Chaperones drive in vitro evolution of uracil glycosylase towards misfolded states

Oran Melanker[1], Pierre Goloubinoff[2], Gideon Schreiber[1].

[1]Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel.
[2]Department of Plant Molecular Biology, Lausanne University, 1015 Lausanne, Switzerland.

Email: gideon.schreiber@weizmann.ac.il

Author Contributions: The study was designed by OM, PG and GS; samples preparations were made by OM. Experiments were performed by OM. Data were analyzed and the manuscript was written by OM, PG, GS.

Competing Interest Statement: none.

Classification: Protein evolution.

Keywords: Protein-protein interaction, protein-evolution, promiscuous binding, chaperones

Abstract

Evolution is driven by random mutations, whose fitness outcome is tested over time. In vitro evolution of a library of a randomly mutated protein mimics this process, however, on a short time-scale, driven by a specific outcome (such as binding to a bait). Here, we used directed in vitro evolution to investigate the role of molecular chaperones in curbing promiscuity in favor of specificity of protein-protein interactions. Using yeast surface display, we generated a random library of the *E. coli* protein Uracil glycosylase (UNG), and selected it against various baits. Those included the purified chaperones GroEL, DnaK+DnaJ+ATP, or total protein extracts from WT or delta DnaK+DnaJ cells. We show that in-vitro evolution differs from natural evolution in cells, both physically and thermodynamically. We found that chaperones, whether purified or as part of the protein-extract, select for, and thus enrich uracil glycosylase (UNG) misfolded species during this in vitro evolution process. In a more general context, our results show that chaperones purge promiscuous misfolded clones from the system, and thereby avoiding their detrimental effects, such as forming wrong interactions with other macromolecules, including proteins, which can harm proteostasis.

Significance

Molecular recognition of proteins is fast and specific, even in the crowded milieu of the cell. This, despite the high probability of mutations to promote promiscuous binding. Therefore, without constant purge, promiscuous interactions should be the norm. Here, we show that misfolded species dominate, when a randomly mutated protein is selected for binding either purified chaperones or chaperone-containing cell extract. This suggests that owing to their ability to specifically bind misfolded protein structures, chaperones are an evolutionary force to purge mutations-destabilized, misfolded proteins and avoid non-specific interactions that may harm proteostasis. Moreover, this shows that misfolding and binding to chaperones, is the most likely outcome of random mutations, rather than the formation of new specific protein-protein interactions.

Introduction

Specificity of interactions is a fundamental aspect of cellular processes. Specific binding is achieved when the bound state is thermodynamically or kinetically favored, over non-specific interactions (1). Proteins have evolved to maintain high degree of specificity, regardless of the crowded cellular milieu in which they operate (2). Whereas specific interactions between few natively folded proteins can synergistically increase their functional efficacy and serve as the basis to evolve new functions, promiscuous (unselective, without a specific purpose) interactions between many misfolded proteins, as in aggregates, can lead to loss of function, cellular toxicity, and in metazoans to degenerative diseases, such as Alzheimer and Parkinson (3–5). As formation of promiscuous interactions is proportional to the protein-concentration and its biophysical properties (such as surface hydrophobicity), it is not surprising that higher abundant proteins have been selected by evolution to be more soluble (6–9). In physical terms, the source of interaction specificity that favors a small set of specific interactions over the multitude of possibilities, is evolutionary favored, but how this drives evolution is still an open question. Even weak, short-lived non-specific interactions can affect function (due to slower motility or occlusion of proteins binding membranes or aggregating) as a result of the sheer number of competing macromolecules in the crowded milieu of the cell (10, 11). With this said, how a protein maintains a high degree of specificity is not yet fully resolved (12–14). Promiscuous interactions are often found in hub-proteins, and are also, occasionally necessary to evolve new functions (15–17). However, non-specificity is usually detrimental as it results in slower diffusion, impeded function, and aggregation. Thus, 'sticky' interfaces are less common on abundant proteins, and they have been shown to evolve slower (18, 19). Furthermore, avoiding promiscuous interactions restrict the used sequence space of proteins. This introduces a simple physical reason why an organism complexity is not only reflected by the number of distinct proteins (20–22). A protein's potential to evolve is positively correlated with its stability, as it expands the allowed mutation space. In this sense, stabilizing mutations may be deleterious to the proteome, as they allow promiscuity to evolve more freely (23–25).

HSP70 (DnaK) and Hsp60 (GroEL) are very ancient, highly conserved classes of molecular chaperones that act as polypeptides unfolding enzymes. Both can use the energy of ATP-hydrolysis to unfold stably misfolded structures in aggregated polypeptides and thereby assist misfolding-prone proteins to reach and to stay in the native state, even under stressful conditions

inauspicious to the native state (26). As such, these chaperones can buffer less stable genotypic variations (27–32). In the case of Hsp70-Hsp40 (DnaK-DnaJ in *E coli*), chaperone-assisted (un)folding occurs via binding motives in the primary amino acid sequence generally consisting of five consecutive hydrophobic residues, flanked by positive charges (33). With the help of such chaperones, proteins can endure more destabilizing mutations, alleviating the evolvability penalty from loss of free energy. Yet, increasing the mutation load of a protein also increases its tendency to develop non-specific counter-productive, promiscuous interactions. Hydrophobic solvent-exposed residues are particularly problematic, and thus are constrained by purifying selection to maintain low hydrophobicity and reduce the likelihood of generating non-specific promiscuous sticky patches (34).

Here, we aimed to understand the relation between random mutations and the evolution of promiscuity and how the molecular chaperones Hsp70 and GroEL influence this process. For the *in vitro* evolution, we used the *E.coli* protein Uracil glycosylase (eUNG). This is a 229 amino acid protein, which excises uracil residues from the DNA as a result of mis-incorporation of eUMP residues (35). eUNG binds with nanomolar affinity to the UNG inhibitor, Ugi, and only a folded eUNG can bind Ugi. Therefore, Ugi-binding can serve as a proxy for the structural integrity of eUNG (assuming the incorporated mutations did not directly compromise binding). In vitro evolution was done by generating a library of randomly mutated Uracil glycosylase (eUNG-RL) protein displayed on yeast surface and by FACS selecting the yeast library against various baits. As baits, we used either whole soluble extracts from *E.coli* cells (CE) either from WT (containing both GroEL and DnaK) or from delta DnaK + DnaJ (ΔKJ) (containing only GroEL albeit 3 times more), or purified proteins: GroEL or DnaK+DnaJ+ATP. Each was used to select for enrichment of mutants that were then analyzed. The binding of Ugi to the selected yeast displayed variants was used to distinguish between eUNG that retain its native conformation and mutant eUNG that lost its Ugi-binding ability, and likely turned into a misfolded species prone to promiscuous interactions (assuming that only folded eUNG will bind Ugi, see Figure 1). We found that selection against chaperones resulted, as expected, in the selection of misfolded eUNG that did not bind Ugi, while selection against cell extracts gave more mixed results. Overall, the modulation of a protein evolutionary landscape orchestrated by the chaperones resulted, with evolutionary cycles, in the increasing accumulation of promiscuous mutants with non-native structures, suggesting that the DnaK and GroEL chaperones can salvage the proteome from the deleterious effects of

mutations that cause misfolding, and thus form non-specific promiscuous protein-protein interactions.

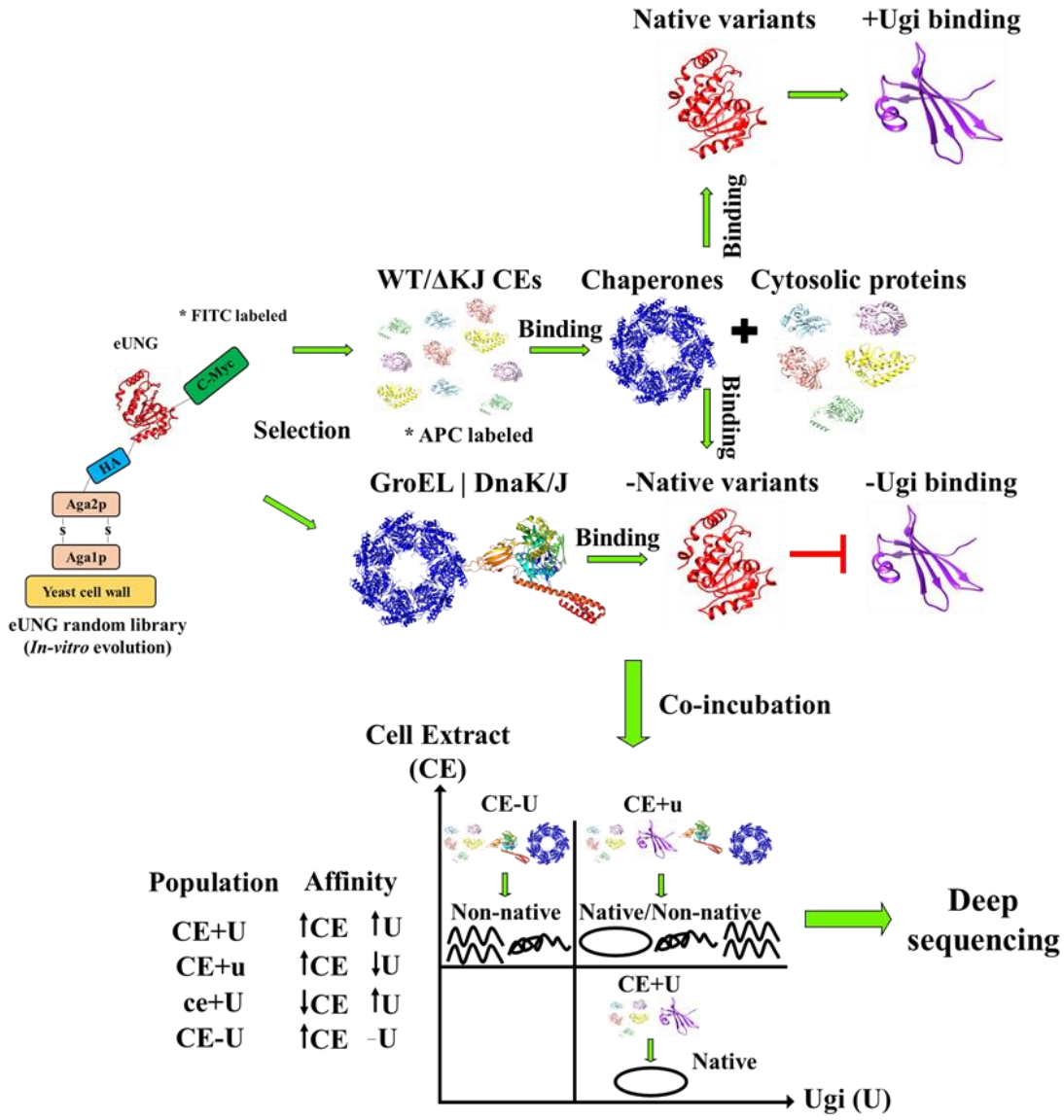

Figure 1. Experimental setup. A yeast surface display library of randomly mutated eUNG was prepared and selected against *E.coli* WT, ΔDnaKJ cell extracts (WT or ΔKJ CEs), purified chaperones: DnaK+DnaJ+ATP (KJ) or GroEL. Native and non-native variants were separated by their ability to bind fluorescently labeled uracil glycosylase inhibitor (Ugi). Selected clones binding Ugi are assumed to be folded, while those not binding Ugi are assumed to be misfolded. Enriched populations were co-incubated with either cell extracts (WT, ΔKJ) or chaperones and Ugi, to distinguish between native and non-native populations. We expect to see three populations, one that binds only the cell extract or the chaperones and two that bind both cell extract and Ugi, with either increased or decreased affinity to Ugi. eUNG variants from each population were sequenced to study the relation between evolution of mutations and the baits used to enrich them.

Results

In vitro evolution of a randomly mutated protein is a powerful tool to study how the bait is dictating the fate of the evolved protein. Most commonly, the bait consists of another protein, where enhanced binding between the two is the outcome. However, in vitro evolution can be applied also to investigate the outcome using any other kind of bait. Here, we used two *E.coli* cell soluble extracts (WT and ΔKJ) as bait (Figure 1). Since knockout of DnaKJ is compensated by elevated levels of GroEL, HtpG, ClpB, IbpB and IbpA (36), the ΔKJ cell extract can be regarded as rich, albeit in the other classes of chaperones (Figure S1). For comparison, we evolved eUNG-RL against purified DnaK+DnaJ (DnaKJ) or GroEL proteins, with and without added ATP, respectively. Evolution against WT and ΔKJ cell extracts was done without ATP. The random eUNG library was constructed by error-prone PCR, incorporating on average ~2-4 nucleotide substitutions per clone. Deep sequencing was used to evaluate the completeness of the eUNG-RL library (Figure S2A), which was used as baseline for the sorted libraries. It is important to note that frequencies of mutations requiring 2 nucleotide changes were, as expected, very low. Thus, only a subset of the 20 aa reachable by single-nucleotide mutations were probed, in line with natural evolution.

Evolving eUNG for chaperone binding

Chaperones have been shown to enhance evolution by increasing the allowed mutation load for marginally stable proteins, preventing their misfolding and aggregation (29). Here, we examined the outcome of in vitro evolution of randomly mutated eUNG using chaperones as baits. First, we validated that in these systems, externally supplied chaperones can specifically recognize and bind a misfolded protein but not its native conformation. Heat pre-misfolded luciferase is known to bind DnaKJ and GroEL (37). Luciferase was inserted into the pCTcon2 yeast display vector C-terminal to Aga2p, where it is exposed to the surface, and incubated with pure GroEL, or DnaK+DnaJ+ATP with and without unlabeled WT *E.coli* cell extract. These were incubated on ice or during a short mild heat treatment (42°C) that is known to cause luciferase to irreversibly misfold (but does not affect the yeast viability). We found that, as expected, the heat-denatured yeast-exposed luciferase bound the purified chaperones (albeit better to DnaKJ), whereas native luciferase on ice did not. However, only DnaKJ, but not GroEL, bound misfolded luciferase when it was mixed with cell extract (Figure S3A). Next, we repeated the experiment using labeled *E.coli* cell extract. Also, in

this case, only the heated-luciferase was found to bind the cell extract, in agreement with the notion that the native luciferase is not engaged in promiscuous interactions, whereas the heat-misfolded luciferase is (Figure S3B). An evident explanation is that the chaperones in the CE also specifically recognize and bind this particularly thermolabile misfolding-prone protein. Chaperones DnaK, DnaJ, GroEL, HtpG, ClpB, IbpB and IpbA constitute 2.8% of the WT *E.coli* proteome mass. If, indeed misfolded luciferase preferentially binds chaperones within the cell extract despite them constituting together ~3% of the protein-mass, it would provide a clear path for misfolded proteins to be purged from coming in way of other macromolecules, such as proteins.

Having established that pure GroEL and DnaKJ bind preferentially to a yeast surface-exposed misfolded, but not to a natively-folded luciferase, provided an incentive to test whether chaperones as bait will also specifically select for misfolded eUNG proteins from a randomly mutated library. Each round of selection was evaluated by in vitro co-incubation of the yeast cells with the chaperones and with Ugi. The progress of selection against DnaKJ and GroEL was found to be accompanied by a loss of binding to Ugi, indicating that the chaperones select mostly for misfolded eUNGs (Figure 2). Strikingly, the binding pattern for Ugi versus DnaKJ appears to be

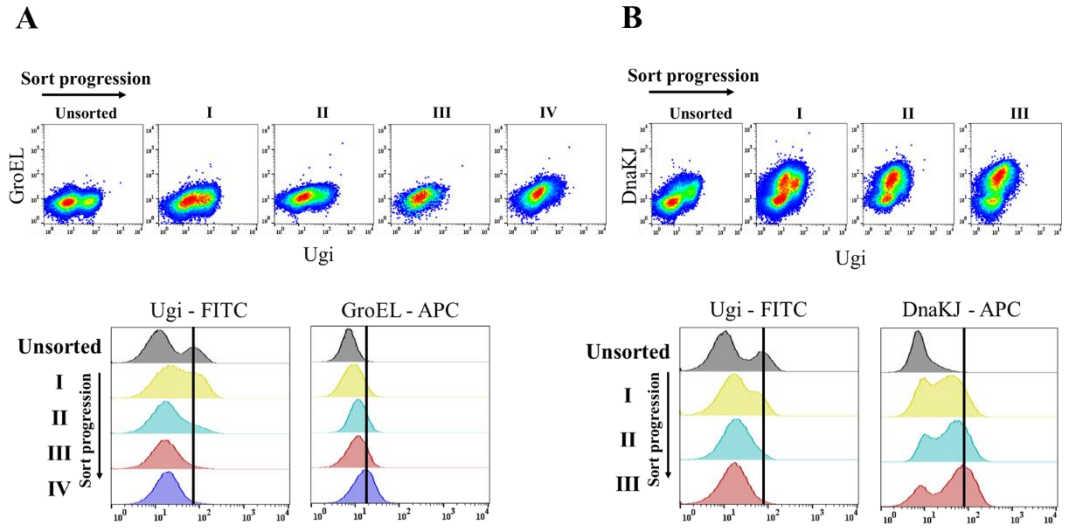

Figure 2. eUNG-RL selection against GroEL or DnaKJ results in loss of Ugi binding. eUNG-RL was selected against either GroEL (A) or DnaKJ (B), monitoring binding of Ugi after each round of selection. Progression of sort is indicated in roman numerals. A. Scatter plots (top) and histogram fit (bottom) of GroEL [1uM] and Ugi [5uM] as bait, binding yeast displayed eUNG-RL at each round of sort is shown. After four rounds of selection, Ugi binding is not observed. B. Scatter plots (top) and histogram fit (bottom) of DnaKJ [2 or 0.4uM] and Ugi [5uM] with eUNG library at each round of sort. The gradual decrease in Ugi binding is visible. Black bar indicate maximum binding intensity for each substrate (chaperones or Ugi). Scatter plots: x axis – Ugi binding (FITC) and y axis – DnaKJ/GroEL binding (APC). Histograms: x axis – FITC/APC channel (530/640nM respectively) and y axis – cell count. DnaKJ was labeled by APC and Ugi was labeled by FITC.

mutually exclusive. Furthermore, tighter binding towards DnaKJ evolved more rapidly (after fewer evolution cycles) than binding towards GroEL (as judged by the increase in the APC signal). This suggests that the random library was selected for presumably misfolded eUNG–RL mutants with higher affinity towards DnaKJ, than towards GroEL.

Misfolded eUNG-RL variants were selected when sorted against *E.coli* cell-extract

Next, we examined the outcome of evolving eUNG-RL using CE as bait. Here, we used two different cell extracts from *E.coli*: 1) WT, containing 1% DnaKJ, 1% GroEL 0.15% HtpG (HSP90) and no IbpA/B, and 2) ΔKJ (containing no DnaKJ), 3.5% GroEL, 1% HtpG and 0.5% IbpA/B (Figure S1). eUNG-RL was subjected to four rounds of selection against WT and ΔKJ *E.coli* CE. Binding to CE and to Ugi was monitored after each round of selection (Figure 3). The increase in Ugi binding after the first round of selection is due to selection for expression, in addition to CE binding. With progressive rounds of sorting, Ugi binding decreased, while CE binding increased (Figure 3). After four successive rounds of selection, the WT CE resulted in a single enriched population of eUNG (Figure 3A, WT CE), while ΔKJ CE was enriched for 3 distinct populations

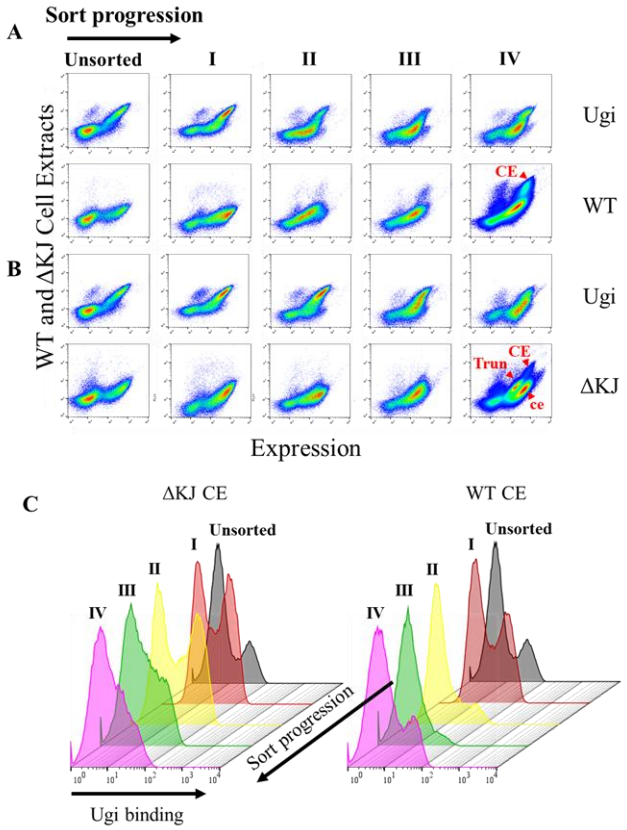

Figure 3. eUNG-RL selection against WT or ΔDnaKJ CE results in partial loss of Ugi binding. A. From top – eUNG-RL after consecutive sorts against WT-CE, binding to Ugi [5uM] or WT-CE [250 ng]. B. as A. but using ΔKJ CE [250 ng]. C. Binding intensity histogram of eUNG-RL selected against WT or ΔKJ cell extract along sorts. Enriched populations were achieved after four rounds, characterized by a reduced population of Ugi binding clones. Sample order is indicated by the black arrow and according to the legend on the C top right panel. Expression labeling was done with FITC and cell extract was labeled with APC. For A and B. x axis – FITC channel (530 nM) and y axis – APC channel (640 nm) and for C. x axis –APC channel (640 nm) and y axis – # of cells. Progression of sort is indicated in roman numerals.

(Figure 3B, ΔKJ CE). The gradual decrease in Ugi binding but increase of CE binding as a function of selection progression raises the possibility that binding of misfolded mutant conformers to the chaperones in the extract was positively selected for. In addition, the following sub-populations were thought to be in each enriched population (Figure 1): clones binding both CE and Ugi and clones binding only CE but not Ugi (CE-U). The former was further divided into a sub-population with minor Ugi binding (CE+u) and the other with high Ugi binding (CE+U). Minor CE binding was denoted by lowercase CE. To separate between these populations, the selected libraries after four rounds of selection shown in Figure 3, and again, in Figure 4A pointing with arrows to the populations used for selection of round five using this time co-incubated with WT-CE and Ugi, revealed an equilibrium-like distribution between CE-U, CE+u and CE+U binding (Figure 4B, panel 1). The CE-U population (Figure 4B, panel 1, CE-U) consisted mostly of non-native eUNG variants, however, we cannot rule out that some clones were mutated at the Ugi binding site. The observation that chaperones select for non-native proteins, raises the possibility that this is the case for the clones within CE-U. The CE+U population consists of more native eUNG variants, as they bind Ugi. The CE+u population (Figure 4B1, CE+u) shares both CE-U and CE+U characteristics. Similarly, selection against ΔKJ CE followed by analysis for ΔKJ CE and Ugi binding (Figure 4B, panels 2-4) exhibited clones that bind CE+u, ce+U and CE-U binding, but without a CE+U population. This suggests that selection against ΔKJ CE results in a population that has a greater tendency for misfolding than after selection against WT CE (as is also clear from Figure 3C). This may be a result of the higher concentration of binding chaperones GroEL, HtpG and IbpAB in the ΔKJ CE.

Thermodynamic characterization of selected eUNG clones as function of CE binding

To further investigate the degree of nativity of eUNG in the different selected populations, twenty single-clones from each population were isolated and sequenced. The mutational profile from each variant was used to calculate the change in free energy of folding of the individual mutations (ΔΔG [kcalmol^{-1}]) (38), from which a ΣΔΔG was estimated for each variant. In addition, using FACS, we evaluated the individual variants for their binding to WT or ΔKJ CE and Ugi. Variants that displayed Ugi binding above or below average (of the WT CE+U population), were considered as +Ugi or –Ugi binding, respectively. Binding to +Ugi can be a good indicator for structural intactness of the individual variants. Conversely, failing to bind Ugi (–Ugi) is most probably a result of the clone to be misfolded, but it cannot be ruled out that the specific mutations hinder Ugi binding. The variants were divided to +Ugi and –Ugi binding, selected against WT or ΔKJ CE, and plotted by their respective ΣΔΔG values (Figure 5). To validate our calculations, we compared ΣΔΔG from WT CE+U to that of GroEL and DnaKJ selected clones. The former eUNG clones are assumed to be folded, while the latter misfolded or unfolded, as they do not bind Ugi but bind chaperones. Indeed, ΣΔΔG of WT CE+U is around zero, while clones binding either GroEL or DnaKJ have negative ΣΔΔG values (destabilized), with no significant difference between the GroEL or DnaKJ selected populations (Figure 5). Next, we compared WT CE to –Ugi selected clones. The latter were significantly less stable (p-value < 0.01), regardless of the CE source (WT or ΔKJ), which clearly shows that –Ugi binding populations contain less stable variants, indicating

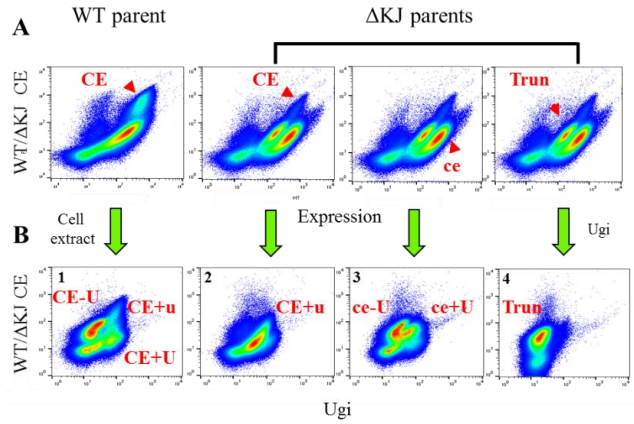

Figure 4. eUNG-RL binding to CE and Ugi after a fifth sort against CE and Ugi. The CE binding populations after four rounds of sort (WT or ΔKJ parents, panel A) were co-incubated with Ugi and CE and sorted against both simultaneously (panel B). The population in A re-sorted in B are marked with red arrows. Panel 1 shows WT CE/Ugi sorted from WT parent. Panels 2-4 show ΔKJ CE/Ugi sorted from ΔKJ parents. WT CE-U, CE+u and CE+U were derived from the WT CE+U parent. Uppercase/lowercase letters indicate higher or lower fluorescence signal against either CE or Ugi, which relates to binding potency of the clonal population. Trun population was neglected as it consists mostly of truncated clones. WT and ΔKJ CEs/Ugi were incubated for 1h on ice prior to selection. WT or ΔKJ cell extracts were labeled by APC and Ugi was labeled by FITC. X axis – FITC channel (530 nM) and y axis – APC channel (640 nm).

their tendency to be misfolded. The ΔKJ ce+U variants appear to be in between the other variants, lower than WT CE+U but higher than the CE-U (Figure 5, purple box).

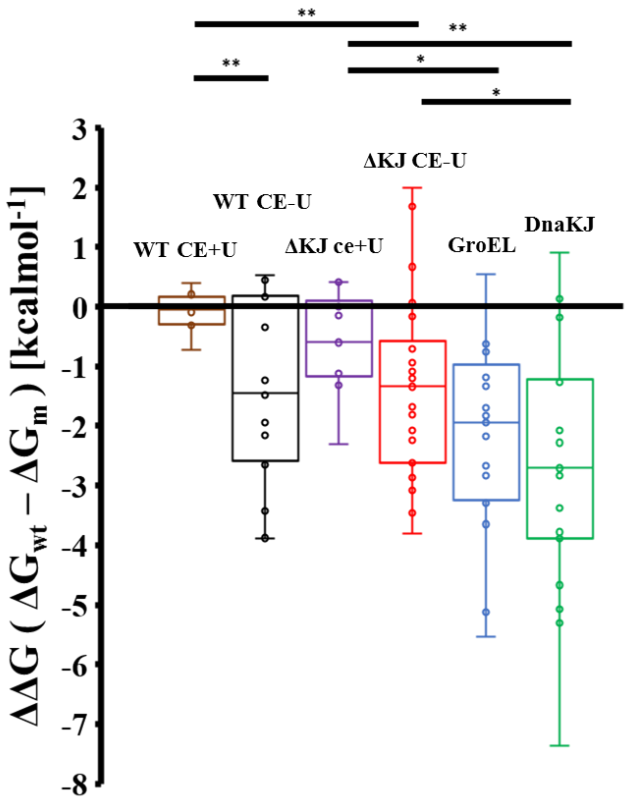

Figure 5. Free energy calculations for different sorted populations of eUNG. Box plot of the ΣΔΔG eUNG variants sorted against WT or ΔKJ cell extract as a function of +/- Ugi binding. Each box is titled with its corresponding CE and whether it is +/–Ugi binding. Uppercase letters indicates at least one variant with enhanced affinity to either cell extract or Ugi. ΔΔG was calculated using SDM and represents the difference in free energy between the eUNG wt and the mutant. Values represents the ΣΔΔG of mutation for a single variant.* p-value < 0.05 and ** p-value < 0.01.

Analyzing the mutational landscape of the differently sorted populations by deep sequencing

Mutations in the core of a protein or in conserved positions often translate in a loss of folding free energy. The degree in which a protein is destabilized changes the equilibrium between locally folded, transiently unfolded, stably unfolded (which is rare for globular proteins because of the tendency of exposed hydrophobic residues to avoid water), and stably misfolded conformations. Here, we examined the relation between the position and conservation of mutations in the different eUNG sorted populations and their tendency to bind Ugi. Each sorted population (see Figure 2 and 4) and the unsorted eUNG-RL, underwent deep mutational scanning. The unsorted eUNG-RL was used to determine the genetic complexity of the starting library. Each selected population was evaluated against the initial genetic heterogeneity of the unsorted library. Using Enrich2 (see materials and methods) and Matlab (R2017a) the number of mutation reads per position were extracted. Output reads were used to calculate the log-ratio of each population (Table S1). The log-

ratio is the natural log of the total reads per position divided by the total number of reads including

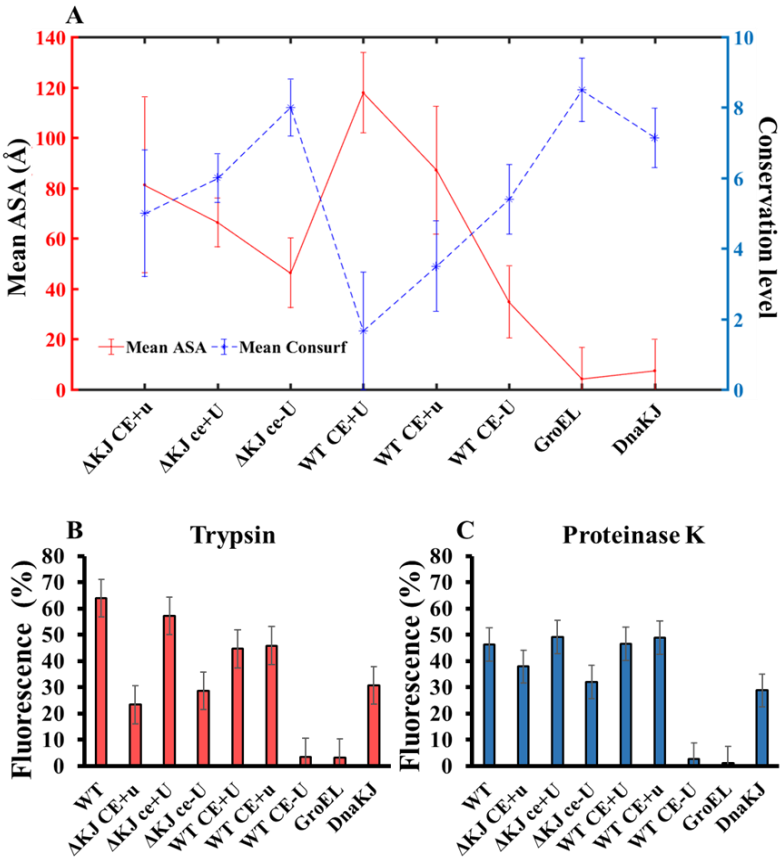

Figure 6. Biophysical properties of chaperones and CEs enriched populations. A. Negative correlation between accessible surface area (ASA) and consurf conservation score for ΔKJ /WT CE, DnaKJ and GroEL selected populations. ASA and conservation scores were calculated using PDBePISA and Consurf servers, respectively. Each population mean log ratio was calculated and positions with values significantly higher (p.value = 0.05 or 0.01) were used to calculate the mean ASA and Consurf values. ASA is plotted with blue broken line and each population is marked with asterisk. Consurf is plotted with red solid line and each population marked by a dot. B. Trypsin (left) or Proteinase K (right) treatment of enriched co-incubated populations. Each population was labeled with FITC and incubated with a set limiting concentration of trypsin or of Proteinase K, such that following 5 minutes at 37˚C over 50% of the WT native eUNG resisted the protease treatment. Controls were incubated only with PBSX1. Change in fluorescence was analyzed using FlowJO and plotted as percentage of remaining fluorescence.

the number of WT reads normalized by the same calculation for the unsorted library using Eq. 1 (see materials and methods). Log ratio > 0 indicates a positive selection and log ratio < 0 a negative selection for a given position. Most selected population's log-ratios show reduced genetic complexity relative to the unsorted library (eUNG-RL). An outlier is the CE+u population in Figure 4B, panel 2, which has almost the same log-ratio pattern as the parent population shown in Figure 4A. Therefore both populations were combined as ΔKJ CE+u parent (Figure S4A).

Additionally, in Figure 4B panel 4 (Trun), the population contained mostly truncated variants and thus, was not further analyzed. For mutations with log-ratio > 0 and which were also significantly higher than the mean of log-ratio > 0, the mean accessible surface area (ASA) and Consurf conservation score (CS) were calculated (Table S2) (39, 40). The respective ASA and CS values of the different selected eUNG populations are shown in Figure 6A. As in Figure 5, WT CE+U and GroEL/DnaKJ point towards the two extremes: The mutations in WT CE+U are located at residues with high ASA (surface exposed) and low CS, while mutations enriched by selection against the chaperones are characterized by having low ASA (buried) and high CS.

Buried and conserved mutations increase the susceptibility of a protein to become non-native, explaining loss of Ugi binding in these populations. Selection of eUNG against WT CE shows a clear order of ASA and CS in respect to binding to CE, Ugi or both. High ASA and low CS was calculated for CE+U, followed by CE+u, which flips towards low ASA and high CS for CE-U. For ΔKJ CE the same order is observed, albeit the ASA values are lower and CS values higher for ΔKJ CE+u and ce+U than observed for WT CE+U (Figure 6A). Overall, these results suggest that the population of eUNG designated as ce+U in selection against ΔKJ cell extract, are more prone to misfolding than CE+u selected from the WT cell extract, in line with the results shown in Figure 5.

Susceptibility to protease cleavage of the sorted populations

Trypsin and Proteinase K are serine proteases that cleave after Arginine/Lysine or hydrophobic amino acids, respectively (41, 42). Cleavage is dependent on the recognized site to be exposed, with buried sites being mostly shielded. Therefore, loss of native structure renders a protein to be more sensitive to digestion by a limiting concentration of a protease. Accordingly, chaperones/CEs-enriched populations should be sensitive to protease digestion in accordance with their respective ASA, CS and their –/+Ugi binding phenotype. This hypothesis was examined by incubating enriched yeast populations with limiting concentrations of trypsin or proteinase K, for 5 min. The conditions were preset such that no more than 50% of the natively folded WT eUNG was cleaved. Protease digestion was measured by FACS as loss of FITC signal (due to yeast surface displayed eUNG being cleaved) and the percentage of remaining fluorescence was plotted. Supporting the lack of native structure in –Ugi binding populations and in line with the other results, WT CE –Ugi binding populations were highly sensitive to a limiting concentration of both

trypsin and proteinase K, similar to the GroEL-selected population. The WT CE+Ugi binding populations were almost as resistant to proteolysis as the WT (Figure 6B and C, respectively). This correlates to the ASA and CS of each population, and the calculated ΣΔΔG values shown in Figure 5. For ΔKJ CE selected populations the picture was again more complex. While partial proteolysis was more extensive for CE+u and CE-U than for ce+U, complete proteolysis was not observed for either population, in contrast with the WT CE-U or GroEL selected populations that were completely digested. These results are in line with those observed for ASA and CS in Figure 6A, suggesting differences in the characteristics of the WT CE and ΔKJ CE selected populations. Interestingly, while the GroEL-selected population was completely digested, the DnaKJ selected population was only partially digested, despite displaying the same reduction in ΣΔΔG (Figure 5) and high ASA and CS as with the GroEL-selected population. Its fluorescence signal resembles the ΔKJ CE+u and ΔKJ ce-U populations, suggesting that although it contains mostly misfolded species, their structural characteristics are different, possibly more compact and native-like.

Residue-dependent enrichment of selected populations

So far, we discussed the general properties of the different evolved populations (Figures 5 and 6). Next, we analyzed the mutation space of the evolved populations on a per-residue basis (Figure 7). The sequence space of the eUNG protein was divided by two parameters: 1. Log-ratio of the frequency to obtain a mutation at a given position of a sorted library divided by the frequency of the unsorted library (R0). 2. Frequency of a mutation in a specific amino-acid divided by the number of reads at the given position (Table S3). The first represents the spatial distribution (SD) of mutations per position in relation to the unselected library (positive or negative). The second, designates as specific spatial distribution – SSD, which describes the frequency of a specific position to be mutated into a specific amino acid (rather than any). The log-ratio of SD of the different selected populations was clustered in hierarchal manner over the sequence of eUNG (Figure 7A). In simple words, it shows the per-residue in vitro evolution as a function of the specific bait. It demonstrates that the sum of the offsprings constitute the parent population (ΔKJ (ce-U + ce+U = [P]ce) and WT (CE+U + CE+u + CE-U = [P]CE)). It also demonstrates that the –Ugi populations differ from those binding Ugi. However, the –Ugi populations selected against ΔKJ and WT differ from one another, and are on different edges of the cluster. Remarkably, some positions (particularly 15 and 31) evolved similarly in the WT and ΔKJ cell extracts, with separated

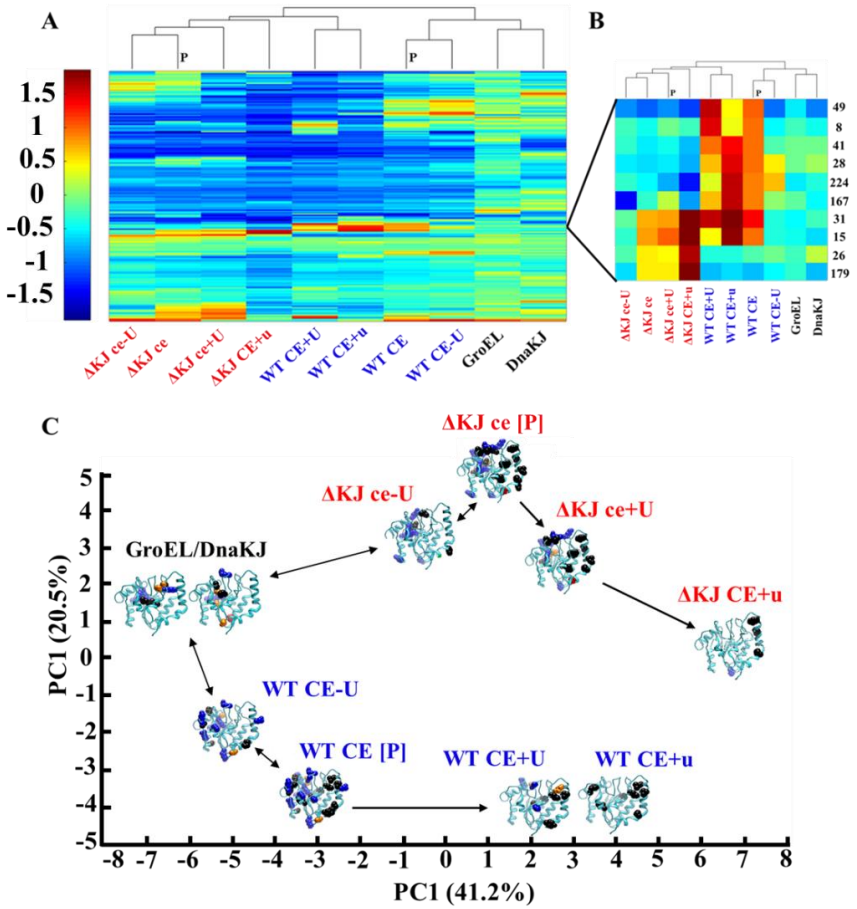

Figure 7. Covariance of enriched chaperones or CEs sorted populations. A. Log-ratio clustering of observed per residue mutations. Data is shown as dendrograms with corresponding heat map for each branch. Each population log-ratio was combined to form an n-by-m matrix where n is the row number of the log-ratio in each position of the eUNG sequence and m is the chaperones/CEs enriched population. Red indicates high variance and blue low variance, per position. Values were standardized to have mean of 0 and SD of 1. B shows a zoom out of an area of interest of the cluster. Numbers on the right indicate variance position. C. Principle component analysis (PCA) of the log-ratio values presented in A. Each arrow represents a different pathway for either WT or ΔKJ cell extract selection. For each population positive log-ratio values were mapped on eUNG (PDB = 1FLZ) structure with their corresponding mutations. Red/Blue colors are used to distinguish between WT and ΔKJ CE selected populations. Axis percentage reflect the explained variance by a given principle component. Parent populations are denoted by P. PCA was calculated using the log ratio data derived using Eq. 1 (see materials and methods).

enrichment for +Ugi and –Ugi binding (Figure 7B). Other positions (for example 8, 41 and 49) evolved differently between WT and ΔKJ cell extract, but similarly in relation to +Ugi or –Ugi binding (Figure 7B). The two populations evolved against purified GroEL and DnaKJ proteins cluster together, and are somewhat different from the populations evolved against cell extract.

Hierarchal clustering is a method to measure similarity between groups. However, it fails to capture the core variance which is shared. Therefore, we used a principle component analysis

(PCA) to correlate population variance. This statistical method transforms a large set of variables into a smaller one, retaining most of the information content. In contrast to hierarchical clustering, PCA discards information which is weakly correlated. Here, PCA was assigned to directly correlate per residue sequence space information, based on population variance (Figure 7C). PCA clearly identified the similarity between selections against GroEL and DnaKJ in the two main principle components. In addition, PCA makes a clear distinction in PC2 between selections against WT versus ΔKJ cell extracts, in what appears as two distinct evolutionary pathways. However, PC1 shows clear similarities between these two, with the order along PC1 of GroEL/DnaKJ ⇒ CE-U/ce-U ⇒ CE+U/ce [P] ⇒ CE+U/ce+U binding maintained for both. The gradual shift from chaperone binding populations to promiscuous populations suggests that after four rounds of CE selection the SD diverged as a function of chaperone recognition. Cell extract from ΔKJ bacteria applies an additional selection pressure on eUNG that is not applied by WT CE. Markedly, the ΔKJ ce+U is closer to ΔKJ CE+u population as the latter SD is a subset of the first (Figure S4B), indicating that the first is actually the pool of solutions from which ΔKJ CE+u has emerged.

Specific spatial distribution (SSD) was the second element that was thought to influence promiscuous evolution. It consists of amino acid residues, which were either negatively or positively selected. Since a single point mutation does not necessarily translate to an amino acid change, due to nonsynonymous mutations, the initial complexity of the unsorted eUNG-RL was first evaluated (Figure S2A, Eq. 4). To evaluate whether the unsorted library contained all possible amino acids reachable by single nucleotide substitution in accordance to the number of codons of each amino-acid, we calculated the frequency of positive, negative, polar, hydrophobic and other amino-acids observed in the library versus their expected number from random codon usage and single nucleotide mutations (which is the expected for random mutagenesis) (Figure S2B). In addition, we calculated the fraction of amino-acid type for eUNG related proteins using consurf. The mutation frequencies in the starting, eUNG-RL library were the same as calculated using the theoretical values from codon usage, confirming that the starting library represents a random distribution of mutations. Interestingly, while WT eUNG and homologous UNG proteins (as retrieved by consurf) also show the same amino-acid type frequencies, both are richer in negatively charged residues and poorer in positively charge residues than the random library. This suggests

evolutionary pressure towards a more neutral charge distribution of UNG than randomly expected, when starting from the eUNG sequence (43).

The unsorted library (R0) sequence space showed bias towards hydrophobic and positively charged residues in accordance to the number of codons translating these amino-acids. Next, we calculated the amino-acid frequencies of the different selected populations and grouped them according to positive, negative, polar, hydrophobic and other amino-acids (Figure 8A). Interestingly, all selections, except WT CE+u, resulted in further enrichment of positively charged amino-acids relative to eUNG-RL, which is in contrast to natural UNG evolution (Figure 8A). The highest enrichment calculated for ΔKJ ce+U and WT CE-U, followed by ΔKJ ce-U, GroEL and DnaKJ. Most selections, except against GroEL and DnaKJ, resulted in fewer polar residues and negatively charged residues, except ΔKJ ce+U and ΔKJ CE+u. To obtain a more detailed picture, the log enrichment of the individual amino-acid mutations of each population was clustered and compared (Figure 8B). Strikingly, while both GroEL and DnaKJ selected population displayed an increase in positive charge, their amino acid distribution were different. Even though both show increase in arginine, only GroEL was enriched with lysines. In addition, GroEL selection favors prolines while DnaKJ does not. These significant disparities motivated our reexamination of the different responses to partial protease digestion of these populations (Figure 6B and C). Since trypsin cleaves after lysine or arginine, and prolines leads to backbone breaks and local unfolding, their enrichment in the GroEL sequence space suggests more disordered eUNG conformations, and thus higher protease sensitivity, compared to DnaKJ selected populations. Furthermore, the increase in positive charge and polarity in the GroEL and DnaKJ selected populations, resembles properties observed in intrinsically disordered proteins (IDPs) (44). Particularly, the ratio between lysine and arginine was shown to influence the conformational equilibria of IDPs. While an increase in arginine over lysine in the DnaKJ selected population promotes compactness in IDPs, higher lysine to arginine ratio in the GroEL selected population promotes weaker phase separation

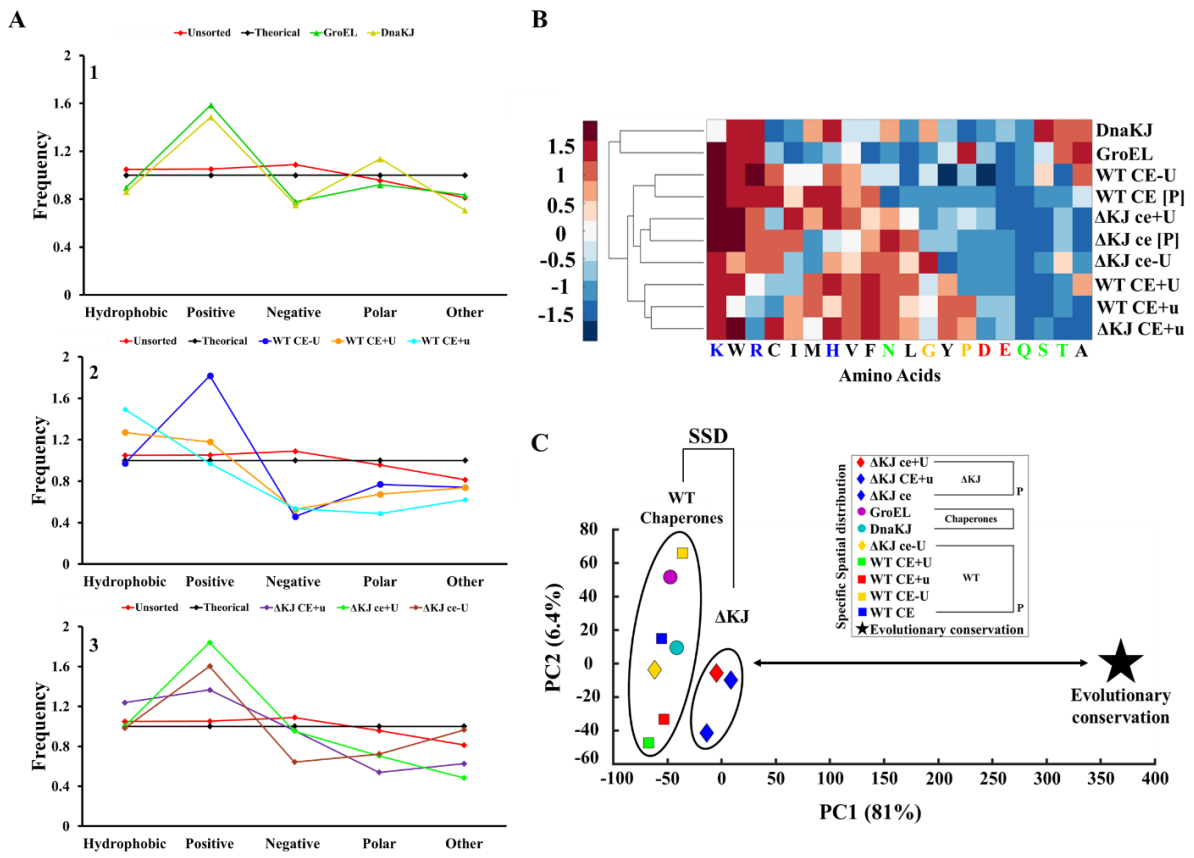

Figure 8. Spatial Specific Distribution of amino-acids of the chaperon or CE selected populations. **A**) Frequency of type of amino acid in relation to their expected frequency by codon usage. The probability of having hydrophobic, positive, negative, polar or other side chain was calculated for selected populations, including the unsorted library (panels 1, 2 and 3 for chaperones, WT and ΔKJ CEs population, respectively). These were normalized by the theoretical frequency as calculated from the codon usage. Y axis depicts enrichment related to the theoretical value and the x axis depicts the different chemical properties of aa side-chains. **B**) Sequence space of enriched promiscuous and chaperone binding populations. The average frequency of each mutation, along protein sequence, was normalized by the unsorted library (see Equation 3). The natural log of these values was plotted for GroEL and DnaKJ binding populations along with WT CE+u, CE+U, ΔKJ CE+u and the natural evolutionary conservation as calculated by Consurf. Blue diamond – ΔKJ CE+u, red/green squares - WT CE+u/CE+U populations, respectively. Purple/teal circles – GroEL and DnaKJ, respectively. Black pentagram –evolutionary conservation (according to Consurf). **C**) Co-variance of enriched population's sequence space. Co-variance was calculated by PCA for the average frequency enrichment relative to the unsorted library (Equation 2 and 3), of each mutation that resulted in WT CE, ΔKJ CE and chaperone selection. Data was compared to Consurf frequency matrix (Table S2). X axis resembles evolutionary distance (marked by two-sided arrow) where clustered dots suggest similar SP. Enriched populations are displayed in the legend; Diamond – ΔKJ CE populations, Square - WT CE populations and Circle – GroEL or DnaKJ. Pentagram – Consurf as evolutionary conservation. Analysis was done using Matlab R2017A). Parent populations are denoted by P.

and results in reduced compactness (45, 46). This implies that the equilibria of conformations,

between GroEL and DnaKJ selected populations was altered, enhancing susceptibility to protease digestion for the first, but less so for the latter.

As observed in Figure 8A, selected eUNG populations were characterized by an increase in both hydrophobicity and positive charge, alongside a decrease in polarity and negative charge. Mapping the log ratio values of the most abundant mutations in the SSD, revealed that promiscuous populations form a hydrophobic patch on the C-terminus of UNG (Figure S5). Alternatively, –Ugi binding populations (including chaperones), accumulate positive charges on the N-terminus. Increase in arginine residues might also lead to solubilization of the misfolded state, as arginine has been shown to enhance the solubility of aggregates or unstructured proteins (47). In addition, positive charges surrounding a hydrophobic core have been shown to characterize DnaK binding motifs (33, 48). As eUNG selection against DnaKJ resulted primarily in an increase in positive charges, it is plausible that, at least partially, they were the driving force for DnaKJ binding. We assumed that the amino-acid variance observed for UNG (evolutionary conservation) as calculated by consurf will differ from the SSD for chaperone or CE evolved populations. PCA was also implemented in this regard to examine the difference between promiscuity and natural evolution. Strikingly, most of the variance between populations can be explained by the PC1 (Figure 8C). It revealed that chaperones, –Ugi binding populations and WT CE binding populations share similar SSDs. Remarkably, the 'Evolutionary Conservation' was most distinct, confirming that in-vivo evolution is completely different from in-vitro evolution against cell extracts. Interestingly, ΔKJ CE binding populations were slightly shifted towards natural evolution, suggesting that the increased levels of GroEL, HtpG and IbpAB chaperones and the total lack of DnaKJ can alter the trajectory of promiscuous evolution. To better understand the observed increased selection for positively-charged residues, observed for selection against CE and pure chaperone proteins, we calculated the electrostatic potentials of eUNG and Ugi (Figure S6). Ugi shows a strong negatively charged surface pointing towards eUNG, which displays a positively charged surface, albeit weaker, towards Ugi. Therefore, while the increased positive charge of selected eUNG populations was not driven by Ugi binding, its observation also in Ugi binding populations suggests that it may also benefit this interaction.

Disscussion

On the role of chaperones in protein evolution

Our findings suggest that the in vitro evolution directed for CE or chaperon binding differs from natural evolution (as seen from Consurf - Figure 8C). While in our experimental setup, evolution was set towards CE or chaperone binding, natural evolution has a functional purpose, with selection pressure towards keeping proteostasis (which include avoiding promiscuity). Thermodynamically, our results further imply that evolution of promiscuous interactions is accompanied with a loss of free energy and destabilization, while natural evolution selects for maintaining stability. In this regard, chaperones promote destabilization of their interaction partners. However, as binding to chaperones leads eventually towards sorting the protein for degradation, during the process of evolution, chaperones, by virtue of their ability to bind new excessively unstable mutant proteins, may actively "skim out" all the excessively destabilizing mutations. This leads proteins to retain only near native mutations that escape constant chaperone binding, to evolve new functions (see Fig 9 above the dashed line).

Our selection was for mutants without active proteases which could have operated with the chaperones, to select, bind and unfold misfolded mutants, and readily degrade them. Even though the CE used for selection contained proteases (HslUV, ClpXP, lon), these are ATP-fueled proteases and our selection was made without ATP. Thus, the protease degradation of selected populations (Figure 6B and C) might mimic in-vivo evolution, because it may resemble what could have been the natural repartition of the work between chaperones and the ATP-fueled proteases in evolution (49). In our assay, we differ from the conditions in the cell, as the proteins, chaperones and proteases are not at their cellular concentrations. However, it is plausible that our in-vitro chaperone selection is not so different than the in-vivo one. In the cell, chaperones work hand in hand with proteases, and the non-promiscuous new mutants that are slightly less stable may indeed be stabilized as previously been shown (29, 50). In contrast, those that are much less stable than the native WT, cannot be stabilized to the native state by ATPase chaperones during their iterative unfolding cycles. Chaperone released species, instead of transiently refolding to the native state, will eventually be hijacked by the AAA+ proteases that are in the same cellular compartment, and will be degraded and lost for evolution.

The evolutionary pathways of ΔKJ CE selected populations is different than WT CE selected populations (Figures 7B and 8C), possibly resulting from the very different composition

and concentrations of the other chaperones (which are not DnaK) between the two. GroEL and DnaKJ chaperones recognize exposed motives at the surface of misfolded proteins that are mostly hydrophobic and positively charged. Noticeably, these can be naturally enriched in mutants through a natural bias in evolution: the codon usage bias which dictates that random mutagenesis should more frequently result in a sticky mutation, mainly hydrophobic but also positive (Figure 7A and B, Figure S5). A protein can acquire more mutations due the chaperones acting as polypeptide unfoldases (51), but paradoxically, because of the codon usage bias, it increases the likelihood of promiscuity-enhancing mutation (52). Chaperones work as a selection pressure against promiscuity, not fortuitously, since they have evolved to precisely detect this bias (33, 37). Indeed chaperones can select against destabilizing mutations, until a new stabilizing solution, which is not further recognized by them, has evolved. The result is a feedback loop in which the component that enhances evolution by mutation load increase, also inhibits the possible enriched promiscuous product of this enhancement, by its destabilization. Inhibiting promiscuity suits chaperones well as they are major abundant organizers of the cellular protein-protein interaction network (53). If chaperones were lacking this feature, then protein stickiness should have been the null hypothesis (2, 17, 23, 52, 54). Any time scale for a solution to emerge would have to be found in a chaperone dependent manner. It also explains why an organism complexity is not reflected by the number of distinct proteins it produces, but rather by the number of new folds (55), the extend of fold combinations and of domain repeats resulting in increasingly long beta-fold enriched proteins, as well as of chaperone-resistant IDPs along evolution (27). Noticeably, our results suggest that the IDPs that did survive chaperone selection in evolution would have to lack typical chaperone-binding motives, such as clusters of hydrophobic residues flanked by positive charges as identified by us (33, 56–58). Furthermore, it might explain why core-chaperones have not evolved proportionally with the increased complexity of proteomes. Rather co-chaperones, such as the J-domain proteins, of which DnaJ in a member, which specifically target the Hsp70 unfoldase machineries to bind with high affinity and use energy of ATP to unfold misfolded protein substrates. This, while avoiding binding to the natively unfolded and the compact native proteins, which are in general the products of the chaperone-mediated unfolding reaction (27). In addition, our results suggest that chaperones, alongside being ATP-fueled unfoldases, may actually act in assisting the proper formation of protein oligomers. This function was first suggested by John Ellis and colleagues, who showed that the chloroplast chaperonin (GroEL) is crucial for

specific assembly of RubisCO L_8S_8 dodecamers (59). It seems that the term chaperones, initially chosen by John Ellis, to describe features of this class of proteins, was not inaccurate.

As such, chaperones are required for specific protein multimerization, thus it is plausible that they inhibit promiscuous interactions, which promote specific interactions in native oligomers and thereby limiting their evolvability (60). In conclusion, we suggest that the evolutionary landscape of protein promiscuity is modulated by molecular chaperones, performing as the molecular "police" of cells (61) to shield promiscuous, "criminal" cytotoxic protein-protein interactions, carried out primarily by misfolding and aggregating proteins. This passively drives specificity, as only binding interfaces which are not recognized by chaperones are positively selected for, in the course of protein evolution (Figure 9).

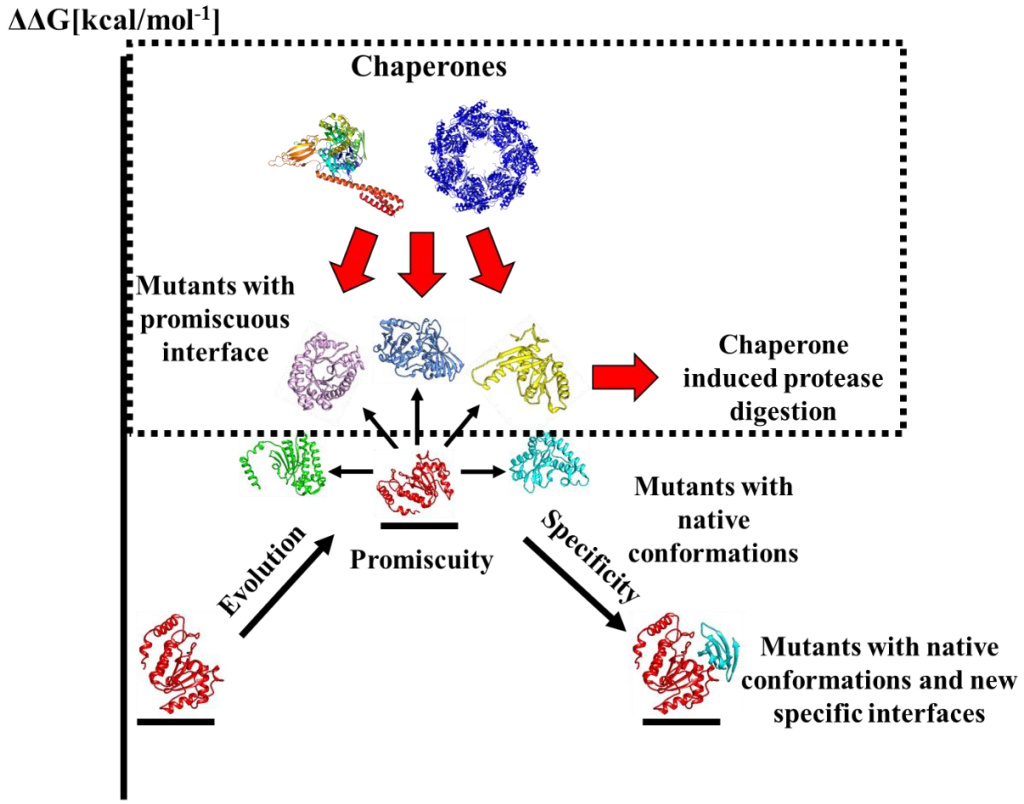

Figure 9. Chaperones modulate the evolutionary landscape of protein protein interactions. When a protein evolves, it most likely results in promiscuous interactions with non-target proteins. Chaperones promotes evolution by buffering mutations, expanding the sequence space in which a protein can evolve. However, expanding the number of mutations a protein endures increases the likelihood it will mis-interact. Accumulation of more mutations also increases its free energy. Chaperones selects against unstable promiscuous sequences (red arrows), until a near native solution is achieved, which is no longer recognized by chaperones.

Materials and Methods

Yeast library formation

The library was performed on *E.coli* Uracil glycosylase (eUNG). EBY100 yeast that were used for the library creation were formed by following the procedure described by Benatuil (62). DNA mutagenesis was performed with mutazyme (GeneMorph II random mutagenesis kit catalog # 200550) resulting in an average of ~4 mutations per gene. Amplification of eUNG gene was done with Taq DNA-polymerase as described by Chao (63). Library size was estimated to be ~10^7 by plating serial dilutions on YPD selection plates lacking Tryptophan. Library was further confirmed for correct gene insertion and average mutation by sequencing 20 single clones.

Co-incubation experiments

After enrichment was achieved against chaperones or WT or ΔKJ cell extracts, enriched population were collected and incubated both with chaperones or WT, ΔKJ cell extracts and Ugi. Chaperones and WT or ΔKJ cell extracts were labeled with APC (Biotium # 92108) while Ugi was labeled with FITC (Biotium # 92103) as described by the manufacturer. Co-incubated populations were analyzed by Bio-Rad FACS sorter and according to population distribution were sorted out into new vials with SDCCA selection media (63).

Uracil glycosylase inhibitor production

Urcail glycosylase inhibitor (Ugi) gene was ordered from TWIST and later cloned into the pet28 vector.(64) This vector was modified to have a longer linker on the N' terminus for better protease cleavage. Ugi was expressed in BL21(DE3) bacteria, grown at 37˚C until $O.D_{600}$ = 0.6-0.8 and then incubated with 0.1 mM IPTG at 16˚C overnight. Bacteria was pelleted and sonicated and suspended in SUMO protease buffer (50mM Tris-HCL pH 8, 200mM NaCl 10% glycerol 1mM DTT and 1:200 sumo protease both of which were added fresh). Ugi was purified from Ni beads binding 6xHis tag as described in (64) and further purified using ion exchange chromatography. After cleavage Ugi was dialyzed into labeling buffer (0.1M $NaHCO_3$ pH = ~8.3).

Cell extracts production

W3110 *E.coli* (WT and ΔKJ) cell extract were prepared by first inoculating at 1:1000 ratio into 2YT followed by incubation at 37˚C until $O.D_{600}$ = 0.6-0.8 was achieved. Cells were then harvested

by centrifugation at 4˚C and resuspended in lysis buffer (Tris HCl pH = 8, 200 mM NaCl, 10 mM Imidazole and 10% glycerol). The bacteria were sonicated and cytosolic fraction was collected by 17000 rpm 4˚C for 30 minutes. Supernatant was filtered using 0.4 μM filters and was dialyzed in labeling buffer (0.1M $NaHCO_3$ pH = ~8.3) in double distilled water and in 3.5 MWCO dialysis membrane.

Sorting

The library was sorted 5 times against 5mg/ml cell extract labeled with CF640R succinimidyl ester (APC) through amine-coupling. APC labeled cell extract was filtered from excess dye before each sort. Protein expression levels on the yeast surface was detected by labeling the c-myc tag found at the C' terminal of expressed eUNG with FITC coloring. The selection continued until notable binding enrichment was detected by FACS. The selected libraries were then plated and 20 clones were sequenced and evaluated for binding cell extracts or chaperones.

Deep Sequencing

Deep sequencing was done on all the populations. Plasmids encoding the selected populations were extracted using Lyticase (Sigma # L2524). The eUNG gene (636bp) was amplified as 4 fragments of 250 bp each. Amplified plasmid flanks, which are not part of the eUNG gene were discarded in the analysis. Deep sequencing was done using Illumina MiSeq V2 (500 cycles) with paired-ends, 250bp read length and 12M reads. Deep sequencing reads were extracted using Enrich2 (65) and were further processed using Matlab R2017a. Sequence logos images were done using Seq2Logo (66).

Matlab analysis

Deep sequencing raw data was extract using Enrich2. The reads were organized in excel files containing reads per mutation along eUNG sequence. Reads were used to write scripts (Extended data) which calculates three parameters; Log-Ratio – SD (Eq. 1), amino acid frequency per position and from it the SSD (Eq. 2 and 3) and the initial amino acid variance, denoted by $V_{AA,}$ of the unsorted library (Eq. 4):

$$\text{Eq. 1: SD} = \text{Log10}\left(\frac{\Sigma R_P / \Sigma R_{\Sigma P+\,wt}}{\Sigma R_{P-U} / \Sigma R_{\Sigma P\,-\,U+\,wt}}\right)$$

$$\text{Eq. 2: } f_{M/P} = \frac{R_{M/P}}{\Sigma R_P}$$

$$\text{Eq. 3: SSD} = 100*\left(\bar{x}\left(\frac{f_{M/\Sigma P}}{f_{M/\Sigma P-U}}\right) - 100\right.$$

$$\text{Eq. 4: } V_{AA} = \frac{R_M}{\Sigma R_M}$$

ΣR_P – total reads per position, $\Sigma R_{\Sigma P\,+\,WT}$ – total reads over all positions including WT sequences, $\Sigma R_{P\text{-}U}$ – unsorted library total reads per position, $\Sigma R_{\Sigma P\text{-}U\,+\,WT}$ – unsorted library total reads over all positions including WT sequences, $f_{M/P}$ – mutation frequency per position, $R_{M/P}$ – reads of a mutation per position, $\bar{x}$ – mean, $f_{M/\Sigma P\,-}$ frequency of a mutation over all positions. $f_{M/\Sigma P\,-\,U}$ - unsorted library frequency of a mutation over all positions, V_{AA} – initial amino acid variance (the frequency of a possible mutation for an amino acid) displayed as a fraction of: R_M – reads of a specific mutation X of amino acid Y and ΣR_M – sum of all reads of possible mutations for an amino acid Y.

Funding Sources. This research was supported by the Israel Science Foundation grant No. 1268/18 (GS) and Grant 31003A_175453 from the Swiss National Fund (PG)

Acknowledgement. We thank Satyam Tiwari and Tatiana Fomekong T. Mbefo from the University of Lausanne (UNIL), for providing us with purified GroEL, DnaK and DnaJ proteins and Pierre Gevevaux from the University of Toulouse III for giving us the delta DnaJK-DnaJ *E. coli* stain.

References

1. G. Schreiber, G. Haran, H.-X. Zhou, Fundamental aspects of protein− protein association kinetics. *Chem. Rev.* **109**, 839–60 (2009).
2. G. Schreiber, A. E. Keating, Protein binding specificity versus promiscuity. *Curr. Opin. Struct. Biol.* **21**, 50–61 (2011).
3. C. A. Ross, M. A. Poirier, Protein aggregation and neurodegenerative disease. *Nat. Med.* **10**, S10 (2004).
4. G. Gat-Yablonski, *et al.*, Quantitative proteomics of rat livers shows that unrestricted feeding is stressful for proteostasis with implications on life span. *Aging (Albany. NY).* **8**, 1735–1758 (2016).
5. O. Chakrabarti, R. S. Hegde, Functional Depletion of Mahogunin by Cytosolically Exposed Prion Protein Contributes to Neurodegeneration. *Cell* **137**, 1136–1147 (2009).
6. Q. Wang, A. Zhuravleva, L. M. Gierasch, Exploring weak, transient protein-protein interactions in crowded in vivo environments by in-cell nuclear magnetic resonance spectroscopy. *Biochemistry* **50**, 9225–9236 (2011).
7. X. Mu, *et al.*, Physicochemical code for quinary protein interactions in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E4556–E4563 (2017).
8. G. Rivas, A. P. Minton, Toward an understanding of biochemical equilibria within living cells. *Biophys. Rev.* **10**, 241–253 (2018).
9. 1. M. Heo, S. Maslov, E. Shakhnovich, Topology of protein interaction network shapes protein abundances and strengths of their functional and nonspecific interactions. *Proc. Natl. Acad. Sci.* **108**, 4258–4263 (2011).
10. O. Peleg, J. M. Choi, E. I. Shakhnovich, Evolution of specificity in protein-protein interactions. *Biophys. J.* **107**, 1686–1696 (2014).
11. M. H. Huntley, A. Murugan, M. P. Brenner, Information capacity of specific interactions. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 5841–5846 (2016).
12. J. L. Ross, The Dark Matter of Biology. *Biophys. J.* **111**, 909–916 (2016).
13. S. Bhattacharyya, *et al.*, Transient protein-protein interactions perturb E. coli metabolome and cause gene dosage toxicity. *Elife* **5**, 1–22 (2016).

14. M. A. Siddiq, G. K. Hochberg, J. W. Thornton, Evolution of protein specificity: insights from ancestral protein reconstruction. *Curr. Opin. Struct. Biol.* **47**, 113–122 (2017).

15. A. S. Pillai, *et al.*, Origin of complexity in haemoglobin evolution. *Nature* **581**, 480–485 (2020).

16. A. Aharoni, *et al.*, The "evolvability" of promiscuous protein functions. *Nat. Genet.* **37**, 73–76 (2005).

17. I. Nobeli, A. D. Favia, J. M. Thornton, Protein promiscuity and its implications for biotechnology. *Nat. Biotechnol.* **27**, 157–167 (2009).

18. E. D. Levy, S. De, S. A. Teichmann, Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. *Proc. Natl. Acad. Sci.* **109**, 20461–20466 (2012).

19. J.-R. Yang, B.-Y. Liao, S.-M. Zhuang, J. Zhang, Protein misinteraction avoidance causes highly expressed proteins to evolve slowly. *Proc. Natl. Acad. Sci.* **109**, E831–E840 (2012).

20. J. Zhang, S. Maslov, E. I. Shakhnovich, Constraints imposed by non-functional protein-protein interactions on gene expression and proteome size. *Mol. Syst. Biol.* **4**, 1–11 (2008).

21. F. Etoc, *et al.*, Non-specific interactions govern cytosolic diffusion of nanosized objects in mammalian cells. *Nat. Mater.* **17**, 740–746 (2018).

22. M. E. Johnson, G. Hummer, Nonspecific binding limits the number of proteins in a cell and shapes their interaction networks. *Proc. Natl. Acad. Sci.* **108**, 603–608 (2011).

23. R. Cohen-Khait, O. Dym, S. Hamer-Rogotner, G. Schreiber, Promiscuous Protein Binding as a Function of Protein Stability. *Structure* **25**, 1867-1874.e3 (2017).

24. O. Ashenberg, L. I. Gong, J. D. Bloom, Mutational effects on stability are largely conserved during protein evolution. *Proc. Natl. Acad. Sci.* **110**, 21071–21076 (2013).

25. J. D. Bloom, S. T. Labthavikul, C. R. Otey, F. H. Arnold, Protein stability promotes evolvability. *Proc. Natl. Acad. Sci.* **103**, 5869–5874 (2006).

26. P. Goloubinoff, A. S. Sassi, B. Fauvet, A. Barducci, P. De Los Rios, Chaperones convert the energy from ATP into the nonequilibrium stabilization of native proteins article. *Nat. Chem. Biol.* **14**, 388–395 (2018).

27. M. E. Rebeaud, S. Mallik, P. Goloubinoff, D. S. Tawfik, On the evolution of chaperones and cochaperones and the expansion of proteomes across the Tree of Life. *Proc. Natl. Acad. Sci. U. S. A.* **118** (2021).

28. S. L. Rutherford, Between genotype and phenotype: Protein chaperones and evolvability. *Nat. Rev. Genet.* **4**, 263–274 (2003).

29. D. S. Libich, V. Tugarinov, G. M. Clore, Intrinsic unfoldase/foldase activity of the chaperonin GroEL directly demonstrated using multinuclear relaxation-based NMR. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8817–8823 (2015).

30. N. Tokuriki, D. S. Tawfik, Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature* **459**, 668–673 (2009).

31. A. Finka, R. U. H. Mattoo, P. Goloubinoff, Experimental Milestones in the Discovery of Molecular Chaperones as Polypeptide Unfolding Enzymes. *Annu. Rev. Biochem.* **85**, 715–742 (2016).

32. L. Agozzino, K. A. Dill, Protein evolution speed depends on its stability and abundance and on chaperone concentrations. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 9092–9097 (2018).

33. S. Rüdiger, A. Buchberger, B. Bukau, Interaction of Hsp70 chaperones with substrates. *Nat. Struct. Biol.* **4**, 342–349 (1997).

34. M. A. Stiffler, D. R. Hekstra, R. Ranganathan, Evolvability as a Function of Purifying Selection in TEM-1 β-Lactamase. *Cell* **160**, 882–892 (2015).

35. L. H. Pearl, Structure and function in the uracil-DNA glycosylase superfamily. *Mutat. Res. - DNA Repair* **460**, 165–181 (2000).

36. B. Fauvet, *et al.*, Bacterial Hsp90 Facilitates the Degradation of Aggregation-Prone Hsp70–Hsp40 Substrates. *Front. Mol. Biosci.* **8** (2021).

37. A. Buchberger, H. Schröder, T. Hesterkamp, H. J. Schönfeld, B. Bukau, Substrate shuttling between the DnaK and GroEL systems indicates a chaperone network promoting protein folding. *J. Mol. Biol.* **261**, 328–333 (1996).

38. C. L. Worth, R. Preissner, T. L. Blundell, SDM - A server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res.* **39**, 215–222 (2011).

39. E. Krissinel, K. Henrick, Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).

40. H. Ashkenazy, *et al.*, ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).

41. K. A. Walsh, D. L. Kauffman, K. S. Kumar, H. Neurath, On the Structure and Function of Bovine Trypsinogen and Trypsin. *Proc. Natl. Acad. Sci. United States* **51**, 301–308 (1964).

42. W. Ebeling, *et al.*, Proteinase K from Tritirachium album Limber. *Eur. J. Biochem.* **47**, 91–97 (1974).

43. R. Cohen-Khait, G. Schreiber, Low-stringency selection of TEM1 for BLIP shows interface plasticity and selection for faster binders. *Proc. Natl. Acad. Sci. U. S. A.* **113**, 14982–14987 (2016).

44. V. N. Uversky, What does it mean to be natively unfolded? *Eur. J. Biochem.* **269**, 2–12 (2002).

45. M. J. Fossat, X. Zeng, R. V. Pappu, Uncovering Differences in Hydration Free Energies and Structures for Model Compound Mimics of Charged Side Chains of Amino Acids. *J. Phys. Chem. B* **125**, 4148–4161 (2021).

46. C. S. Sørensena, M. Kjaergaarda, Effective concentrations enforced by intrinsically disordered linkers are governed by polymer physics. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 23124–23131 (2019).

47. J. I. Austerberry, *et al.*, Arginine to Lysine Mutations Increase the Aggregation Stability of a Single-Chain Variable Fragment through Unfolded-State Interactions. *Biochemistry* **58**, 3413–3421 (2019).

48. E. M. Clerico, J. M. Tilitsky, W. Meng, L. M. Gierasch, How Hsp70 molecular machines interact with their substrates to mediate diverse physiological functions. *J. Mol. Biol.* **427**, 1575–1588 (2015).

49. Fauvet, B., Rebeaud, M. E., Tiwari, S., De Los Rios, P. & Goloubinoff, P. Repair or Degrade: the Thermodynamic Dilemma of Cellular Protein Quality-Control. *Front. Mol. Biosci.* **8**, 1–11 (2021).

50. D. F. Jarosz, M. Taipale, S. Lindquist, Protein homeostasis and the phenotypic manifestation of genetic diversity: Principles and mechanisms. *Annu. Rev. Genet.* **44**, 189–216 (2010).

51. S. Priya, S. K. Sharma, P. Goloubinoff, Molecular chaperones as enzymes that catalytically unfold misfolded polypeptides. *FEBS Lett.* **587**, 1981–1987 (2013).

52. G. Calloni, *et al.*, DnaK Functions as a Central Hub in the E. coli Chaperone Network. *Cell Rep.* **1**, 251–264 (2012).

53. C. D. Aakre, *et al.*, Evolving New Protein-Protein Interaction Specificity through Promiscuous Intermediates. *Cell* **163**, 594–606 (2015).

54. G. K. A. Hochberg, *et al.*, A hydrophobic ratchet entrenches molecular complexes. *Nature* **588**, 503–508 (2020).

55. M. Taipale, *et al.*, A quantitative chaperone interaction network reveals the architecture of cellular protein homeostasis pathways. *Cell* **158**, 434–448 (2014).

56. P. Goloubinoff, J. T. Christeller, A. A. Gatenby, G. H. Lorimer, Reconstitution of active dimeric ribulose bisphosphate carboxylase from an unfolded state depends on two chaperonin proteins and Mg-ATP. *Nature* **342**, 884–889 (1989).

57. U. Jakob, M. Gaestel, K. Engel, J. Buchner, Small heat shock proteins are molecular chaperones. *J. Biol. Chem.* **268**, 1517–1520 (1993).

58. P. Koldewey, F. Stull, S. Horowitz, R. Martin, J. C. A. Bardwell, Forces Driving Chaperone Action. *Cell* **166**, 369–379 (2016).

59. S. M. Hemmingsen, *et al.*, Homologous plant and bacterial proteins chaperone oligomeric protein assembly. *Nature* **333**, 330–334 (1988).

60. P. Duraõ, *et al.*, Opposing effects of folding and assembly chaperones on evolvability of Rubisco. *Nat. Chem. Biol.* **11**, 148–155 (2015).

61. M.-P. Hinault, P. Goloubinoff, Molecular Crime and Cellular Punishment. *Mol. Asp. Stress Response Chaperones, Membr. Networks*, 47–54 (2007).

62. L. Benatuil, J. M. Perez, J. Belk, C. M. Hsieh, An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* **23**, 155–159 (2010).

63. G. Chao, *et al.*, Isolating and engineering human antibodies using yeast surface display. *Nat. Protoc.* **1**, 755–768 (2006).

64. J. Zahradník, *et al.*, Flexible regions govern promiscuous binding of IL-24 to receptors IL-20R1 and IL-22R1. *FEBS J.* **286**, 3858–3873 (2019).

65. A. F. Rubin, *et al.*, A statistical framework for analyzing deep mutational scanning data. *Genome Biol.* **18**, 1–15 (2017).

66. M. C. F. Thomsen, M. Nielsen, Seq2Logo: A method for construction and visualization of amino acid binding motifs and sequence profiles including sequence weighting, pseudo counts and two-sided representation of amino acid enrichment and depletion. *Nucleic Acids Res.* **40**, 281–287 (2012).