

RESEARCH ARTICLE SUMMARY

PROTEOMICS

Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability

Pascal Leuenberger, Stefan Ganscha, Abdullah Kahraman, Valentina Cappelletti, Paul J. Boersema, Christian von Mering, Manfred Claassen, Paola Picotti*

INTRODUCTION: Temperature is crucially important to life. Small temperature changes can differentiate optimal and lethal growth conditions of living organisms. Because of the higher abundance and lower stability of proteins as compared with those of other biological macromolecules, thermally induced cell death is thought to be due to protein denaturation, but the determinants of thermal sensitivity of proteomes remain largely uncharacterized.

RATIONALE: To determine the thermal stability of proteins on a proteome-wide scale and with domain-level resolution, we developed a structural proteomic approach that relies on limited proteolysis (LiP) and mass spectrometry (MS) applied over a range of temperatures.

RESULTS: Our LiP-MS strategy was validated through analysis of purified proteins in the presence and absence of a biologically relevant

matrix. We then obtained proteome-wide thermal denaturation profiles for *Escherichia coli*, *Saccharomyces cerevisiae*, *Thermus thermophilus*, and human cells. In contrast to previous predictions that proteome instability derives from the simultaneous and generalized loss of hundreds of proteins, we observed that at a temperature at which cells experience temperature-induced physiological impairment, a subset of essential proteins undergoes denaturation.

Confirming results of previous studies on the basis of comparison of genomes of thermophilic and mesophilic bacteria, we observed enrichment for lysine residues and β -sheet structures in thermostable proteins. We also found that unstable proteins have

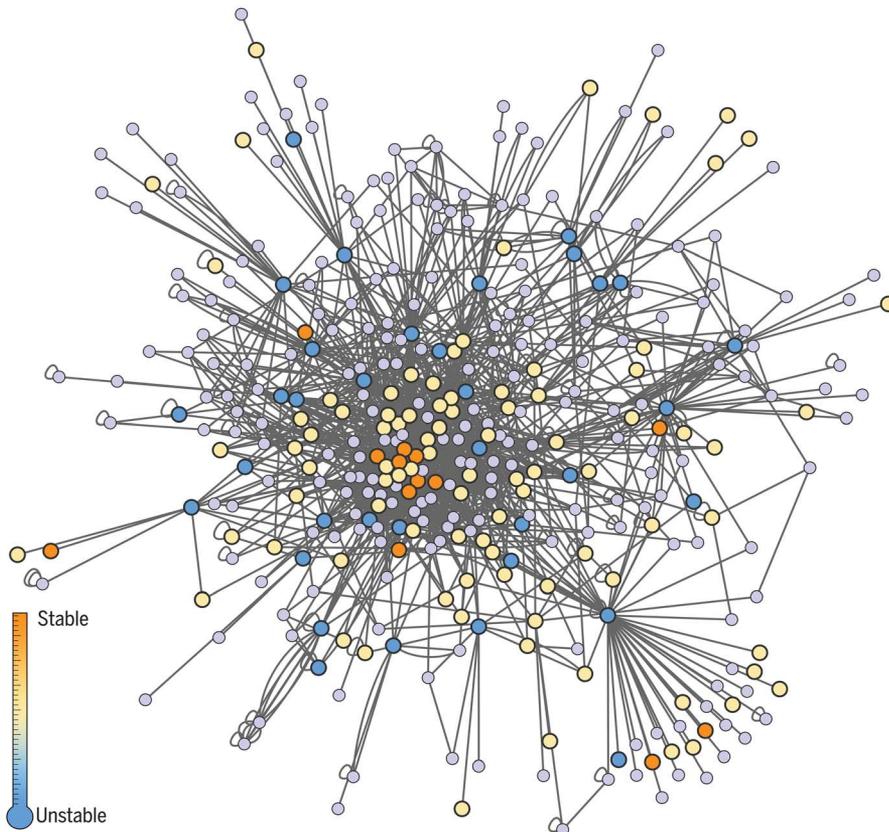
a higher content of aspartic acid than that of stable proteins and observed an inverse correlation between protein length and thermal stability. Further, thermostable proteins are substantially less prone to thermal aggregation than unstable proteins.

Relative domain thermostability was conserved both within species and across organisms. Thermal stability was not generally similar for proteins encoded by orthologous genes. This suggests that the melting temperatures of proteins are affected by the reshuffling of protein domains, despite the conservation of domain stability.

According to the “translational robustness” theory, highly expressed proteins must tolerate translational errors that can lead to the accumulation of toxic misfolded species. Our data show a clear direct relationship between protein thermal stability and intracellular abundance and an inverse relationship between protein stability and aggregation or local unfolding. Increasing the thermodynamic stabilities of the folds of abundant proteins will broaden the range of amino acid replacements that a protein can tolerate before misfolding. Our findings suggest that over the course of evolution, the burden of intracellular misfolding has been reduced by increasing the thermodynamic stability of abundant proteins. Last, although up to 30% of proteomes have been predicted to consist of intrinsically disordered proteins (IDPs), our data revealed that about half of these proteins showed two-state denaturation profiles in the cellular matrix. This suggests that many IDPs are globally or locally structured in cells.

CONCLUSION: Our study contributes insight into the molecular and evolutionary bases of protein and proteome thermostability and provides a blueprint for future studies on the stability of proteomes and thermal denaturation. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: paola.picotti@bc.biol.ethz.ch
Cite this article as P. Leuenberger et al., *Science* 355, eaai7825 (2017). DOI: 10.1126/science.aai7825



Protein-protein interaction network of *E. coli*. Node color indicates protein thermostability. Blue, unstable; yellow, medium-stable; orange, stable; gray, not measured. At the temperature of thermal cell death of *E. coli*, a subset of highly connected protein nodes involved in key cellular processes undergoes temperature-induced denaturation.

RESEARCH ARTICLE

PROTEOMICS

Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability

Pascal Leuenberger,^{1,2} Stefan Ganscha,^{2,3} Abdullah Kahraman,⁴ Valentina Cappelletti,¹ Paul J. Boersema,¹ Christian von Mering,⁴ Manfred Claassen,³ Paola Picotti^{1*}

Temperature-induced cell death is thought to be due to protein denaturation, but the determinants of thermal sensitivity of proteomes remain largely uncharacterized. We developed a structural proteomic strategy to measure protein thermostability on a proteome-wide scale and with domain-level resolution. We applied it to *Escherichia coli*, *Saccharomyces cerevisiae*, *Thermus thermophilus*, and human cells, yielding thermostability data for more than 8000 proteins. Our results (i) indicate that temperature-induced cellular collapse is due to the loss of a subset of proteins with key functions, (ii) shed light on the evolutionary conservation of protein and domain stability, and (iii) suggest that natively disordered proteins in a cell are less prevalent than predicted and (iv) that highly expressed proteins are stable because they are designed to tolerate translational errors that would lead to the accumulation of toxic misfolded species.

Temperature has a profound effect on cell physiology. Organisms grow optimally in narrow temperature windows. Heat is also an important epigenetic factor because expression of sex determinants in amphibians is influenced by the temperature of egg incubation (1). In humans, spermatogenesis ceases at body temperatures above 37°C (2). Fever in response to pathogen infection is a conserved biological trait in vertebrates associated with resolution of infections (3), and high temperature can selectively injure cancer cells or sensitize them to treatment modalities (4). In industrial and clinical applications, heat treatment is one of the most widely used methods for the inactivation of bacteria, and the understanding of thermotolerance has important implications for the development of industrial biocatalysts, vaccines, and probiotics (5, 6).

Because of its biological, clinical, and biotechnological implications, the response of cells to thermal stress has been the subject of extensive studies dating back to the early 1970s (7). Because proteins are both the least stable and the most abundant biomolecules, and cell death can be triggered by heat and other chemical or physical agents that alter protein structure, it has been

proposed that thermal sensitivity of a cell depends on protein denaturation and loss of proteome function (8). Consistently, the evolutionarily conserved response of cells to heat stress involves the expression of molecular chaperones that act to maintain protein folding (9). Studies based on purified proteins and sequence analyses of thermophilic organisms and their mesophilic counterparts have revealed that enrichment in specific amino acids or types of secondary structure enhance thermostability (10–12). Further, bioinformatic predictions indicate that ~30% of proteins in a proteome are intrinsically disordered, indicating that the folding state of this subset should be relatively insensitive to temperature changes (13, 14).

Despite these advances, the determinants of thermal sensitivity of proteomes remain uncharacterized. Cell death at high temperature may result from global loss of protein structure or may be caused by loss of specific proteins with key functional roles (8, 15). Distinguishing between these two scenarios and identifying the most and least stable proteins in a proteome would require a proteome-wide analysis of protein thermal stability under physiologically relevant conditions. Large-scale stability analyses thus far have used data from purified proteins from different studies (8, 15, 16). These data sets are very heterogeneous and include measurements conducted under various experimental conditions, which makes comparison difficult. Further, the cellular matrix likely influences protein stability because chaperones, molecular crowding, and binding events have profound effects on the structures of proteins (17). These factors could also affect current estimates of the prevalence

of intrinsically disordered proteins (IDPs) because these polypeptides could fold upon interaction with cellular components.

We developed an approach to evaluate the thermal stability of a multitude of proteins in a proteome simultaneously and directly in their cellular matrix. The approach relies on the combination of limited proteolysis (LiP) and mass spectrometry (MS), applied over a range of temperatures. Using our strategy, we obtained protein thermal denaturation profiles on a proteome-wide scale for *Escherichia coli*, *Saccharomyces cerevisiae*, the thermophilic bacterium *Thermus thermophilus*, and human cells. We used the data set generated to evaluate the in vivo determinants of protein and proteome thermostability, the prevalence of IDPs in cells, and the evolutionary conservation of protein thermal stability.

Systematic profiling of protein thermal stability in complex matrices

To evaluate the thermal stabilities of proteins on a proteome-wide scale directly in a biological matrix, we adapted a method recently developed by our group that measures protein structural changes in complex cell extracts, which we called LiP-coupled mass spectrometry (LiP-MS) (18). The LiP step includes treatment of native proteome extracts with a broad-specificity protease in order to generate structure-specific proteolytic patterns for each protein in the proteome. Structure-specific protein fragments are then denatured and reduced in size through further digestion with the specific protease trypsin so as to generate peptides compatible with bottom-up MS analysis (Fig. 1A). As a control, an aliquot of the same proteome taken before the LiP step is subjected only to trypsinization. The resulting peptide mixtures are analyzed by means of liquid-chromatography-coupled tandem MS (LC-MS/MS). Protein regions embedding LiP sites are identified by a decrease in abundance of the associated tryptic peptides or by appearance of peptides with nontryptic termini (half-tryptic peptides) in the native sample compared with the denatured control sample. Comparison of LiP patterns from proteomes subjected to different conditions enables the identification of protein regions that underwent perturbation-induced structural changes. Analysis of the sample subjected only to trypsinization controls for changes in protein concentration or endogenous protease activity due to the perturbations.

LiP cleavages preferably occur in locally unfolded protein segments (19). Therefore, we reasoned that LiP-MS could be used to monitor thermally induced unfolding of proteins with peptide-level resolution and on a proteome-wide scale. In our workflow (Fig. 1A), cells are first mechanically lysed, and proteomes are extracted under conditions that preserve native protein structures. Next, aliquots of these protein extracts are subjected to incubation at increasing temperatures in order to induce protein denaturation. At each temperature, the proteome is subjected to a LiP step with proteinase K (PK), followed by denaturation, complete trypsin digestion, and shotgun proteomic analysis (Fig. 1, A and B).

¹Institute of Biochemistry, Department of Biology, ETH Zurich (ETHZ), CH-8093 Zurich, Switzerland. ²Systems Biology Graduate School PhD Program, ETHZ and University of Zurich, CH-8093 Zurich, Switzerland. ³Institute of Molecular Systems Biology, Department of Biology, ETHZ, CH-8093 Zurich, Switzerland. ⁴Institute of Molecular Life Sciences and Swiss Institute of Bioinformatics, University of Zurich, CH-8057 Zurich, Switzerland.

*Corresponding author. Email: paola.picotti@bc.biol.ethz.ch

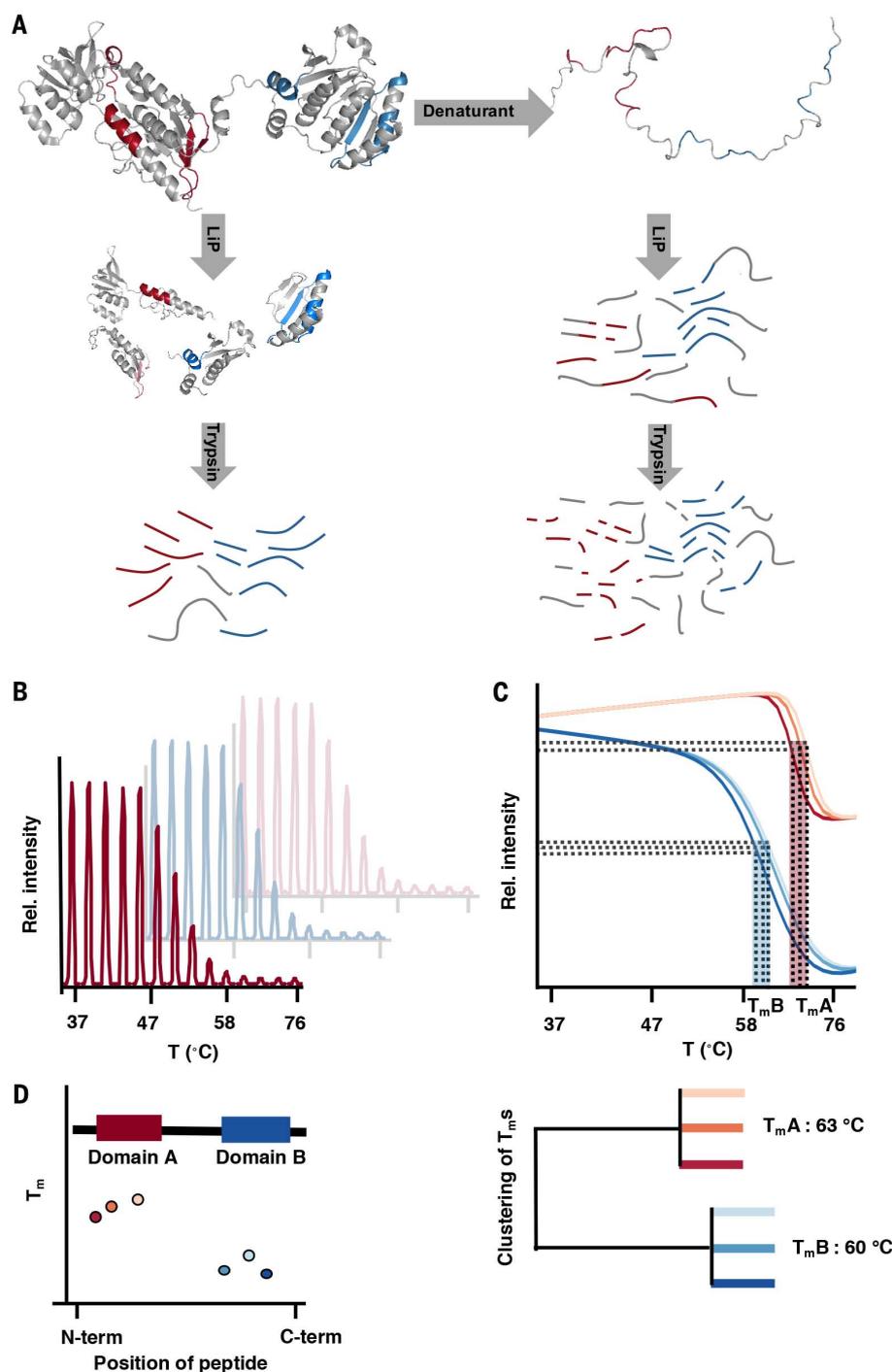


Fig. 1. Workflow of the LiP-MS-based analysis of protein stability. (A) Left and right images are illustrations of a folded protein and a protein subjected to denaturing conditions, respectively. In red and blue are tryptic peptides from domain A and B, respectively. At each temperature, the proteome is subjected to a LiP step by use of PK. Trypsin digestion leads to tryptic peptides (for example, red and blue stretches) and semitryptic peptides that arise from the LiP step. In the folded form, specific sites are protected from LiP. (B) Samples from each of the 14 temperature conditions are analyzed by means of shotgun MS, and peptides are quantified. (C) Peptide quantification data are fit to a two-state model from which T_m s are extracted. (D) Different domains may exhibit different thermal stabilities. In this example, domain A is more stable than domain B, and thus the curve obtained from quantification of the peptides from domain A and B are indicative of a higher T_m for domain A. Hierarchical clustering of peptide profiles allows identification of domains associated with different T_m s.

Proteolytic patterns obtained at the different temperatures are quantified and compared by use of a label-free MS approach. The signal intensities of distinct, fully tryptic peptides from each protein are plotted as a function of temperature, and the data are analyzed by fitting a sigmoidal model of a two-state denaturation curve. For peptides with sigmoidal intensity profiles, the temperature at the inflection point is extracted (Fig. 1C). We term this temperature the “apparent melting temperature,” referred to here as T_m because this parameter reflects the melting of a protein, but equilibrium may not have been reached at the considered temperatures.

Our approach results in peptide-level T_m values and thus in a variable number of T_m s for each protein. We expect that T_m s of peptides that map to the same folding unit or domain should coincide. Thus, for multidomain proteins, peptide measurements should result in a number of T_m s matching the number of domains, whereas peptides from single-domain proteins should have similar T_m s. We therefore applied hierarchical clustering to the T_m s extracted from multiple peptides from the same protein so as to reduce the number of T_m s for that protein to the number of domains (Fig. 1D). In order to compare T_m s of different proteins, we assigned to each protein the T_m of the domain with the lowest thermal stability, under the assumption that the function of the protein will be lost when the least stable domain is denatured.

Validation of the approach

To validate the proposed workflow for the analysis of protein thermal unfolding in a complex matrix, we selected four well-characterized single-domain proteins with different thermal stabilities: α -synuclein (α -Syn), bovine serum albumin (BSA), triose phosphate isomerase (TPI), and ubiquitin (Ub). First, we analyzed the thermal unfolding of the purified recombinant proteins using far-ultraviolet (UV) circular dichroism (CD) measurements (Fig. 2, A to D). TPI and BSA displayed sigmoidal denaturation profiles with T_m s of 61°C [coefficient of determination (R^2) = 0.995] and 65°C (R^2 = 0.995), respectively. No temperature-induced denaturation was observed for Ub or α -Syn in the considered temperature range, which is in agreement with the fact that they are very stable (20) and natively unfolded (fig. S1A) (21), respectively.

We then used LiP-MS as a readout for thermal unfolding of the pure proteins. We plotted the MS intensity of each detected tryptic peptide as a function of temperature (three replicates) (Fig. 2, E to H). Peptides from TPI and BSA displayed sigmoidal intensity profiles with T_m s of 59.5°C (R^2 = 0.950) and 63.9°C (R^2 = 0.905). The proteolytic peptides obtained from Ub and α -Syn did not show a sigmoidal decay along the temperature gradient. However, whereas α -Syn was heavily digested by PK at 37°C , Ub was not, according to the starting point of the thermal profiles, which is in agreement with the natively unfolded and very stable structures of the two proteins, respectively. For TPI and BSA, T_m s extracted

Fig. 2. Experimental validation of the approach and proteome-wide measurement of protein stability.

(A to D) Circular dichroism analyses of purified (A) α -Syn, (B) TPI, (C) BSA, and (D) Ub as a function of temperature. The data were normalized so that the highest value measured is 1 and baseline corrected so that the lowest value is 0.

α -Syn and Ub data could not be fit to a sigmoidal curve; hence, a linear model was chosen (dashed blue line).

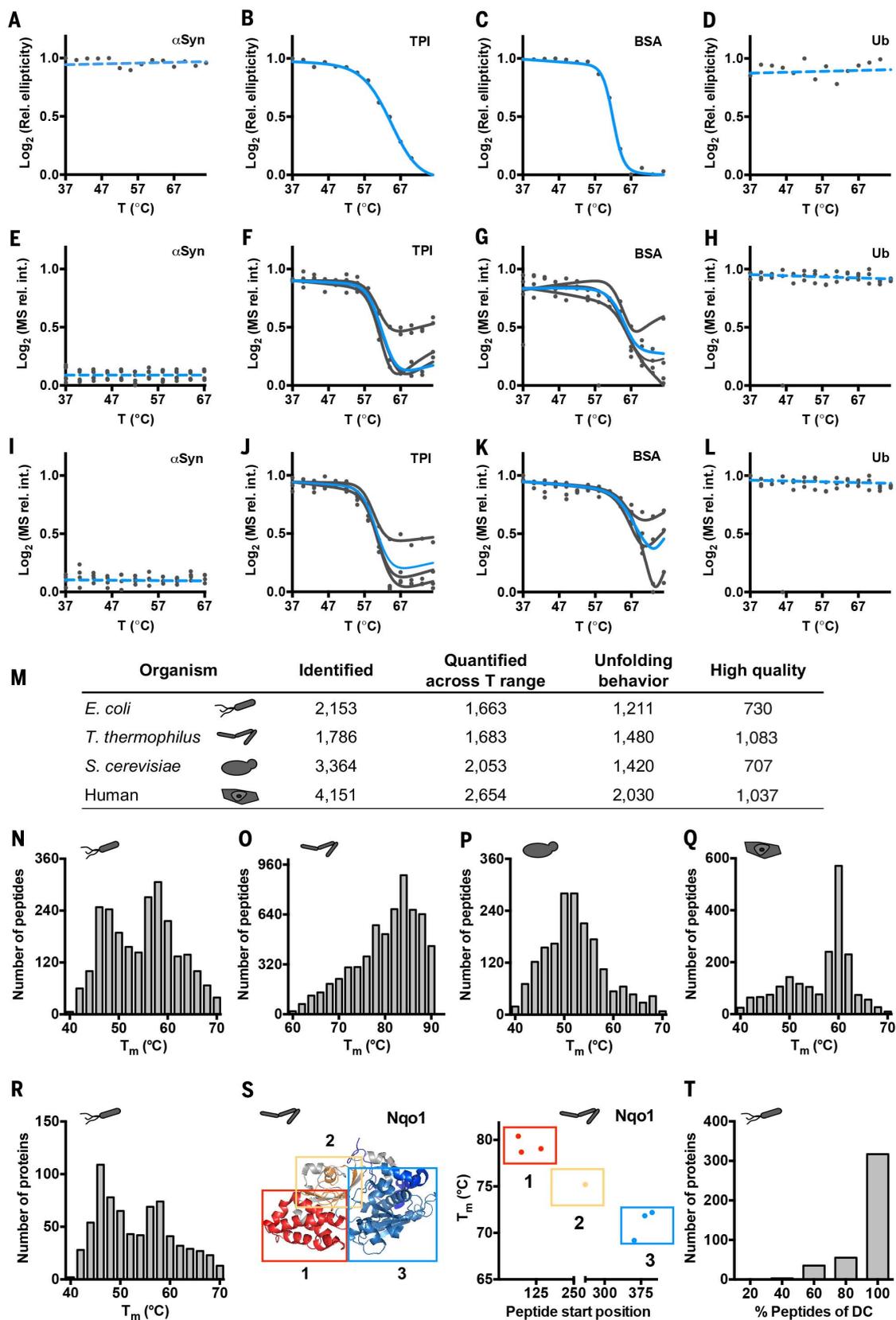
(E to H) LiP-MS analyses of purified (E) α -Syn, (F) TPI, (G) BSA, and (H) Ub, three replicates. The gray lines represent the fits of the three replicates (gray dots) for a single peptide. The blue line represents the average of all measured peptides for one protein.

(I to L) LiP-MS analyses of purified (I) α -Syn, (J) TPI, (K) BSA, and (L) Ub spiked into an *E. coli* background.

(M) Number of proteins analyzed for each species. "Identified," numbers of distinct proteins identified with MS (from the 37°C conditions, digested with trypsin only, all replicates); "Quantified across T-range," numbers of proteins consistently quantified across all 14 temperature points; "Unfolding behavior," numbers of proteins with sigmoidal melting curves; "High quality," numbers of proteins fulfilling our filtering criteria for high-quality fits (see text).

(N to Q) Distribution of T_m s of all peptides passing the high-quality filter in (N) *E. coli*, (O) *T. thermophilus*, (P) *S. cerevisiae*, and (Q) HeLa cells. (R) Distribution of protein T_m s for *E. coli*.

(S) Representation of the *T. thermophilus* protein Nqo1 (Protein Data Bank structure 3I9V) with the complex 151- K_d domain in red, the SLBB domain in yellow, and the NADH $^4Fe-^4S$ domain in blue (<http://pfam.xfam.org>) and T_m s of peptides mapped to the three domains. (T) For 410 *E. coli* proteins for which domain information was available in Pfam, frequency with which peptides mapping to the same domain were associated with the same T_m cluster (DC, domain cluster) are plotted.



from intensity profiles of different tryptic peptides were similar and resulted in one cluster per protein (tables S1 and S2); this is in line with TPI and BSA being single-domain proteins. The activity of PK was considered constant in the chosen temperature range, on the basis of analysis of a mixture of short synthetic peptides subjected to PK digestion (fig. S1B and table S1).

To evaluate whether our approach derived correct T_{ms} for proteins embedded in a complex biological matrix, we added the recombinant proteins to *E. coli* proteome extracts. LiP-MS-based melting curves were similar in the presence and absence of this background. TPI and BSA had T_{ms} of 58.5°C ($R^2 = 0.921$) and 66.1°C ($R^2 = 0.976$), respectively, and no sigmoidal decay was observed for Ub and α -Syn (Fig. 2, I to L). The differences in T_{ms} measured for TPI and BSA in the presence and absence of the cellular matrix were 1.0° and 2.2°C, respectively, which is comparable with differences observed when the same pure proteins were analyzed with far-UV CD and LiP-MS. Data from the ProTherm database (www.abren.net/protherm) (16) of published protein thermostability measurements indicate that the T_{m} of BSA varies by 3°C when different approaches are used to analyze thermostability, and even larger T_{m} variations are observed across experiments performed in different buffers. Overall, these data indicate that our approach enables extraction of T_{ms} of proteins in the presence of complex biological matrices and that these T_{ms} are similar to those obtained with standard methods typically applied to purified proteins.

Measurement of protein thermal stability on a proteome-wide scale

We applied our LiP-MS approach to the analysis of thermal unfolding of proteomes from organisms with different complexities and optimal growth temperatures (OGTs): *E. coli* (OGT, 37°C), *T. thermophilus* (OGT, 60°C), *S. cerevisiae* (OGT, 30°C), and human-derived HeLa cells (OGT, 37°C). We exposed each proteome (three replicates) to 14 temperatures and applied the LiP-MS workflow to each sample. To bracket the OGT of *T. thermophilus*, we adjusted the applied temperature range and replaced PK with thermolysin, a protease that remains active at high temperatures (22).

The numbers of proteins measured for each species are summarized in Fig. 2M. The total number of proteins detected ranged from 1786 (for *T. thermophilus*) to 4151 (for human cells). As the temperature was increased, the number of detected proteins decreased because at higher temperatures there is a progressive decrease in tryptic peptide abundance and an increase in half-tryptic peptide fragments and overall sample complexity. To increase the quality of our data and avoid quantification of noise, we retained only peptides that could be quantified across the complete temperature gradient. We further filtered our data by retaining only peptide profiles that corresponded to sigmoidal decay. We also excluded that our unfolding profiles were affected by temperature-induced changes in the activity of endogenous proteases in the lysates (fig. S1,

C to E). We were able to monitor the complete unfolding profiles of 2030 *E. coli* proteins, 1211 thermophile proteins, 1420 yeast proteins, and 1480 human proteins (table S2). We further

selected a subset of high-quality measurements for proteome-wide comparative analyses, based on the following criteria: (i) quality of fit of the selected sigmoidal model ($R^2 > 0.7$); (ii) extracted

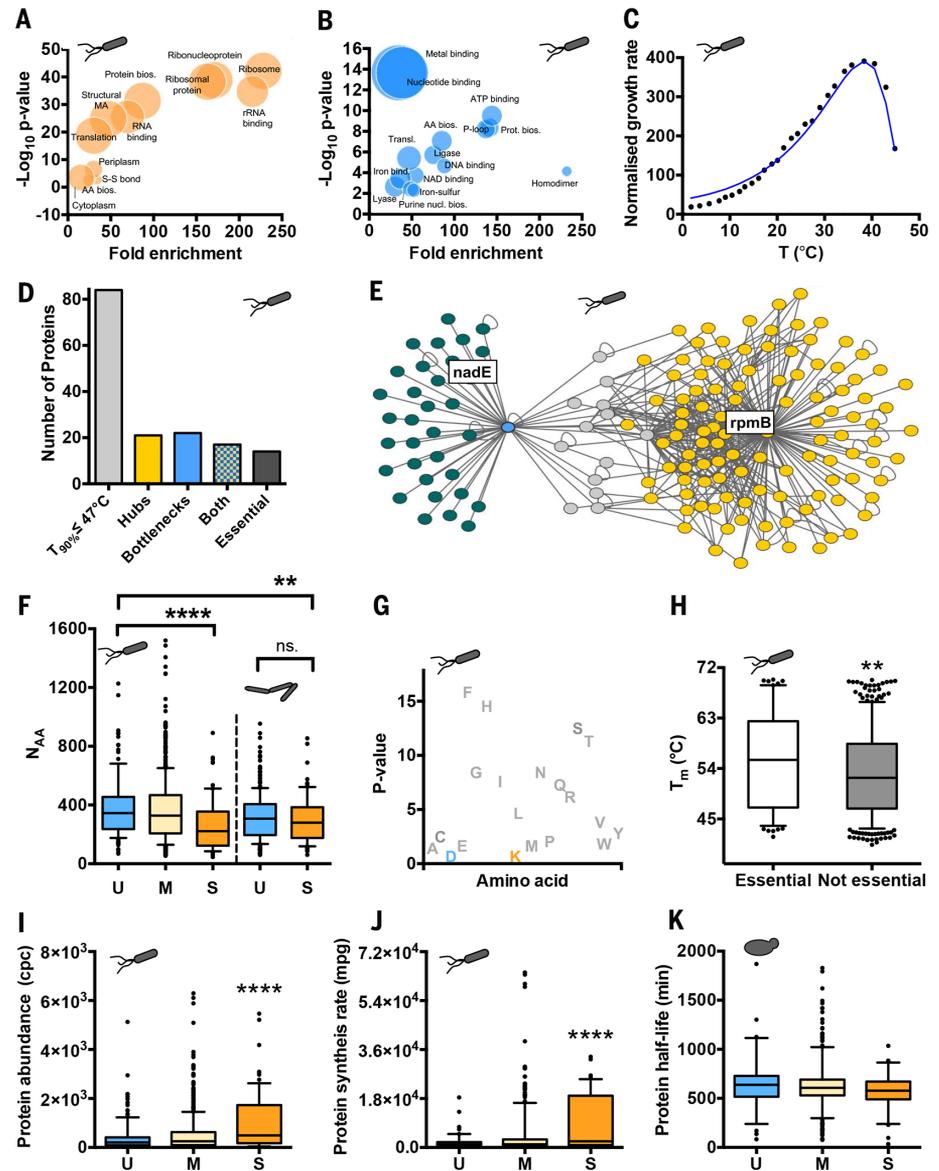


Fig. 3. Comparison of protein stabilities with structural and biological features in *E. coli*. (A and B) Functional enrichment analysis for (A) stable and (B) unstable proteins in *E. coli*. Fold enrichment of a specific functional term is plotted versus statistical significance. The circle size reflects the number of proteins matching a given term. Functional enrichment was performed with the tool DAVID (<https://david.ncicrf.gov>) by using GO and Swiss-Prot Protein Information Resource terms. (C) Growth rate of *E. coli* as a function of temperature [adapted from (11)]. (D) Graphical representation of the *nadE* and *rpmB* sub-network highlighting bottlenecks (*nadE*) and hubs (*nadE* and *rpmB*). (E) Frequency of hubs, bottlenecks, and essential genes in the set of *E. coli* proteins with $T_{90\%}$ values <47°C. (F) Comparison number of amino acids and stability (U, unstable; M, medium; S, stable proteins) in *E. coli* ($P = 7.9 \times 10^{-6}$, Wilcoxon test) and *T. thermophilus* proteomes (comparison across species, $P = 0.0024$, Wilcoxon test). (G) Plot of P values corrected for multiple testing of correlating amino acid composition with stability (Wilcoxon test; gray, no significant difference; blue, more abundant in unstable proteins; orange, more abundant in stable proteins). (H) T_m values for essential and nonessential proteins in *E. coli* ($P = 0.0079$, Wilcoxon test). (I) Protein abundance in copies per cell (cpc) in stability classes of *E. coli* ($P = 0.0001$, Wilcoxon test). (J) Protein synthesis rates of *E. coli* stable, medium, and unstable proteins. Rate of synthesis of stable proteins is significantly higher than that for unstable proteins ($P = 0.0017$, Wilcoxon test). $**P \leq 0.01$; $****P \leq 0.0001$. (K) *S. cerevisiae* protein half-lives versus stability class.

T_m s at least 6°C higher and 6°C lower than the lowest and highest temperature conditions, respectively, in order to exclude T_m s based on fewer than three measurement points; and (iii) decreasing signal over the sigmoidal region of the function. The resulting high-quality data subset included T_m s for 730 *E. coli* proteins, 1083 thermophile proteins, 707 yeast proteins, and 1037 human proteins (table S3).

On the basis of this data set, we then analyzed the distributions of peptide-level T_m s across the temperature range for each species. *E. coli* peptides showed a double bell-shaped T_m distribution (Fig. 2N). Distributions of *T. thermophilus* and yeast peptides were slightly skewed bell-shapes, with mean T_m value around 25°C above the OGT (Fig. 2, O and P). For both proteomes, the mean T_m value was around 25°C above the OGT. The T_m distribution of human peptides had a peak at around 60°C (Fig. 2Q), likely because of the under-sampling of the human proteome; the fraction of the proteome analyzed was considerably smaller for human than for other organisms. The distributions of protein-level T_m s were similar to those observed at the peptide level (Fig. 2, N and R, and fig. S2, A to D).

T_m clusters identify protein domains

To evaluate our interpretation of peptide clusters as independent folding units of a protein, we compared our results with available domain

annotation data from the Pfam database (<http://pfam.xfam.org>). We first focused on a protein for which three-dimensional structure and domain information is available, nicotinamide adenine dinucleotide (reduced form) (NADH)-quinone oxidoreductase 1 (Nqo1) from *T. thermophilus*. Our approach identified seven peptides from this protein with high-quality sigmoidal unfolding profiles, resulting in three different T_m clusters. The peptides from each of the three T_m clusters fell within structural boundaries of the three domains of the protein (fig. 2S). Thus, our approach correctly predicted the number and position of domains for this protein and resulted in measured T_m s for each domain.

We next used proteome-wide domain information in Pfam to evaluate the reliability of our approach on a large scale. We focused on the *E. coli* proteome and excluded proteins from our analysis for which only one peptide was identified and those for which no domain information was available. For each protein, we extracted domain information from Pfam and identified peptides in our data set matching each domain. For these 410 proteins, we evaluated how frequently peptides mapping to the same domain were associated with the same cluster. For 284 proteins (70%), all peptides in a given Pfam-annotated domain were associated with the same T_m cluster. For an additional 73 proteins (18%), 80% of the peptides from a given domain were

in the same T_m cluster (Fig. 2T). This indicates that our approach can be used to identify domains in proteins for which a three-dimensional structure is not available, although it cannot identify boundaries to amino acid-level resolution.

Evaluating the determinants of thermal sensitivity of a proteome

To identify the determinants of thermal sensitivity of a proteome, we asked whether thermal collapse of a cell could be attributed to specific classes of proteins. We classified the 10% of proteins with the lowest T_m s as unstable, the 10% of proteins with the highest T_m s as stable, and the remaining proteins as having medium stability. For the *E. coli* proteome, which had the largest coverage, we performed functional enrichment analyses (Fig. 3, A and B). We found a clear separation of ontology terms. Stable proteins were enriched for ribosomal, RNA-binding, and protein biosynthesis processes (Fig. 3A). Unstable proteins were enriched for cofactor and DNA-binding proteins (Fig. 3B). This might be due to the fact that proteins that bind to other biomolecules may require a high degree of conformational flexibility to enable these interactions.

Next, we evaluated the relationship between the temperature-induced loss of cell and proteome function. We used growth rate data for *E. coli* from a previous study (11) to determine the temperature at which cells show signs of heat-induced physiological impairment (Fig. 3C). *E. coli* cells undergo a heat-induced increase in growth rate, then the growth rate drops precipitously with a transition midpoint at 47°C. To identify those proteins that lose their functionality at this transition temperature because of thermal denaturation, we calculated the temperature at which 90% of each protein was denatured ($T_{90\%}$) and function presumably lost. Compared with that of T_m s, the distribution of $T_{90\%}$ values (fig. S2, A to D) was shifted toward higher temperatures by a few degrees celsius. At 10°C above the OGT, 83 of the measured *E. coli* proteins were denatured. These included 14 proteins encoded by essential genes involved in processes such as protein, nucleic acid, and fatty acid biosynthesis and nucleotide and cofactor binding (table S5).

To evaluate whether the proteins with function lost at 47°C included additional crucial nodes, we assessed whether any of the 83 proteins were hubs or bottlenecks in the *E. coli* protein-protein interaction network (Fig. 3D, fig. S3A, and supplementary text). Hubs are nodes with a large number of edges and typically have crucial functionalities for the survival and fitness of an organism (23). Bottlenecks are defined as necessary bridges between portions of a biological network, through which interactions must pass to connect distant network branches (supplementary text). Our list of *E. coli* proteins with $T_{90\%}$ values less than 47°C contained 21 hubs (such as rpmB) (Fig. 3E) and 22 bottlenecks; 17 proteins, such as the enzyme catalyzing NAD biosynthesis (nadE), were both hubs and bottlenecks (Fig. 3E and table S5). The distribution of

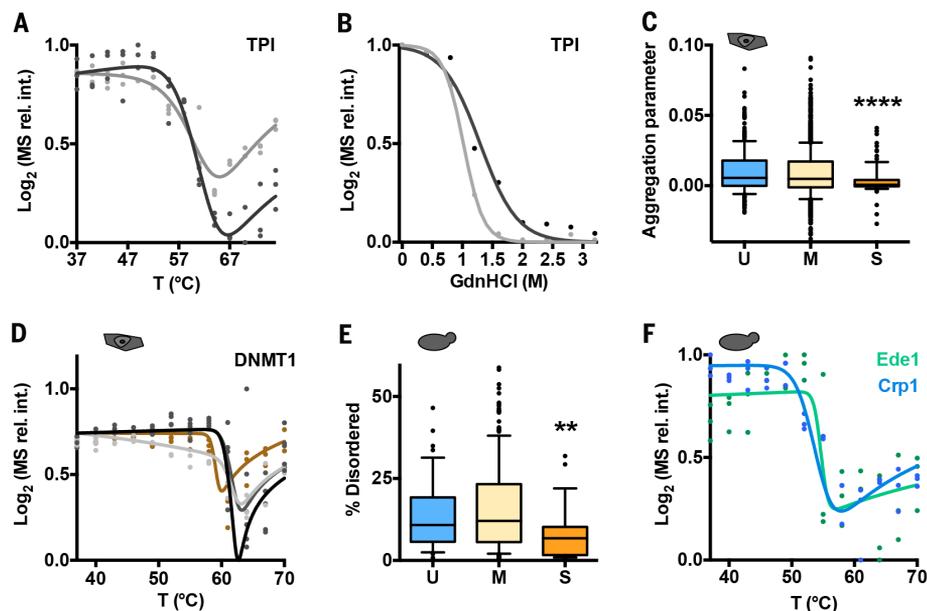


Fig. 4. Thermal profiles to detect aggregation and intrinsic disorder. (A) TPI was spiked into an *E. coli* lysate, and two peptides (SNVSDAVAQSTR and IYGGSVTGATCK, different from those shown in Fig. 2J) were analyzed in triplicate in a thermal denaturation experiment. (B) The same two peptides from TPI spiked into an *E. coli* lysate analyzed in triplicate over a GdnHCl denaturation gradient. The concentration of PK was adjusted to correct for GdnHCl-induced loss of PK activity (fig. S1F). (C) Aggregation parameter from the thermal denaturation experiment versus stability classes (U, unstable; M, medium; S, stable) of human proteins indicates a highly significant reduction in aggregation propensity in stable human proteins compared to unstable proteins ($P = 1.64 \times 10^{-4}$, Wilcoxon test). (D) Thermal profiles (three replicates) of peptides from the aggregation-prone human protein DNMT1. (E) Percent structural disorder for the different stability classes in *S. cerevisiae* ($P = 0.018$, Wilcoxon test). (F) Thermal profiles of *S. cerevisiae* proteins Ede1 (green) and Crp1 (blue) from triplicate experiments. ** $P \leq 0.01$; **** $P \leq 0.0001$.

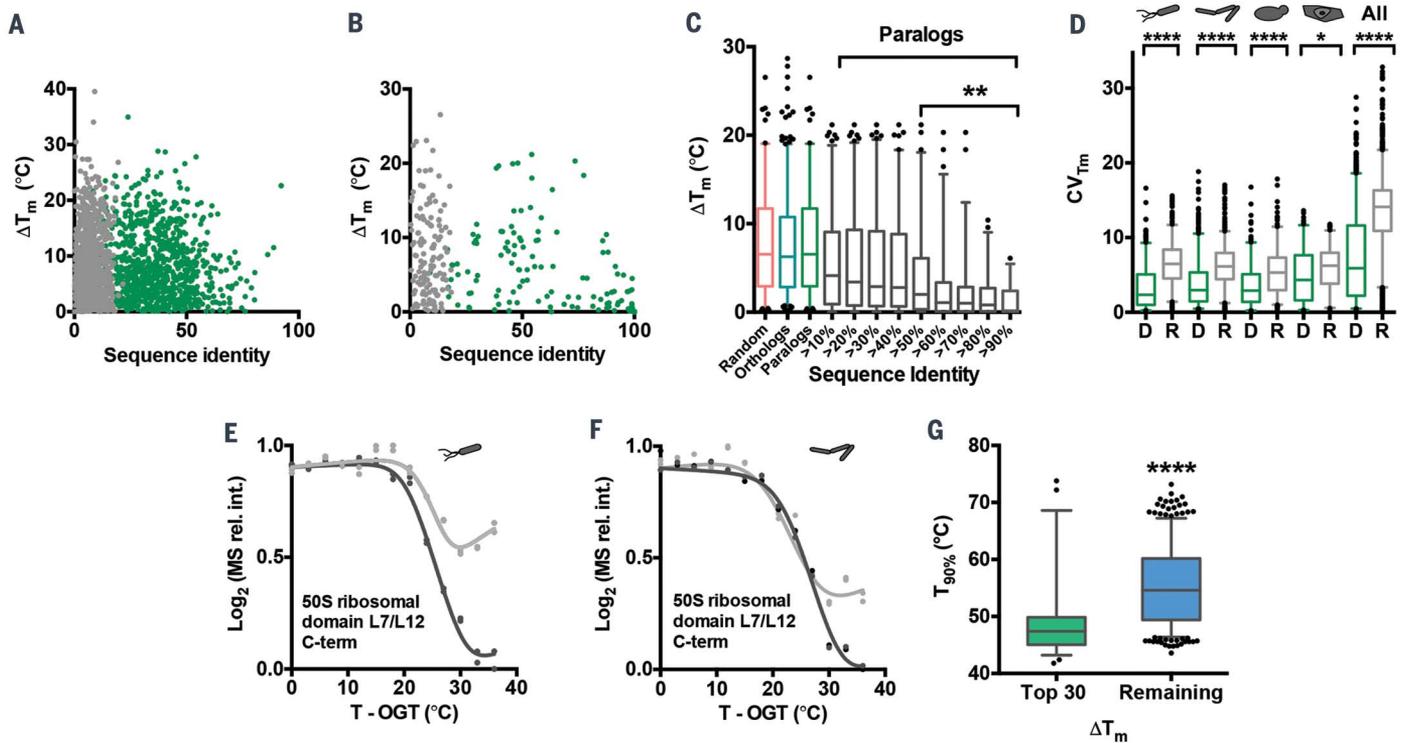


Fig. 5. Evolutionary conservation of protein and domain stability. (A and B) Plot of ΔT_m versus sequence identity for (A) orthologous and (B) paralogous protein pairs from all proteomes in our data set (green) and randomly assigned pairs (gray). (C) Plot of ΔT_m versus sequence identity for random (red), orthologous (blue), and paralogous (green and gray) protein pairs. Gray bars show paralogous protein pairs classified according to sequence identity cutoffs indicated on the x axis. (D) Plot of the coefficient of variation (CV) of T_m s from all peptides associated to homologous domain pairs, D, and from the same number of randomly assigned peptides, R, for each organism and for all organisms combined. (E and F) Signal intensity versus

temperature plots for the 50S ribosomal C-terminal domain L7/L12 from (E) *E. coli* and (F) *T. thermophilus* from triplicate experiments. To enable comparison, the temperature was adjusted through subtraction of the OGT. (G) Analysis of T_m differences between *E. coli* and *T. thermophilus* ortholog pairs. Orthologous proteins were ranked according to their ΔT_m after normalization by OGTs. The 30 ortholog pairs displaying the largest T_m differences were compared with the remaining set of ortholog pairs. T_m differences between the two species are plotted against $T_{90\%}$ s of the associated *E. coli* proteins. Proteins displaying the largest T_m differences were significantly enriched for *E. coli* proteins losing their fold below 47°C ($P = 3.4 \times 10^{-7}$). * $P \leq 0.05$; **** $P \leq 0.0001$.

$T_{90\%}$ of hubs and bottlenecks showed a pronounced peak precisely at 47°C, the temperature of *E. coli* cell death (fig. S3B). Overall, this suggests that the sensitivity of *E. coli* cells to temperature is due to the loss of a small set of crucial effectors and regulators of biological processes.

Molecular and biological bases of protein thermostability

We next evaluated whether our conclusions on the influence of physicochemical and biological properties on protein stability differed from those previously derived from studies of purified proteins or genome-wide sequence analyses. We again chose *E. coli* as a paradigm because of the high proteome coverage we achieved for this species and the availability of genome-wide studies and protein stability data for *E. coli*. We first compared our data with thermal stability data from purified recombinant proteins in ProTherm. The two data sets showed a poor correlation ($R^2 = 0.1322$) (fig. S4A), suggesting that protein stability is substantially influenced by the presence of the cellular background.

Next, we evaluated the influence of specific protein structural features on protein thermal

stability. We analyzed the relationship between protein stability and protein length and found that stable proteins are significantly shorter than unstable proteins ($P = 7.9 \times 10^{-6}$, Wilcoxon test). Stable protein domains were also significantly shorter than unstable domains, indicating that this relationship is not affected by our use of T_m s of the most unstable domains (fig. S4C). We also compared the lengths of *E. coli* and *T. thermophilus* proteins and found that unstable and medium-stability proteins from *E. coli* were significantly longer than were *T. thermophilus* proteins from the corresponding stability groups ($P = 0.0024$, Wilcoxon test) (Fig. 3F). Stable *E. coli* proteins were of comparable length with stable *T. thermophilus* proteins, and no significant difference in protein length was observed across *T. thermophilus* proteins from the different stability groups. Overall, this suggests that protein length is a determinant of protein thermal stability. Comparison of amino acid frequencies in stable and unstable *E. coli* proteins revealed slight but significant enrichments in lysine ($P = 0.002$, Wilcoxon test, Bonferroni corrected) and aspartic acid ($P = 1.216 \times 10^{-4}$, Wilcoxon test, Bonferroni corrected) in stable and unstable proteins, respectively (Fig. 3G).

Next, we studied the secondary structure content of proteins from different stability classes. We predicted the α -helical, β -sheet, and coil contents of *E. coli* proteins as previously described (24). Stable proteins were significantly enriched in β -sheet elements ($P = 0.0327$, Wilcoxon test) (fig. S4D) and unstable proteins in α -helical structure ($P = 0.047$, Wilcoxon test) (fig. S4E). Coil content did not significantly differ across stability classes.

We then tested for a relationship between gene essentiality and thermostability of the protein product. We found that proteins encoded by genes shown to be essential for *E. coli* growth were more stable than were proteins encoded by nonessential genes ($P = 0.0079$, Wilcoxon test) (Fig. 3H).

We found that protein abundance was significantly higher for stable proteins than for unstable proteins ($P = 2.5 \times 10^{-5}$, Wilcoxon test) (Fig. 3I). According to data from previous ribosome profiling experiments (25), stable proteins were synthesized at significantly higher rates than those of unstable proteins ($P = 0.0017$, Wilcoxon test) (Fig. 3J). Protein turnover data are not available for the *E. coli* proteome. For the *S. cerevisiae* proteome (26, 27), we found no correlation between protein

turnover rate and protein thermal stability (Fig. 3K and fig. S4F). This is in line with the fact that enriched Gene Ontology (GO) terms for stable and unstable proteins (Fig. 3, A and B) differ from functional terms enriched for proteins with low and high turnover rates (28), with the exception of a common enrichment in ribosomal and translation-related proteins for thermally stable and low-turnover proteins. This in turn suggests that the determinants of protein thermostability differ from those of susceptibility to proteolysis.

Overall, our proteome-wide analysis of thermally induced protein unfolding enabled the identification of new relationships between protein stability and the function and fate of cellular proteins. None of our conclusions changed when we chose alternative criteria to classify stable and unstable proteins or used alternative statistical tests (supplementary text).

Temperature-induced aggregation

From inspection of our proteome-wide denaturation curves, we noticed that specific peptides from some proteins showed an increase in proteolytic resistance after the main unfolding transition. For example, TPI peptides SNVSDAVAQSTR and IYGGSVTGATCK became PK-resistant at high temperatures (Fig. 4A and fig. S5A). We reasoned that this could be due to aggregation events that occur during or after unfolding that result in proteolytic protection of residues located at aggregation interfaces.

To evaluate this possibility, we compared thermal denaturation curves of specific proteins with denaturation curves obtained for the same proteins by progressively raising the concentration of guanidinium hydrochloride (GdnHCl), a chemical denaturant that inhibits aggregation (29). Posttransition slopes obtained as the GdnHCl concentration was increased were substantially lower than slopes of thermal transitions for all peptides measured (Fig. 4B and fig. S5A). This indicates that posttransition slopes reflect the extent of unfolding-induced aggregation. We term this slope the aggregation parameter. In agreement with our interpretation, no aggregates of TPI were detectable at 37° or 55°C but clearly formed at 76°C, according to electron microscopy analysis (fig. S5B), as previously reported (30).

Our aggregation parameter significantly correlated with the propensity to ordered β -sheet aggregation (fig. S5D), as predicted with the TANGO tool (31). However, this relationship was weak, most likely because the propensity to form heat-induced amorphous aggregates does not necessarily reflect the propensity to spontaneous β -sheet aggregation and amyloidogenesis. Indeed, α -Syn, a highly amyloidogenic protein, did not form aggregates at high temperature (Fig. 2), which is in line with previous reports (32).

Stable proteins from human cells were significantly less prone to thermal aggregation than were unstable proteins ($P = 0.0001$, Wilcoxon test) (Fig. 4C and table S3). Similar trends were observed for the two bacteria, but not for yeast (fig. S5C). As an example, four peptides from human DNMT1, a protein shown to form intra-

cellular aggregates (33), displayed high aggregation propensity at high temperatures (Fig. 4D).

T. thermophilus proteins were largely less prone to aggregation than were *E. coli* proteins (fig. S5E). This suggests that a reduction in thermal aggregation is among the determinants of protein thermal stability.

Intrinsically disordered proteins

According to recent predictions, up to 30% of cellular proteomes consist of IDPs (13, 14). Because of the challenges in analyzing protein folding in cells, this estimate has not been experimentally validated for a cellular proteome. We used our thermal unfolding data set to evaluate the prevalence of IDPs. Theoretically, natively unfolded proteins should not display sigmoidal thermal unfolding profiles because they should not undergo two-state unfolding transitions as temperature is increased, as shown for the IDP α -Syn (Fig. 2, E and I). Similarly, proteins with large locally unfolded regions should appear to be more unstable than highly structured proteins or should be more prone to temperature-induced aggregation events.

We compared our results with a recent analysis of protein disorder in *S. cerevisiae* (14). We found that thermostable proteins were significantly less disordered than were unstable proteins (Fig. 4E).

Next, we evaluated the unfolding profile of *S. cerevisiae* proteins predicted to be highly unstructured and for which we could quantify an MS signal across the whole temperature range. Of these 333 proteins, 153 (46%) showed a sigmoidal unfolding behavior; 74 of these were from the high-quality data set and had melting temperatures between 40.4° and 60.9°C. We repeated the same analysis focusing on the top 30 most unstructured proteins identified in the previous study, and the results were similar: For 47% of these proteins, sigmoidal melting curves were observed. As an example, Fig. 4F shows the unfolding profiles of Ede1 and Crp1, lipid- and DNA-binding protein, respectively, both predicted to be IDPs. Thus, about half of proteins predicted to be highly disordered undergo local or global transitions from folded to unfolded states (34). In cells, these proteins may acquire structure upon interaction with molecular chaperones or binding to other molecules. In support of this, of the predicted IDPs that had sigmoidal melting curves, 54% had GO molecular functions associated with binding of ligands. Among the most populated protein classes were nucleic acid-binding (31%), enzyme-modulating proteins (13%), and proteins involved in membrane trafficking (8%). IDPs that showed two-state unfolding transitions were enriched in RNA-binding proteins ($P = 6.1 \times 10^{-6}$, after Bonferroni correction), ribonucleoprotein complexes ($P = 2.7 \times 10^{-4}$), and proteins interacting with lipid vesicles ($P = 0.0053$).

Of the remaining predicted IDPs, 81 proteins (24%) showed an increasing degree of structural protection along the temperature gradient, with sigmoidal transition midpoints between 36.5° and 76.1°C. This is suggestive of temperature-induced aggregation phenomena involving proteins in their native state. The high aggregation

propensity of unstructured proteins has been reported previously (35). For 99 of the remaining proteins, our peptide data are indicative of the simultaneous unfolding of certain regions and aggregation of others, suggesting the coexistence of folded and locally unfolded, aggregation-prone domains. For two proteins, the respective peptides showed a constant relationship between PK cleavage and temperature, compatible with natively unfolded structures, invariant along the temperature gradient. Thus, only about half of proteins predicted to be highly unstructured displayed temperature-induced unfolding profiles indicative of intrinsic disorder. For 37 proteins, no clear trend could be derived. Thus, in a cellular context the fraction of IDPs may be substantially lower than originally anticipated, and many of the predicted IDPs are likely to locally or globally acquire structure upon binding of specific ligands.

Evolutionary components of protein thermal stability

Last, we asked whether protein stability is an evolutionary conserved feature. We first identified orthologous and paralogous genes using the eggNOG database (<http://eggnogdb.embl.de>) (36). Orthologs are genes from different species that evolved from a common ancestral gene through speciation and typically retain the same function. Paralogs arise from genomic duplication within the same species and typically evolve different functions. In eggNOG, the age of orthologous proteins is equal to the time of the speciation event. Paralogous proteins are almost always younger and arise from gene duplication events that occur well after speciation. Comparison of the T_{ms} of proteins from orthologous gene pairs and randomly selected pairs mostly showed no significant differences ($P > 0.01$, Wilcoxon test) for sequence identities below 60% (Fig. 5A and fig. S6, A and C). Recently diverged paralogs with high sequence identity (above 60%) had more similar T_{ms} than random protein pairs ($P < 1 \times 10^{-9}$, Wilcoxon test) (Fig. 5, B and C, and fig. S6, B and D). The same test for orthologs was inconclusive because of a lack of orthologs with suitably high sequence identity (Fig. 5A and fig. S6C). The conservation of T_{ms} for paralogous proteins could be the result of their younger evolutionary age as compared with orthologous proteins. The lack of orthologs at comparable levels of sequence identity prevents drawing conclusions on whether the mode of inheritance matters for the conservation of protein stability.

Next, we asked whether homologous domains had similar stabilities. We compared the differences in T_{ms} of the same domain within an organism and across all organisms measured. When the differences in T_{ms} of the same domains in different proteins were compared with differences between randomly selected domain pairs, we found that homologous domains both within an organism (*E. coli*, $P < 2.2 \times 10^{-16}$; *T. thermophilus*, $P < 2.2 \times 10^{-16}$; *S. cerevisiae*, $P = 4.872 \times 10^{-13}$; human cells, $P = 0.015$) and across all organisms ($P < 2.2 \times 10^{-16}$, Wilcoxon test) had smaller ΔT_{ms} than those of random pairs (Fig. 5D). As an

example, after correcting for the OGT, the C-terminal domains of the L7/L12 50S ribosomal proteins from *E. coli* and *T. thermophilus* have similar T_{ms} s (Fig. 5, E and F). This indicates that OGT-corrected domain stability is highly conserved and is not substantially influenced by the structural features of the flanking protein regions.

Last, we used our analysis of ortholog pairs from *E. coli* and *T. thermophilus* to evaluate whether protein thermostability is uniformly increased in the proteome of thermophiles or whether specific classes of proteins are preferentially stabilized. Analysis of ΔT_{ms} s of orthologous protein pairs revealed that protein thermal stability was enhanced to different extents across the proteome of *T. thermophilus* (fig. S6E). Proteins displaying the lowest degree of thermal adaptation were enriched in ribosomal proteins ($P = 6.8 \times 10^{-11}$, after Bonferroni correction). Proteins subjected to the most pronounced thermal adaptation were strongly enriched in orthologs of the 83 *E. coli* proteins that denature below the *E. coli* OGT ($P = 3.395 \times 10^{-7}$) (Fig. 5G). This suggests that cells achieve thermal adaptation through the preferential stabilization of a specific set of physiologically crucial proteins.

Discussion

Reductionist approaches focused on single proteins in dilute solutions have substantially contributed to our understanding of protein biophysics and function; however, it is unknown whether the behavior of proteins in vitro resembles that in cells (37). Protein stability is likely to be strongly influenced by the cellular milieu because of molecular crowding, intermolecular interactions, chaperones, cotranslational folding, and variable translation rates (37, 38). Identifying the most stable and labile proteins in proteomes and comparing proteome stabilities across species pose additional challenges because a multitude of proteins must be evaluated simultaneously.

These considerations have prompted researchers to search for methods that enable evaluation of folding and stability of multiple proteins directly in a cellular matrix. Nuclear magnetic resonance and fluorescence-tagging approaches (39, 40) have allowed analysis of protein folding in cells, but these methods require protein labeling, which may affect protein stability and limits multiplexing. Computational analyses of protein sequences and simulations of a crowded cytoplasm have analyzed the stability of multiple proteins (8) and the effects of a cellular matrix (38), but different approaches have reached conflicting conclusions (38, 41, 42).

In this study, we presented a proteomic approach to measure the thermal stability of proteins directly in a cellular matrix on a proteome-wide scale. The approach can be applied to evaluate the effects of heat or other denaturants on proteome stability. The peptide-level resolution allows analysis of the stabilities of individual domains within a protein. Application of our approach to proteomes from different species resulted in the largest data set on protein stability to date and contributed insights into the stability of proteins, domains, and pro-

teomes. Comparison of our data on endogenous proteins with thermal stability measurements from pure recombinant proteins (16) revealed that the cellular matrix substantially influences protein stability. T_{ms} s extracted from our measurements deviated as much as 22°C from those in ProTherm. Both stabilization and destabilization of proteins relative to measurements in isolation were observed (fig. S4, A and B), suggesting that the effect of the cellular matrix on protein stability cannot be generalized. This conclusion is not in conflict with the results of our validation experiments in which the T_{ms} s of recombinant proteins (TPI, BSA, α -Syn, or Ub) produced in bacteria, purified, and added to the proteome extract of a species that did not contain the protein matched those of the pure protein in isolation. Under these conditions, we do not expect that the background provided by the foreign cell lysate plays a major role in influencing protein stability because cellular events such as cotranslational folding, molecular chaperoning, or posttranslational modifications are not or much less likely to occur in the lysate for a spiked-in protein produced elsewhere.

Our distribution of protein thermal stabilities for *E. coli* differs from that previously predicted (8). The previous simulation predicted a broad distribution of protein stabilities with a pronounced tail toward high temperatures. That distribution suggested that a multitude of proteins are less stable than the average protein and that proteome instability derives from the generalized, simultaneous loss of hundreds of proteins, described as a proteome catastrophe. Further, according to the previous model, the shortest proteins are the most unstable and are the determinants of cellular collapse. We observed that at a temperature at which cells experience temperature-induced physiological impairment, a subset of essential proteins and proteins mediating crucial cellular interactions undergoes loss of function. This subset did not show sequence length biases; in fact, shorter proteins were found to be generally more stable. These proteins were also subjected to selective thermal adaptation in the thermophilic bacterium, indicating that stabilizing key temperature-sensitive proteins is more cost-effective than a uniform thermal modulation of the whole proteome. We thus propose that the temperature-induced loss of a relatively small set of proteins that play key physiological roles (table S5) is sufficient to explain the thermal sensitivity of cells.

We cannot exclude the possibility that our data are affected by the issue of proteome undersampling. However, supplementation of the functions of specific proteins predicted to be unstable metabolic hubs was sufficient to alleviate temperature-induced growth defects (15), supporting the idea that loss of a specific protein subset mediates cellular dysfunction at increased temperature. The experimental identification of labile proteins might guide engineering of microorganisms with increased thermostability to be used in industrial processes operating at high temperature.

Our analysis also revealed the structural and biological bases of protein thermostability. Confirming results of previous studies based on the

comparison of genomes of thermophilic and mesophilic bacteria, we observed enrichment for lysine residues and β -sheet structures in thermostable proteins (43). Our experimental stability data do not support the previously reported correlation between enhanced thermostability and increased occurrence of negatively charged residues (44). We rather found that unstable proteins have a higher content of aspartic acid residues. Previous genome sequence analyses led to conflicting results on a possible relationship between protein stability and protein length (10, 11, 45, 46). We experimentally show an inverse correlation between protein length and thermal stability. Thus, the structural determinants of protein stability extracted from bioinformatic analyses match only partly those derived from our proteome-wide measurements in the cellular matrix. Our comparison of stable and unstable proteins within the same species may have eliminated analytical biases due to specific evolutionary trends of previous analyses based on cross-species comparisons. We also showed that thermostable proteins are substantially less prone than unstable proteins to thermal aggregation and to local unfolding, suggesting that a low aggregation and unfolding propensity (43) are among the determinants of protein thermostability.

Domain thermal stability, normalized by OGT where needed, was conserved, both for proteins of the same species and across organisms. This suggests that domain stability is under high selection constraint. We also found conservation of thermal stability for proteins synthesized from paralogous genes, which typically evolve to have different functions. In contrast, thermal stability was not generally similar for proteins from orthologous genes, although these genes typically retain similar functions. This result might seem in conflict with the observed conservation of domain stabilities across species because orthologs of multidomain proteins typically retain functionally important domains (47, 48). This apparent discrepancy is explained by the fact that both gains and losses of domains from functionally related proteins occurred frequently during evolution, with eukaryotic proteins including substantially more domains than bacterial proteins, and bacterial proteins comprising more domains than archaeal proteins (49–51). Because the overall stability of a protein depends on its domain composition, protein T_{ms} s are affected by reshuffling of protein domains, despite the conservation of domain stability.

A clear correlation was observed in our study between protein stability and abundance. We found that stable proteins are substantially more abundant and less prone to aggregation. These findings are of particular importance in light of the recent “translational robustness” theory, formulated on the basis of genome-sequence analyses, gene expression data, and evolutionary simulations (52–54). According to this theory, highly expressed proteins should tolerate translational errors that may lead to the accumulation of toxic misfolded species. According to the law of mass action and the known high rate of mistranslation, misfolding of highly abundant proteins due to

translation errors is potentially more deleterious to the cell than misfolding of low-abundance proteins. Mistranslation-induced misfolding can be reduced by increasing translational accuracy or by increasing tolerance of protein folds to missense mutations (54). Our data show a clear direct relationship between protein-folding stability and intracellular abundance and an inverse relationship between protein stability and aggregation or local unfolding. Increasing the thermodynamic stability of abundant proteins will broaden the range of amino acid replacements that a protein can tolerate before misfolding (55). This suggests that selection reduces the burden of intracellular misfolding by increasing the thermodynamic stability of abundant proteins, thus expanding the theory of translational robustness. It also suggests that small alterations of the fine balance between protein concentration and aggregation propensity (for example, by an increase in the concentration of unstable, aggregation-prone proteins or by oxidation of abundant proteins that increases aggregation) may cause a misfolding catastrophe, potentially leading to protein-misfolding diseases, as postulated in the “edge” theory by Tartaglia *et al.* (56).

Our data revealed that about half of predicted IDPs (13, 14) showed two-state denaturation profiles in the cellular matrix and suggested that many predicted IDPs are structured in cells or undergo global or local disorder-to-order transitions upon ligand binding (13, 34).

Limitations of our approach include the need for cell lysis before analysis, which may result in dilution of the cellular medium and loss of compartmentalization. However, compartmentalization effects should be minimal for *E. coli*, the organism chosen for most of our analyses on thermostability determinants, because of its low degree of subcellular organization (57). Another limitation is incomplete proteome coverage, a typical issue of proteomics approaches (58), complicated here by the complexity of LiP digests under unfolding-inducing conditions. Proteome coverage could be increased by combining our method with an approach to drug target identification termed the cellular thermal shift assay (CETSA) (45). CETSA is based on the measurement of protein precipitation in cell lysates upon an increase in temperature. Shifts in the obtained thermal profiles are indicative of drug target engagement. The approach achieves high proteome coverage, but whether measurement of protein aggregation by CETSA can be used as a reasonable readout for protein unfolding would have to be experimentally validated. In addition, CETSA does not achieve peptide-level resolution.

Our study contributes insight into the molecular and evolutionary bases of protein and proteome thermostability and provides a blueprint for future studies on the stability of proteomes and thermal denaturation.

Materials and methods

Cell culture

E. coli cells (strain BW25113 K12) were grown at 37°C in 5-L shaker flasks in M9 minimal medium

(Fischer *et al.* 2003) containing 5 g/L glucose to an OD600 of ~0.8. Cells were harvested by centrifugation at 4000 g for 5 min at 4°C and washed twice with an ice-cold lysis buffer (20 mM HEPES, 150 mM KCl, 10 mM CaCl₂, pH 7.5). Cells were resuspended in the same buffer and disrupted by vortexing in the presence of acid-washed glass beads in three consecutive rounds of 30 s of beating and 4 min of incubation at 4°C. The lysates were centrifuged at 16,200 g for 15 min at 4°C to remove cellular debris, the supernatants were transferred to a fresh tube, and the protein concentration in the extracts was determined by the bicinchoninic acid assay (BCA Protein Assay Kit, Thermo Scientific). The protein extracts were stored at -80°C until use.

T. thermophilus cells (a kind gift of Prof. Hans Martin Fischer, ETH Zurich) were obtained as frozen 30%-glycerol stocks in 0.8% Bacto-Peptone, 0.4% yeast extract, 0.3% NaCl, pH 7.5. Th-A medium (0.5% Bacto-Peptone, 0.2% yeast extract, 0.2% NaCl, pH 7.0–7.2) was inoculated with a single colony from an agar plate, and cells were grown overnight at 60°C under vigorous shaking. Cells were harvested as described above for *E. coli*.

S. cerevisiae (strain BY4741) cells were grown at 30°C in synthetic complete (SC) liquid medium containing 2% glucose (w/v) to an OD600 of ~0.8 and harvested as described above for *E. coli*.

HeLa FRTTO T-Rex cells (a kind gift of Prof. Matthias Peter, ETH Zurich) were grown at 37°C in DMEM supplemented with 10% (v/v) fetal bovine serum, 100 U/ml penicillin, 100 µg/ml streptomycin, and 1% L-glutamate (GIBCO, Invitrogen). To harvest the cells, the cell culture medium was removed, and cells were washed with PBS. Cells were then mechanically removed from the plate, washed with 20 mM HEPES, 150 mM KCl, 10 mM CaCl₂, pH 7.5 and lysed by freeze milling. Subsequently, the lysates were treated as described for *E. coli* lysates.

Purified proteins

Triosephosphate isomerase (TPI) purified and lyophilized from rabbit muscle (Sigma), bovine serum albumin (BSA) from purified and lyophilized bovine plasma (Sigma) and ubiquitin (Ub) from bovine erythrocytes purified and lyophilized (Sigma) were used in the analyses with purified protein.

Circular dichroism

Circular dichroism spectra were recorded on a J-710 spectropolarimeter (Jasco) in a 1-mm path-length quartz cell. The protein concentration was 0.2 mg/ml. The mean residue ellipticity $[\theta]$ (deg·cm²·dmol⁻¹) was calculated from the formula $[\theta] = (\theta_{\text{obs}}/10) \cdot [\text{MRW}/(l \cdot c)]$, where θ_{obs} is the observed ellipticity in degrees, MRW is the mean residue molecular weight of the protein, l the optical pathlength in cm, and c the protein concentration in g/mL. Spectra were recorded in PBS, pH 7.4 and normalized by subtracting the lowest value measured followed by dividing all measurements by the largest value measured.

Limited proteolysis

Protein concentrations in different cell extracts were adjusted to 1 mg/ml with 20 mM HEPES, 150 mM KCl, 10 mM MgCl₂. The pH was adjusted

with 5 M NaOH to approximately 7.6. Cell extracts were pre-equilibrated at a given temperature for 5 min. Proteinase K from Tritirachium album (Sigma) was added to the protein extract at an enzyme:substrate ratio of 1:30 and incubated for 1 min at a temperature from 37° to 76°C (in 3°C increments, corresponding to 14 temperature points). Thus, cell extracts were exposed to each of the chosen temperatures for a total of 6 min, considering the preincubation time. The digestion was stopped by adding sodium deoxycholate (DOC) to a final concentration of 5% and by boiling immediately for 3 min. The samples were then subjected to complete tryptic digestion as described below.

LiP experiments of *T. thermophilus* samples were done using thermolysin from *Bacillus thermoproteolyticus rokko* (Sigma) at an enzyme:substrate ratio of 1:30. Samples were incubated for 1 min at a temperature from 55° to 96°C (in 3°C increments, corresponding to 14 temperature points). The LiP reaction was stopped by adding DOC to a final concentration of 5% and EDTA to a final concentration of 25 mM. The digestion mixtures were then subjected to complete tryptic digestion as described below. For the *T. thermophilus* samples, 64 µmol CaCl₂ was added to the samples after the LiP step to compensate for the addition of 5 µmol EDTA and allow for an efficient tryptic digestion.

Tryptic digestion

Proteins were reduced by incubation of samples with tris(2-carboxyethyl)phosphine (final concentration of 5 mM) for 30 min at 37°C and were alkylated by incubation with iodoacetamide (final concentration of 40 mM) for 30 min at 25°C in the dark. Samples were diluted with 0.1 M NH₄HCO₃ to a final concentration of 1% sodium deoxycholate. Samples containing guanidinium chloride were diluted to less than 1 M guanidinium chloride and predigested with lysyl endopeptidase (Wako Chemicals) at an enzyme:substrate ratio of 1:100. After 2 hours at 37°C, sequencing-grade porcine trypsin (Promega) was added to a final enzyme:substrate ratio of 1:100, and samples were incubated overnight at 37°C. The digestion was stopped by addition of formic acid to a final pH less than 3. The peptide mixtures were loaded onto Sep-Pak tC18 cartridges (Waters), desalted, and eluted with 80% acetonitrile. All peptide samples were evaporated on a vacuum centrifuge to dryness, resolubilized in 0.1% formic acid, and immediately analyzed by mass spectrometry.

Proteomic analysis

Peptide samples were analyzed on an Orbitrap mass spectrometer (Q Exactive Plus, Thermo Scientific) equipped with a nano-electrospray ion source and a nano-flow LC system (EASY-nLC 1000, Thermo Scientific). Peptides were separated on a 40 cm × 0.75 µm i.d. column packed in-house with 1.9 µm C18 beads (Dr. Maisch Reprosil-Pur 120) using a linear LC gradient from 2% to 35% acetonitrile over 120 min and a flow rate of 300 nL/min. For each MS scan at a resolution of 70,000 from 350 to 1500 m/z, twenty MS-MS spectra were acquired at a resolution of 17,500. All multiply

charged ions were used for triggering MS-MS attempts; singly charged ions and ions of undefinable charge state were excluded. A threshold of 1300 ion counts was used to trigger MS-MS scans followed by a dynamic exclusion for 30 s.

The collected spectra were searched against the *S. cerevisiae*, *T. thermophilus*, or *H. sapiens* Uniprot protein database (downloaded December 2015) using the Sorcerer-SEQUEST database search engine (Thermo Electron). We allowed up to two missed cleavages, and excluded cleavage of KP and RP peptide bonds. Cysteine carboxyamido-methylation (+57.0214 Da) and methionine oxidation (+15.99492) were allowed as fixed and variable modifications, respectively. Monoisotopic peptide tolerance was set to 10 ppm, and fragment mass tolerance was set to 0.02 Da. The identified proteins were filtered using the high peptide confidence setting in Protein Discoverer (version 1.4, Thermo Fisher Scientific). Label-free MSI-based quantification was performed using the software Progenesis (2.0, Nonlinear Systems). For quantification purposes, only fully tryptic peptides were considered. Additionally, only peptides unique to a protein were considered.

Melting curves

Peptide abundance profiles as a function of temperature were fit to the following formula (59) $S = (S_f + m_f \cdot T) + (S_u + m_u \cdot T) \cdot e^{(\Delta H_m/RT) \cdot [(T - T_m)/T_m]} / (1 + e^{(\Delta H_m/RT) \cdot [(T - T_m)/T_m]})$ where S is the observed spectroscopic signal in arbitrary units (a.u.); S_f and S_u are the spectroscopic signals of the folded and unfolded states at zero K, respectively, in arbitrary units (a.u.); m_f and m_u are the linear dependencies of S_f and S_u , respectively, on temperature (i.e., slopes of the pre- and posttransition baselines, respectively) in a.u. K⁻¹; T is the temperature in K; T_m is the melting temperature in K; ΔH_m is the unfolding enthalpy at the T_m in J mol⁻¹; and R is the gas constant, 8.314 J K⁻¹ mol⁻¹.

We imported the peptide abundance data from Excel into Matlab (v2015a), eliminated data points with exactly equal abundance values occurring more than once per species, converted the temperature to K and the abundances to log₂ scale (assuming normality on this scale). We fit the model parameters (determining the mean of the data) and the Gaussian variance by numerically computing the maximum a posteriori probability estimate of the log likelihood via the Matlab function mle with all parameters in linear scale except for ΔH_m , which was in log₁₀ scale. For each peptide, we performed 100 multistarts with parameters initialized according to Latin hypercube sampling. Outlier detection was performed by removing data points outside four standard deviations from the estimated mean and repeating the parameter inference once with the reduced data set. For further analysis we computed the gradient of $S(T = T_m)$ w.r.t. T , confidence intervals for all parameters, and fit statistics (R^2 , SSE, F -statistic). The parameters' prior distributions were either uninformative $N(0, 5)$ for log₂(S_f), $N(0, 10)$ for log₂(S_u), $N(338, 30)$ for T_m , or defined from a priori available empirically determined quantities, i.e., $N(5.5, 1)$ for log₁₀(ΔH), or defined in an empirical

Bayes fashion by visual inspection of selected high-quality fits i.e., $N(0, 0.02)$ for m_f , $N(0, 0.2)$ for m_u , where $N(m, s)$ is a normal distribution with mean m and standard deviation s . The results of the melting curve estimation were complemented with additional information (e.g., peptide position in protein) and together served as input for further analyses.

Selection of high-quality data

Only peptides with ΔH greater than 100 kJ mol⁻¹ and positive S_u and S_f values were considered. Further, only proteins with peptides with T_m s in the range of 43° to 70°C for proteinase K-treated samples and in the range of 64° and 88°C for the thermolysin-treated samples were considered. In order to display the unfolding traces equally, we normalised the signals relative to the set of acquired data for each individual protein.

Statistical analyses and signal normalization

All statistical analyses and string processing after the estimation of biophysical properties were performed using R (version 3.2.3), and data were visualized using Prism 6.0. P values from Wilcoxon tests were calculated using Prism 6.0. Significance levels of P values are graphically represented by: ns. for $P > 0.05$, * for $P \leq 0.05$, ** for $P \leq 0.01$, *** for $P \leq 0.001$, and **** for $P \leq 0.0001$.

Signal intensities from MS or CD were divided by the maximum value shown in the plots and baseline corrected by the smallest value in order to present all data as relative intensities (rel. int.) between 0 and 1.

Clustering and assignment of protein T_m s

Melting temperatures of all peptides of a single protein were subjected to hierarchical clustering using the built in "hclust" function of R (version 3.2.3) and starting with one cluster. The maximum difference in T_m of peptides in a cluster was then calculated. If that value was smaller than 3°C, the average T_m of that cluster was calculated and taken as the T_m of the protein. If, however, the value was larger than 3°C, another cluster of T_m values was added, and the procedure was repeated until the maximum difference in T_m was less than 3°C. Finally, the cluster with the lowest T_m was assigned as the T_m of the protein, unless there was a cluster that was at least 3 times more populated than that cluster (to avoid outliers), in which case the next higher cluster was considered. A maximum of five clusters per protein was allowed.

Protein domain data

Protein domain information was downloaded from the Pfam database (<http://pfam.xfam.org/>, version 30.0). We considered a peptide in a domain when its entire sequence matched the annotation of the domain provided by Pfam (from the value provided as "alignment start to alignment end").

Functional enrichment and analysis of ProTherm data

Functional enrichment analysis of stable and unstable proteins in *E. coli*, based on GO terms

«Biological Process» and «Molecular Function» and SP-PIR Keywords, was performed with the tool DAVID (<https://david.ncifcrf.gov/>). Bonferroni-corrected P values were used. Redundant terms were collapsed, retaining the lowest P values associated. Results were filtered based on a P value < 0.01 and a fold enrichment > 5. Only terms and keywords matching to more than 5 proteins in the data set were retained. Bubble size is proportional to the number of proteins matching a given term. Functional classification and enrichment analysis of IDPs from *S. cerevisiae* was performed with the tools Panther (<http://pantherdb.org>) and FunSpec (<http://funspec.med.utoronto.ca/>), respectively. Data from ProTherm were downloaded from www.abren.net/protherm/. Only measurements performed at pH values between 6 and 9 were considered.

Protein network analysis of E. coli

The network centrality parameters degree and betweenness were calculated for all nodes within the network using the Cytoscape plug-in CentiScaPe (60). Degree is a topological index corresponding to the number of edges incident on a node; betweenness is a node centrality index corresponding to the total number of shortest paths passing through a certain node. We define "hubs" and "bottlenecks" proteins falling in the top 20% of the degree and betweenness distributions, respectively (23)

Estimation of $T_{90\%}$

Fitted curves of peptide profiles were used to extract the signal value at a given temperature for each peptide. Using 0.001°C steps the relative signal at each temperature relative to the signal at the first temperature point was calculated. We termed $T_{90\%}$ the temperature at which the relative signal loss was 90% with respect to the signal at the first temperature point measured. The above described clustering algorithm was used to derive protein-level $T_{90\%}$ from peptide level data.

Estimation of aggregation

The aggregation parameter was the slope of the linear regression characterizing the unfolded state on the high-temperature side of the denaturation transition (m_u). To estimate the extent of post-unfolding aggregation, we calculated the relative signal gain after the sigmoidal decay as a function of temperature.

Electron microscopy

A 5 μ L aliquot of 1 mg/mL TPI in 20 mM HEPES 150 mM KCl and 10 mM MgCl₂ was deposited for 1 min on a glow-discharged EM copper grid. To stain the samples, 2% w/v uranyl acetate was applied for 15 s. Electron microscopy images were acquired using a FEI Morgagni 268 transmission electron microscope operated at 100 kV.

Prediction of disorder and secondary structure

Disorder and secondary structure were predicted from protein FASTA sequences. To calculate disorder we used the previously described

software tool DISOPRED3 and calculated the relative degree of disorder from that output by counting the amino acid residues predicted to be in a disordered segment of the protein divided by the total number of residues of the protein (67). We calculated the percentage of secondary structure types using PSIPRED (24).

Selected reaction monitoring (SRM) measurements

Shotgun proteomic data were used to select up to 5 intense singly charged fragment ions of the y-series from doubly or triply charged precursor ions to perform SRM analysis. This set of transitions was then experimentally tested in SRM mode. Matching of retention times and relative fragment ion intensities observed in SRM and shotgun experiments was confirmed, after realignment of the gradients used. Transitions associated with obvious interference based on visual inspection using the analysis software Skyline (version 2.5 MacCoss Lab Software) were discarded. In the same experiment, exact peptide retention times from the LC-SRM platform were annotated in order to perform scheduled SRM acquisition, using a 300-s retention time window. All SRM analyses were performed on a triple quadrupole/ion trap mass spectrometer (5500 QTrap, ABSciex) equipped with a nano-electrospray ion source and operated in SRM mode. Peptides were separated using an on-line Eksigent ID-plus Nano liquid chromatography system (Eksigent/ABSciex), equipped with an 18-cm fused silica column with 75- μ m inner diameter (New Objective). Columns were packed in-house using Magic C18 AQ 5- μ m beads (Michrom Bioresources). A cooled (4°C), auto-sampler (Eksigent/ABSciex) was used to load the dissolved peptides (~3 μ g), and then the peptide mixture was separated using a linear gradient from 5 to 35% acetonitrile over 30 min. Q1 and Q3 were operated at unit resolution (0.7 m/z half maximum peak width) aiming at a dwell time from 15 to 30 ms and a cycle time of less than 2.5 s (100 transitions per run).

Collision energies (CE) were calculated according to the formulas: CE = 0.044 · m/z + 5.5 for doubly charged precursor ions or CE = 0.055 · m/z + 0.55 for triply charged precursor ions. We confirmed coelution and peak shape similarity of the transitions monitored for a given peptide and used the respective areas under the curve for quantification. Relative quantification of all peptides across conditions was obtained by dividing each peptide's peak area from each of the 14 conditions by the highest value and subtracting the lowest value individually for each peptide. Outlier SRM transitions (e.g., noisy or shouldered transition traces from potentially similar peptides in Q1/Q3 of the background proteome) were not considered in the final calculations.

Evolutionary analysis

Proteins from *E. coli*, *T. thermophilus*, *S. cerevisiae*, and *H. sapiens* with an assigned melting temperature based on high-quality peptide measurements were analyzed for their orthologs and paralogues using eggNOG (version 4.5; <http://eggnogetdb.embl>

<http://eggnogetdb.embl>). To align protein sequences the Needleman-Wunsch algorithm was used.

Enzyme activity

Thermolysin activity in the temperature range considered was considered constant based on (62). To evaluate the activity of proteinase K across the temperature range, we used the set of synthetic peptides (Thermo Scientific) listed in fig. S1A. Peptides were spiked into the proteome extracts at a concentration of 33 μ g per mg of protein before the LiP step. The activity of proteinase K in guanidinium hydrochloride was measured by a standard protease activity assay described previously (63). We used the p-nitroanillin peptide (AAPF) at 1 mM with 340 nM proteinase K in 50 mM HEPES buffer. The change in absorbance was measured at 405 nm, and the slope was calculated for the linear range of the signal during 60 s (fig. S1B). Based on the results of this activity test, we adjusted the concentration of PK used at the different guanidinium hydrochloride concentrations to obtain a constant PK activity along the guanidinium hydrochloride gradient.

REFERENCES AND NOTES

- M. W. Ferguson, T. Joanen, Temperature of egg incubation determines sex in *Alligator mississippiensis*. *Nature* **296**, 850–853 (1982). doi: [10.1038/296850a0](https://doi.org/10.1038/296850a0); pmid: [7070524](https://pubmed.ncbi.nlm.nih.gov/7070524/)
- D. Durairajanayagam, A. Agarwal, C. Ong, Causes, effects and molecular mechanisms of testicular heat stress. *Reprod. Biomed. Online* **30**, 14–27 (2015). doi: [10.1016/j.rbmo.2014.09.018](https://doi.org/10.1016/j.rbmo.2014.09.018); pmid: [25456164](https://pubmed.ncbi.nlm.nih.gov/25456164/)
- S. S. Evans, E. A. Repasky, D. T. Fisher, Fever and the thermal regulation of immunity: The immune system feels the heat. *Nat. Rev. Immunol.* **15**, 335–349 (2015). doi: [10.1038/nri3843](https://doi.org/10.1038/nri3843); pmid: [25976513](https://pubmed.ncbi.nlm.nih.gov/25976513/)
- D. S. Coffey, R. H. Getzenberg, T. L. DeWeese, Hyperthermic biology and cancer therapies: A hypothesis for the “Lance Armstrong effect”. *JAMA* **296**, 445–448 (2006). doi: [10.1001/jama.296.4.445](https://doi.org/10.1001/jama.296.4.445); pmid: [16868303](https://pubmed.ncbi.nlm.nih.gov/16868303/)
- K. M. Polizzi, A. S. Bommaris, J. M. Broering, J. F. Chaparro-Riggers, Stability of biocatalysts. *Curr. Opin. Chem. Biol.* **11**, 220–225 (2007). doi: [10.1016/j.cobpa.2007.01.685](https://doi.org/10.1016/j.cobpa.2007.01.685); pmid: [17307381](https://pubmed.ncbi.nlm.nih.gov/17307381/)
- M. Berovic, Sterilisation in biotechnology. *Biotechnol. Annu. Rev.* **11**, 257–279 (2005). doi: [10.1016/S1387-2656\(05\)11008-4](https://doi.org/10.1016/S1387-2656(05)11008-4); pmid: [16216780](https://pubmed.ncbi.nlm.nih.gov/16216780/)
- M. N. Moore, Cytochemical demonstration of latency of lysosomal hydrolases in digestive cells of the common mussel, *Mytilus edulis*, and changes induced by thermal stress. *Cell Tissue Res.* **175**, 279–287 (1976). doi: [10.1007/BF00218706](https://doi.org/10.1007/BF00218706); pmid: [1000605](https://pubmed.ncbi.nlm.nih.gov/1000605/)
- K. Ghosh, K. Dill, Cellular proteomes have broad distributions of protein stability. *Biophys. J.* **99**, 3996–4002 (2010). doi: [10.1016/j.bpj.2010.10.036](https://doi.org/10.1016/j.bpj.2010.10.036); pmid: [21156142](https://pubmed.ncbi.nlm.nih.gov/21156142/)
- Z. Li, P. Srivastava, Heat-shock proteins. *Curr. Protoc. Immunol. Append.* **1**, IT (2004). doi: [10.1002/0471142735.ima01ts58](https://doi.org/10.1002/0471142735.ima01ts58); pmid: [18432918](https://pubmed.ncbi.nlm.nih.gov/18432918/)
- R. Das, M. Gerstein, The stability of thermophilic proteins: A study based on comprehensive genome comparison. *Funct. Integr. Genomics* **1**, 76–88 (2000). doi: [10.1007/s101420000003](https://doi.org/10.1007/s101420000003); pmid: [11793224](https://pubmed.ncbi.nlm.nih.gov/11793224/)
- K. A. Dill, K. Ghosh, J. D. Schmit, Physical limits of cells and proteomes. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 17876–17882 (2011). doi: [10.1073/pnas.1114477108](https://doi.org/10.1073/pnas.1114477108); pmid: [22006304](https://pubmed.ncbi.nlm.nih.gov/22006304/)
- A. Razvi, J. M. Scholtz, Lessons in stability from the thermophilic proteins. *Protein Sci.* **15**, 1569–1578 (2006). doi: [10.1110/ps.062130306](https://doi.org/10.1110/ps.062130306); pmid: [16815912](https://pubmed.ncbi.nlm.nih.gov/16815912/)
- C. J. Oldfield et al., Comparing and combining predictors of mostly disordered proteins. *Biochemistry* **44**, 1989–2000 (2005). doi: [10.1021/bi047993o](https://doi.org/10.1021/bi047993o); pmid: [15697224](https://pubmed.ncbi.nlm.nih.gov/15697224/)
- J. Gspöner, M. E. Futschik, S. A. Teichmann, M. M. Babu, Tight regulation of unstructured proteins: From transcript synthesis to protein degradation. *Science* **322**, 1365–1368 (2008). doi: [10.1126/science.1163581](https://doi.org/10.1126/science.1163581); pmid: [19039133](https://pubmed.ncbi.nlm.nih.gov/19039133/)
- R. L. Chang et al., Structural systems biology evaluation of metabolic thermotolerance in *Escherichia coli*. *Science* **340**,

- 1220–1223 (2013). doi: [10.1126/science.1234012](https://doi.org/10.1126/science.1234012); pmid: [23744946](https://pubmed.ncbi.nlm.nih.gov/23744946/)
- K. A. Bava, M. M. Gromiha, H. Uedaira, K. Kitajima, A. Sarai, ProTherm, version 4.0: Thermodynamic database for proteins and mutants. *Nucleic Acids Res.* **32**, D120–D121 (2004). doi: [10.1093/nar/gkh082](https://doi.org/10.1093/nar/gkh082); pmid: [14681373](https://pubmed.ncbi.nlm.nih.gov/14681373/)
- H. X. Zhou, Influence of crowded cellular environments on protein folding, binding, and oligomerization: Biological consequences and potentials of atomistic modeling. *FEBS Lett.* **587**, 1053–1061 (2013). doi: [10.1016/j.febslet.2013.01.064](https://doi.org/10.1016/j.febslet.2013.01.064); pmid: [23395796](https://pubmed.ncbi.nlm.nih.gov/23395796/)
- Y. Feng et al., Global analysis of protein structural changes in complex proteomes. *Nat. Biotechnol.* **32**, 1036–1044 (2014). doi: [10.1038/nbt.2999](https://doi.org/10.1038/nbt.2999); pmid: [25218519](https://pubmed.ncbi.nlm.nih.gov/25218519/)
- A. Fontana et al., Probing protein structure by limited proteolysis. *Acta Biochim. Pol.* **51**, 299–321 (2004). pmid: [15218531](https://pubmed.ncbi.nlm.nih.gov/15218531/)
- G. I. Makhatadze, M. M. Lopez, J. M. Richardson 3rd, S. T. Thomas, Anion binding to the ubiquitin molecule. *Protein Sci.* **7**, 689–697 (1998). doi: [10.1002/pro.5560070318](https://doi.org/10.1002/pro.5560070318); pmid: [9541401](https://pubmed.ncbi.nlm.nih.gov/9541401/)
- V. N. Uversky, α -synuclein misfolding and neurodegenerative diseases. *Curr. Protein Pept. Sci.* **9**, 507–540 (2008). doi: [10.2174/138920308785915218](https://doi.org/10.2174/138920308785915218); pmid: [18855701](https://pubmed.ncbi.nlm.nih.gov/18855701/)
- A. Fontana, P. Polverino de Lauro, V. De Filippis, E. Scaramella, M. Zamboni, Probing the partly folded states of proteins by limited proteolysis. *Fold. Des.* **2**, R17–R26 (1997). doi: [10.1016/S1359-0278\(97\)00010-2](https://doi.org/10.1016/S1359-0278(97)00010-2); pmid: [9135978](https://pubmed.ncbi.nlm.nih.gov/9135978/)
- H. Yu, P. M. Kim, E. Sprecher, V. Trifonov, M. Gerstein, The importance of bottlenecks in protein networks: Correlation with gene essentiality and expression dynamics. *PLOS Comput. Biol.* **3**, e59 (2007). doi: [10.1371/journal.pcbi.0030059](https://doi.org/10.1371/journal.pcbi.0030059); pmid: [17447836](https://pubmed.ncbi.nlm.nih.gov/17447836/)
- D. T. Jones, Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* **292**, 195–202 (1999). doi: [10.1006/jmbi.1999.3091](https://doi.org/10.1006/jmbi.1999.3091); pmid: [10493868](https://pubmed.ncbi.nlm.nih.gov/10493868/)
- 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). doi: [10.1016/j.cell.2014.02.033](https://doi.org/10.1016/j.cell.2014.02.033); pmid: [24766808](https://pubmed.ncbi.nlm.nih.gov/24766808/)
- R. Christiano, N. Nagaraj, F. Fröhlich, T. C. Walther, Global proteome turnover analyses of the yeasts *S. cerevisiae* and *S. pombe*. *Cell Reports* **9**, 1959–1965 (2014). doi: [10.1016/j.celrep.2014.10.065](https://doi.org/10.1016/j.celrep.2014.10.065); pmid: [25466257](https://pubmed.ncbi.nlm.nih.gov/25466257/)
- M. Martin-Perez, J. Villén, Feasibility of protein turnover studies in prototroph *Saccharomyces cerevisiae* strains. *Anal. Chem.* **87**, 4008–4014 (2015). doi: [10.1021/acs.analchem.5b00264](https://doi.org/10.1021/acs.analchem.5b00264); pmid: [25767917](https://pubmed.ncbi.nlm.nih.gov/25767917/)
- D. Gawron, E. Ndah, K. Gevaert, P. Van Damme, Positional proteomics reveals differences in N-terminal proteoform stability. *Mol. Syst. Biol.* **12**, 858 (2016). doi: [10.15252/msb.20156662](https://doi.org/10.15252/msb.20156662); pmid: [26893308](https://pubmed.ncbi.nlm.nih.gov/26893308/)
- H. Lilie, E. Schwarz, R. Rudolph, Advances in refolding of proteins produced in *E. coli*. *Curr. Opin. Biotechnol.* **9**, 497–501 (1998). doi: [10.1016/S0958-1669\(98\)80035-9](https://doi.org/10.1016/S0958-1669(98)80035-9); pmid: [9821278](https://pubmed.ncbi.nlm.nih.gov/9821278/)
- E. N. Carcamo-Noriega, G. Saab-Rincon, Identification of fibrillogenic regions in human triosephosphate isomerase. *PeerJ* **4**, e1676 (2016). doi: [10.7717/peerj.1676](https://doi.org/10.7717/peerj.1676); pmid: [26870617](https://pubmed.ncbi.nlm.nih.gov/26870617/)
- A. M. Fernandez-Escamilla, F. Rousseau, J. Schymkowitz, L. Serrano, Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. *Nat. Biotechnol.* **22**, 1302–1306 (2004). doi: [10.1038/nbt1012](https://doi.org/10.1038/nbt1012); pmid: [15361882](https://pubmed.ncbi.nlm.nih.gov/15361882/)
- W. Ariessandi, C. F. Chang, T. E. Chen, Y. R. Chen, Temperature-dependent structural changes of Parkinson's alpha-synuclein fibrillation. *PLOS ONE* **8**, e53487 (2013). doi: [10.1371/journal.pone.0053487](https://doi.org/10.1371/journal.pone.0053487); pmid: [23349712](https://pubmed.ncbi.nlm.nih.gov/23349712/)
- J. Baets et al., Defects of mutant DNMT1 are linked to a spectrum of neurological disorders. *Brain* **138**, 845–861 (2015). doi: [10.1093/brain/awv010](https://doi.org/10.1093/brain/awv010); pmid: [25678562](https://pubmed.ncbi.nlm.nih.gov/25678562/)
- V. N. Uversky, Functional roles of transiently and intrinsically disordered regions within proteins. *FEBS J.* **282**, 1182–1189 (2015). doi: [10.1111/febs.13202](https://doi.org/10.1111/febs.13202); pmid: [25631540](https://pubmed.ncbi.nlm.nih.gov/25631540/)
- J. Gspöner, M. M. Babu, Cellular strategies for regulating functional and nonfunctional protein aggregation. *Cell Reports* **2**, 1425–1437 (2012). doi: [10.1016/j.celrep.2012.09.036](https://doi.org/10.1016/j.celrep.2012.09.036); pmid: [23168257](https://pubmed.ncbi.nlm.nih.gov/23168257/)
- L. J. Jensen et al., eggNOG: Automated construction and annotation of orthologous groups of genes. *Nucleic Acids Res.* **36**, D250–D254 (2008). doi: [10.1093/nar/gkm796](https://doi.org/10.1093/nar/gkm796); pmid: [17942413](https://pubmed.ncbi.nlm.nih.gov/17942413/)

37. A. Gershenson, L. M. Gierasch, Protein folding in the cell: Challenges and progress. *Curr. Opin. Struct. Biol.* **21**, 32–41 (2011). doi: [10.1016/j.sbi.2010.11.001](https://doi.org/10.1016/j.sbi.2010.11.001); pmid: [21112769](https://pubmed.ncbi.nlm.nih.gov/21112769/)
38. S. R. McGuffee, A. H. Elcock, Diffusion, crowding & protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput. Biol.* **6**, e1000694 (2010). doi: [10.1371/journal.pcbi.1000694](https://doi.org/10.1371/journal.pcbi.1000694); pmid: [20221255](https://pubmed.ncbi.nlm.nih.gov/20221255/)
39. Z. Ignatova, L. M. Gierasch, Monitoring protein stability and aggregation in vivo by real-time fluorescent labeling. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 523–528 (2004). doi: [10.1073/pnas.0304533101](https://doi.org/10.1073/pnas.0304533101); pmid: [14701904](https://pubmed.ncbi.nlm.nih.gov/14701904/)
40. M. Guo, Y. Xu, M. Gruebele, Temperature dependence of protein folding kinetics in living cells. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17863–17867 (2012). doi: [10.1073/pnas.1201797109](https://doi.org/10.1073/pnas.1201797109); pmid: [22665776](https://pubmed.ncbi.nlm.nih.gov/22665776/)
41. D. Tsao, N. V. Dokholyan, Macromolecular crowding induces polypeptide compaction and decreases folding cooperativity. *Phys. Chem. Chem. Phys.* **12**, 3491–3500 (2010). doi: [10.1039/b924236h](https://doi.org/10.1039/b924236h); pmid: [20355290](https://pubmed.ncbi.nlm.nih.gov/20355290/)
42. J. Hong, L. M. Gierasch, Macromolecular crowding remodels the energy landscape of a protein by favoring a more compact unfolded state. *J. Am. Chem. Soc.* **132**, 10445–10452 (2010). doi: [10.1021/ja103166y](https://doi.org/10.1021/ja103166y); pmid: [20662522](https://pubmed.ncbi.nlm.nih.gov/20662522/)
43. K. A. Luke, C. L. Higgins, P. Wittung-Stafshede, Thermodynamic stability and folding of proteins from hyperthermophilic organisms. *FEBS J.* **274**, 4023–4033 (2007). doi: [10.1111/j.1742-4658.2007.05955.x](https://doi.org/10.1111/j.1742-4658.2007.05955.x); pmid: [17683332](https://pubmed.ncbi.nlm.nih.gov/17683332/)
44. R. B. Greaves, J. Warwicker, Mechanisms for stabilisation and the maintenance of solubility in proteins from thermophiles. *BMC Struct. Biol.* **7**, 18 (2007). doi: [10.1186/1472-6807-7-18](https://doi.org/10.1186/1472-6807-7-18); pmid: [17394655](https://pubmed.ncbi.nlm.nih.gov/17394655/)
45. M. M. Savitski *et al.*, Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* **346**, 1255784 (2014). doi: [10.1126/science.1255784](https://doi.org/10.1126/science.1255784); pmid: [25278616](https://pubmed.ncbi.nlm.nih.gov/25278616/)
46. K. Ghosh, K. A. Dill, Computing protein stabilities from their chain lengths. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10649–10654 (2009). doi: [10.1073/pnas.0903995106](https://doi.org/10.1073/pnas.0903995106); pmid: [19541647](https://pubmed.ncbi.nlm.nih.gov/19541647/)
47. H. Hegyi, M. Gerstein, Annotation transfer for genomics: Measuring functional divergence in multi-domain proteins. *Genome Res.* **11**, 1632–1640 (2001). doi: [10.1101/gr.183801](https://doi.org/10.1101/gr.183801); pmid: [11591640](https://pubmed.ncbi.nlm.nih.gov/11591640/)
48. C. Vogel, C. Berzuini, M. Bashton, J. Gough, S. A. Teichmann, Supra-domains: Evolutionary units larger than single protein domains. *J. Mol. Biol.* **336**, 809–823 (2004). doi: [10.1016/j.jmb.2003.12.026](https://doi.org/10.1016/j.jmb.2003.12.026); pmid: [15095989](https://pubmed.ncbi.nlm.nih.gov/15095989/)
49. L. Brocchieri, S. Karlin, Protein length in eukaryotic and prokaryotic proteomes. *Nucleic Acids Res.* **33**, 3390–3400 (2005). doi: [10.1093/nar/gki615](https://doi.org/10.1093/nar/gki615); pmid: [15951512](https://pubmed.ncbi.nlm.nih.gov/15951512/)
50. A. Nasir, K. M. Kim, G. Caetano-Anollés, Global patterns of protein domain gain and loss in superkingdoms. *PLoS Comput. Biol.* **10**, e1003452 (2014). doi: [10.1371/journal.pcbi.1003452](https://doi.org/10.1371/journal.pcbi.1003452); pmid: [24499935](https://pubmed.ncbi.nlm.nih.gov/24499935/)
51. S. Yang, P. E. Bourne, The evolutionary history of protein domains viewed by species phylogeny. *PLoS ONE* **4**, e8378 (2009). doi: [10.1371/journal.pone.0008378](https://doi.org/10.1371/journal.pone.0008378); pmid: [20041107](https://pubmed.ncbi.nlm.nih.gov/20041107/)
52. G. G. Tartaglia, S. Pechmann, C. M. Dobson, M. Vendruscolo, A relationship between mRNA expression levels and protein solubility in *E. coli*. *J. Mol. Biol.* **388**, 381–389 (2009). doi: [10.1016/j.jmb.2009.03.002](https://doi.org/10.1016/j.jmb.2009.03.002); pmid: [19281824](https://pubmed.ncbi.nlm.nih.gov/19281824/)
53. D. A. Drummond, J. D. Bloom, C. Adami, C. O. Wilke, F. H. Arnold, Why highly expressed proteins evolve slowly. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14338–14343 (2005). doi: [10.1073/pnas.0504070102](https://doi.org/10.1073/pnas.0504070102); pmid: [16176987](https://pubmed.ncbi.nlm.nih.gov/16176987/)
54. D. A. Drummond, C. O. Wilke, Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. *Cell* **134**, 341–352 (2008). doi: [10.1016/j.cell.2008.05.042](https://doi.org/10.1016/j.cell.2008.05.042); pmid: [18662548](https://pubmed.ncbi.nlm.nih.gov/18662548/)
55. J. D. Bloom *et al.*, Thermodynamic prediction of protein neutrality. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 606–611 (2005). doi: [10.1073/pnas.0406744102](https://doi.org/10.1073/pnas.0406744102); pmid: [15644440](https://pubmed.ncbi.nlm.nih.gov/15644440/)
56. G. G. Tartaglia, S. Pechmann, C. M. Dobson, M. Vendruscolo, Life on the edge: A link between gene expression levels and aggregation rates of human proteins. *Trends Biochem. Sci.* **32**, 204–206 (2007). doi: [10.1016/j.tibs.2007.03.005](https://doi.org/10.1016/j.tibs.2007.03.005); pmid: [17419062](https://pubmed.ncbi.nlm.nih.gov/17419062/)
57. D. Z. Rudner, R. Losick, Protein subcellular localization in bacteria. *Cold Spring Harb. Perspect. Biol.* **2**, a000307 (2010). doi: [10.1101/cshperspect.a000307](https://doi.org/10.1101/cshperspect.a000307); pmid: [20452938](https://pubmed.ncbi.nlm.nih.gov/20452938/)
58. R. Aebersold, M. Mann, Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003). doi: [10.1038/nature01511](https://doi.org/10.1038/nature01511); pmid: [12634793](https://pubmed.ncbi.nlm.nih.gov/12634793/)
59. M. M. Santoro, D. W. Bolen, Unfolding free energy changes determined by the linear extrapolation method. I. Unfolding of phenylmethanesulfonyl α -chymotrypsin using different denaturants. *Biochemistry* **27**, 8063–8068 (1988). doi: [10.1021/bi00421a014](https://doi.org/10.1021/bi00421a014); pmid: [3233195](https://pubmed.ncbi.nlm.nih.gov/3233195/)
60. G. Scardoni, M. Petteerlini, C. Laudanna, Analyzing biological network parameters with CentiScaPe. *Bioinformatics* **25**, 2857–2859 (2009). doi: [10.1093/bioinformatics/btp517](https://doi.org/10.1093/bioinformatics/btp517); pmid: [19729372](https://pubmed.ncbi.nlm.nih.gov/19729372/)
61. D. T. Jones, D. Cozzetto, DISOPRED3: Precise disordered region predictions with annotated protein-binding activity. *Bioinformatics* **31**, 857–863 (2015). doi: [10.1093/bioinformatics/btu744](https://doi.org/10.1093/bioinformatics/btu744); pmid: [25391399](https://pubmed.ncbi.nlm.nih.gov/25391399/)
62. D. P. Minde, M. M. Maurice, S. G. Rüdiger, Determining biophysical protein stability in lysates by a fast proteolysis assay, FASTpp. *PLoS ONE* **7**, e46147 (2012). doi: [10.1371/journal.pone.0046147](https://doi.org/10.1371/journal.pone.0046147); pmid: [23056252](https://pubmed.ncbi.nlm.nih.gov/23056252/)
63. N. Hennrich *et al.*, Aminopeptidase K from *Tritirachium album* Limber. I. Isolation and some properties. *Hoppe Seylers Z. Physiol. Chem.* **354**, 1339–1346 (1973). doi: [10.1515/bchm2.1973.354.2.1339](https://doi.org/10.1515/bchm2.1973.354.2.1339); pmid: [4373364](https://pubmed.ncbi.nlm.nih.gov/4373364/)

ACKNOWLEDGMENTS

P.P. is supported by a European Union Seventh Framework Program (FP7)–European Research Council (ERC) Starting Grant (FP7-ERC-StG-337965), by a “Foerderungsprofessur” grant from the Swiss National Science Foundation (grant PP00P3_133670), and by a Promedica Stiftung (grant 2-70669-11). V.C. is supported by a short-term European Molecular Biology Organization fellowship. C.v.M. and A.K. acknowledge SystemX.ch for funding. We are grateful to R. Glockshuber, R. Riek, J. Greenwald (ETHZ), J. Gsponer (University of British Columbia), and C. Vogel (New York University) for insightful discussions. We thank H. M. Fischer (ETHZ) for the *T. thermophilus* strain used in the study. P.L. and P.P. designed the experiments. P.L. performed mass spectrometry data analysis. P.J.B. contributed to mass spectrometry measurements. S.G. and M.C. designed and implemented the analysis of peptide denaturation profiles resulting from MS measurements. P.L., A.K., C.v.M., and P.P. wrote the manuscript. P.P. conceived and supervised the project. P.P. is a scientific advisor for the company Biognosys AG (Zurich, Switzerland). P.P. is an inventor on a patent licensed by Biognosys AG that covers the LiP method used in this manuscript as a general tool to study protein structural changes. Thermal unfolding curves and fitting parameters have been deposited at Harvard Dataverse, doi:[10.7910/DVN/Q4SQZL](https://doi.org/10.7910/DVN/Q4SQZL) and doi:[10.7910/DVN/ZT1EZG](https://doi.org/10.7910/DVN/ZT1EZG).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/355/6327/eaai7825/suppl/DC1
Materials and Methods
Figs. S1 to S6
Tables S1 to S5
References (64–67)

18 August 2016; accepted 12 January 2017
[10.1126/science.aai7825](https://doi.org/10.1126/science.aai7825)



Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability

Pascal Leuenberger, Stefan Ganscha, Abdullah Kahraman, Valentina Cappelletti, Paul J. Boersema, Christian von Mering, Manfred Claassen and Paola Picotti (February 23, 2017)
Science **355** (6327), . [doi: 10.1126/science.aai7825]

Editor's Summary

How proteomes take the heat

Living organisms are very sensitive to temperature, and much of this is attributed to its effect on the structure and function of proteins. Leuenberger *et al.* explored thermostability on a proteome-wide scale in bacteria, yeast, and human cells by using a combination of limited proteolysis and mass spectrometry (see the Perspective by Vogel). Their results suggest that temperature-induced cell death is caused by the loss of a subset of proteins with key functions. The study also provides insight into the molecular and evolutionary bases of protein and proteome stability.

Science, this issue p. eaai7825; see also p. 794

This copy is for your personal, non-commercial use only.

- Article Tools** Visit the online version of this article to access the personalization and article tools:
<http://science.sciencemag.org/content/355/6327/eaai7825>
- Permissions** Obtain information about reproducing this article:
<http://www.sciencemag.org/about/permissions.dtl>

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright 2016 by the American Association for the Advancement of Science; all rights reserved. The title *Science* is a registered trademark of AAAS.