Protein surface chemistry encodes an adaptive resistance to desiccation

Paulette Sofía Romero-Pérez[1], Haley M. Moran[4], Azeem Horani[2], Alexander Truong[1], Edgar Manriquez-Sandoval[5], John F. Ramirez[3], Alec Martinez[1], Edith Gollub[1], Kara Hunter[1], Jeffrey M. Lotthammer[6,7], Ryan J. Emenecker[6,7], Thomas C. Boothby[3], Alex S. Holehouse[6,7,✉], Stephen D. Fried[4,5,✉], Shahar Sukenik[1,2,✉]

1 Dept of Chemistry and Biochemistry, University of California Merced, Merced, CA 95343, USA
2 Quantitative and Systems Biology Program, University of California Merced, Merced, CA 95343, USA
3 Department of Molecular Biology, University of Wyoming, Laramie, WY 82071, USA
4 Department of Chemistry, Johns Hopkins University, Baltimore, Maryland 21218, USA
5 T. C. Jenkins Department of Biophysics, Johns Hopkins University, Baltimore, Maryland 21218, USA
6 Department of Biochemistry and Molecular Biophysics, Washington University School of Medicine, St. Louis, MO 63110, USA
7. Center for Biomolecular Condensates (CBC), Washington University in St. Louis, St. Louis, MO 63130, USA

Abstract: Cellular desiccation - the loss of nearly all water from the cell - is a recurring stress in an increasing number of ecosystems that can drive proteome-wide protein unfolding and aggregation. For cells to survive this stress, at least some of the proteome must disaggregate and resume function upon rehydration. The molecular determinants that underlie the ability of proteins to do this remain largely unknown. Here, we apply quantitative and structural proteomic mass spectrometry to desiccated and rehydrated yeast extracts to show that some proteins possess an innate capacity to survive extreme water loss. Structural analysis correlates the ability of proteins to resist desiccation with their surface chemistry. Remarkably, highly resistant proteins are responsible for the production of the cell's building blocks - amino acids, metabolites, and sugars. Conversely, those proteins that are most desiccation-sensitive are involved in ribosome biogenesis and other energy consuming processes. As a result, the rehydrated proteome is preferentially enriched with metabolite and small molecule producers and depleted of some of the cell's heaviest consumers. We propose this functional bias enables cells to kickstart their metabolism and promote cell survival following desiccation and rehydration.

✉ correspondence: alex.holehouse@wustl.edu, sdfreid@jhu.edu, ssukenik@ucmerced.edu

Introduction

In a desiccating cell proteins, ions, and small molecules experience a dramatic increase in concentration. Many proteins exist at the edge of their solubility limits, even under well-hydrated conditions (*1*). As this limit is reached, proteins become super-saturated leading to their aggregation and inactivation, often irreversibly. For an organism to survive desiccation, at least some of its proteome must continue to function upon rehydration (*2*). This can be facilitated by "priming," the process by which cells take up or synthesize high levels of protective molecules during the onset of water loss. These include small polyols such as trehalose, sorbitol, and sucrose (*3*, *4*), as well as larger proteins such as heat shock proteins (e.g., Hsp20s (*5*)) and specialized disordered proteins (e.g., CAHS proteins in tardigrades and LEA proteins in plants (*6*, *7*)). However, some organisms can survive desiccation even without priming, and the intrinsic ability of proteomes to resist desiccation damage is poorly characterized (*8*, *9*).

Our understanding of the molecular consequences of desiccation is largely based on reconstitution assays, examining the impact on specific proteins under well-controlled conditions (*10*). Biochemical assays examining enzymatic activity before and after desiccation have provided evidence that many proteins lose their function via aggregation or misfolding (*11–14*). However, these findings are anecdotal and focus on the loss of activity for a small number of proteins, often in purified form and under dilute buffer conditions. While these experiments benefit from being well-controlled, they are distinct from the heterogeneous cellular milieu (*15*). To uncover the molecular mechanisms behind successful protein rehydration, we considered an assay that could provide a holistic examination of the entire proteome.

Here, we use mass spectrometry to study how an entire proteome withstands desiccation and rehydration. Using neat extracts obtained by cryogenic pulverization of *Saccharomyces cerevisiae*, we quantified the resolubility of thousands of proteins. This was done by measuring the change in abundance of soluble lysate proteins following desiccation and subsequent rehydration (desiccation-rehydration; D-R). We find that, on average, proteins partition 1:3 between the supernatant (resoluble) and pellet (non-resoluble) fractions, with a wide variance of partitioning across the proteome. Additionally, using limited proteolysis mass spectrometry (LiP-MS) (*16*, *17*), we determined that resolubility and retention of structural fidelity are highly correlated; proteins with high resolubility also demonstrate high levels of structural retention following D-R.

To understand the chemical and structural determinants of desiccation resistance, we correlated our results with structural predictions of the *S. cerevisiae* proteome (*18*). These analyses revealed that desiccation resistance is common in smaller, well-folded proteins with fewer disordered regions and little or no known interactions with other proteins. Desiccation resistance also correlated strongly with specific surface area chemistry: resistant proteins were depleted of surface-exposed aromatic and hydrophobic amino acids and enriched in negative (but not positive) surface charges.

Strikingly, the attributes that promote desiccation resistance correlate with specific functions and cellular processes: resistant proteins are largely responsible for biosynthesis of small molecules and metabolites. Conversely, non-resistant proteins are those that consume large portions of the

cell's resources, primarily related to ribosome biogenesis (*19*). We propose that this enrichment of resource producers and depletion of resource consumers offers an adaptive mechanism to kickstart cellular metabolism and facilitate cell survival.

The unprimed yeast proteome is poorly resoluble

To understand proteomic resistance to desiccation, we profiled the intrinsic resistance to desiccation without the help of protectants - molecules that are upregulated in response to the stressor, such as chaperones, heat-shock proteins, or cosolutes that confer desiccation resistance(*20*, *21*). During logarithmic growth phase, *S. cerevisiae* cells do not biosynthesize protective proteins or solutes (*22*), significantly reducing their resistance to desiccation compared to during stationary phase (**Fig. 1A**) (*3*, *22*). We subjected *S. cerevisiae* cells, grown to mid-log phase (**Fig. 1B**), to cryogenic pulverization without any lysis buffer, which yielded a "neat" extract with cellular solutes and proteins (130 ± 40 mg/mL) near their physiological concentrations in the cell. The lysate was subjected to a single cycle of desiccation-rehydration (D-R) (**Fig. 1C)**. The total neat lysate sample (T) was vacuum dried in a desiccator until sample weight plateaued (**Fig. 1D**). Thermogravimetric measurements revealed that the desiccated lysate's water content was 11.2%, in line with previous definitions for a desiccated state (*23*) (**Fig. 1D**).

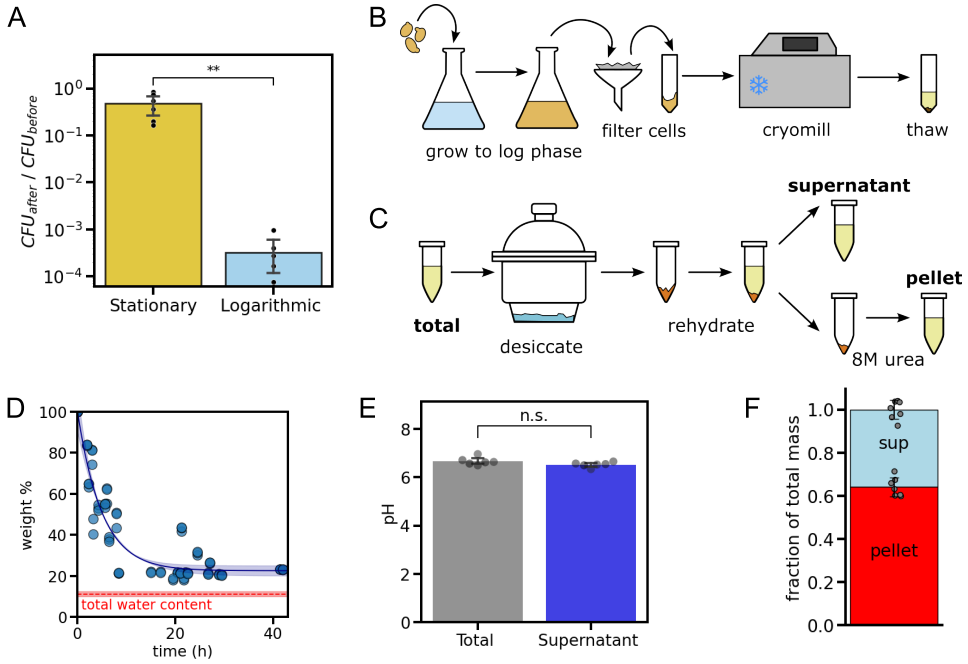

Figure 1. Neat lysates characterization before and after desiccation. (**A**) Surviving fraction of colony-forming units (CFUs) of stationary and log-phase yeast based on dilution plating from liquid cultures subjected to desiccation-rehydration (D-R). Error bars are the standard deviation of 6 independent repeats. (**B**) Schematic of the experimental procedure to prepare neat yeast extracts (**total**). (**C**) Schematic describing the experimental procedure to isolate fractions that remain soluble following D-R (**supernatant**) or not (**pellet**). (**D**) Lysate weight loss (in %) during the desiccation process. Blue points are taken from three independent treatments, with a line showing an exponential decay fit to the data. The dashed red line shows the total water content as determined by TGA. (**E**) pH of the soluble lysate before and after desiccation. Points are obtained from three independent samples, and error bars

are standard deviations. (**F**) Protein mass distribution between the supernatant and pellet following desiccation and rehydration. Points are obtained from eight independent measurements, and error bars are standard deviations.

Following desiccation, dried lysates were rehydrated such that all water lost was replaced, and the supernatant (S) was separated from the pellet (P) via centrifugation. The supernatant showed no change in pH compared to the initial lysate (**Fig. 1E**). Protein quantification using a bicinchoninic acid (BCA) assay revealed that less than 35% of the protein mass remained soluble, while the remainder aggregated (**Fig. 1F**).

Protein resolubility correlated with structural fidelity

Next, we used liquid chromatography tandem mass spectrometry (LC-MS/MS) to evaluate the capacity of each protein to resolubilize following D-R by quantifying relative protein abundances in the total, supernatant, and pellet fractions across three biological replicates. We were able to reliably quantify 4468 proteins in all three biological replicates, representing about 75% of the whole *S. cerevisiae* proteome (**Fig. 2A, Table S1**). Protein abundances in the total, supernatant, and pellet fractions and their rank orders were highly similar across the replicates (**Fig. S1**). Protein resolubilities were calculated by dividing protein abundance in the supernatant by the abundance in the total sample (S/T), which also showed a high degree of reproducibility (**Fig. S2**). An independent metric of the same measure, S/(S+P), strongly correlated with S/T, suggesting that the mass balance for most proteins in our experiment approaches 100% (**Fig. 2B**). From here on, we utilize the S/T ratio as a measure of a protein's resolubility.

Overall, proteins partition primarily into the pellet following desiccation. Less than 2% of the proteins identified (68/4468) partition preferentially into the supernatant significantly, while over 85% (3881/4468) partition preferentially into the pellet (**Fig. 2A**). In line with our BCA assay (**Fig. 1G**), the median resolubility amongst the 4468 proteins is 0.24, with a long tail towards higher values (**Fig. 2D**). However, this quantitative proteomics study does not reveal if the resolubilized proteins also retain their native structure after rehydration - which for many proteins is prerequisite for proper function.

To assess structural changes in rehydrated proteins following desiccation and rehydration, we employed limited proteolysis mass spectrometry (LiP-MS) on the resolubilized proteins (fraction **S**) and compared the proteolysis profile to extracts that were not subject to the D-R stress (fraction **T**). Briefly, T and S samples were incubated for one minute with the non-specific protease Proteinase K. This protease preferentially cuts the protein at solvent-exposed and disordered protein regions, resulting in structural information that can be read out by sequencing the resulting peptide fragments with LC-MS/MS (*16*, *24*). Afterwards, the peptides are subjected to a full trypsin digest and prepared for shotgun proteomics (**Fig. 2E**). The half-tryptic fragments – in which one cut arises from trypsin and another from proteinase K – indicate sites that were susceptible to proteolysis; ipso facto, the full-tryptic fragments indicate regions in which proteinase K did not cut. By identifying peptides with significant changes in abundance between S and T samples, we can assign sites within proteins that were structurally perturbed by the D-R cycle (**Fig. 2E**) (*16*). Because D-R treatment occurs *ex vivo*, other potential non-structural factors that could lead to changes in proteolytic susceptibility (such as changes in post-translational modifications) are not likely to be relevant.

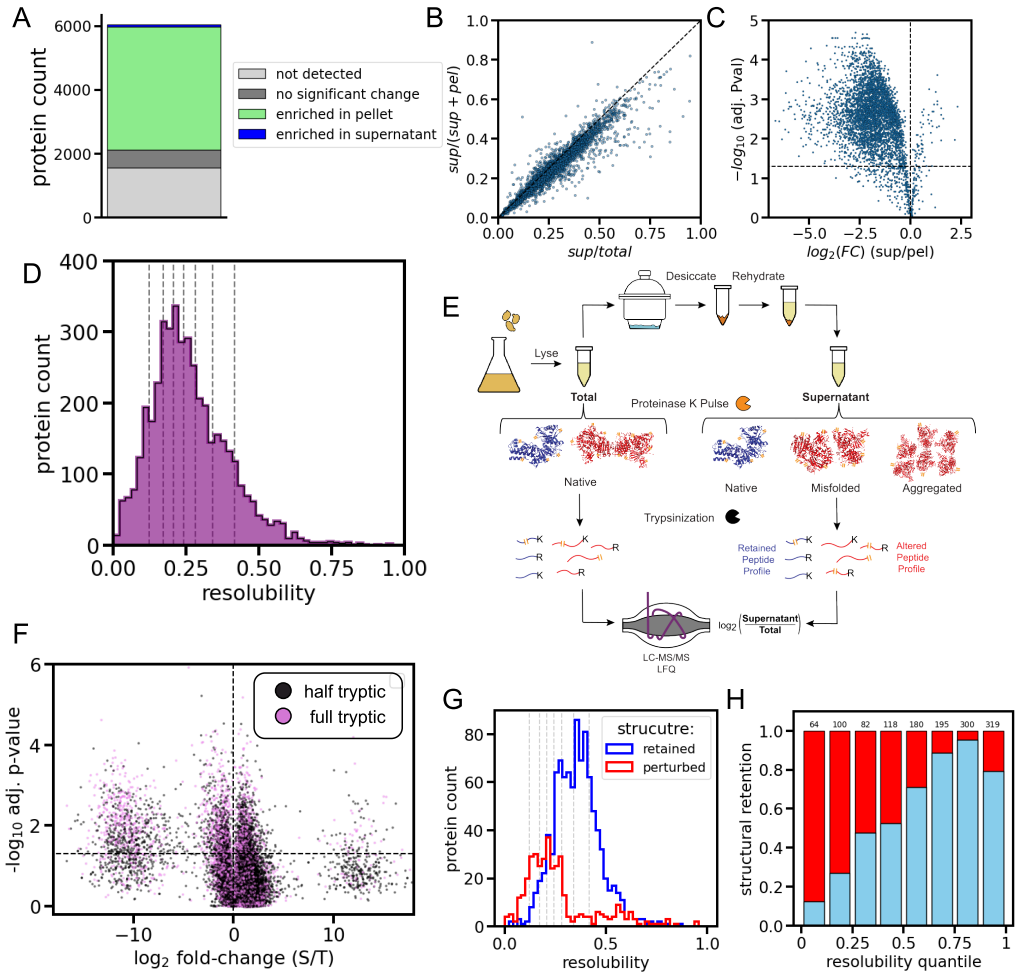

Figure 2. Proteome-wide resolubility and structure retention of the rehydrated proteome. (A) Proteome coverage of LC-MS profiling of lysates from *Saccharomyces cerevisiae* (UniProt UP000002311). (**B**) Correlation between S/(S+T) and S/P, two independent metrics for protein resolubility. Pearson's r = 0.98. (**C**) Volcano plot of proteome partitioning between supernatant and pellet following D-R. Dashed horizontal lines represent statistical significance (FD-R adjusted p-value). (**D**) Histogram of proteome resolubility fractions. Dashed lines represent 1/8th quantiles (~ 550 proteins in each quantile). Median resolubility is 0.24. (**E**) Schematic of experimental workflow for limited proteolysis mass spectrometry (LiP-MS). (**F**) Volcano plot showing the relative abundance of full-tryptic (black) and half-tryptic (purple) peptide abundance fold-change before and after D-R and the adjusted p-values (t-test with Welch's correction, Benjami-Hochberg adjusted) against null hypothesis peptide is present in similar abundance between S and T. Horizontal line represents cutoff for significance ($p < 0.05$). (**G**) Distribution of resolubility percentages for proteins whose structure was retained or perturbed. Dashed vertical lines are the same as (D). (**H**) Fraction of proteins that retained their structure for each resolubility quantile. Numbers on top indicate the number of proteins identified in each quantile by LiP-MS.

Adopting a recently published computational workflow for LiP-MS data analysis (*25*), these experiments assessed the structural status for 1750 proteins, or 39% of the proteins with assigned resolubility (**Fig. S3**). Many full-tryptic and half-tryptic peptide fragments were enriched in the total sample compared to the supernatant (**Fig. 2F, Table S2**). Proteins were labeled structurally perturbed if they possessed two or more peptides with significantly different abundance (>2 fold, $p < 0.05$ by t-test with Welch's correction after adjusting for multiple

hypothesis testing) in the rehydrated supernatant. Using this criteria, we found that the resolubility distribution of proteins whose structure is perturbed is sharply different from that of proteins where the structure has been retained (p-value < 10^{-29} by t-test) (**Fig. 2G**). The majority of proteins that are structurally perturbed after rehydration were poorly resolubilized in the first place (median resolubility 0.21), whereas those that retain native-like structure after rehydration also resolubilize to a much greater extent (median resolubility 0.35). A quantile analysis of the entire dataset demonstrates a strong correlation (Pearson's r = 0.94) between structural retention and resolubility (**Fig. 2H**). These data suggest that proteins which can resolubilize after D-R also retain their native structures in the dry-state.

Desiccation resistance is encoded in the surface area of folded regions

To understand the chemical and structural determinants of D-R resolubility, we first annotated proteins as membrane or non-membrane proteins, revealing 25% of proteins are annotated as being integral or peripheral membrane proteins (**Fig. 3A**). This revealed a significant difference: predictably, non-membrane proteins are more resoluble than membrane proteins. Because the fate of membrane proteins is potentially more dependent on the lysis (which lacked detergents), our subsequent analysis focused exclusively on the 3226 non-membrane proteins.

We next compared resolubility percentages with protein abundances (*26*). Dividing the detected, non-membrane proteins into eight S/T quantiles (shown in **Fig. S4**), we found that resoluble proteins are generally more abundant than non-resoluble proteins (**Fig. 3B**). We also saw a systematic decrease in protein length associated with the more resoluble proteins (**Fig. 3C**) but a weak correlation with the number of distinct folded domains (**Fig. 3D**). Intriguingly, we saw a strong correlation between resolubility and disorder content (that is, the fraction of residues predicted to reside in intrinsically disordered regions). The least resoluble proteins were enriched in, while the most resoluble proteins were depleted from disordered regions (**Fig. 3E**). Finally, we used STRING-DB to annotate each protein with its number of interactors in the yeast interactome (*27*). This analysis revealed that resoluble proteins tend to have fewer interactors than non-resoluble proteins (**Fig. 3F**). Overall, our analysis suggests that resolubility (i.e. D-R resistance) is highest for yeast proteins with high cellular abundance, low disorder content, shorter length, and a low number of interactors.

Given how substantial of an effect disordered regions appear to have on resolubility, we extracted all the disordered regions from all the non-membrane proteins and asked whether various amino acid compositions of these regions would explain differences in resolubility. To our surprise, we could not find a correlation between any sequence features investigated across disordered regions (**Fig. S5**). This suggests that the sequence chemistry of disordered regions does not play a major role in determining D-R resistance, and simply that the absence of disordered regions improves D-R resistance while their presence undermines it.

We also examined two-dimensional spaces described by pairs of attributes and found that plotting resolubility versus both copy number (i.e., protein abundance) *and* interaction count creates a clean dividing surface that discriminates the most resoluble proteins as those with high copy number and few interactors and the least resoluble proteins as those with low copy number and many interactors (**Fig. 3G**). Finally, we investigated the subcellular localization of

the top vs. bottom 20% of proteins as ranked by resolubility. In general, cytoplasmic proteins tended to be more resoluble, while nuclear proteins (and nucleolar proteins especially) tended to be less resoluble (**Fig. 3H**).

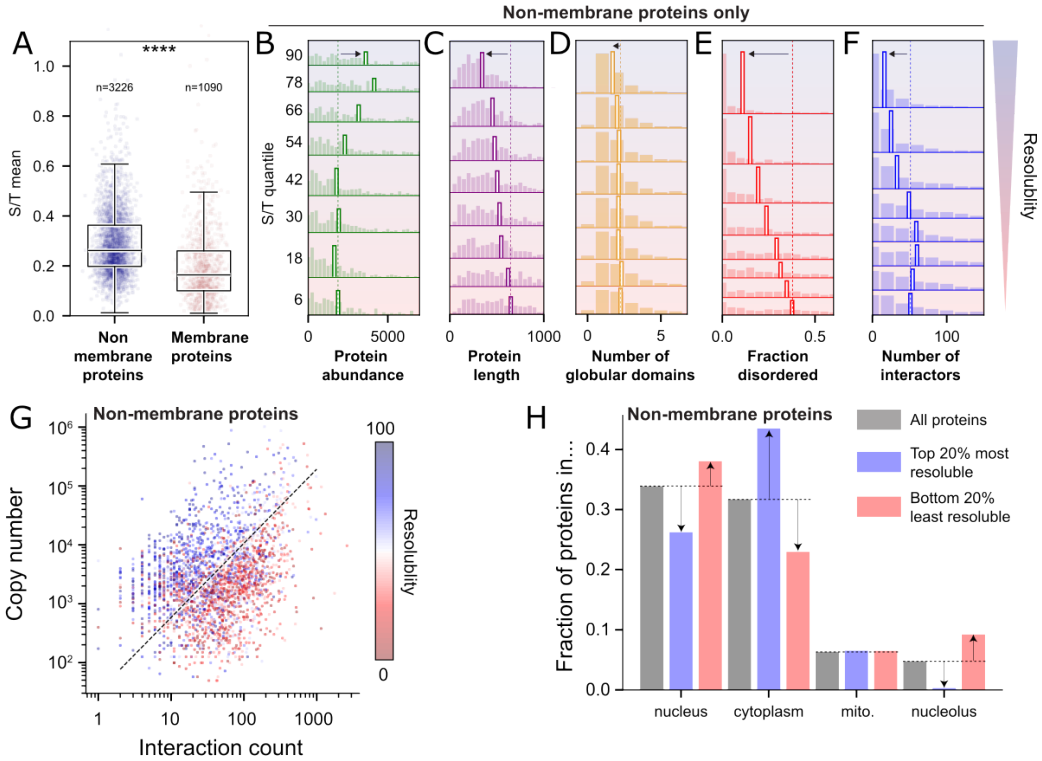

Fig. 3. Correlates of protein resolubility. (A) Of the 4468 proteins in our dataset, 75% are non-membrane proteins, and 25% are membrane proteins. Membrane proteins are, on average, less resoluble than non-membrane proteins. **(B)** Distribution of abundance (copy number per cell) for non-membrane proteins in different resolubility quantiles shows protein abundance is well correlated with resolublity. **(C)** Distribution of protein length for non-membrane proteins in different resolubility quantiles shows length anti-correlated with resolubility; shorter proteins are more resoluble than longer proteins. **(D)** The number of globular domains shows a weak correlation, with less resoluble proteins tending to contain more independent globular domains. **(E)** Distribution of disorder content (measured as fraction of residues predicted to be disordered) for non-membrane proteins in different resolubility quantiles shows disorder content is strongly anti-correlated with resolubility; proteins with less disorder are more resoluble. **(F)** Distribution of the number of interactors for non-membrane proteins in different resolubility quantiles shows interactor count is anti-correlated with resolubility; proteins with fewer interactors are more resoluble than those with many. **(G)** Proteins projected into interaction count vs. copy number space, colored based on their resolubility. **(H)** Subcellular localization of all proteins (gray) vs. the top 20% most resoluble proteins (blue) vs. the bottom 20% least resoluble proteins (red).

Our analysis thus far is correlative; it illustrates attributes that tend to co-occur with resistance to D-R but fails to offer mechanistic insight into why these properties endow resistance. To address this, we initiated a structural bioinformatic analysis to investigate the physicochemical origins of resistance.

Given the lack of explanatory power offered by sequence features in disordered regions, and because most resoluble proteins retained their structure, we next hypothesized the core determinants of D-R resistance may be encoded in the globular domains. Using the AlphaFold2 structural models for all *S. cerevisiae* proteins, we excised out all globular domains using two independent approaches (**Fig. 4A**) (*18*, *28*). One approach used Chainsaw, a trained, multi-track convolutional neural network for decomposing distinct subdomains (*29*). The other used DODO (https://github.com/idptools/dodo), a structure-based domain decomposition approach. These approaches reveal different numbers of globular domains (**Fig. 4B,C**), and have distinct strengths and weaknesses (see *Methods*). We used a threshold of 90 amino acids to minimize the analysis of subdomains or fragments that do not reflect autonomously folding units. Having excised folded domains, we identified the solvent-accessible residues on a per-domain level.

Using individual globular domains, we identified the solvent-accessible residues (see *Methods*) and investigated the relationship between globular domain surface chemistry and resolubility (**Fig. 4D**). All results presented in Fig. 4 use domains defined via DODO, although comparable results are obtained if the analysis is repeated using domains defined by Chainsaw (**Figs. S6-8**). We first investigated gross chemical properties, revealing that more resoluble proteins tend to have a lower surface hydrophobicity, higher surface negative charge, but curiously, lower surface positive charge (**Fig. 4E**). To obtain finer-grain insight into the role of surface chemistry, we calculated normalized distributions of surface-accessible amino acid fractions across distinct S/T quantiles in order to assess if specific residue types are more commonly found on the surface of resoluble vs. non-resoluble proteins. Striking, this analysis effectively revealed two distinct classes of residues (**Fig. 4F**). Positively charged residues, aliphatic residues, aromatic residues, and most polar residues are enriched on the surface of non-resoluble domains. In contrast, negatively charged residues and the small amino acids (proline, glycine, and alanine) were all enriched on the surface of domains from highly resoluble proteins. Taken together, these results imply the surface of globular domains may be tuned to enable resolubility.

While this analysis identified strong trends in domain surface chemistry, the molecular implications of these trends remain opaque. To understand the consequences of the observed trends in surface chemistry, we calculated predicted mean-field interactions between each domain's surface residues and a set of thirty-six previously identified dipeptide repeats (see *Methods*) (**Fig. 4G**) (*30*). These test peptides encapsulate a chemically diverse set of chemistries and allow us to test how different globular domain surfaces may interact with various types of putative partners. In total, we computed ~500,000 domain:peptide interactions to construct a systematic map of the chemical biases encoded by every globular domain in the *S. cerevisiae* proteome (**Fig. 4H**).

Having calculated all possible globular domain:peptide attractive interactions, we calculated normalized interaction scores over the range observed for a given test peptide, and assessed how different test peptides correlated with resolubility. This analysis revealed a strong and systematic trend that was virtually invariant to the nature of the test peptide used: resolubility is strongly correlated with weak surface interactions with acidic, polar, and hydrophobic peptides (**Fig. 4I**). Strikingly, non-resoluble proteins contain globular domains that show more attractive

interactions with hydrophobic, polar, and acidic test peptides. In contrast, the most resoluble proteins contain globular domains with weak attractive interactions for those same test peptides. Hence, this analysis implies that the most resoluble proteins possess globular domains with chemically inert surfaces. This observation, rooted in physical chemistry, is in line with our previous observation that the most resoluble proteins tend to have fewer interactors in the actual yeast interactome as well (**Fig. 3F, G**)

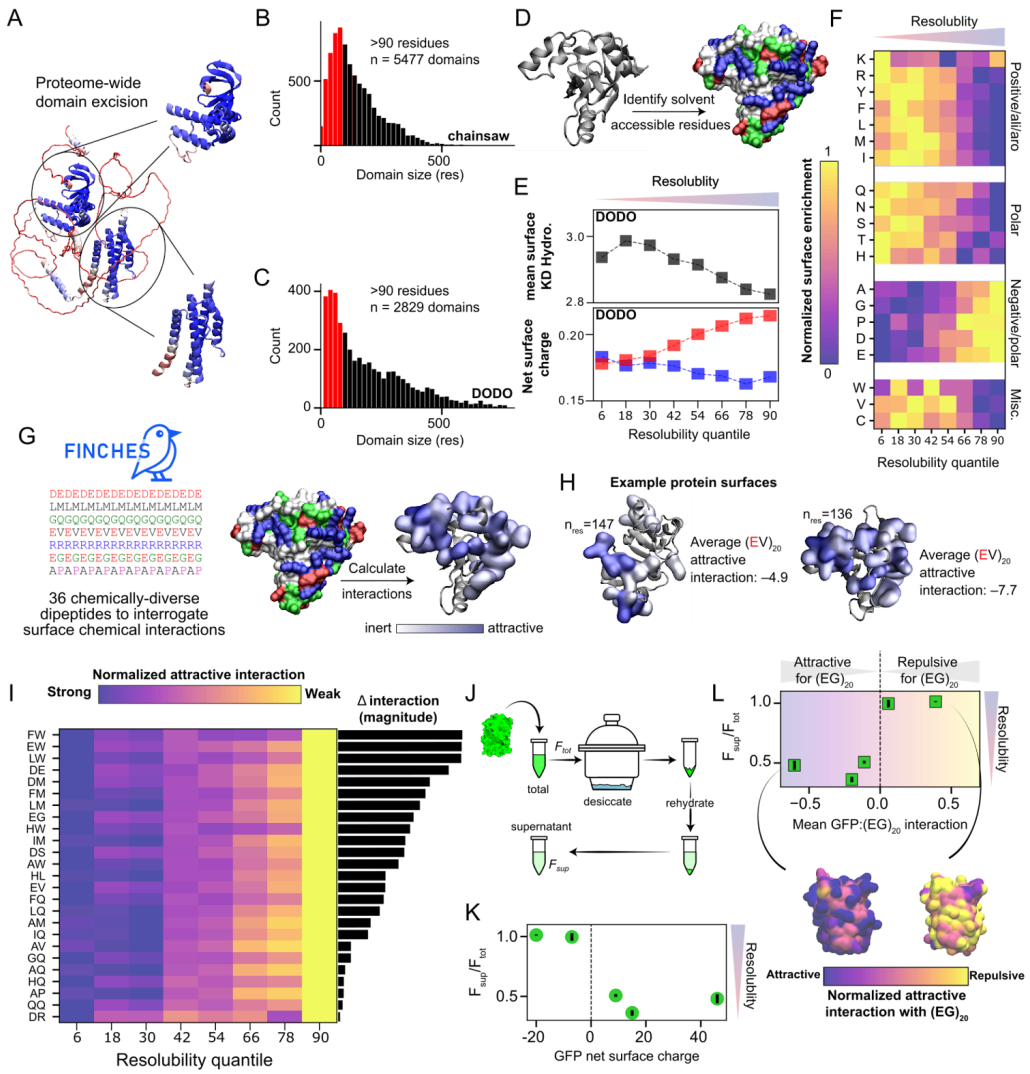

Fig. 4. Surface chemistries in folded regions facilitate resistance to desiccation. **(A)** Globular domains from AlphaFold2 predicted structures can be excised. **(B,C)** Histogram of domain size (in residues) for the two methods. **(D)** Schematic representation of domain surface decomposition. **(E)** Overall hydrophobicity (top) and net charge (bottom; red symbols represent negative charge, blue symbols represent positive charge) for surface-exposed residues in different resolubility quantiles reveal trends with respect to surface chemistry. **(F)** Analogous analysis for normalized amino acid fractions across different S/T quantiles. Each row is internally normalized to enable amino acids with different mean abundance to be visualized on the same axes. **(G)** Chemically orthogonal test peptides that were identified previously using the computational method FINCHES. **(H)** Examples of globular domains with

equivalent numbers of residues but different attractive interactions with the glutamic acid/valine (EV) test peptide, with surface regions that have attractive interactions with $(EV)_{20}$ highlighted in blue. **(I)** Overall trends for 25 test peptides show a decrease in attractive interactions as a function of S/T quantile. Normalized changes in predicted interaction strength are shown in the heat map to enable interactions of different strengths to be visualized on the same axes. The absolute magnitude of the maximum change in interaction strength is shown in the bars on the right. **(J)** Schematic of GFP spike-in experiment. **(K)** Ratio of supernatant fluorescence over total fluorescence for GFP variants vs. net surface charge. **(L)** Ratio of supernatant fluorescence over total fluorescence for GFP variants vs. mean attractive/repulsive interaction with the $(EG)_{20}$ guest peptide.

To test the inferences obtained from this analysis, we took rationally designed GFP variants with a range of surface charges (*31*). These variants were spiked into lysates, and their fluorescence was measured before and after D-R (**Fig. 4J**). As predicted, these results showed negatively charged GFPs were almost entirely resoluble, whereas positive charges did not endow the protein with high resolubility (**Fig. 4K**). This behavior is recapitulated when the average attractive/repulsive interaction between the GFP surface and an acidic test peptide were calculated; when the overall GFP:peptide interaction is predicted to be attractive, GFP resolubility is minimal. In contrast, when the overall GFP:peptide interaction is predicted to be repulsive, the GFP is almost entirely resoluble (**Fig. 4L**). Taken together, our results imply the surface chemistry of globular domains is a key determinant governing resolubility after desiccation, with interpretable chemical rules elucidated via proteome-wide informatic analysis.

Resistant proteins are producers of metabolites and small molecules

We next wanted to see if resistant proteins have common functions. For this, we took the 20% most and least resoluble proteins (~800 each) and examined them for functional enrichment in gene ontology (GO) and KEGG pathway association (*32*–*34*). Enrichment is calculated against the full set of proteins identified in the total (T) samples prior to desiccation as a reference set.

We found that the most resoluble proteins are involved in small molecule biosynthesis. A significant enrichment of amino-acid and secondary metabolism synthesis was observed in KEGG, and a similar fold-change was observed for biosynthetic processes and enzymatic functions in GO (**Figs. 5A**, left). The least resistant proteins showed different functionalities, with significant enrichment of ribosome biogenesis in KEGG and related functions such as RNA processing and helicase activity in GO (**Fig 5A**, right). This indicates that the supernatant following D-R is enriched in enzymes that process small molecules, but depleted from ribosome biogenesis machinery, arguably the greatest consumer of cellular resources (*19*). Effectively, this means that the ability of the cell to generate new proteins is severely dampened following D-R.

As an independent test of enrichment for small molecule producers, we looked for domain-types (fold topologies) enriched in resistant and non-resistant proteins. Fold topologies are often associated with high-level biochemical functions(*35*). We assigned fold topologies using ECOD to each of the proteins in the dataset shown in **Fig. 4**, and identified 6 families with over 50 domains (*35*). Comparing against the entire dataset, we found that TIM β/α barrels (*36*, *37*) and Rossmann-like folds (*38*, *39*) show a significantly higher resolubility. Both of these

motifs are common in metabolic and glycolytic enzymes. Conversely, proteins that have no assigned structural folds show significantly lower resolubility, which is expected given such proteins are generally disordered (**Fig. 5B**).

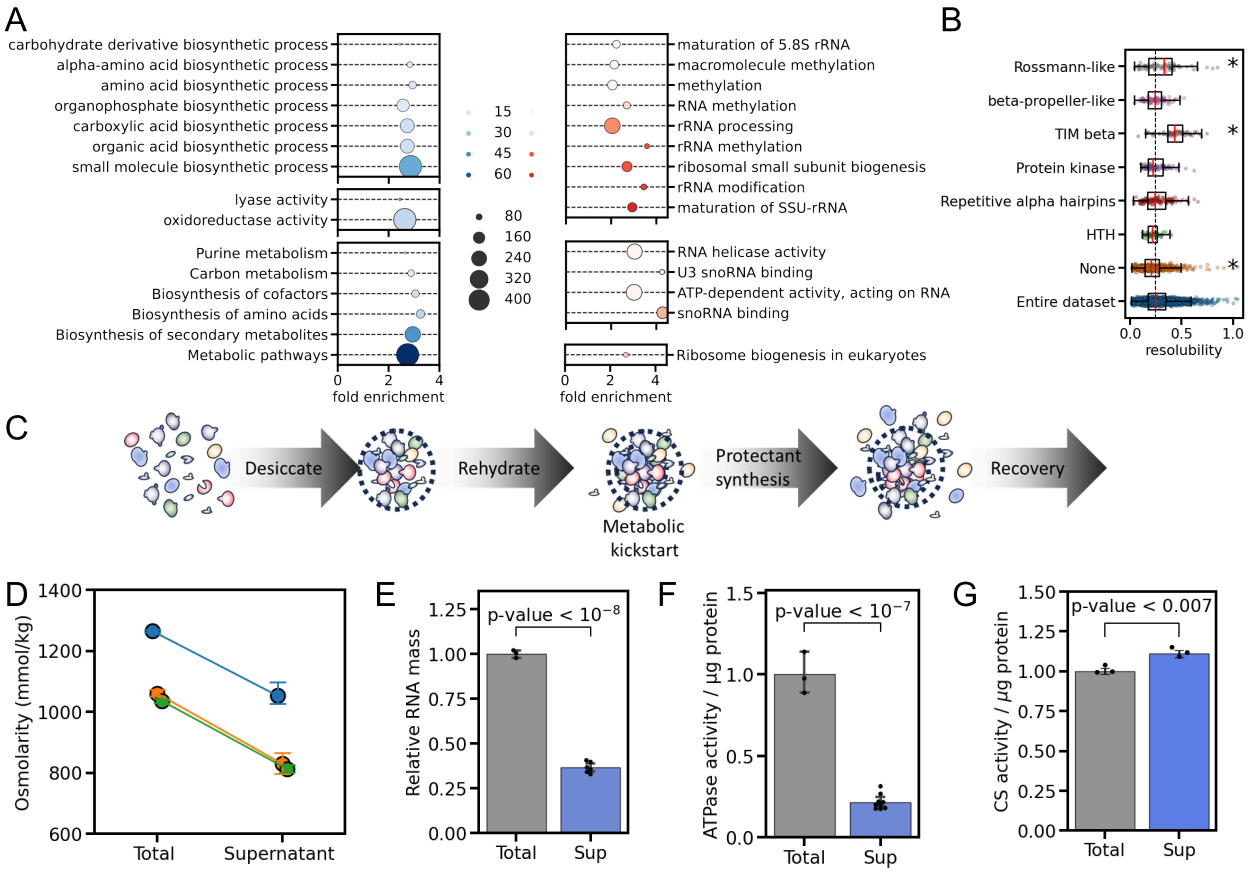

Figure 5. Functional enrichment of most and least resoluble proteins. (**A**) Enrichment of molecular function terms in 20% most (blue) and least (red) resoluble proteins. Enrichments are for molecular processes (top), biological function (middle), and KEGG pathways (bottom). (**B**) Quantile-based enrichment of ECOD domains. Asterisk denotes statistical significance ($P < 0.01$) by t-test. (**C**) Schematic of the metabolic kickstarting process. (**D**) Comparison of soluble lysate osmolarity before and after D-R. Colors represent different performances of the experiments, and each point represents an average of 3 biological replicates. (**E**) Soluble RNA mass before (Total) and after (Sup) D-R. (**F**) ATPase activity of lysates normalized to total protein mass before (Total) and after (Sup) D-R. (**G**) Citrate synthase activity normalized to total protein mass before (Total) and after (Sup) D-R.

This analysis indicates there is a unique functional composition in the proteome following D-R. We hypothesized that this composition lets the proteome restart cellular function following rehydration. If, upon rehydration, the cell resumed the constitutive function of biosynthesizing ribosomes, it would quickly be left without the molecular building blocks and metabolites required for survival. This higher resistance associated with metabolic enzymes to D-R may allow cells to focus on metabolism key to survival rather than growth (**Fig. 5C**) - a process we refer to as 'metabolic kickstarting.'

Metabolic kickstarting can assist the initial stages of desiccation recovery

The ‘metabolic kickstarting’ hypothesis makes several predictions regarding metabolic activity after D-R. First, that upon rehydration the supernatant should be depleted from both small molecules and ribosomes. Second, that the supernatant should be depleted of biomolecules that convert ATP to ADP; if energy consumers are less resoluble, they will be enriched in the pellet. Third, that the supernatant should be enriched for biomolecules associated with cellular energy production; if energy producers are more resoluble, they will be depleted from the pellet. We tested these hypotheses directly using independent biochemical assays.

We first wanted to assess the depletion of small molecules following D-R. To do this, we measured the osmotic pressure of lysate before and after D-R. Our analysis revealed that there is a ca. 300 mOsm drop in osmotic pressure (**Fig. 5D**). With the proteome itself contributing only 1-5 mOsm of this (the total proteome concentration in yeast is ~ 1-2 mM (*40*)), we conclude that the bulk of the loss results from small molecules and ions that are trapped in the desiccated pellet. This explains the functional importance of the rehydrated proteome being enriched for small molecule producers.

Next, we tested the depletion of ribosomes following D-R. We reasoned that a depletion in ribosome biogenesis proteins will only be useful if a majority of the cell’s ribosomes are no longer active following D-R. To measure this, we quantified the total RNA content in our lysate before and after desiccation/rehydration. Given that 80% of RNA in the cell originates from ribosomes, total RNA acts as a proxy for ribosome concentration. We find that while initially lysates contain a significant amount of RNA (> 8 μg/μL), there is a significant drop of over 50% in the total RNA amount following desiccation and rehydration (**Fig. 5E**).

To test depletion of energy consumers, we examined the ATPase activity change in the proteome following D-R. To do this, we used an assay that measures the turnover of ATP into ADP from the lysate before D-R and in the supernatant after. This value is then normalized by protein mass, giving us a value that reflects the per-protein ATP utilization. In agreement with our hypothesis, ATPase activity is reduced by nearly 75% in the supernatant, indicating that energy consumers are typically non-resoluble (**Fig. 5F**).

Finally, to test for the enrichment of producers in the supernatant we followed the activity of the enzyme citrate synthase (CS). Beyond being a key component of the Krebs cycle and critical for cellular energy production, it has historically been employed as a model enzyme to assess desiccation protection (*41*). Measuring lysate CS activity before and following D-R, and normalizing this activity value per total protein mass, we find that the rehydrated lysate is significantly enriched in active CS (**Fig. 5G**). This result confirms that – at least for citrate synthase – the supernatant is enriched for functionally active energy producers. This result is also in line with our LC-MS results, which showed CS S/T at 0.34 (compared to the median of 0.24 for the entire proteome)

Taken together, these experiments corroborate a metabolic kickstarting hypothesis: The rehydrated proteome is depleted of small molecules, ribosomes, and ATP consumers. On the other hand, enzymes that biosynthesize small molecules have more inert surface chemistries and generally low disorder content which enable them to solubilize more efficiently following D-R, thereby helping yeast survive this severe stress.

Discussion

The resistance of the proteome to desiccation events is crucial for understanding cellular resilience and survival under extreme conditions. By examining the proteome of a well-established model organism in the absence of protection afforded by priming, we aimed to uncover the basal resistance encoded in protein sequence and structure following desiccation/rehydration. The neat extracts we produced for this work resemble the composition, pH, and solute concentration conditions encountered *in vivo*. This allows us to observe protein behavior in a complex environment that is more physiologically relevant than afforded by most *in vitro* studies.

We specifically analyzed the proteome of *S. cerevisiae* in the log phase, in which cells do not exhibit desiccation resistance because they have not upregulated chaperones or produced protective osmolytes (*20*). This intentional choice provides a baseline for understanding the intrinsic resistance of each protein when exposed to desiccation in the context of the cellular environment. This methodology also establishes an experimental platform capable of exploring how various molecules, including protective cosolutes and less understood proteins such as intrinsically disordered proteins known as hydrophilins, contribute to desiccation resistance (*21*).

Our results revealed that approximately 35% of the protein mass remained soluble after a desiccation and subsequent rehydration cycle. While this finding aligns with expectations (*42*), proteome-wide analysis breaks down this metric to reveal a complex landscape of desiccation resistance across the *S. cerevisiae* proteome, with a gamut of partitioning values. While some proteins were found to partition nearly entirely into the supernatant, nearly all proteins were present at detectable amounts in the rehydrated supernatant. This indicates that even poorly resistant proteins have some copies solvated following rehydration. However, rehydration is not sufficient: to function, proteins must also retain their structure. To test this, we used Limited Proteolysis (LiP) to profile protein structures. Our results show a high degree of correlation between resolubility and structural retention following desiccation/rehydration. This correlation is observed here for the first time on a proteomic scale, linking together two methods for dysfunction following desiccation that were previously thought to be largely decoupled (*43*, *44*).

The most resistant proteins are small, bereft of disordered regions, and abundant. They also tend to have smaller interactomes and less attractive interactions with many “test” peptides based on computational predictions.. This indicates that resolubility is a more complex feature than mere solubility:, a lack of strongly interacting surfaces is important to help proteins escape the complex matrix of a desiccated cytoplasm. We were surprised to find that disordered regions are actually counterproductive for D-R resistance - in marked contrast to what was observed for their role in rescuing proteins from denaturant-induced unfolding (*17*). However, desiccation places the proteome in a profoundly different state than heat shock or denaturant (*10*, *15*). In the desiccated state where proteins become ultra-concentrated, the higher potential for disordered regions to interact with other species may render them a greater liability than asse.

As described, the key factors that govern resolubility are the average surface chemistry of folded domains. Globular domain surface composition has already been shown to correlate with

stress resistance in extremophiles (*45*, *46*). We specifically show that increasing negative charges alone is enough to dramatically change the resolubility of fluorescent proteins in rehydrated lysates. This finding suggests that proteins can be made resistant to desiccation through modification of their surface charges, with little effect on their activity. This observation is logical: surface-exposed residues are subject to higher evolutionary rates than core residues (*47*, *48*), and for monomeric enzymes with small interactomes, they can be tuned with relatively few restrictions. This observation is also in line with the bioengineering practice of "supercharging" proteins by mutating many surface residues, which has been found to render a number of proteins more robust (*31*, *49*).

Finally, we speculate the bias for negative charge over positive charge reflects the nature of the counterion these surface residues will likely embrace in the desiccated state. The most abundant cation in yeast is K+ whereas the most abundant anion is glutamate (*48*, *50*). In the desiccated form, negative surfaces will become sheathed with K+ that easily repartition into solution during rehydration. Positive surfaces would become coated with glutamate, fatty acids, nucleic acids, and other potentially "stickier" interactions that are more challenging to resolubilize.

Perhaps most surprising is the correlation between protein function and their desiccation survival. We find that small molecule producing enzymes are highly enriched in the rehydrated supernatant, while ribosome biogenesis is highly enriched in the least resistant fraction. Depleting the rehydrated supernatant of ribosomes acts to reduce the synthesis of new protein machinery, allowing the cell to focus on rebuilding its reserve of building blocks, and slowly recuperate metabolism following rehydration. As desiccation is an environmentally relevant stressor yeast must face, we propose that abundant enzymes responsible for small molecule production have evolved to survive D-R by adapting their surface chemistry in specific ways. It is also possible that resistance to D-R provided negative selective pressure that prevented metabolic enzymes from acquiring disordered regions (*51*).

Overall, our results point to a molecular mechanism, rooted in protein surface chemistry, that helps a specific functional subset of proteins survive desiccation and rehydration - a critical stressor in many ecosystems. It remains for future studies to uncover how the introduction of protectant solutes and proteins may affect D-R resistance globally. Finally, these findings suggest that protein surfaces can be rationally engineered to facilitate survival from desiccation stress.

Acknowledgements

Support for this project came from an NSF Collaborative Proposal via the Integrative Research in Biology (IntBIO) program under awards 2128069 to TCB, 2128068 to ASH, and 2128067 to SS. SS and SDF thank the support of the Sloan Foundation. JML was supported by the National Science Foundation via grant number DGE-2139839. HMM thanks the Chemistry-Biology Interface Program training grant (NIH-T32GM080189-13). EMS thanks the Program in Molecular Biophysics training grant (NIH-T32GM135131). SDF thanks the NSF Division of Molecular and Cellular Biology for a CAREER grant (MCB-2045844) as well as support from a Camille Dreyfus Teacher-Scholar Award. We thank members of the Water and Life Interface

Institute (WALII), supported by NSF DBI grant #2213983, for helpful discussions. We thank the "Extreme Biophysics - The Molecular Limits of Life" research coordination Network (NSF award 1817845) in which some of the collaborations were established. Special thanks to Bruker Proteomics Applications Scientists Matthew Willetts and Diego Assis for their crucial assistance with LC-MS Data Acquisition and analysis.

Methods

Saccharomyces cerevisiae culture

A single colony of *Saccharomyces cerevisiae* strain BY4742 (S288C) (*52*), picked from a fresh YPD plate, was used to inoculate 250 mL of YPD medium. Yeast was grown at 30 °C and 200 rpm overnight. The overnight culture was then used to inoculate 2 X 2L flasks containing 1 L of YPD medium. Cultures were initiated with an optical density at 600 nm (OD_{600}) of 0.07 and collected by centrifugation when the OD_{600} reached the mid-log phase (0.6).

Neat lysate preparation

Yeast cells grown to mid-log phase were resuspended in 2 mL of water and subsequently vacuum filtered on a 1.2 µm pore cellulose membrane to eliminate media residues. Approximately 4 g of cells were obtained from a 2 L log-phase yeast culture. The filtered cells were promptly submerged in liquid nitrogen and then lysed via mechanical grinding using a Spex 6850 Freezer Mill. The cryogrinding protocol involved 4 minutes of rest followed by 9 cycles of 1 minute of grinding at a rate of 9 and 1 minute of rest. A measured mass of the resulting fine powder was then transferred to a fresh 50 mL falcon tube containing Protease Inhibitors (PI). The required volume of PI 100x stock solutions (0.5 mM PMSF. 0.015 mM E-64, and 0.05 mM Bestatin in DMSO) was determined using a yeast cell density of 1.15 g/mL as the reference point. The ground cells were allowed to thaw with gentle shaking at 4°C. Following thawing, cell debris was removed by centrifugation at 3,220 x g for 20 minutes in a refrigerated tabletop centrifuge set at 4°C. The resulting supernatant was then subjected to a second spin at 16,000 x g for 15 minutes at 4°C. The supernatant, referred to as the neat-lysate or total (T) fraction, was carefully transferred to a fresh tube. The protein concentration in the neat-lysates was determined using the bicinchoninic acid assay (Pierce™ Rapid Gold BCA Protein Assay Kit, Thermo Scientific).

Desiccation-rehydration treatment

A microfuge tube holding an Initial Volume of 0.1 mL of neat lysate was vacuum-dried in a desiccator for ~30 hours until it attained a stable weight. The rehydration process involved adding water to double the sample's initial mass. The rehydrated sample was gently agitated on a rotisserie at 4°C for 10 minutes, subjected to pipetting up and down for 5 minutes and subsequently re-incubated at 4°C for 1 hour using the rotisserie. The resoluble fraction, also known as the Supernatant (S), was obtained by centrifuging the rehydrated sample at 16,000xg for 15 minutes at 4°C. The supernatant fraction was then meticulously collected, and its volume was recorded to ensure accurate mass balance calculations. The remaining Pellet (P) fraction was resuspended in 1 mL of 8M urea solution. Subsequently, the protein concentration in both

the Supernatant (S) and Pellet (P) fractions was determined using the bicinchoninic acid assay (BCA Protein Assay Kit, Thermo Scientific).

Mass balance

The volume and concentration of each fraction were meticulously documented to evaluate the conservation of protein mass. Utilizing this information, we estimated the distribution of proteins between the S and P fractions and employed partitioning factors to scale the individual protein intensities recorded by MS.

Thermogravimetric analysis

Samples were run on a TA TGA5500 instrument in 0.1 mL platinum crucibles (TA 952108.906). Crucibles were tared before sample loading. Crucibles were loaded with between 5 mg and 10 mg of sample. Each sample was heated from 30 °C to 250 °C at a 10 °C per minute ramp.

Determination of water loss was conducted using TA's Trios Software (TA Instruments TRIOS version # 5.0.0.44608). The Trios software "Smart Analysis" tool was used to identify the inflection points along the Weight (%) curve and the 1st derivative Weight (%) curve. The Trios software "Step transition" tool was then used to select the area of the Weight (%) curve from the start of the run to the rightmost region of the curve that corresponded with the last flat region of the 1st derivative Weight (%) curve. The Trios software "Onset" and "Endset" tools were then used to identify the onset and endset of water loss.

Lysate characterization

The osmolality of neat lysates was measured using a Wescor VAPRO 5520 (Wescor, Logan, UT), calibrated with manufacturer-provided standard solutions of known osmotic potentials (100, 290, and 1000 mmol/kg). For each trial, 10 µL of the total fraction and 10 ul of the supernatant fraction lysate were placed on a filter paper disk and measured. Triplicate measurements were taken for each sample. The pH was measured in triplicate for each of the total and supernatant fractions. For each measurement, 50 µL of the sample was used. A Mettler Toledo InLAB Micro pH probe attached to a Mettler Toledo SevenCompact Duo pH/Conductivity meter were employed for the pH measurements

GFP Spike Experiments

Neat lysates were spiked with green fluorescence protein (-30 GFP, +9 GFP, +15 GFP, +36 GFP, mEGFP) to a final 1 µM. To assess the active concentration of the protein, Fluorescence was measured using 488±25 nm excitation light, and read between 490 and 600 nm emission using a BMG LabTech ClarioSTAR. Fluorescence values before and after D-R were measured in triplicates from the spiked total sample and after D-R from the supernatant sample. The total fluorescence is reported as the ratio between the supernatant and total, and taken from the average emission between 505 nm to 515 nm. Errors are standard deviations of the average between triplicates.

Proteome-wide Resolubility Sample Preparation

Total(T), Supernatant(S), and Pellet(P) fractions were measured for 200 µg protein diluted in 200 µL of Native Buffer (20 mM HEPES-KOH pH 7.4, 2 mM MgCl2). Samples were then transferred to fresh microfuge tubes containing 152 mg urea. Next, 4.5 µL of a freshly prepared 700 mM stock of DTT was added to each sample, and the mixture was incubated at 37 °C for 30 minutes at 700 rpm on a thermomixer to reduce cysteine residues. Following reduction, an 18 µL portion of a freshly prepared 700 mM stock of iodoacetamide (IAA) was added to each sample, and the mixture was then incubated at room temperature in the dark for 45 minutes to alkylate the reduced cysteine residues. After alkylation of the cysteines, 942 µL of 100 mM ammonium bicarbonate was added to each sample to dilute the urea. Subsequently, a 1 µL portion of a 1 mg/mL stock of trypsin (NEB Typsin-ultra™, Mass Spectrometry Grade) was added to each sample and the mixtures were incubated overnight at 25 °C at 700 rpm.

Sep-Pak C18 1cc Vac Cartridges (Waters) were employed using a vacuum manifold to desalt the peptides. The clean-up procedure, outlined in (*53*) began with acidification of the peptides using 12.6 µL of trifluoroacetic acid (TFA). Subsequently, the cartridges were washed with 1 mL of Buffer B (comprising 80% ACN and 0.5% TFA), followed by equilibration with 4 × 1 mL of Buffer A (0.5% TFA). Peptides were then loaded onto the cartridges and subjected to a wash step using 4 × 1 mL of Buffer A. Elution of peptides was achieved by adding 1 mL of Buffer B into the columns. Vacuum cartridges were placed inside 15 mL conical tubes and centrifuged at 350 g for 8 minutes. The eluted peptides were transferred to microfuge tubes and vacuum-dried before being stored at −80 °C until MS analysis.

Proteome-wide Resolubility LC-MS Data Acquisition

LC–MS was performed on a NanoElute (Bruker Daltonik) system coupled online to a hybrid TIMS-quadrupole TOF mass spectrometer (*54*) (Bruker Daltonik timsTOF Pro2, Germany) via a nano-electrospray ion source (Bruker Daltonik Captive Spray). Approximately 200 ng of peptides were separated on an Aurora column 25 cm × 75 µm ID, 1.7 um reversed-phase column (Ion Opticks) at a flow rate of 250 nL min-1 in an oven compartment heated to 50 °C. To analyze samples from whole-proteome digests, we used a gradient starting with a linear increase from 5% B to 25% B over 25 min, followed by further linear increases to 37% B in 4 min and to 90% B in 4min, which was held constant for 5 min. The column was equilibrated using 4 volumes of solvent A.

The timsTOF was operated in dia-PASEF mode (*55*) with mass range from 100 to 1700 m/z, 1/k0 Start 0.75 V s/cm2, End 1.30 V s/cm2, ramp and accumulation times set to 50ms, Capillary Voltage was 1600V, dry gas 3 l/min, and dry temp 200 °C.

dia-PASEF settings were set to an optimized acquisition scheme using py_diAID software (*56*) with dia-windows isolation starting from 7 to 95 Da according to the multiple charge ion cloud density map. Each cycle consisted of 1x MS1 full scan and 60x MS2 windows covering 350 – 1250 m/z and 0.7 - 1.30 1/k0 (see supplementary material). The cycle time was 1.12 seconds. CID collision energy was 20 eV (0.60 1/k0) to 59 eV (1.6 1/k0) as a function of the inverse mobility of the precursor.

Proteome-wide Resolubility LC-MS Data Analysis

The Spectronaut18 proteomics analyzer (*57*) was utilized with the directDIA+ (Deep) workflow to analyze spectra and perform Label-Free Quantification (LFQ) of detected peptides with a 1% FDR. A tryptic search allowing up to 2 missed cleavages for peptides ranging 7-52 residues was conducted against the *S. cerevisiae* (UP000002311, UniProt) reference proteome database. Methionine oxidation and N-terminus acetylation were allowed as dynamic modifications. All ratios reported here (S/T, S/P, and S/S+T) were determined from protein-level intensity measurements (as calculated from areas under the peak) averaged across three replicates, which were then normalized by protein mass as determined by BCA assay. The p-value is calculated from the standard deviation between replicates, and corrected for FDR.

Limited Proteolysis Mass Spectrometry (LiP-MS) Sample Preparation

The LiP assay was conducted according to the procedure outlined in the referenced literature (*58*). In summary, 2 µL of a PK stock solution (0.116 mg/mL PK in a 1:1 mixture of Native Buffer and 20% glycerol) was dispensed into a sterile 1.5 mL microcentrifuge tube. Total (T) and Supernatant (S) samples, each containing 400 µg of protein diluted in 400 µL of Native Buffer, were prepared. These samples were divided into two sets: one set was kept on ice as a non-LiP control, while the other set was subjected to LiP treatment. To initiate LiP, the samples were combined with the PK-containing microcentrifuge tube and vigorously mixed by rapid vortexing, followed by immediate centrifugation to sediment the liquids at the bottom of the tube. Samples were incubated for exactly 1 min at 25°C. Afterwards, the microfuge tube containing the LiP samples was transferred to a mineral oil bath pre-equilibrated at 100°C for 5 min to quench PK activity. Boiled samples were then flash centrifuged. LiP and non-LiP control samples were transferred to 152 mg of urea. Next, 4.5 µL of a freshly prepared 700 mM stock of DTT was added to each sample, the mixture was incubated at 37 °C for 30 minutes at 700 rpm on a thermomixer to reduce cysteine residues. Following reduction, an 18 µL portion of a freshly prepared 700 mM stock of iodoacetamide (IAA) was added to each sample, and the mixture was then incubated at room temperature in the dark for 45 minutes to alkylate the reduced cysteine residues. After alkylation of the cysteines, 942 µL of 100 mM ammonium bicarbonate was added to each sample to dilute the urea. Subsequently, a 1 µL portion of a 1 mg/mL stock of trypsin (New England Biolabs, Inc. Typsin-ultraTM, Mass Spectrometry Grade) was added to each sample, the mixtures were incubated overnight at 25 °C at 700 rpm.

LiP-MS Data Acquisition

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

residual peptides between runs in a sequence. A Thermo Q-Exactive HF-X Orbitrap mass spectrometer was used to analyze protein digests. A full MS scan in positive ion mode was followed by 20 data-dependent MS scans. The full MS scan was collected using a resolution of 120000 (@ m/z 200), an AGC target of 3E6, a maximum injection time of 64 ms, and a scan range from 350 to 1500 m/z. The data-dependent scans were collected with a resolution of 15000 (@ m/z 200), an AGC target of 1E5, a minimum AGC target of 8E3, a maximum injection time of 55 ms, and an isolation window of 1.4 m/z units. To dissociate precursors prior to their reanalysis by MS2, peptides were subjected to an HCD of 28% normalized collision energies. Fragments with charges of 1, 6, 7, or higher and unassigned were excluded from analysis, and a dynamic exclusion window of 30.0 s was used for the data dependent scans.

LiP-MS Data Analysis

The FragPipe v20.0 proteomics pipeline with IonQuant v1.9.8 with a match between runs (MBR) false discovery rate (FDR) of 5% was utilized to analyze spectra and perform Label-Free Quantification (LFQ) of detected peptides (*59*, *60*). Using MSFragger v3.8 (*61*) and Philosopher v5.0 (*62*), a semitryptic search allowing up to 2 missed cleavages was conducted against the *S. cerevisiae* (UP00000231, UniProt) reference proteome database, and identifications were filtered to a 1% FDR. An MS1 precursor mass tolerance of 10 ppm and an MS2 fragmentation tolerance of 20 ppm was used. Methionine oxidation and N-terminus acetylation were allowed as dynamic modifications, while carbamidomethylation on cysteines was defined as a static modification. Raw ion intensity data for identified peptides were exported and processed utilizing FLiPPR (v0.1.4) (*25*). Data were merged from the ion to the peptide level (**Table S4**) and normalized using the log2 (S/T Protein Abundance Ratio) as outlined in the resolubility study. In all cases, missing data imputation, filtering, and Benjami-Hochberg FDR correction were carried out per-protein, as implemented in FLiPPR . In proteome-wide analyses, peptides were labeled significantly perturbed by desiccation if their normalized abundance changed by more than 2-fold between S and T samples and adjusted P-values were less than 0.05. Metadata were compiled from various sources (Saccharomyces Genome Database (*63*); ECOD (*35*)), and assembled together with the FLiPPR outputs using custom Python scripts and DomainMapper (*64*).

ATPase activity assays

ATPase activity in the Total and Supernatant fractions was assessed using the ATPase Assay Kit (abcam ab270551), which detects inorganic phosphate (Pi) colorimetrically as a direct product of ATPase activity. Given the intrinsic presence of ATP and Pi in neat lysates, the T and S fractions were diluted with Native Buffer to a concentration of approximately 2 µg/µL. Subsequently, 100 µL of the diluted fractions were processed through Zeba™ Spin Desalting Columns with a 7K MWCO to eliminate small solutes, including Pi and ATP. The ATPase assay followed the kit instructions, utilizing protein quantities ranging from 0.25 to 5 µg from each fraction. The ATPase reaction was performed at 30°C for 1 hour. We prepared assay blanks that lacked protein from either fraction to serve as controls. The signal from these blanks was subtracted from the samples' signal to ensure accurate measurements. Results were expressed as relative ATPase Activity, calculated by dividing the absorbance at 624 nm by µg of protein in

the Supernatant fraction by the corresponding change in the Total fraction. This ratio represents the fold change or percentage relative to the Total sample.

Citrate Synthase Activity

The Citrate Synthase Activity in the Total and Supernatant fractions was evaluated using the Citrate Synthase Activity Assay Kit (Sigma-Aldrich MAK193). This assay detects -SH groups produced during the enzymatic conversion of oxaloacetate and acetyl-CoA to citrate by citrate synthase. The reaction progress was monitored at 412 nm over 90 minutes. Following the kit instructions, we measured citrate synthase activity using protein quantities ranging from 100 to 400 µg from each fraction in triplicates. Due to significant background levels in the samples, a sample blank was included for each sample by omitting the substrate in the reaction mix. The readings from the sample blanks were subtracted from all sample readings to account for background interference and ensure accurate measurement of the enzymatic activity.

The results were expressed as Relative Citrate Synthase Activity by dividing the change in absorbance at 412 nm per minute per microgram of protein in the Supernatant fraction by the corresponding change in the Total fraction. This ratio represents the fold change or percentage relative to the Total sample and was averaged between the triplicates. The errors are standard deviations of the average between triplicates.

RNA purification and Quantification

4-20 uL of T and S fraction was used for RNA purification. The RNeasy Plus Mini Kit was used following the manufacturer's procedure. RNA quantification was done using the Qubit™ RNA Broad Range (BR) kit. Measurements were taken in triplicates and averaged. Errors are standard deviations of the averages.

Proteome-wide structural bioinformatics

Proteome-wide structural bioinformatics was based on the AlphaFold2 *S. cerevisiae* proteome (UP000002485) obtained from https://alphafold.ebi.ac.uk/download#proteomes-section in March 2024 (*18*, *28*). This initial dataset contained 6039 yeast proteins. From those proteins, we excised out individual intrinsically disordered regions using metapredict (V2-FF) for disorder analysis (see below) (*65*, *66*). Disordered regions were identified using the `predict_disorder_domains()` function in metapredict (V2-FF). All informatics used SHEPHARD to organize, parse, and sanity-check protein sequence information and annotations, while all sequence-based analysis was performed using sparrow or FINCHES (*30*, *67*) (https://github.com/idptools/sparrow). All code and scripts for the proteome-wide informatic analysis are at https://github.com/holehouse-lab/supportingdata/tree/master/2024/romero_2024.

Membrane proteins vs. non-membrane proteins were segmented using UniProt annotations based on subcellular localization. Annotations were obtained on June 14th, 2024. This identified 1775 membrane proteins and 4264 non-membrane proteins across the yeast proteome. After cross-referencing with proteins recovered in our mass spectrometry experiments, this left 1090 membrane proteins and 3226 non-membrane proteins. Membrane proteins here include both integral and peripheral membrane proteins.

All analyses that correlate an observable with resolubility are based on dividing the 3226 non-membrane proteins into eight quantiles based on ranking their S/T. A higher S/T value implies the protein is more resoluble. The set of proteins in each quantile is given in **Table S3**.

Protein abundance data was obtained from Ghaemmaghami, as described previously (*26*, *67*). Protein length was based on the total number of amino acids in each protein sequence. The number of globular domains used in Fig. 3D was based on the DODO decomposition (described below). The fraction disordered is calculated as the fraction of amino acids in each protein that falls within a contiguous disordered domain, as identified by metapredict V2-FF. The number of interactors was determined by taking physical interaction partners reported in the STRING database (specifically, we used the data from **`4932.protein.physical.links.v12.0.txt`**) and mapping interprotein interaction partners (*27*). Subcellular localization annotations were taken from UniProt, and only the top four most commonly observed annotations shown.

We also excised out globular folded domains using two distinct bioinformatic approaches. One approach used chainsaw, a supervised learning approach trained to predict whether a pair of residues comes from the same contiguous domain or not(*29*). The other used DODO, a structural informatics approach that approximates folded domain boundaries by quantifying the number of atoms within a specific threshold distance of all other atoms in a given structure. For chainsaw, we performed domain segmentation on the structure predictions of the S. cerevisiae proteome obtained from the AlphaFold Database. Chainsaw predictions were performed with default parameters - e.g., `--remove_disordered_domain_threshold` left at 0.35 and `--min_domain_length` was left at 30 residues and `--min_ss_components` left at 2. For DODO, we used the default settings and the `build.pdb_from_pdb(input_file, output_file, just_fds=True)` command.

Broadly speaking, two types of error can occur in the context of domain decomposition. One type occurs when an algorithm is overly aggressive, segmenting domains that should not be segmented and creating artificial interfaces that expose residues which should not be exposed. Another type occurs if an algorithm is insufficiently sensitive, grouping multiple domains together into a single large domain. These two types of errors are expected to have distinct impacts on the types of residues that emerge as solvent accessible. We therefore sought to apply two different approaches for domain decomposition in which each of the two tended to bias towards each of these types of errors. We reasoned that if both approaches led us to the same conclusion, this would confirm our conclusions were robust to the details and biases of the domain decomposition approach.

We opted to use chainsaw in a mode whereby discontinuous domains were segmented apart, effectively ensuring that it would behave in an overly-aggressive way. In contrast, given DODO was developed for the reconstruction of disordered regions around folded domains, as opposed to domain decomposition, we noticed that DODO often erred on the side of failing to decompose multi-domain subunits into distinct globular domains. We therefore decomposed the yeast proteome and performed all analyses discussed here in these two ways, with the goal of providing confidence that any conclusions drawn at were not due to idiosyncrasies of a specific algorithmic approach to domain decomposition. Finally, we note that all domain-centric analysis

is focussed on those non-membrane proteins that have at least one domain between 90 and 1000 amino acids in length. This filtering leaves us with 3347 domains (DODO) or 5477 domains (chainsaw) spread over 2667 proteins (DODO) or 2884 proteins (chainsaw). The quantisation of proteins for the analysis in figure 3 and 4 is then based on just those proteins.

Solvent-accessible residues were identified using individual globular domains with FINCHES, which in turn uses the Shrake Rupley algorithm implemented within MDTraj (*30*, *68*). Solvent accessibility of each residue was determined in A^2 using a probe radius of 0.14 nm on a per-residue basis. Whether a residue was solvent accessible or not was defined as to whether the total residue SASA was over 40% accessible of the maximum possible accessibility for that residue type. Maximum accessibility was determined by performing all-atom simulations of GXG tripeptides and computing the average accessibility of the guest peptide. Various accessibility percentages were compared, and 40% offered a reasonable tradeoff between being sufficiently permissive to capture sidechain conformers that may occlude their accessibility in the specific AlphaFold2 structure being used and being sufficiently restrictive to prevent residues deep in crevices from being included. We also repeated our entire pipeline using a 10% threshold and found no difference in any of the conclusions, giving us confidence that for the questions here, the trends are robust to the specific details regarding solvent accessibility. **Fig. S8** compares the fraction of residues that are solvent accessible vs. domain size (in amino acids).

The solvent-accessible amino acid composition was computed in two different ways: by the overall surface area of a given residue and the overall count of a given residue. In both cases, the overall approach involved taking each individual globular domain, identifying solvent-accessible residues, and then either counting the number of each residue type that was solvent accessible or taking the summed surface area for each different residue type. These values were added to a per-residue tally, which was updated for each domain. For surface hydrophobicity, the average Kyte-Doolitle value of all surface-exposed residues was computed (**Fig. 4E**). For surface net charge, the average fraction of arginine + lysine (positive) or aspartic acid + glutamic acid (negative) was computed (**Fig. 4E**). For individual amino acids, the fraction of surface residues made of each amino acid was calculated (**Fig. 4F**). We also calculated the fraction of the overall globular domain surface area made up of each amino acid (**Fig. S9**). In both cases, the same overarching conclusions emerge.

Intermolecular interaction prediction

Surface-dependent intermolecular interactions were calculated using FINCHES, as described previously. FINCHES uses the physical chemistry encoded into molecular forcefields and repurposes that physical chemistry for bioinformatic analysis. We previously performed systematic clustering of all possible peptides to identify a minimal set of 36 distinct peptides that are chemically orthogonal to one another (see Fig. S11 in Ginell et al(*30*)). For each individual globular domain, we calculated the average mean-field attractive interaction between the surface of the domain and each of these 36 peptides, enabling us to determine domains where more or less attractive for different types of chemistry.

References

1. G. Vecchi, P. Sormanni, B. Mannini, A. Vandelli, G. G. Tartaglia, C. M. Dobson, F. U. Hartl, M. Vendruscolo, Proteome-wide observation of the phenomenon of life on the edge of solubility. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 1015–1020 (2020).

2. J. D. Hibshman, J. S. Clegg, B. Goldstein, Mechanisms of desiccation tolerance: Themes and variations in brine shrimp, roundworms, and tardigrades. *Front. Physiol.* **11**, 592016 (2020).

3. H. Tapia, L. Young, D. Fox, C. R. Bertozzi, D. Koshland, Increasing intracellular trehalose is sufficient to confer desiccation tolerance to Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 6122–6127 (2015).

4. D. K. Hincha, E. Zuther, A. V. Popova, Stabilization of dry sucrose glasses by four LEA_4 proteins from Arabidopsis thaliana. *Biomolecules* **11**, 615 (2021).

5. J. D. Hibshman, S. Carra, B. Goldstein, Tardigrade small heat shock proteins can limit desiccation-induced protein aggregation. *Commun Biol* **6**, 121 (2023).

6. T. C. Boothby, H. Tapia, A. H. Brozena, S. Piszkiewicz, A. E. Smith, I. Giovannini, L. Rebecchi, G. J. Pielak, D. Koshland, B. Goldstein, Tardigrades Use Intrinsically Disordered Proteins to Survive Desiccation. *Mol. Cell* **65**, 975–984.e5 (2017).

7. M. Battaglia, Y. Olvera-Carrillo, A. Garciarrubio, F. Campos, A. A. Covarrubias, The enigmatic LEA proteins and other hydrophilins. *Plant Physiol.* **148**, 6–24 (2008).

8. Y. Yang, S.-I. Yokobori, A. Yamagishi, Bacterial survival in response to desiccation and high humidity at above zero and subzero temperatures. *Adv. Space Res.* **43**, 1285–1290 (2009).

9. M. Iwaya-Inoue, M. Sakurai, M. Uemura, *Survival Strategies in Extreme Cold and Desiccation: Adaptation Mechanisms and Their Applications* (Springer, 2018).

10. G. I. Olgenblum, B. O. Hutcheson, G. J. Pielak, D. Harries, Protecting Proteins from Desiccation Stress Using Molecular Glasses and Gels. *Chem. Rev.* **124**, 5668–5694 (2024).

11. K. Goyal, L. J. Walton, A. Tunnacliffe, LEA proteins prevent protein aggregation due to water stress. *Biochem. J* **388**, 151–157 (2005).

12. S. Chakrabortee, C. Boschetti, L. J. Walton, S. Sarkar, D. C. Rubinsztein, A. Tunnacliffe, Hydrophilic protein associated with desiccation tolerance exhibits broad protein stabilization function. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 18073–18078 (2007).

13. C. J. Crilly, J. A. Brom, O. Warmuth, H. J. Esterly, G. J. Pielak, Protection by desiccation-tolerance proteins probed at the residue level. *Protein Sci.*, doi: 10.1002/pro.4231 (2021).

14. J. Crowe, A. E. Oliver, F. Tablin, Is there a single biochemical adaptation to anhydrobiosis?1. *Integr. Comp. Biol.* **42**, 497–503 (2002).

15. P. S. Romero-Perez, Y. Dorone, E. Flores, S. Sukenik, S. Boeynaems, When Phased without Water: Biophysics of Cellular Desiccation, from Biomolecules to Condensates. *Chem. Rev.* **123**, 9010–9035 (2023).

16. S. Schopper, A. Kahraman, P. Leuenberger, Y. Feng, I. Piazza, O. Müller, P. J. Boersema, P. Picotti, Measuring protein structural changes on a proteome-wide scale using limited proteolysis-coupled mass spectrometry. *Nat. Protoc.* **12**, 2391–2410 (2017).

17. P. To, A. M. Bhagwat, H. E. Tarbox, A. Ecer, H. Wendorff, Z. Jamieson, T. Trcek, S. D. Fried, Intrinsically Disordered Regions Promote Protein Refoldability and Facilitate Retrieval from Biomolecular Condensates, *bioRxiv* (2023)p. 2023.06.25.546465.

18. J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

19. D. Shore, B. Albert, Ribosome biogenesis and the cellular energy economy. *Curr. Biol.* **32**, R611–R617 (2022).

20. D. Koshland, H. Tapia, Desiccation tolerance: an unusual window into stress biology. *Mol. Biol. Cell* **30**, 737–741 (2019).

21. A. Garay-Arroyo, J. M. Colmenero-Flores, A. Garciarrubio, A. A. Covarrubias, Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *J. Biol. Chem.* **275**, 5668–5674 (2000).

22. D. Calahan, M. Dunham, C. DeSevo, D. E. Koshland, Genetic Analysis of Desiccation Tolerance in Saccharomyces cerevisiae. *Genetics* **189**, 507–519 (2011).

23. P. Alpert, The limits and frontiers of desiccation-tolerant life. *Integr. Comp. Biol.* **45**, 685–695 (2005).

24. Y. Feng, G. De Franceschi, A. Kahraman, M. Soste, A. Melnik, P. J. Boersema, P. P. de Laureto, Y. Nikolaev, A. P. Oliveira, P. Picotti, Global analysis of protein structural changes in complex proteomes. *Nat. Biotechnol.* **32**, 1036–1044 (2014).

25. E. Manriquez-Sandoval, J. Brewer, G. Lule, S. Lopez, S. D. Fried, FLiPPR: A Processor for Limited Proteolysis (LiP) Mass Spectrometry Data Sets Built on FragPipe. *J. Proteome Res.* **23**, 2332–2342 (2024).

26. S. Ghaemmaghami, W.-K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea, J. S. Weissman, Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).

27. D. Szklarczyk, R. Kirsch, M. Koutrouli, K. Nastou, F. Mehryary, R. Hachilif, A. L. Gable, T. Fang, N. T. Doncheva, S. Pyysalo, P. Bork, L. J. Jensen, C. von Mering, The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.* **51**, D638–D646 (2023).

28. K. Tunyasuvunakool, J. Adler, Z. Wu, T. Green, M. Zielinski, A. Žídek, A. Bridgland, A. Cowie, C. Meyer, A. Laydon, S. Velankar, G. J. Kleywegt, A. Bateman, R. Evans, A. Pritzel, M. Figurnov, O. Ronneberger, R. Bates, S. A. A. Kohl, A. Potapenko, A. J. Ballard, B. Romera-Paredes, S. Nikolov, R. Jain, E. Clancy, D. Reiman, S. Petersen, A. W. Senior, K.

Kavukcuoglu, E. Birney, P. Kohli, J. Jumper, D. Hassabis, Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590–596 (2021).

29. J. Wells, A. Hawkins-Hooker, N. Bordin, I. Sillitoe, B. Paige, C. Orengo, Chainsaw: protein domain segmentation with fully convolutional neural networks. *Bioinformatics* **40**, btae296 (2024).

30. G. Ginell, R. J. Emenecker, J. M. Lotthammer, E. T. Usher, A. Holehouse, Direct prediction of intermolecular interactions driven by disordered regions. *bioRxiv*, doi: 10.1101/2024.06.03.597104 (2024).

31. M. S. Lawrence, K. J. Phillips, D. R. Liu, Supercharging proteins can impart unusual resilience. *J. Am. Chem. Soc.* **129**, 10110–10112 (2007).

32. L. Kolberg, U. Raudvere, I. Kuzmin, P. Adler, J. Vilo, H. Peterson, g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* **51**, W207–W212 (2023).

33. M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44**, D457–62 (2016).

34. M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).

35. R. D. Schaeffer, L. N. Kinch, Y. Liao, N. V. Grishin, Classification of proteins with shared motifs and internal repeats in the ECOD database. *Protein Sci.* **25**, 1188–1203 (2016).

36. A. D. Goldman, J. T. Beatty, L. F. Landweber, The TIM Barrel Architecture Facilitated the Early Evolution of Protein-Mediated Metabolism. *J. Mol. Evol.* **82**, 17–26 (2016).

37. R. R. Copley, P. Bork, Homology among (betaalpha)(8) barrels: implications for the evolution of metabolic pathways. *J. Mol. Biol.* **303**, 627–641 (2000).

38. K. E. Medvedev, L. N. Kinch, R. Dustin Schaeffer, J. Pei, N. V. Grishin, A Fifth of the Protein World: Rossmann-like Proteins as an Evolutionarily Successful Structural unit. *J. Mol. Biol.* **433**, 166788 (2021).

39. K. E. Medvedev, L. N. Kinch, R. D. Schaeffer, N. V. Grishin, Functional analysis of Rossmann-like domains reveals convergent evolution of topology and reaction pathways. *PLoS Comput. Biol.* **15**, e1007569 (2019).

40. R. Milo, P. Jorgensen, U. Moran, G. Weber, M. Springer, BioNumbers—the database of key numbers in molecular and cell biology. *Nucleic Acids Res.* **38**, D750–D753 (2010).

41. K. C. Shraddha, K. Nguyen, V. Nicholson, A. Walgren, T. Trent, E. Gollub, S. Romero, A. S. Holehouse, S. Sukenik, T. C. Boothby, Disordered proteins interact with the chemical environment to tune their protective function during drying. *Elife* **13** (2024).

42. X. Sui, D. E. V. Pires, A. R. Ormsby, D. Cox, S. Nie, G. Vecchi, M. Vendruscolo, D. B. Ascher, G. E. Reid, D. M. Hatters, Widespread remodeling of proteome solubility in

response to different protein homeostasis stresses. *Proc. Natl. Acad. Sci. U. S. A.*, doi: 10.1073/pnas.1912897117 (2020).

43. S. Sanchez-Martinez, K. Nguyen, S. Biswas, V. Nicholson, A. V. Romanyuk, J. Ramirez, S. Kc, A. Akter, C. Childs, E. K. Meese, E. T. Usher, G. M. Ginell, F. Yu, E. Gollub, M. Malferrari, F. Francia, G. Venturoli, E. W. Martin, F. Caporaletti, G. Giubertoni, S. Woutersen, S. Sukenik, D. N. Woolfson, A. S. Holehouse, T. C. Boothby, Labile assembly of a tardigrade protein induces biostasis. *Protein Sci.* **33**, e4941 (2024).

44. T. Furuki, M. Sakurai, Group 3 LEA protein model peptides protect enzymes against desiccation stress. *Biochim. Biophys. Acta* **1864**, 1237–1243 (2016).

45. A. H. Elcock, J. A. McCammon, Electrostatic contributions to the stability of halophilic proteins. *J. Mol. Biol.* **280**, 731–748 (1998).

46. A. Szilágyi, P. Závodszky, Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure* **8**, 493–504 (2000).

47. J. Echave, S. J. Spielman, C. O. Wilke, Causes of evolutionary rate variation among protein sites. *Nat. Rev. Genet.* **17**, 109–121 (2016).

48. H. Wennerström, E. Vallina Estrada, J. Danielsson, M. Oliveberg, Colloidal stability of the living cell. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 10113–10121 (2020).

49. C. Ma, A. Malessa, A. J. Boersma, K. Liu, A. Herrmann, Supercharged Proteins and Polypeptides. *Adv. Mater.* **32**, e1905309 (2020).

50. C. M. Davis, M. Gruebele, S. Sukenik, How does solvation in the cell affect protein folding and binding? *Curr. Opin. Struct. Biol.* **48**, 23–29 (2018).

51. J. H. Fong, B. A. Shoemaker, A. R. Panchenko, Intrinsic protein disorder in human pathways. *Mol. Biosyst.* **8**, 320–326 (2012).

52. R. K. Mortimer, J. R. Johnston, Genealogy of principal strains of the yeast genetic stock center. *Genetics* **113**, 35–43 (1986).

53. P. To, Y. Xia, S. O. Lee, T. Devlin, K. G. Fleming, S. D. Fried, A proteome-wide map of chaperone-assisted protein refolding in a cytosol-like milieu. *Proc. Natl. Acad. Sci. U. S. A.* **119**, e2210536119 (2022).

54. F. Meier, A.-D. Brunner, S. Koch, H. Koch, M. Lubeck, M. Krause, N. Goedecke, J. Decker, T. Kosinski, M. A. Park, N. Bache, O. Hoerning, J. Cox, O. Räther, M. Mann, Online Parallel Accumulation-Serial Fragmentation (PASEF) with a Novel Trapped Ion Mobility Mass Spectrometer. *Mol. Cell. Proteomics* **17**, 2534–2545 (2018).

55. F. Meier, A.-D. Brunner, M. Frank, A. Ha, I. Bludau, E. Voytik, S. Kaspar-Schoenefeld, M. Lubeck, O. Raether, N. Bache, R. Aebersold, B. C. Collins, H. L. Röst, M. Mann, diaPASEF: parallel accumulation-serial fragmentation combined with data-independent acquisition. *Nat. Methods* **17**, 1229–1236 (2020).

56. P. Skowronek, M. Thielert, E. Voytik, M. C. Tanzer, F. M. Hansen, S. Willems, O. Karayel, A.-D. Brunner, F. Meier, M. Mann, Rapid and In-Depth Coverage of the

(Phospho-)Proteome With Deep Libraries and Optimal Window Design for dia-PASEF. *Mol. Cell. Proteomics* **21**, 100279 (2022).

57. O. M. Bernhardt, N. Selevsek, L. C. Gillet, O. Rinner, P. Picotti, R. Aebersold, L. Reiter, Spectronaut A fast and efficient algorithm for MRM-like processing of data independent acquisition (SWATH-MS) data. *F1000Res.* **5** (2012).

58. P. To, B. Whitehead, H. E. Tarbox, S. D. Fried, Nonrefoldability is Pervasive Across the E. coli Proteome. *J. Am. Chem. Soc.* **143**, 11435–11448 (2021).

59. F. Yu, S. E. Haynes, A. I. Nesvizhskii, IonQuant Enables Accurate and Sensitive Label-Free Quantification With FDR-Controlled Match-Between-Runs. *Mol. Cell. Proteomics* **20**, 100077 (2021).

60. C. Rojas Ramírez, J. A. Espino, L. M. Jones, D. A. Polasky, A. I. Nesvizhskii, Efficient Analysis of Proteome-Wide FPOP Data by FragPipe. *Anal. Chem.* **95**, 16131–16137 (2023).

61. A. T. Kong, F. V. Leprevost, D. M. Avtonomov, D. Mellacheruvu, A. I. Nesvizhskii, MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. *Nat. Methods* **14**, 513–520 (2017).

62. F. da Veiga Leprevost, S. E. Haynes, D. M. Avtonomov, H.-Y. Chang, A. K. Shanmugam, D. Mellacheruvu, A. T. Kong, A. I. Nesvizhskii, Philosopher: a versatile toolkit for shotgun proteomics data analysis. *Nat. Methods* **17**, 869–870 (2020).

63. J. M. Cherry, C. Adler, C. Ball, S. A. Chervitz, S. S. Dwight, E. T. Hester, Y. Jia, G. Juvik, T. Roe, M. Schroeder, S. Weng, D. Botstein, SGD: Saccharomyces Genome Database. *Nucleic Acids Res.* **26**, 73–79 (1998).

64. E. Manriquez-Sandoval, S. D. Fried, DomainMapper: Accurate domain structure annotation including those with non-contiguous topologies. *Protein Sci.* **31**, e4465 (2022).

65. R. J. Emenecker, D. Griffith, A. S. Holehouse, Metapredict: a fast, accurate, and easy-to-use predictor of consensus disorder and structure. *Biophys. J.* **120**, 4312–4319 (2021).

66. J. M. Lotthammer, G. M. Ginell, D. Griffith, R. J. Emenecker, A. S. Holehouse, Direct prediction of intrinsically disordered protein conformational properties from sequence. *Nat. Methods* **21**, 465–476 (2024).

67. G. M. Ginell, A. J. Flynn, A. S. Holehouse, SHEPHARD: a modular and extensible software architecture for analyzing and annotating large protein datasets. *Bioinformatics* **39** (2023).

68. R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane, V. S. Pande, MDTraj: a modern, open library for the analysis of molecular dynamics trajectories. *Biophys. J.* **109**, 1528–1532 (2015).