. Bioinformatics. Ecocyc database (7) was used to obtain information about cellular compartment (cytosol, inner membrane, periplasmic space, outer membrane, ribosome, cell projection), subunit composition, essentiality, copy number, cofactors, and molecular weight (from nucleotide sequence) for each protein. When the information was available, we used Ecocyc's Component Of category to obtain the full constitutive composition of the protomer within a complex.

Copy number information predominantly comes from a single ribosome profiling study by Li and co-workers (8). We used copy number in Neidhardt EZ rich defined medium because of its similarity to the growth medium used in these studies.

Domain information was based on the SCOP hierarchy and obtained through the Superfamily database (http://supfam.org) (9). We used custom scripts to edit the "raw" file available from supfam. org into a format more usable for our purposes (including the switch from a Uniprot identifier to the gene symbol identifier). This database was used to count the number of domains per protein, and to perform the domain-level analysis in which peptides are mapped to individual domains within proteins based on residue ranges. Domains are categorized by their 'fold.' Note that in SCOP, folds correspond to collections of superfamilies with similar topologies, and in most situations (but not always) correspond to deep evolutionary relationships (10).

Gene ontology analysis was conducted using PantherDB (11). The set of 105 chaperone-nonrefolders was entered as the test set, and the *E. coli* proteome used as the reference set. Statistical overrepresentation tests were selected using the complete set of GO biological processes.

Isoelectric effects were obtained from the isoelectric database (12). We downloaded the file corresponding to E. coli K-12 MG1655 and took an average of the isoelectric points calculated by all the algorithms available for each protein. Chaperonin classes were obtained from Kerner et al. (13). Specifically, we examined Table S3, manually identified the current Uniprot accession code for each of the proteins identified by Kerner et al. and transferred this information into a file that contains the gene symbol, the current Uniprot accession code, and the class assignment. We also compiled information from Fujiwara et al. which breaks down class III proteins into class III− and class IV (14).

5d. Quantification and statistical analysis. All analyses of aggregation were conducted on independent refolding reactions from independent biological replicates (n = 3). Raw values shown for pelleting assay and significance by *t* test with Welch's correction for unequal population variances. Analytical ultracentrifugation and mass photometry data shown from representative examples from among replicates. Standard target-decoy based approaches were used to filter protein identifcations to an FDR < 1%, as implemented by Percolator (when searching with Sequest), or Philosopher (when searching with MSFragger). All mass spectrometry experiments were conducted on three biological replicates used to generate three native samples and three independent refolding reactions from the same biological replicates. For each peptide group, abundance difference in refolded relative to native was judged by the *t* test with Welch's correction for unequal population variances. Fisher's method was used to combine P-values when there were multiple quantifiable features per peptide group. P-values less than 0.01 were used as a requirement to consider a region structurally distinct in the refolded form. Differences in means of distributions are assessed with the Mann-Whitney rank-sum test. To test whether particular categories are enriched with (or de-enriched with) (non)refoldable proteins, the chi-square test or Fisher's exact test is used.

Table S1: Reagent and Resources

Reagent of Resource	Source	Identifier
Bacterial Strains		
E. coli (Strain K12)	New England BioLabs	ER2738
Chemicals and Reagents		
Adenosine diphosphate (ADP)	Sigma-Aldrich	A2765
Adenosine Triphosphate (ATP) 10 mM	New England BioLabs	P0756S
Ammonium Bicarbonate	Acros Organics	393210050
Dithiothreitol (DTT)	Sigma-Aldrich	D06632
Glycerol	Fisher	G33-1
Guanidium Chloride (GdmCl)	Acros Organics	50-01-1
Hydrochloric Acid (HCl)	Fisher Scientific	A142-212
Iodoacetamide (IAA)	Acros organics	122271000
KCl	Fisher	BP366-1
MgCl2	Fisher	BP214-500
NaCl	Fisher	BP358-212
NAD+	Sigma-Aldrich	481911-5GM
Sodium dodecyl sulfate (SDS)	Fisher	BP166-500
Tris-Base	Sigma-Aldrich	RD008
Tris-HCl	Sigma	RD009
Tryptone	Fisher	BP1421-2
Urea	Fisher	U17-212
Yeast Extract	Fisher	BP1422-500
Components to Make Protease Inhibitor Cocktail		
AEBSF.HCl	Thermo Scientific	78431
Aprotonin from bovine lung	Sigma Aldrich	10820
Bestatin	Thermo Scientific	78433
E-64	Thermo Scientific	78434
Leupeptin	Sigma Aldrich	EI8
Pepstatin A	Thermo Scientific	78436
Components to Make MOPS MEDIA		
13C6 L-ARGININE-HCL	Fisher Scientific	PI88433
13C6 L-LYSINE-2HCL	Fisher Scientific	PI89988
Adenine	Alfa Aesar	A14906
Ammonium Chloride	EMD	AX1270-1

Ammonium Molybdate		
Boric Acid	Fisher Scientific	BP-168-1
Calcium Chloride	Fisher	C77-500
Calcium Pantothenate	Acros organics	24330100000
Cobalt Chloride	Acros	B42357-1000
Cupric Sulfate		
Cytosine	TCI America	71-30-7
di Hydroxy Benzoic Acid		
Glucose (Dextrose) (D)	Fisher	D16-3
Guanine	Alfa Aesar	A12024
Iron Sulfate Stock		
L-Alanine	Acros organics	102831000
L-Arginine HCl	Alfa Aesar	A14730
L-Asparigine	Alfa Aesar	b21473
L-Aspartic Acid, Potassium Salt	Alfa Aesar	A13520
L-Cysteine HCl	Alfa Aesar	I06328
L-Glutamic Acid, Potassium Salt	Alfa Aesar	A15031
L-Glutamine	Alfa Aesar	A14201
L-Glycine	J T Baker	4059-02
L-Histidine HCl H2O	Acros Organics	166151000
L-Isoleucine	Fisher	bp384-100
L-Leucine	Alfa Aesar	A12311
L-Lysine	Alfa Aesar	J62225
L-Methionine	Alfa Aesar	A10318
L-Phenylalinine	Alfa Aesar	A13238
L-Proline	Fisher	BP-392-100
L-Serine	Acros Organics	132665000
L-Threonine	Acros Organics	138931000
L-Tryptophan	Acros Organics	140590250
L-Tyrosine	Alfa Aesar	A11141
L-Valine	Acros organics	140811000
Magnesium Chloride	Fisher	BP214-500
Manganese Chloride		
MOPS (MW 209.3)	Sigma LS	RDD0018
para-Amino Benzoic Acid	Sigma Aldrich	1973
para-Hydroxy benzoic Acid	Alfa Aesar	A13700
Potassium Hydroxide	Fisher	P250-3
Potassium Phosphate Dibasic	Fisher	BP363-1
Potassium Sulfate		
Sodium Chloride	Fisher	BP358-212
Thiamine	Fisher	BP892-100

Tricine (MW 179.2)	Sigma LS	RDD024
Uracil	Alfa Aesar	A15570
Zinc Sulfate	Alfa Aesar	33399
Recombinant Proteins and Enzymes		
DnaK Mix	Cosmo-Biology	GFK-PF003-0.5-EX
DNase I from Bovine Pancreas	Sigma Aldrich	31136
GroEL	Sigma Aldrich	C7688
GroES	Sigma Aldrich	C7438
Endoproteinase LysC	New England BioLabs	P8109S
Proteinase K	Fisher Scientific	BP1700
Trypsin -ultra, Mass Spectrometry Grade	New England BioLabs	P8101S
Phosphomannose Isomerase (*E. coli*)	Sigma-Aldrich	P2621
D-Mannose 6-phosphate	Sigma-Aldrich	**M3655**
Commercial Assays		
Rapid Gold BCA Assay (Pierce)	Thermo Scientific	A53225
ATP Determination Kit (Invitrogen)	Thermo Scientific	A22066
Phosphoglucose Isomerase Colorimetric Assay Kit (Sigma)	Millipore Sigma	MAK103-1KT
LC-MS Reagents		
Acetonitrile (HPLC Grade)	Sigma-Aldrich	34851
Formic Acid (Optima LC/MS Grade)	Fisher Scientific	A117-50
Trifluoroacetic acid (TFA) (HPLC Grade)	Alfa Aesar	44630
Water (Optima LC-MS Grade)	Fisher Scientific	W6500
Databases		
Chaperonin Class Dataset	Kerner et al. 2005	N/A
DnaK Enrichment Dataset	Calloni et al. 2012	N/A
EcoCyc	Keseler et al. 2017	N/A
Fujiwara Data Set	Fujiwara et al. 2010	N/A
SCOP Database	Gough et al., 2001; Pandurangan et al., 2019	N/A
Instrumentation		
1/4 Inch Probe	QSonica	N/A
1/8 Inch Probe	QSonica	N/A

AN60 Ti Rotor	Beckman Coulter	N/A
Eppendorf Centrifuge (5910R)	Eppendorf	N/A
OneMP (Mass Photometry)	Refeyn	N/A
pH Meter	Mettler-Toledo	N/A
Plate Reader (ID3)	Molecular Devices	N/A
Plate Reader (Spark)	Tecan	N/A
Probe Sonicator (Q700R)	QSonica	N/A
Q-Exactive HF-X Orbitrap Mass Spectrometer	Thermo Scientific	N/A
Rheometer (MC92)	Anton-Parr	N/A
SPEX™ Sample Prep Dual Freezer/Mill	SPEX	N/A
SW55 Ti Rotor	Beckman Coulter	N/A
UltiMate3000 UHPLC	Thermo Scientific	N/A
Ultracentrifuge (Optima XL-A)	Beckman Coulter	N/A
Ultracentrifuge (Optima-XPN)	Beckman Coulter	N/A
Other		
3 mL *Konical Tubes*	Beckman Coulter	N/A
5 mL ultra-clear ultrancentrifugation tubes	Beckman Coulter	N/A
JL20 Tubes	Beckman Coulter	N/A
Mass Spec Vials	Thermo Scientific	N/A
Pierce™ 96-Well Polystyrene Plates, White Opaque	Thermo Scientific	15042
Sep Pak C18 1 cc 50 mg Resin	Waters	186000308
Viva-Spin 15R 2K MWCO Hydrosart	Sartorius	VS15RH92
25 mm plate	Anton-Parr	N/A
Software and Algorithms		
AcquireMP	Refeyn	N/A
AnalyzerV18	In House	https://github.com/FriedLabJHU/Refoldability-Tools
DCDT+	John Philo	N/A
DiscoverMP	Refeyn	N/A
MSFragger Node in PD	Nesvizhskii Lab	N/A
Proteome Discoverer v 2.4 (PD)	Thermo Scientific	N/A
RheoCompass	Anton-Parr	N/A
SEDNTERP	John Philo	N/A
XCalibur	Thermo Scientific	N/A

Table S2: Resource Availability Table

Resource	Source	Location
Domain Summary Files	In House	Available Upon Request
Protein Metadata Files	In House	https://github.com/FriedLabJHU/Refoldability-Tools
PD v2.4 LFQ Output Files (.pdresult)	In House	PRIDE (PXD030869)
Peptide Quantifications (*_out.txt)	In House	Dryad (10.5061/dryad.bnzs7h4dg)
Protein Summary Data	In House	This Paper
Raw Mass Spectra Files (.raw)	In House	PRIDE (PXD030869)

Table S3: Proteome Discoverer Peptide Identification Node Parameters

	SequestHT	**MSFragger**
Peak Matching and Outputting		
Precursor Mass Tolerance	10 ppm	10 ppm
Fragment Mass Tolerance	20 ppm	20 ppm
Use Average Precursor Mass	FALSE	N/A
Use Average Fragment Mass	FALSE	N/A
Isotope Error	N/A	0/1/2
Localized Delta Mass	N/A	0
Delta Mass Exclude Ranges	N/A	(-1.5,3,5)
Max. Precursor Charge	N/A	4
Digestion		
Enzyme	Trypsin (semi)	Trypsin (semi)
Min. Peptide Length	6	7
Max. Peptide Length	144	50
Max. Missed Cleavage Sites	2	2
Min. Peptide Mass	N/A	500
Max. Peptide Mass	N/A	5000
Maximum Charge State for Theoretical Fragments to Match	N/A	2
Clip n Term M	N/A	TRUE
Spectrum Matching		
Use of Neutral Loss a Ions	TRUE	N/A
Use of Neutral Loss b Ions	TRUE	N/A
Use of Neutral Loss y Ions	TRUE	N/A
Use Ranking Ions	TRUE	N/A
Weight of a Ions	0	N/A
Weight of a Ions	1	N/A
Weight of a Ions	0	N/A
Weight of a Ions	0	N/A
Weight of a Ions	1	N/A
Weight of a Ions	0	N/A
Spectral Processing		
Min. Peaks	N/A	15
Min. Fragments Modeling	N/A	2
Min. Ratio	N/A	0.01
Use Top N Peaks	N/A	150
Min. Matched Fragments	N/A	3
Min. Clear m/z Range	N/A	0
Max. Clear m/z Range	N/A	0
Mass Callibration	N/A	On and find optimal parameters
Modifications		

Max. Equal Modifications Per Peptide	3	N/A
Multiple Variable Mods on Residue	N/A	TRUE
Max. Variable Mods per Peptide	N/A	3
Max. Variable Mods Combinations	N/A	5000
Mass Offsets	N/A	0
Dynamic Modifications		
Oxidation	+15.995 Da (M)	+15.995 Da (M)
Label: 13C6	N/A	+6.020 Da (K,R)
Acetyl	+42.011 Da (Any N-Terminus)	+42.011 Da (Any N-Terminus)
Static Modifications		
Carbamidomethyl	+57.021 Da (C)	+57.021 Da (C)

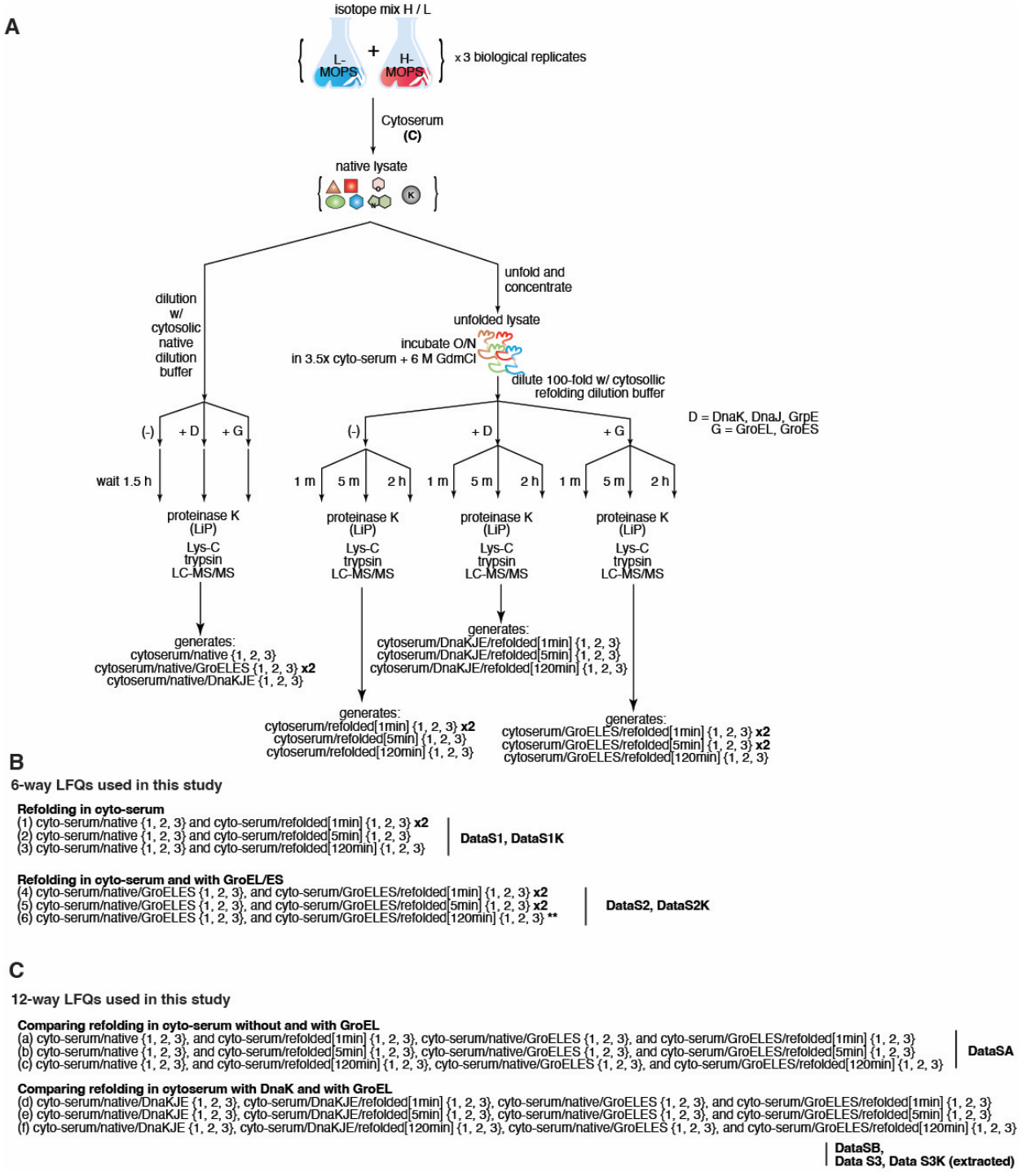


Fig. S1. Summary of all samples prepared for LC-MS/MS, and their combinations to perform label-free quantification (LFQ) analyses. (A) Experimental workflow to prepare the 51 samples for LC-MS/MS used in the final experiments published in this study. See Methods for more details. In brief: three *E. coli* cultures are grown in light MOPS media and three *E. coli* cultures are grown in heavy MOPS media. Pairs are mixed together, and cells are gathered by centrifugation. Pellets are resuspended in cyto-serum lysis buffer. The native samples are probed by limited proteolysis (LiP) with proteinase K (PK) after equilibration. The refolded

samples are probed similarly, but at 3 different timepoints following initiation of refolding by dilution (1 min, 5 min, and 120 min). The cyto-serum-lysed samples are either diluted in cyto-serum native dilution buffer to generate cyto-serum/native samples, or diluted in cyto-serum native dilution buffers supplemented with GroEL/ES or DnaK/J/E. Following equilibration, they are probed with proteinase K. Alternatively, cyto-serum lysates are unfolded into 6 M GdmCl, and refolded by 100-fold dilution into cyto-serum refolding buffer, either supplemented with GroEL/ES, DnaK/J/E, or neither, and given either 1 min, 5 min, or 120 min to refold prior to interrogation with PK. In all cases, immediately following 1 min of LiP, samples are quenched by boiling, fully trypsinized with LysC and trypsin, and prepared for LC-MS/MS. **(B)** Summary of the six 6-way LFQs used in this study, and which set of six samples are analyzed together to generate the peptide refolded/native quantifications. Fig. 2A (**1**); Fig. 2B (**4**); Fig. 2C (**2, 5**); Fig. 2D,E (**2** left, **5** right); Fig. 2F,G (**2**, **5**, (15)); Fig. 4A,C,E,G (**1**–both reps., **2**, **3**); Fig. 4B,D,F,H (**4**–both reps., **5**–both reps.); Fig. 4I (**2**, **5**); Fig. 5B (**1–6** & see below); Fig. 5C (combination of **4**, **5**). **The 2 h timepoint for GroEL/ES refolding was generally not used. **(C)** Summary of the six 12-way LFQs used in this study, and which set of twelve samples are analyzed together to generate peptide refolded/native quantifications. Fig. 3A-H **(b)**; Fig. 5A (**e**– extracting out the DnaK subexperiment); Fig. 5B (**d**, **e**, **f**, & see above); Fig. 5D (combination of **d**, **e**); Fig. 5E-H **(e)**; Fig. 5J **(b)**.

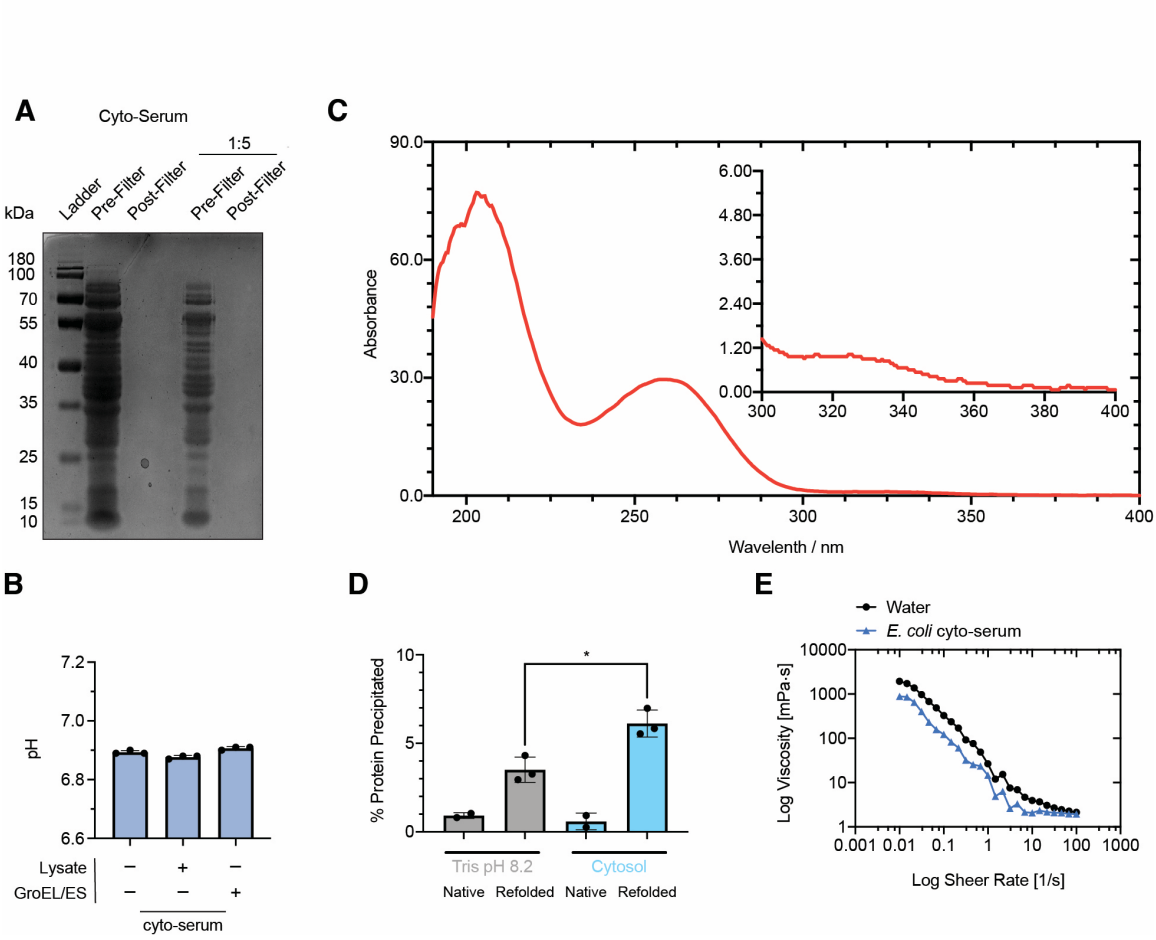


Fig. S2. Characterization of cyto-serum. (A) Coomassie staining of SDS-PAGE of cyto-serum pre and post filtration with Viva-Spin 15R 2K MWCO Filter (Sartorius) as a 1x and a 1:5 dilution. **(B)** Bar chart showing the pH readings of 3 independent preparations of cyto-serum with the addition of either *E. coli* lysate or GroEL/ES. **(C)** UV-VIS spectra obtained of 1x cyto-serum. **(D)** Bar charts showing the quantification of protein aggregation of native and refolded samples in Tris pH 8.2 and cyto-serum using BCA Assay. Refolding in cyto-serum resulted in a small but significant increase in detected protein precipitation upon refolding ($P < 0.05$ by Welch's t-test) when compared to refolding in Tris pH 8.2. **(E)** Log-log diagram showing viscosity of cyto-serum and water as a function of sheer rate. *E. coli* cyto-serum is a non-viscous fluid with rheometric properties similar to water.

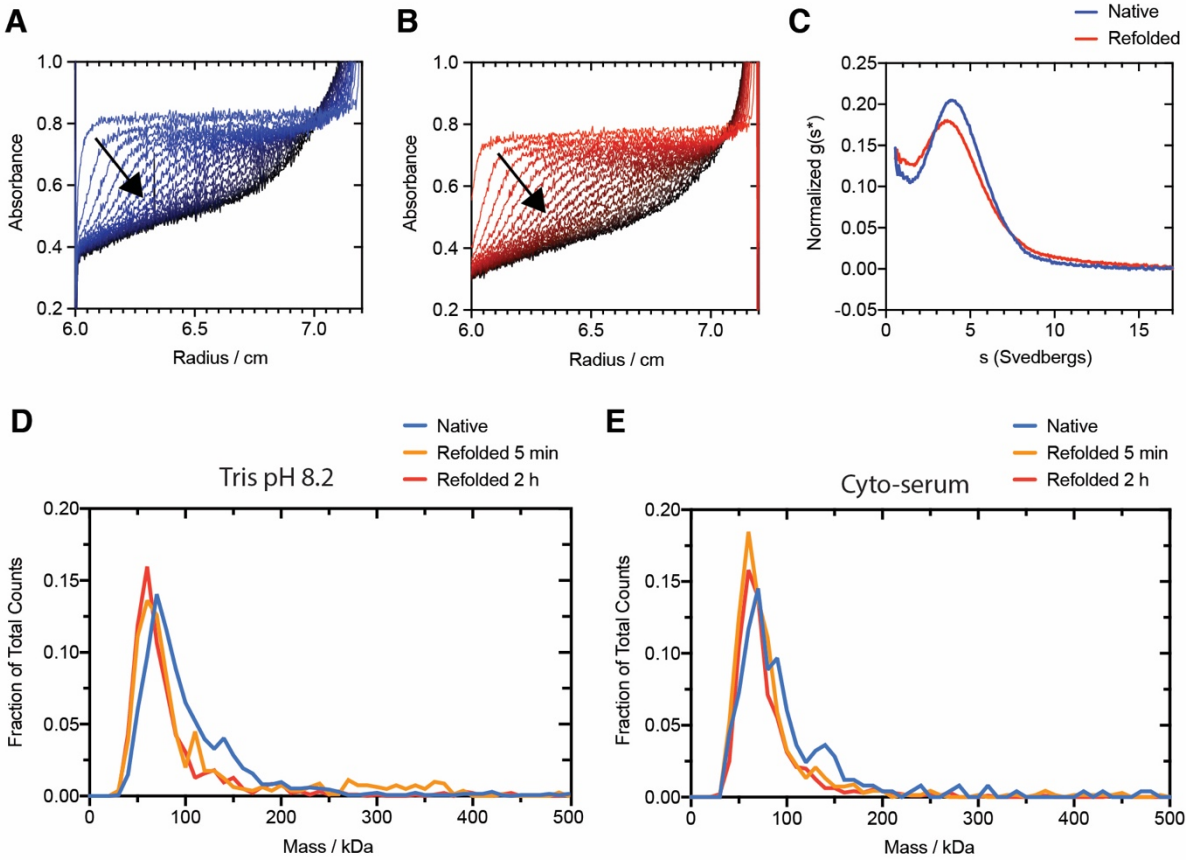


Fig. S3. Low aggregation during global refolding reactions. (A-B) Absorbance at 280 nm as a function of radius along the rotor during sedimentation velocity analytical ultracentrifugation of native **(A)** and refolded **(B)** *E. coli* lysates in Tris pH 8.2. Data from first 100 scans are shown, with each subsequent line representing every 5th scan in directionality of arrow. These datasets are representative of two independent spins on two separately prepared native and refolded lysates. **(C)** Calculated sedimentation coefficient distributions of native and refolded E. coli lysates in Tris pH 8.2 determined using dcdt+. Sedimentation coefficients were corrected to 20 ˚C in water using density, viscosity, and partial specific volume values calculated in SEDNTERP (3, 4). These datasets are representative of two independent spins on two separately prepared native and refolded lysates. **(D-E)** Normalized mass distributions of native and refolded *E. coli* lysates (5 min and 2 h) in Tris pH 8.2 **(D)** and cyto-serum **(E)** as determined by Mass Photometry (MP; 5, 6). All three sample types show overlaying mass distributions in both refolding buffers (Tris or cyto-serum), indicative that there are minimal differences in soluble aggregation between native and refolded samples.

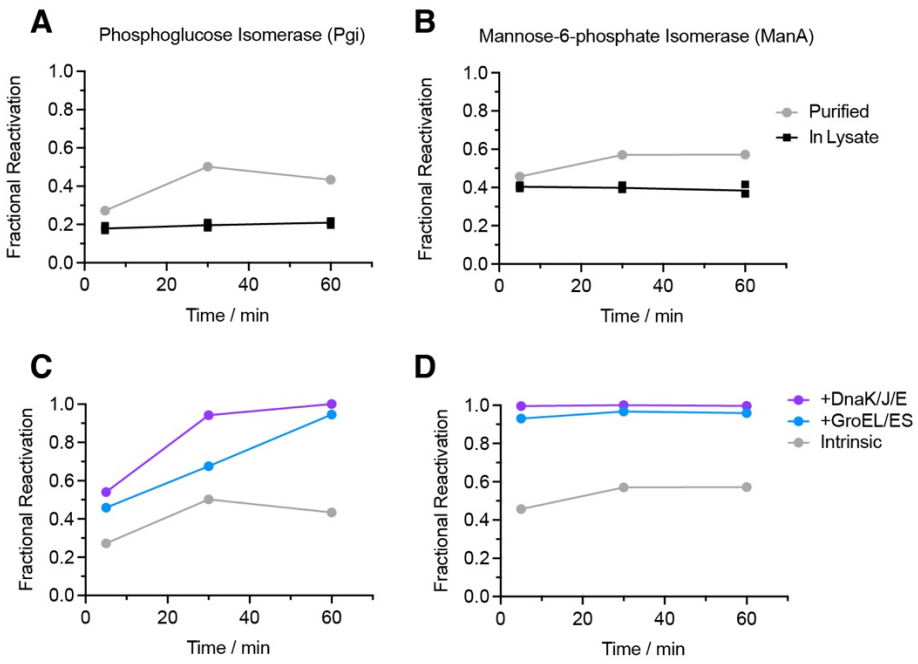


Fig. S4. Enzyme Reactivation Assays on Purified Enzymes and in the *E. coli* Lysate. Activity was measured by visible absorption (450 nm) of a coupled enzyme-assay whose signal is proportional to NADH concentration (see *SI Methods*). **(A)** Fractional activity of Phosphoglucose isomerase (Pgi) recovered after unfolding in 6 M GdmCl (either purified enzyme (gray) or as a component within total *E. coli* lysates (black)) and diluting 50-fold (100-fold in lysate). **(B)** Fractional activity of Mannose-6-phosphate isomerase (ManA) recovered after unfolding in 6 M GdmCl (either purified enzyme (gray) or as a component within total *E. coli* lysates (black)) and diluting 50-fold (100-fold in lysate). **(C)** Activity of purified Pgi recovered after diluting without chaperone (gray), or supplemented with GroEL/ES (blue; 4 μM and 8 μM final concentrations of protomers) or DnaK/J/E (violet; 5 μM, 1 μM, and 1 μM final concentrations of protomers). **(D)** As in panel **C**, except for ManA.

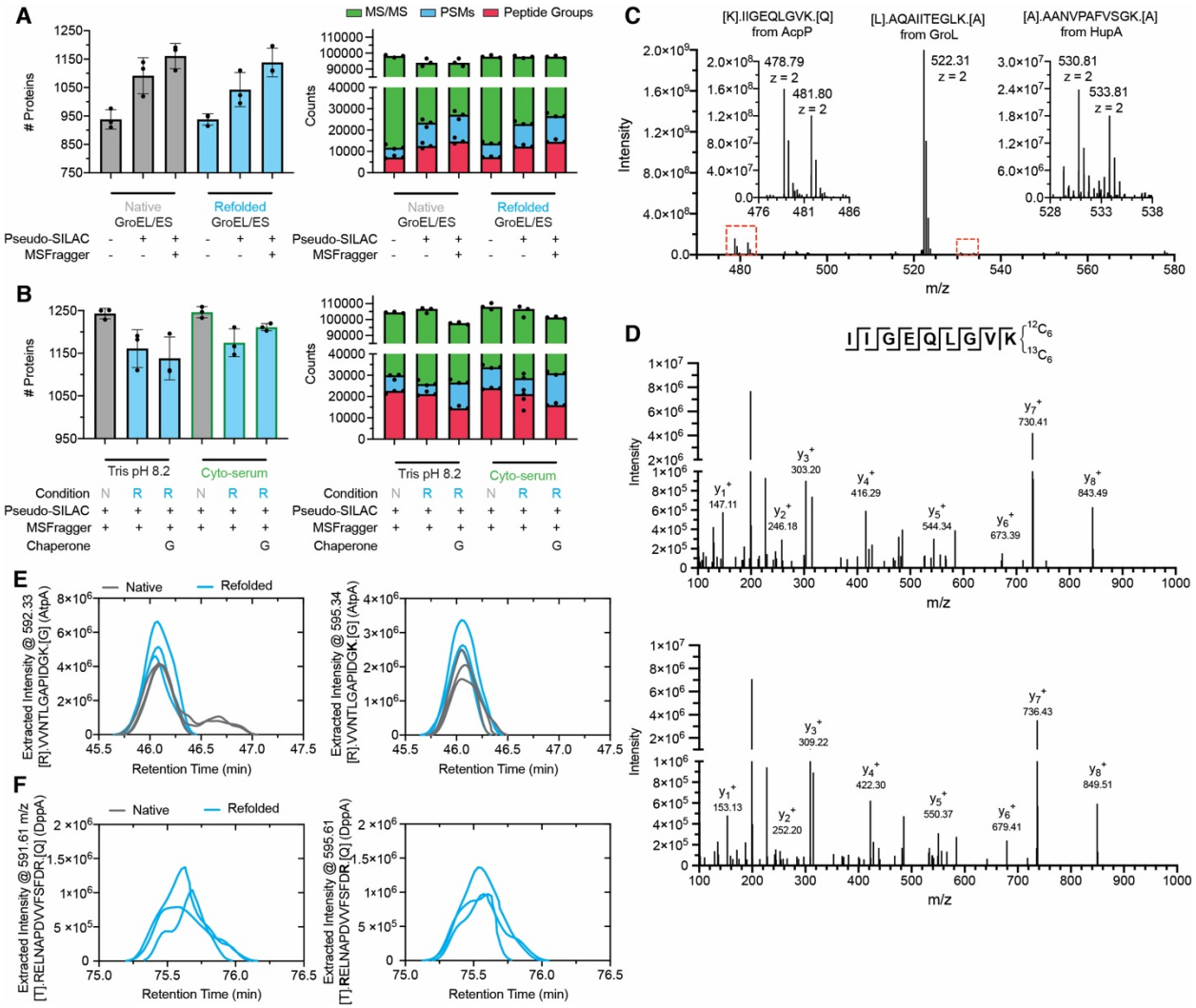


Fig. S5. Pseudo-SILAC to distinguish client-derived peptides from chaperone-derived peptides. **(A)** Bar charts showing the number of proteins and peptide groups identified, peptide-spectrum matches (PSM), and total MS/MS spectra obtained on individual runs, testing the effect of the pseudo-SILAC method and the MSFragger search algorithm. The number of MS/MS spectra acquired per run is similar, but pseudo-SILAC increases the number of unique PSMs and peptide groups. **(B)** Bar charts showing that the combination of pseudo-SILAC and MSFragger (16) results in no significant loss in coverage in experiments conducted in cyto-serum and with GroEL/ES. **(C)** A sample MS1 spectrum from a refolding experiment with GroEL/ES in cyto-serum. Peptides derived from refolded proteins, but not from chaperone, display twin-peaks separated by 3 Th. **(D)** Sample MS2 fragmentation spectra from two co-eluting peptides that differ only by the isotopic composition of the C-terminal lysine. The y-ions (indicated) are all displaced by 6 Th. **(E)** Extracted ion chromatograms for the peptide indicated (from AtpA) in three replicate native samples and in three replicate refolded samples, at two m/z's corresponding to the light and heavy-substituted isotopes. The abundance of the peptide is similar in the native and refolded forms, in both isotope states, implying that this region of AtpA properly refolded; i.e., had the same PK susceptibility in both forms. **(F)** Similar to panel E but for an all-or-nothing peptide from DppA which is not detected in any of the three native replicates at the m/z for both the light- and heavy-substituted isotope.

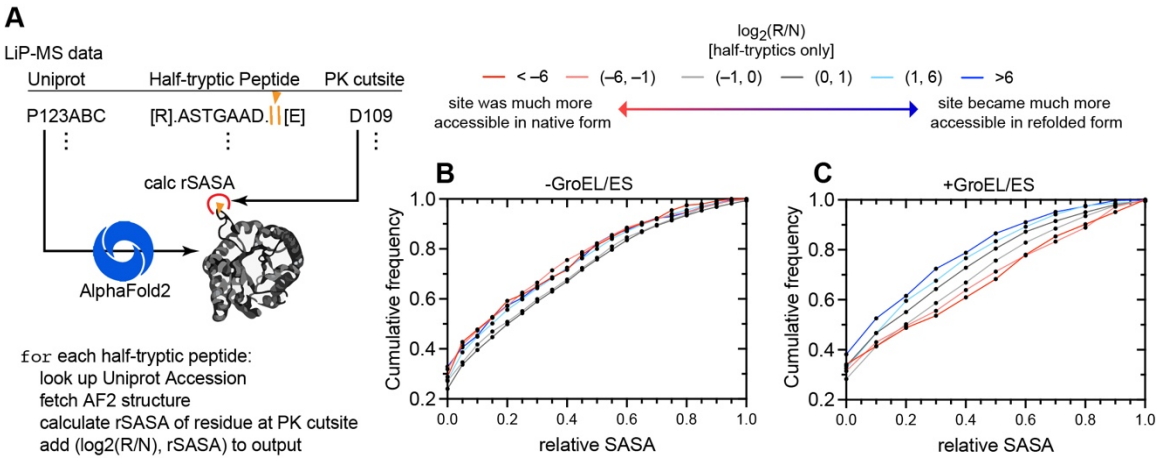
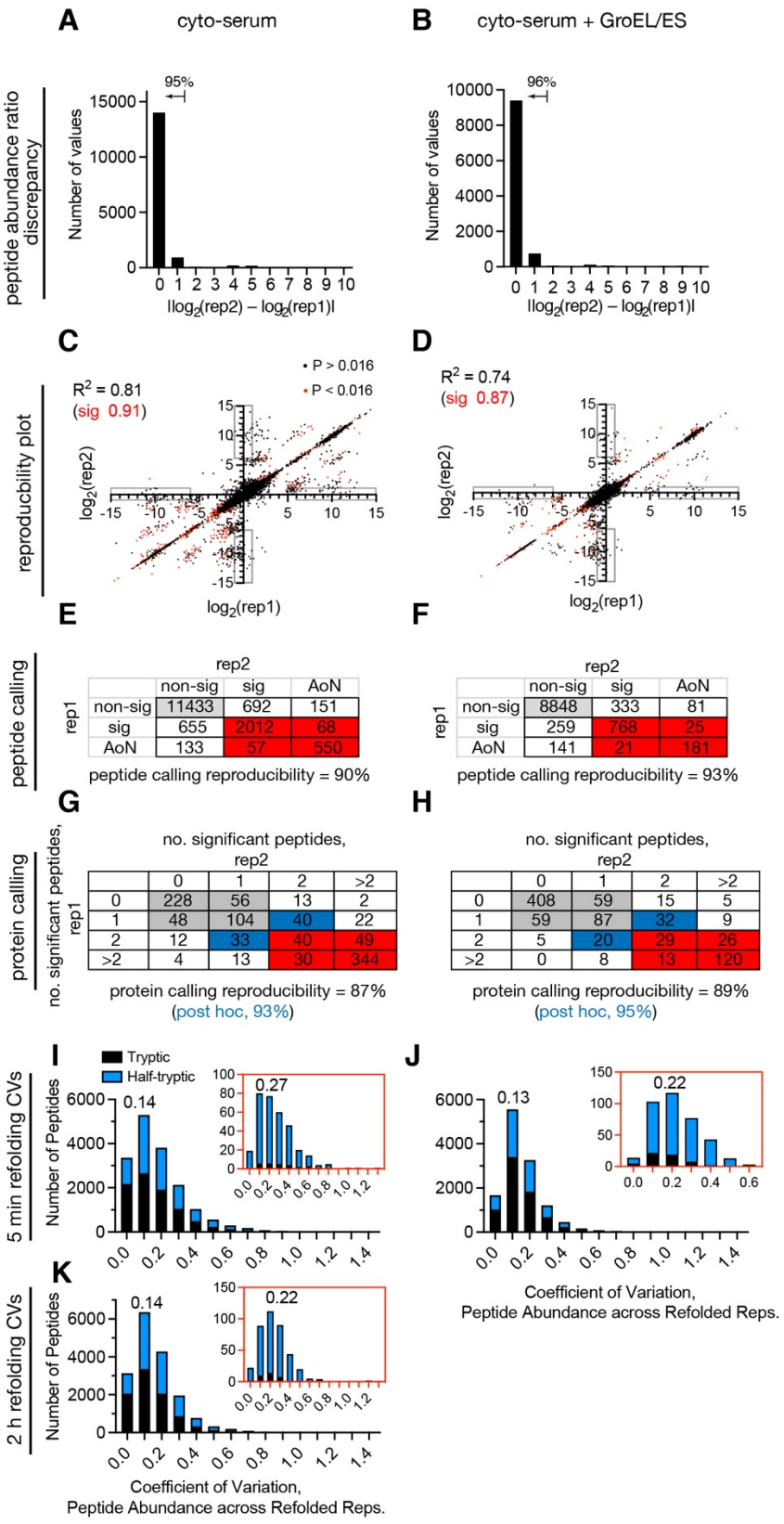


Fig. S6. Half-tryptic Sites Associated with Large Changes in Proteolytic Susceptibility are More Buried in Native Protein Structures. (A) Summary of computational workflow for this analysis. For each half-tryptic peptide sequenced, the relative solvent accessible surface area (rSASA) was calculated at the inferred Proteinase K cut-site in the context of the native protein structure, as predicted by AlphaFold2. **(B, C)** rSASA cumulative frequency distributions for six categories of half-tryptic peptides, separated by their observed change in proteolytic susceptibility from the native to the refolded form, for **(B)** refolding in cyto-serum without chaperones (5 min time-point) and **(C)** refolding in cyto-serum with GroEL/ES (5 min time-point). Particularly with GroEL/ES, sites that became much more accessible in the refolded form were very buried in their native structures (dark blue trace; median rSASA is 0.133), whilst sites that were more accessible in the native form were more exposed to begin with (dark red trace; median rSASA 0.273).



E	rep2 non-sig	rep2 sig	rep2 AoN
rep1 non-sig	11433	692	151
rep1 sig	655	2012	68
rep1 AoN	133	57	550

F	rep2 non-sig	rep2 sig	rep2 AoN
rep1 non-sig	8848	333	81
rep1 sig	259	768	25
rep1 AoN	141	21	181

G	0	1	2	>2
0	228	56	13	2
1	48	104	40	22
2	12	33	40	49
>2	4	13	30	344

H	0	1	2	>2
0	408	59	15	5
1	59	87	32	9
2	5	20	29	26
>2	0	8	13	120

Fig. S7. Reproducibility analysis. (A-B) Histograms showing the peptide quantification discrepancies between two replicates of the experiment in which proteins were refolded for 1 min in **(A)** cyto-serum, or **(B)** cyto-serum with GroEL/ES. These correspond to the two replicates of LFQ 1 and 4 from Fig. S1B. Note that each of these replicates of the experiment involved three separate biological replicates of native and refolded. Peptides that were identified in both experiments were collected and the refolded/native ratio in each replicate was compared to each other. Histograms show the absolute value of the difference of the log2 quantifications. **(A)** 15751 peptides were identified in common, of which 89% were within 1.4-fold and 95% were within 3.8-fold. **(B)** 10564 peptides were identified in common, of which 89% were within 1.4-fold and 96% were within 3.8-fold. **(C-D)** Scatter plots showing the relationship between the peptide log2(refolded/native) quantification in one replicate versus its value in the other replicate for two replicates of the experiment in which proteins were refolded for 1 min in **(C)** cyto-serum, or (**D**) cyto-serum with GroEL/ES. Points in red were considered significant ($P < 0.016$ by Welch's t-test) in both experiments. The coefficients of determination (R^2) are given first for all points in common (black), and then for the subset of points that were considered significant in both replicates of the experiment (red). In all cases, R^2 is greater when only significant peptides are considered (which are the only ones used to call a protein non-refoldable). The gray boxes demarcate regions in which upon separate performances of the experiment, an all-or-nothing peptide is categorized as nonsignificant in the other. Importantly, these boxes have very few red points. **(E-F)** Calling reproducibility of peptides (classified as either non-significant, significant, or all-or-nothing (AoN)) between two replicates of the experiment in which proteins were refolded for 1 min in **(E)** cyto-serum, or **(F)** cyto-serum with GroEL/ES. **(G-H)** Calling reproducibility of proteins between two replicates of the experiment in which proteins were refolded for 1 min in **(G)** cyto-serum, or **(H)** cyto-serum with GroEL/ES. Rows correspond to the number of peptides that were significantly different between native and refolded samples in the first replicate of the experiment, and columns correspond to the number of peptides that were significantly different in the duplication. Numbers in the table correspond to the number of proteins with that many significant peptides in each replicate. Gray cells correspond to proteins that would be called refoldable in both iterations. Red cells correspond to proteins that would be called nonrefoldable in both iterations. Cells in white would have been called differently, resulting in reproducibility from 87–89%. In all comparisons, we exclude proteins that only differ by one significant peptide at the cut-off, shown as blue cells. With these proteins removed *post hoc*, reproducibility increases to 93–95%. **(I-J)** Histograms of the coefficients of variation (CV) for the peptide abundances in refolded samples, from 3 independent refolding reactions, after 5 min of refolding for experiments in which cells were lysed and refolded in either **(I)** cyto-serum, or **(J)** cyto-serum with GroEL/ES. Insets in red correspond to the CV histograms for the peptides detected only in the refolded samples (which are almost all half-tryptic). Numbers represent medians of distributions. **(K)** Same as panels I, except for refolding after 2 h.

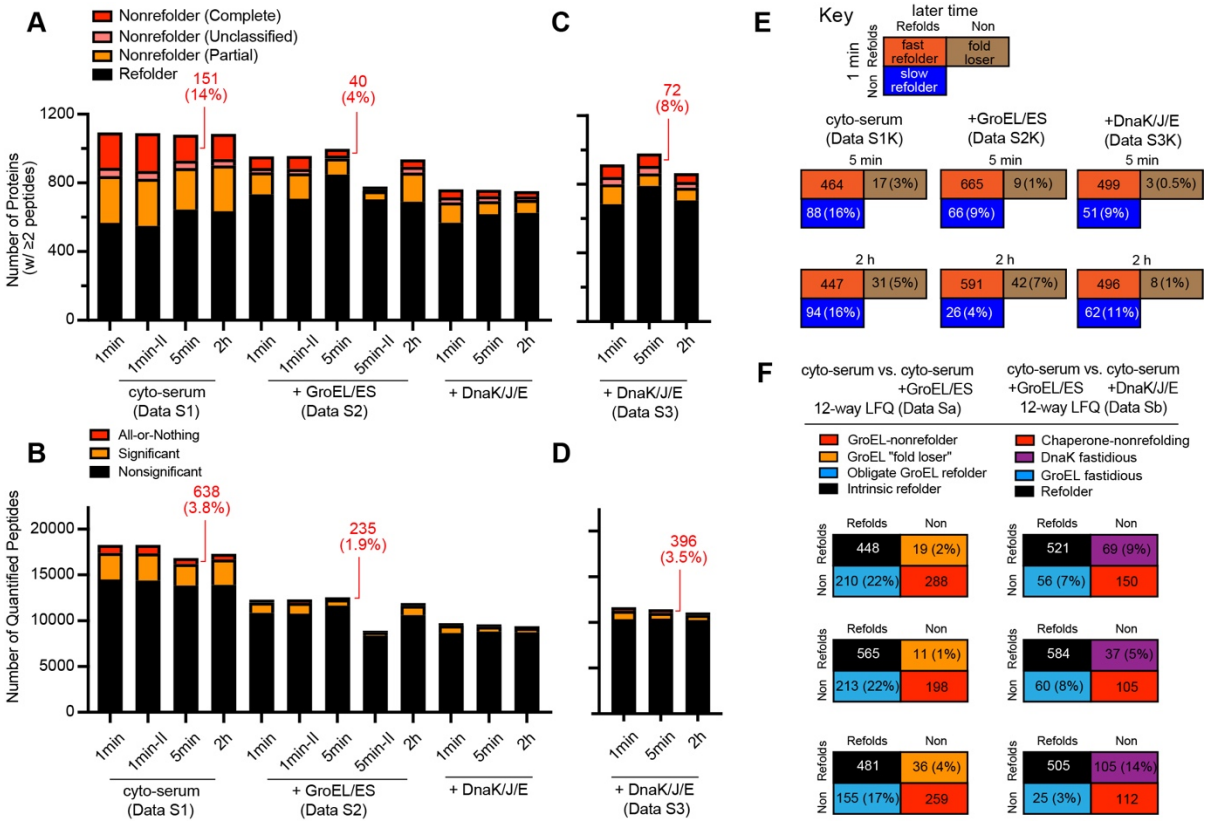


Fig. S8. Summary statistics of all 6-way LFQs, kinetic comparisons, and truth tables for condition comparisons based on 12-way LFQs. (A) The number of proteins assessed in each 6-way LFQ, categorized as either refolding (0 or 1 peptide quantified with significantly different abundance between native and refolded), partial nonrefolding (2 or more peptides quantified with significantly different abundance between native and refolded but fewer than 2 all-or-nothing peptides), and complete nonrefolding (2 or more all-or-nothing peptides). Some nonrefolding proteins do not classify between the subcategories (if they have 1 significant and 1 all-or-nothing peptide). Proteins with only 1 peptide quantified are not included. Bars correspond, in order, to 6-way LFQs labeled **#1–6** in Fig. S1B. 6-way LFQs for DnaK/J/E are not used for analysis (see main text and methods). In red, are number (and percentage) of proteins that are judged complete nonrefolding. **(B)** The number of peptides confidently quantified in each 6-way LFQ, categorized as either nonsignificant, significant, or all-or-nothing. Bars correspond, in order, to the 6-way LFQs labeled **#1–6** in Fig. S1B. In red, are number (and percentage) of peptides that are all-or-nothing. **(C)** The number of proteins assessed in each 12-way LFQ (**#d–f** in Fig. S1C), with the identifications and quantifications for the DnaK channels extracted out, done to increase coverage in the DnaK experiments (see main text). Categorizations same as panel **A**. **(D)** The number of peptides confidently quantified in each 12-way LFQ (**#d–f** in Fig. S1C). Categorizations same as panel **B**. **(E)** Summary of all kinetics experiments. To assess kinetics, we perform a comparison of two 6-way LFQs that correspond to distinct refolding timepoints but for otherwise identical conditions. To be included, a protein must have two or more confidently quantified peptides at both timepoints, be assessed as refoldable in one of the two time points, and cannot differ by only one significant peptide between the two timepoints. Each protein is designated as either a fast refolder, slow refolder, or fold loser; the number of such proteins is given for each kinetic comparison, according to the key. For the top row, from left to right, the data used for each comparison correspond to: **#1** & **2**; **#4** & **5** (Fig. S1B) and **#d** & **e** (Fig. S1C). For the bottom row, from left to right, the data used for each comparison correspond to **#1** & **3**; **#4**

& **6** (Fig. S1B) and **#d** & **f** (Fig. S1C)**. (F)** Summary of all condition comparison experiments. To assess the effect of changing refolding condition, we perform 12-way LFQs that merge the native and refolded (at a given timepoint) replicates for the two conditions being compared. To be included, the protein must have two or more confidently quantified peptides in both conditions, and cannot differ by only one significant peptide between the two conditions. Two types of comparisons were performed (columns): cyto-serum with and without GroEL/ES, GroEL/ES vs. DnaK/J/E in cyto-serum. The color code for the designations associated with each comparison are given, and the truth tables give the number of proteins in each designation. Each comparison was conducted at three timepoints. The data used correspond to **#a–f** (Fig. S1C).

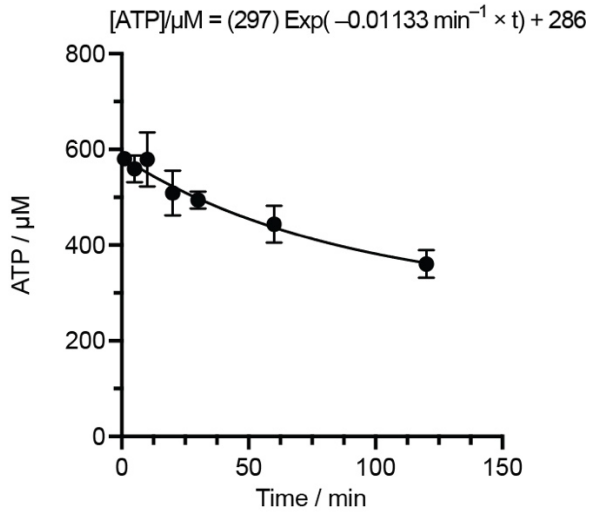


Fig S9. ATP Hydrolysis in global refolding reactions conducted in cyto-serum, supplemented with GroEL/ES (4 μM, 8 μM protomer concentrations). ATP measured by luciferase assay.

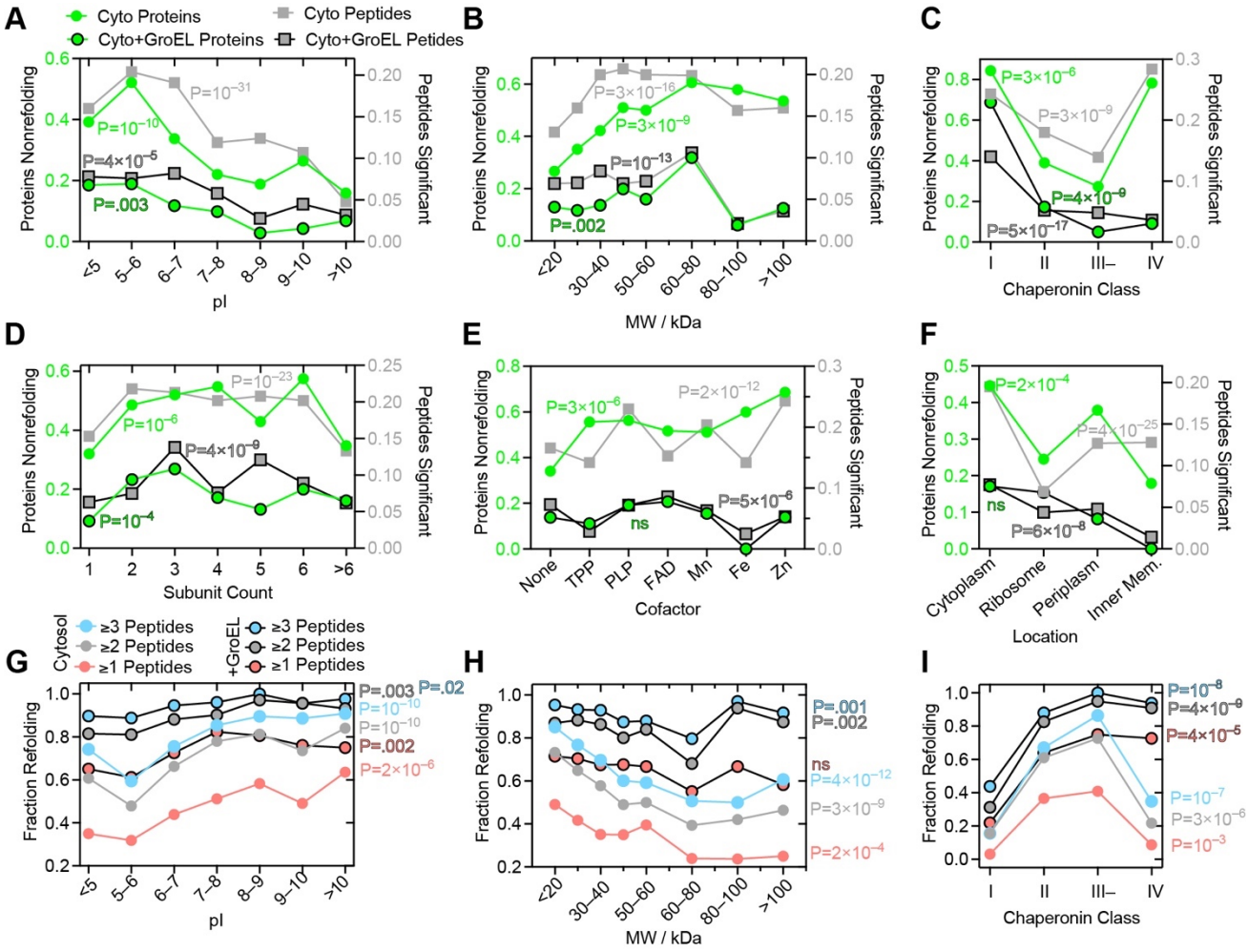


Fig. S10. Peptide-level analyses and comparison to the protein-level. (A) Fraction of proteins that do **NOT** refold in either cyto-serum (green circles), or cyto-serum with GroEL/ES (green circles, black border), separated based on individual proteins' isoelectric point (pI) (left y-axis). Additionally shown is fraction of peptides that have significantly different abundance in refolded samples, lumped together for all proteins within the given pI tranche (in either the cyto-serum experiment (gray boxes) or the cyto-serum with GroEL/ES experiment (gray boxes, black border)) (right y-axis). P-values according to the chi-square test are given on the protein frequencies in green (black border for GroEL/ES) and on the peptide frequencies in gray (black border for GroEL/ES). Proteins with low pI tend to be more nonrefoldable *and* tend to generate significant peptides at a much higher frequency; the trend prevails across the series with protein nonrefoldability fraction tracking closely with the peptide significance rate. This implies that the trend is *robust*, and not a coverage artefact. Data come from #**2**, **5** in Fig. S1B. **(B)** Same as panel **A**, except proteins and peptides are separated based on the protein's molecular weight (MW). Trends associated with MW are robust. Data come from same 6-way LFQs as panel **A**. **(C)** Same as panel **A**, except proteins and peptides are separated based on the protein's chaperonin class (13, 14). Trends associated with chaperonin class are robust. Data come from same 6-way LFQs as panel **A**. **(D)** Same as panel **A**, except proteins and peptides are separated based on the protein's subunit count. In the cytosol, monomeric and large complexes refold the most efficiently, a robust trend. With GroEL/ES, overall trends with respect to subunit count are less substantial, though it appears to have an outsized importance on tetrameric and hexameric proteins. Data come from the same 6-way LFQs as panel **A**. **(E)** Same as panel **A**, except proteins and peptides are separated based on the which cofactors the protein harbours. Data come from the same 6-way LFQs as panel **A**. **(F)** Same as panel **A**, except proteins are peptides

are separated based on the protein's cellular localization. Trends associated with location are robust. GroEL/ES is effective on proteins in all locations, except ribosomal proteins. **(G-I)** Sensitivity analyses showing the fraction of proteins refolding in either cyto-serum (solid circles) or in cyto-serum with GroEL/ES (solid circles, black borders), as a function of the number of significant peptides required to call a protein nonrefoldable (≥1, red; ≥2, gray (the standard cutoff); ≥3, blue). P-values according to the chi-square test are given in matching colors (for the various cutoff schemes) and with black borders for GroEL/ES. Whilst all trends are maintained irrespective of cutoff, statistical significances generally fall with the ≥1 cutoff (red), because it assigns too much weight to a single significant peptide.

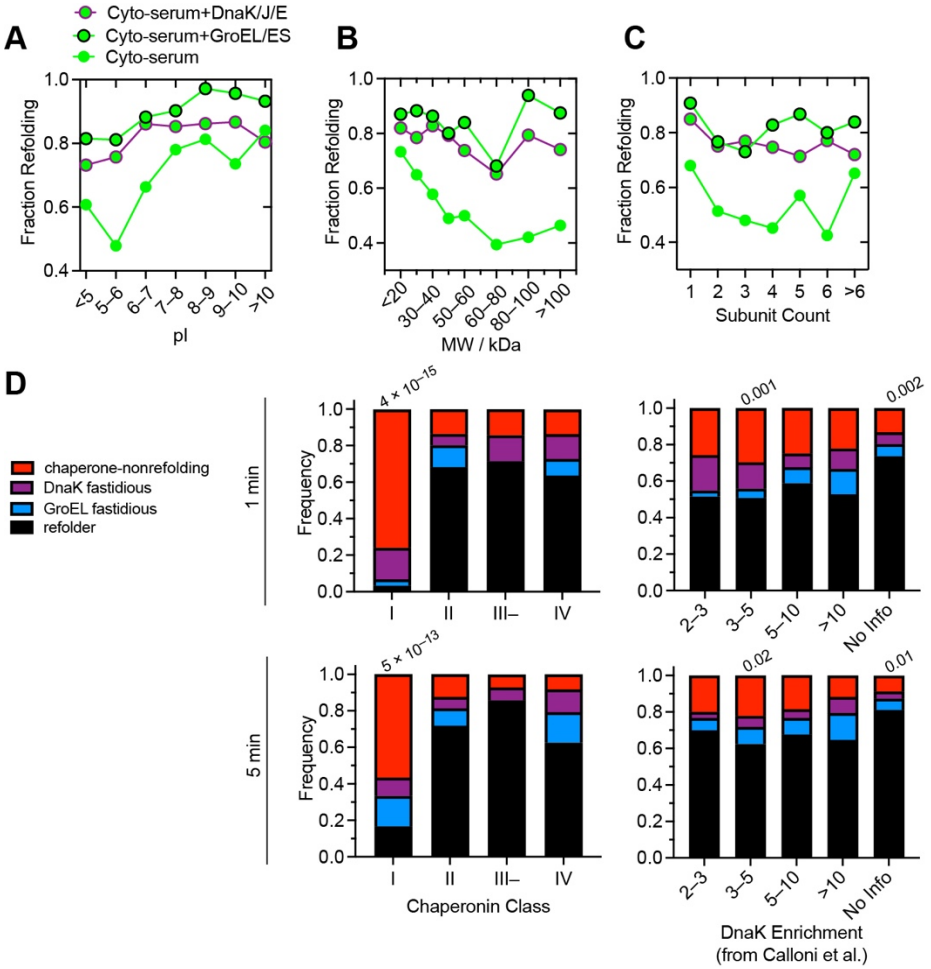


Fig. S11. Further properties of DnaK refolders. (A) Fraction of proteins that refold in either cyto-serum (green), cyto-serum with GroEL/ES (green, black border), or cyto-serum with DnaK/J/E (green, purple borer) separated based on proteins' pI. Data come Nissley et al. (13), **#2** in fig. S1B and **#e** in fig. S1C. **(B)** As **A**, except proteins are separated based on proteins' molecular weight (MW). **(C)** As **A**, except proteins are separated based on the number of subunits in the complex to which they are part. **(D)** Frequency of proteins that refolded in both conditions (black), only with GroEL/ES (light blue), only with DnaK/J/E (purple), or did not refold with either (chaperone-nonrefolder; red), separated on the basis of GroEL chaperonin class (from Kerner et al. (13), *Left*) or on the basis of DnaK enrichment level (from Calloni et al. (17), *Right*). Data from the 1 min time-point and 5 min time-point are given on *Top* and *Below*. Data come from 12-way LFQs (**#d, e** in fig. S1C). Numbers indicate P-values according to chi-square test.

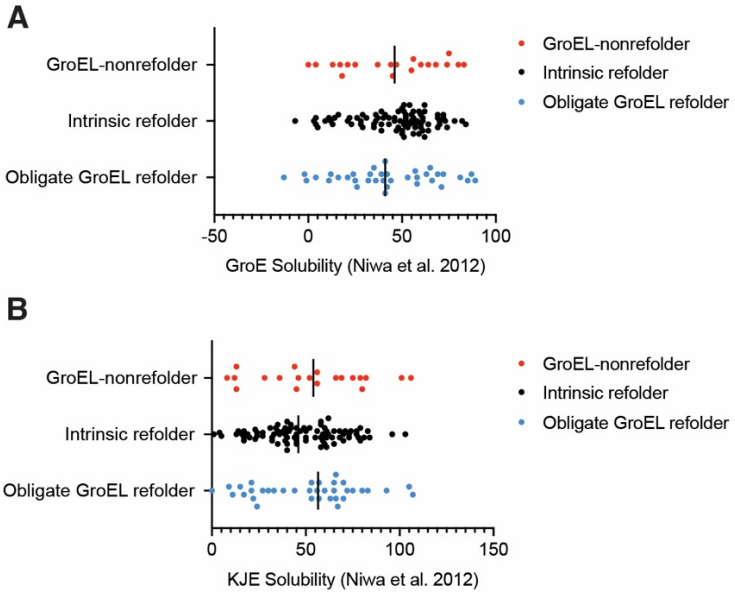


Fig. S12. Comparison of GroEL refoldability to cotranslational aggregation assays. Niwa et al. over-expressed 987 *E. coli* proteins individually by *in vitro* translation and measured the percent of the protein that aggregated, without chaperones, or with GroEL/ES, or with DnaK/J/E (18, 19). 141 of the proteins in **#b** (Fig. S1C, Data SA) were amongst that cohort, and are displayed here with their Δsolubility for GroEL/ES (panel A) and DnaK/J/E (panel B), divided by their refolding category (see Fig. 3; Δsolubility = %soluble(+chaperone) – %soluble(–chaperone)).

Data SA. Cytoserum_v_CytoserumGroEL_12waySummaryData.xlsx
Summary data for 12 way LFQ pooling together replicates from refolding reactions in cytoserum and from refolding reactions in cytoserum supplemented with GroEL/ES. Tabs correspond to different refolding times. See Materials and Methods, Fig. S1C and Fig. S6F for explanation.

Data SB. CytoserumDnaK_v_CytoserumGroEL_12waySummaryData.xlsx
Summary data for 12 way LFQ pooling together replicates from refolding reactions in cytoserum supplemented with DnaK/DnaJ/GrpE and from refolding reactions in cytoserum supplemented with GroEL/ES. Tabs correspond to different refolding times. See Materials and Methods, Fig. S1C and Fig. S6F for explanation.

Data S1. Cytoserum_RefoldingSummaryData.xlsx
Summary data for 6 way LFQ pooling together replicates from refolding reactions in cytoserum. Tabs correspond to different refolding times and technical replicates. See Materials and Methods, Fig. S1B, and Fig. S6A-B for explanation.

Data S2. CytoserumGroEL_RefoldingSummaryData.xlsx
Summary data for 6 way LFQ pooling together replicates from refolding reactions in cytoserum supplemented with GroEL/ES. Tabs correspond to different refolding times and technical replicates. See Materials and Methods, Fig. S1B, and Fig. S6A-B for explanation.

Data S3. CytoserumDnaK_RefoldingSummaryData.xlsx
Summary data for 12 way LFQ pooling together replicates from refolding reactions in cytoserum supplemented with DnaK/DnaJ/GrpE and GroEL/ES, with channels corresponding to DnaK/DnaJ/GrpE samples extracted. Tabs correspond to different refolding times and technical replicates. See Materials and Methods, Fig. S1B, and Fig. S6C-D for explanation.

Data S1K. Cytoserum_RefoldingKineticComparison.xlsx
Kinetic comparison of different time points in Data S1. Slow folders are defined as proteins that are nonrefoldable at 1 min but refoldable at later time points. Tabs correspond to different refolding times and technical replicates. See Materials and Methods, Fig. S1B, and Fig. S6E for explanation.

Data S2K. CytoserumGroEL_RefoldingKineticComparison.xlsx
Kinetic comparison of different time points in Data S2. Slow folders are defined as proteins that are nonrefoldable at 1 min but refoldable at later time points. Tabs correspond to different refolding times and technical replicates. See Materials and Methods, Fig. S1B, and Fig. S6E for explanation.

Data S3K. CytoserumDnaK_RefoldingKineticComparison.xlsx
Kinetic comparison of different time points in Data S3. Slow folders are defined as proteins that are nonrefoldable at 1 min but refoldable at later time points. Tabs correspond to different refolding times and technical replicates. See Materials and Methods, Fig. S1B, and Fig. S6E for explanation.

SI References

1. Y. Feng, G. D. Franceschi, A. Kahraman, M. Soste, A. Melnik, P. J. Boersema, P. P. de Laureto, Y. Nikolaev, A. P. Oliveira, P. Picotti, Global analysis of protein structural changes in complex proteomes. *Nature Biotech*. **32**, 1036–1044 (2014).
2. P. To, B. Whitehead, H. E. Tarbox, S. D. Fried, Nonrefoldability is Pervasive Across the E. coli Proteome. *J. Am. Chem. Soc*. **143**, 11435–11448 (2021).
3. J. S. Philo, Improved methods for fitting sedimentation coefficient distributions derived by time-derivative techniques. *Anal Biochem*. **354**, 238–246 (2006).
4. T. M. Laue, B. D. Shah, T. M. Ridgeway, S. L. Pelletier, Computer-aided interpretation of analytical sedimentation data for proteins. In: S. Harding, A. Rowe, J. Hoarton, editors. Analytical Ultracentrifugation in Biochemistry and Polymer Science. *Royal Society of Chemistry; Cambridge, UK*. 90-125, (1992).
5. D. Cole, G. Young, A. Weigel, A. Sebesta, P. Kukura, Label-Free Single-Molecule Imaging with Numerical-Aperture-Shaped Interferometric Scattering Microscopy. *Acs Photonics*. **4**, 211–216 (2017).
6. G. Young, N. Hundt, D. Cole, A. Fineberg, J. Andrecka, A. Tyler, A. Olerinyova, A. Ansari, E. G. Marklund, M. P. Collier, S. A. Chandler, O. Tkachenko, J. Allen, M. Crispin, N. Billington, Y. Takagi, J. R. Sellers, C. Eichmann, P. Selenko, L. Frey, R. Riek, M. R. Galpin, W. B. Struwe, J. L. P. Benesch, P. Kukura, Quantitative mass imaging of single biological macromolecules. *Science.* **360**, 423–427 (2018)
7. I. M. Keseler, A. Mackie, A. Santos-Zavaleta, R. Billington, C. Bonavides-Martnez, R. Caspi, C. Fulcher, S. Gama-Castro, A. Kothari, M. Krummenacker, M. Latendresse, L. Muiz-Rascado, Q. Ong, S. Paley, M. Peralta-Gil, P. Subhraveti, D. A. Velzquez- Ramrez, D. Weaver, J. Collado-Vides, I. Paulsen, P. D. Karp, The EcoCyc database: reflecting new knowledge about Escherichia coli K-12. *Nucleic Acids Res.* **45**, D543–40 D550 (2017).
8. G.-W. Li, D. Burkhardt, C. Gross, J. S. Weissman, Quantifying Absolute Protein Synthesis Rates Reveals Principles Underlying Allocation of Cellular Resources. *Cell*. **157**, 624–635 (2014).
9. A. P. Pandurangan, J. Stahlhacke, M. E. Oates, B. Smithers, J. Gough, The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. *Nucleic Acids Res*. **47**, gky1130- (2018).
10. H. Cheng, R. D. Schaeffer, Y. Liao, L. N. Kinch, J. Pei, S. Shi, B.-H. Kim, N. V. Grishin, ECOD: An Evolutionary Classification of Protein Domains. *PLoS Comput. Biol*. **10**, e1003926 (2014).
11. H. Mi, A. Muruganujan, X. Huang, D. Ebert, C. Mills, X. Guo, P. D. Thomas, Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat Protoc*. **14**, 703–721 (2019).
12. L. P. Kozlowski, Proteome-pI: proteome isoelectric point database. Nucleic Acids Res. **45**, D1112–D1116 (2017).
13. M. J. Kerner, D. J. Naylor, Y. Ishihama, T. Maier, H.-C. Chang, A. P. Stines, C. Georgopoulos, D. Frishman, M. Hayer-Hartl, M. Mann, F. U. Hartl, Proteome-wide Analysis of Chaperonin-Dependent Protein Folding in Escherichia coli. *Cell*. **122**, 209–220 (2005).
14. K. Fujiwara, Y. Ishihama, K. Nakahigashi, T. Soga, H. Taguchi, A systematic survey of in vivo obligate chaperonin-dependent substrates. *EMBO J*. **29**, 1552–1564 (2010).
15. D. A. Nissley, Y. Jiang, F. Trovato, I. Sitarik, K. B. Narayan, P. To, Y. Xia, S. D. Fried, E. P. O'Brien, Universal protein misfolding intermediates can bypass the proteostasis network and remain soluble and less functional. *Nat Commun*. **13**, 3081 (2022).
16. A. T. Kong, F. V. Leprevost, D. M. Avtonomov, D. Mellacheruvu, A. I. Nesvizhskii, MSFragger: ultrafast and comprehensive peptide identification in shotgun proteomics. *Nature Methods* **14**, 513–520 (2017).
17. G. Calloni, T. Chen, S. M. Schermann, H. Chang, P. Genevaux, F. Agostini, G. G. Tartaglia, M. Hayer-Hartl, F. U. Hartl, DnaK Functions as a Central Hub in the E. coli Chaperone Network. *Cell Reports*. **1**, 251–264 (2012).

18. T. Niwa, B.-W. Ying, K. Saito, W. Jin, S. Takada, T. Ueda, H. Taguchi, 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 (2009).
19. T. Niwa, T. Kanamori, T. Ueda, H. Taguchi, Global analysis of chaperone effects using a reconstituted cell-free translation system. *Proc. Natl. Acad. Sci. USA* **109**, 8937–8942 (2012).