The Proteome Folding Problem and Cellular Proteostasis

Evan T. Powers [1] **and Lila M. Gierasch** [2]

1 - *Department of Chemistry,* The Scripps Research Institute, La Jolla, CA 92037, USA
2 - *Departments of Biochemistry & Molecular Biology and Chemistry,* University of Massachusetts-Amherst, Amherst, MA 01003, USA

Correspondence to : Laboratory of Receptor Biology and Gene Expression, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, United States (Y. Dalal). Biophysics Program, Institute for Physical Science and Technology, University of Maryland, College Park, MD 20742, United States and Department of Chemistry and Biochemistry, University of Maryland, College Park, MD 20742, United States (G.A. Papoian). *epowers@scripps. edu* (E.T. Powers), *gierasch@biochem.umass.edu* (L.M. Gierasch),
https://doi.org/10.1016/j.jmb.2021.167197
Edited by Sheena E. Radford

Abstract

Stunning advances have been achieved in addressing the protein folding problem, providing deeper understanding of the mechanisms by which proteins navigate energy landscapes to reach their native states and enabling powerful algorithms to connect sequence to structure. However, the realities of the in vivo protein folding problem remain a challenge to reckon with. Here, we discuss the concept of the "proteome folding problem"—the problem of how organisms build and maintain a functional proteome—by admitting that folding energy landscapes are characterized by many misfolded states and that cells must deploy a network of chaperones and degradation enzymes to minimize deleterious impacts of these off-pathway species. The resulting proteostasis network is an inextricable part of in vivo protein folding and must be understood in detail if we are to solve the proteome folding problem. We discuss how the development of computational models for the proteostasis network's actions and the relationship to the biophysical properties of the proteome has begun to offer new insights and capabilities.

The Protein Folding Problem

In 1958, Kendrew *et al*. published the first three-dimensional model of the structure of a globular protein, sperm whale myoglobin. Although the resolution was only 6 Å, the complexity of the structure was abundantly clear and led the authors to remark: "The arrangement seems to be almost totally lacking in the kind of regularities which one instinctively anticipates, and it is more complicated than has been predicted by any theory of protein structure".[1] This complexity became even more apparent in later, higher resolution structures of myoglobin.[2] In parallel with these studies, Anfinsen and others were studying the process of protein folding using bovine pancreatic ribonuclease as a model protein. They discovered that the complicated structures reported by Kendrew and others seemed to be the thermodynamically most stable conformations that could be adopted by polypeptide chains. Moreover, they showed that these structures could be attained without assistance from any external biological machinery, despite their complexity.[3–6] These two lines of work immediately suggested that protein primary structures contain all the information needed to produce the rich complexity evident in

protein tertiary structures. And so, the protein folding problem was born.

The protein folding problem has dominated protein science for the past 60 years. There are many definitions of it, but we prefer Dill's three-part formulation[7,8]: (1) how do we predict a protein's three-dimensional tertiary structure from its one-dimensional primary structure? (2) how do the various intra- and intermolecular interactions in the system (protein–protein, protein-solvent, and solvent–solvent) determine this three-dimensional structure? and (3) what is (are) the pathway(s) by which it is formed? Tremendous progress has been made on all three aspects of the protein folding problem, and this progress has been accelerating as more powerful computational and experimental methods have been brought to bear.[9] This past year, however, we witnessed an extraordinary advance toward a solution to the protein folding problem in the performance of AlphaFold, an artificial intelligence program developed by DeepMind to predict protein structures. In CASP14 (the 14th running of the Critical Assessment of Structure Prediction competition), AlphaFold was able to predict the three-dimensional structures of an array of proteins with a median accuracy that was as high as that of models built from experimental data.[10,11] The potential utility of this algorithm was recently powerfully demonstrated by predicting the structures of almost every protein in the human proteome.[12] While AlphaFold cannot be considered a comprehensive solution to the protein folding problem—it addresses only the first of the three questions that make up the protein folding problem as defined above—we believe that it probably will mark a new era in protein science. As such this seems like a suitable moment to pause and consider what challenges might define the next 60 years of protein science the way that the protein folding problem has defined the last 60.

The Proteome Folding Problem

The protein folding problem focuses on one outcome that can occur when an unstructured protein is placed under native conditions: folding to the functional native structure. However, it has long been known that folding is not so straightforward for many (if not most) proteins, and that other, more untoward outcomes are possible.[13] Proteins can become trapped in misfolded conformations that represent local minima on their folding energy landscapes.[14,15] This is especially true for longer proteins with more complicated topologies, which are underrepresented among proteins for which detailed in vitro folding studies have been done.[16] This situation can be visualized with a folding energy landscape that contains more than one deep minimum, as suggested by Clark in 2004.[15] Such a folding energy landscape is depicted in Figure 1. If these misfolded

conformers are abundant enough and persist for long enough, they can self-associate to form aggregates that are virtually never salutary to a cell and are usually toxic, especially in their early stages of formation.[17–20] Even if misfolded proteins did not aggregate, it is very unlikely that they would be functional, and such a loss-of-function could be devastating.[17,19,20] Moreover, many different proteins are present simultaneously in vivo, a feature of the proteome folding problem that introduces major complexity. To mitigate the effects of protein misfolding, all organisms from the time of the last universal common ancestor have had a network of cellular components that is tasked with maintaining protein homeostasis, or "proteostasis", which is the condition of an organism having enough properly folded proteins to carry out the functions essential to life but not so much misfolded/aggregated protein as to interfere with these functions.[21,22]

Recognizing that protein folding is not an ideal process and that there is biological machinery that exists to help it along raises a different problem that has been gaining recognition over the past two decades[14–19,23,24] and that we believe will become one of the dominant questions in protein science going forward: the *proteome folding problem*. In our view, the proteome folding problem can be broken into two questions. The first is still intrinsic to protein sequences under a given set of conditions: what alternative states are competitive with the native state, either thermodynamically or kinetically or both, on the protein folding energy landscape as it exists in living cells? This question subsumes the original protein folding problem, but it also requires an understanding of the energetics of non-native states, which would have been considered decoys or incorrect solutions in the context of the original protein folding problem. In addition, it includes issues relating to the environment within cells. Many environmental variables that can profoundly influence protein folding vary not only between cells but also within cells—from organelle to organelle and potentially even between liquid–liquid phase separated domains in a single organelle.[25–29] These variables include but are not limited to: the pH; the concentrations of ions and osmolytes[30–32]; and macromolecular crowding.[33–35]

The second question in the proteome folding problem is largely extrinsic to protein sequences: how do components of the biological folding machinery, the "proteostasis network", interact with proteins to optimize protein folding outcomes, or when folding fails, how do they mitigate the impact of protein folding failures? We examine these questions in the following sections.

Non-Native States of Proteins and the Cellular Folding Environment

A useful example of the non-native states of proteins comes from work by the Anfinsen lab on
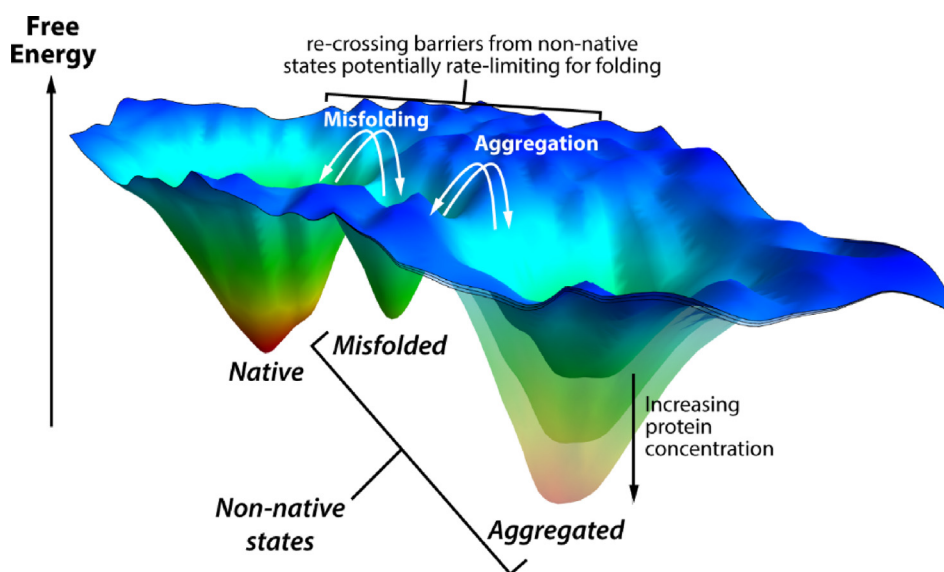
Figure 1. Protein folding energy landscape with multiple minima, following the depiction by Clark.[15] The blue areas represent the unfolded state, while the three energy wells shown represent the native state and two non-native states (a misfolded monomer and an aggregated state, as labeled). The aggregated state becomes more stable as the protein concentration increases. Protein that enters the non-native states may have to cross significant energy barriers to reach the native state, making these steps potentially rate-limiting for folding.

ribonuclease folding.[4,15] Ribonuclease has eight cysteine residues that can form four disulfide bonds. There are 105 possible pairings of these cysteines as disulfides, only one of which allows native folding and function. Re-folding reduced and denatured ribonuclease under conditions that permit cysteine oxidation to disulfides, but not subsequent disulfide shuffling, produces a mixture in which many, or perhaps even most, of the 104 mispaired combinations of disulfides can be found; these all represent non-native states.

Non-native states would not be a problem for biological protein folding if they were transient. They become a problem only when substantial amounts of protein become trapped in non-native states on time scales comparable to other biological processes like protein synthesis, protein trafficking, cell division, etc. In the case of ribonuclease, even when refolding is performed in a redox buffer so that disulfide interchange is possible, non-native states with mispaired disulfides form quickly but resolve relatively slowly to the native state.[3,4,36] As a result, a substantial fraction of ribonuclease is non-native as it reacquires the native state over the course of more than an hour. As Anfinsen observed at the time,[37] this situation seems biologically untenable.

Ribonuclease becomes trapped in non-native states because these states are kinetically stabilized by the disulfide bonds between mispaired cysteine residues. Such covalent linkages, however, are not necessary to create long-lived non-native states. The same forces that stabilize protein native states can stabilize misfolded states, perhaps not as efficiently since

misfolded states should not be as well-packed as native states, but the energy barrier to "un-misfold" a non-native state could cause a substantial slowing of the overall rate of native folding and lead to a persistent population of non-native protein.[15]

Non-native states can also be stabilized by aggregation. In this case, parts of a protein that normally interact intramolecularly in the native state end up on the surface of a non-native state where they can drive self-association via intermolecular interactions.[15,18,38,39] The stability of aggregates formed in this way is concentration dependent. If the aggregation-prone non-native state is below a "critical concentration", then aggregation will not occur. Above that concentration, aggregation is possible.[18,39,40] In addition, aggregation will be faster as the concentration increases further.[39,41] Native folding is likely to be rate limited by the off-rate for protein molecules dissociating from aggregates, which could be very slow indeed (Figure 1). Even worse, the aggregated state can overtake the native state to become the thermodynamically most stable state, especially if the native state stability is compromised by an environmental factor (temperature, pH, etc.) or mutation, or if the protein is highly concentrated as could occur if it was, for example, preferentially partitioned into a phase separated domain.[25]

Answering the first question of the proteome folding problem—what alternative states compete with the native state on the protein folding energy landscape?—requires the ability to predict the properties of non-native states in the cellular environment (i.e., accounting for the pH, solute

composition, crowding, etc. inside the cell) from sequence information. This is like the protein folding problem, but non-native states of proteins are much more difficult to study than native states. Non-native states are heterogeneous, and/or amorphous, and/or insoluble; all these qualities interfere with the usual techniques that are used to study proteins. Nevertheless, substantial inroads have been made in the experimental characterization of non-native states, especially the class of fibrillar, and often disease-associated, protein aggregates known as amyloid.[42] Detailed structures of amyloid fibrils have now been obtained by solid-state NMR,[43,44] X-ray and electron diffraction of microcrystals,[45,46] and cryo-electron microscopy.[47–48] Less structured non-native states have been characterized as well (albeit at much lower resolution) using, for example, single-molecule methods,[49–52] small-angle X-ray scattering,[53,54] NMR,[55,56] and other methods.[57]

Real progress in determining the sequence-structure-energy relationships in the cellular milieu will of course require much more experimental data of the sort described in the preceding paragraph, but in addition this data will have to be wedded to computational methods so that the underlying physics can be understood. Unfortunately, it seems unlikely that algorithms like AlphaFold, which rely primarily on structural bioinformatics-based methods for native structure prediction, will be useful for this endeavor. The massive databases of non-native protein structures that would be needed do not exist, and probably cannot exist because so many non-native states are heterogeneous and dynamic. Computationally exploring non-native regions of the protein folding energy landscape is likely to be the province of computational molecular physics methods.[9] Such methods operate from first principles and therefore do not need auxiliary databases. As these methods become faster and more reliable, they will become better able to predict protein misfolding as well as protein folding and fill a central role in solving the proteome folding problem.

The Proteostasis Network

If the proteome folding problem were just about characterizing non-native states, it would simply be an extension of the protein folding problem. There would be nothing "proteomic" about such a problem because the structures and energetics of non-native states are still inherent in the primary sequence and the cellular environment, complex as it is. However, the non-native states of cellular proteins must compete for attention from the proteostasis network, and through this competition their folding energy landscapes communicate with each other. Misfolding by one protein means that the proteostasis network has less bandwidth to handle other proteins, which leads to the other

proteins misfolding as well; this effect has been compellingly demonstrated in multiple systems.[58–61] Thus, protein misfolding to non-native states and/or aggregates is truly a proteomic problem in a way that native folding is not.

As we mentioned in the previous section, Anfinsen recognized at the time of his original work on ribonuclease folding that ribonuclease's unassisted folding rate was too slow to be biologically reasonable. To address this, Anfinsen and others searched for, and found, a catalyst for ribonuclease folding in cell extracts: protein disulfide isomerase (PDI).[15,37,62–65] Although this result did not get as much attention as the finding that ribonuclease's native state was its thermodynamically most stable conformation, we believe it was equally profound. It showed that, although protein folding can spontaneously yield the native state, it nevertheless requires biological intervention to operate reliably enough in living, evolving organisms with the inherent challenges of the in vivo environment and the multitude of simultaneous folding/unfolding/misfolding events.

After the discovery of PDI it still took decades before there was a more-or-less complete accounting of the types of components that make up the proteostasis network. Now we recognize that perhaps the most prominent components of the proteostasis network are the chaperones, including the Hsp70,[66] Hsp60[67] and Hsp90 systems,[68] the small Hsps,[69] and the Hsp100s[70] (Hsp = heat shock protein). Also prominent are the ATP-driven proteases that degrade non-native proteins, including AAA + ATPases like Lon in prokaryotes[71] and larger systems like the ubiquitin–proteasome system[72] and the autophagy-lysosome pathway[73] in eukaryotes. Together, the chaperones and the degradation machinery, along with proto-chaperones like polyphosphate[74]; post-translational modifications like glycosylation and their associated machinery[75]; the effect of co-translational folding[76–79]; folding enzymes like the PDIs[80] and the *cis–trans* peptidyl-prolyl isomerases[81]; and the signaling pathways that regulate the concentrations of all these components[82–84] act to promote folding, discourage misfolding, disperse aggregates, or degrade non-native proteins to recycle their constituent amino acids.

Answering the second question of the proteome folding problem—how do proteostasis network components optimize protein folding and mitigate the consequences of folding failures?—requires a comprehensive understanding of how proteostasis network components act on their substrates. This includes the kinetics, thermodynamics, and sequence selectivity of substrate binding by proteostasis network components. Some sequence-based predictive algorithms for the energetics of such interactions have already been proposed for some of the more prominent chaperones, like Hsp70.[85–87] Fundamental under-

standing of chaperone action also requires an understanding of how substrate conformations are affected by being shuttled through the (frequently ATP-driven) conformational cycles. These cycles have been intensively studied for all the major ATP-driven chaperones and much is known about them; nevertheless, some issues remain unresolved. For example, for Hsp70 chaperones the structural basis for their selective promiscuity of substrate binding is a puzzle: they discriminate incompletely folded proteins from folded but still their clients represent a large fraction of the proteome.[88] How do Hsp70s alter the folding landscape of their substrates? The Hsp70 system includes the Hsp70 itself plus two co-chaperones, a nucleotide exchange factor and a so-called J-domain protein that activates the Hsp70 ATPase and often helps deliver clients.[66] In addition, they frequently hand off their substrates to downstream chaperones such as Hsp90s.[89,90] Details of these processes are lacking. Lastly, while many properties and functions of Hsp70s are conserved across cellular compartments, species, and kingdoms, the generality of their allosteric properties and functions remains unclear. Efforts to address these questions are being accelerated by the availability of increasingly powerful techniques based on, for example, cryo-electron microscopy and NMR.

The Proteome Folding Problem Can Only Be Solved with Systems Models for In Vivo Protein Folding

The proteome folding problem is not simply the next frontier in protein science. Failures of proteostasis have been implicated in many human diseases, including gain-of-toxicity (Alzheimer's disease, Parkinson's disease, the transthyretin amyloidoses, etc.) and loss-of-function diseases (cystic fibrosis, Gaucher's disease, Fabry's disease, etc.).[17,19,20,91] Thus, understanding proteostasis is important at both a fundamental and a practical level. At a fundamental level, understanding proteostasis informs us about the challenges involved in maintaining the thousands to tens of thousands of proteins in typical organismal proteomes in a functional state, which is a clear prerequisite to understanding how organisms survive and evolve in the face of environmental challenges. At a practical level, many proposed therapies for protein misfolding diseases involve promoting folding or inhibiting aggregation by improving protein folding energetics (through kinetic stabilization or pharmacologic chaperoning of proteins with destabilizing mutations)[92–95] or by manipulating the proteostasis network (through modulating the concentrations or activities of proteostasis network components).[17,19,91,96–101] Both endeavors require a systems-level appreciation of in vivo protein folding, but how can such an appreciation be achieved?

None of the proteostasis network components operates in isolation. Many of them bind to similar features on unfolded and misfolded proteins (for example, exposed hydrophobic residues) so they likely compete for substrates.[85–87,102–106] This competition can be especially important when components that favor folding (like chaperones) compete with the degradation machinery. Different chaperones can also collaborate with each other by binding to substrates and acting on them sequentially.[89,90,107,108] We believe that the simultaneous presence of the full array of proteostasis network components and any emergent properties that arise from their joint activities are too complex to be understood by intuition alone. They can only be understood by developing models that account for this complexity,[109–121] an example of which is FoldEco,[110] a model for proteostasis in *E. coli* (Figure 2).

These models consist of systems of coupled differential rate equations, one for each species in the system, in which the variables are species concentrations, and each term represents a process in model. The coefficients of the terms in the rate equations are the rate constants for the corresponding process. Although analytical solutions of simplified systems and steady states are often illuminating, these sets of rate equations are generally solved numerically.[110,111] This process requires that every parameter in the system be given a numerical value, including the initial concentrations of each species and the rate constants for every process, from the unimolecular conformational transitions involved in protein folding/misfolding, to the bimolecular chaperone-substrate recognition events, among many other processes (Figure 2). Thus, the primary challenge with developing a model for in vivo protein folding is parameterizing it. We have argued that literature estimates of these parameters based on in vitro experiments are suitable for processes like associations between chaperones and co-chaperones that do not directly involve the substrate.[110] The parameters for processes that involve individual members of the proteomes, however, should differ from protein to protein, in principle requiring thousands or tens of thousands of parameters for the model.

A computational model of the sort described in the preceding section that could accurately predict these parameters may one day solve this parameterization problem. In the meantime, we are fortunate that near-proteome-wide estimates of many biophysical properties have been made for *E. coli* proteins using either experimental or computational methods, including native state stability,[122,123] folding kinetics,[124] and solubility/aggregation[125]—which together define much of a protein's folding energy landscape—as well as chaperone efficacy for the Hsp70 and Hsp60 systems.[126] We recently used these estimates to create a coarse-grained representation of the entire
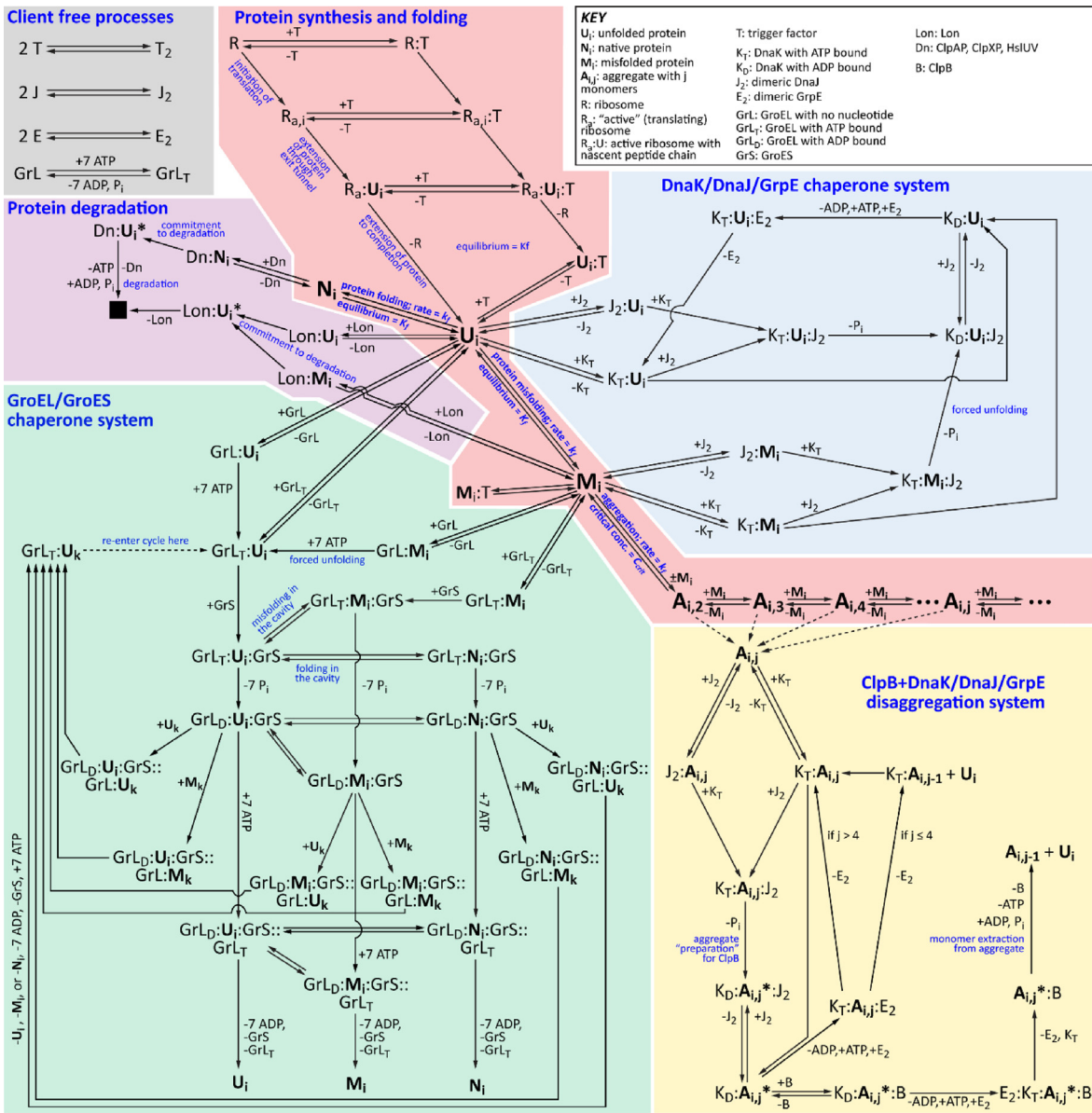
Figure 2. The FoldEco Model for Protein Folding in *E. coli*. The five subsystems in FoldEco are as follows: protein synthesis and folding (light red), the DnaK/DnaJ/GrpE chaperone system (light blue), the GroEL/GroES chaperone system (light green), the ClpB + DnaK/DnaJ/GrpE disaggregation system (light yellow), and protein degradation (light purple). Also included are several client-free processes, such as DnaJ dimerization (gray). The free unfolded (U_i), misfolded (M_i), native (N_i), and aggregated ($A_{i,j}$) states of the client protein are shown in large bold font. Complexes between proteostasis network components and the client protein are denoted by abbreviations for the species in the complex (see key in upper right) separated by colons. When the states of the cis and trans cavities of GroEL are both defined, they are separated by two colons. The subscripts "T" and "D" refer to the ATP- and ADP-bound states of chaperones. Figure and legend reproduced from Powers *et al.*[110]

proteome for use with a new version of FoldEco called FoldEcoSlim.[111] FoldEcoSlim is simpler than FoldEco because it does not represent the internal steps of chaperone cycles in as much detail as the original, but more complicated in that it accounts for the background proteome, the dynamic response of chaperone concentrations to protein folding stress, and cell growth.[111]

These models of proteome folding have already proven valuable for understanding, rationalizing, and/or predicting experimental results about proteostasis. For example, they have been used to analyze how organisms maintain (or fail to maintain) their proteomes under heat[119] or oxidative stress,[116] or as they age,[117] or as their proteomes evolve.[111,112] In addition, we have been able to use proteome folding models to understand the behaviors of a variety of specific proteins. For example, we used a very simple model to understand the effects of ligand binding on dihydrofolate

reductase expression in *E. coli* and α-galactosidase A trafficking in human cells.[121] We used FoldEco to understand the effect of the synthesis rate and co-translational folding on luciferase in *E. coli*.[110] We used FoldEcoSlim to relate the expression efficiencies of cellular retinoic acid binding protein 1 variants to their folding kinetics and thermodynamics. Finally, we used FoldEco in a reverse sense to understand protein folding biophysics based on in vivo protein folding outcomes, rather than understanding in vivo protein folding outcomes based on folding biophysics. In this case, we examined how the extent of aggregation of several proteins changed as the concentrations of proteostasis network components were varied and used this information to infer the proteins' folding and misfolding energetics.[127]

Outlook

Our efforts to solve the proteome folding problem are in their infancy. The models that have been developed for this purpose are still crude with respect to their representation of the processes involved in proteostasis and underdetermined with respect to their parameterizations. These problems will only become more acute as organisms more complicated than single-celled bacteria are addressed. Much more complicated representations of chaperone systems will be needed for proteome folding models of higher organisms since the numbers of chaperone genes scale linearly with genome size,[21] to say nothing of the sprawling ubiquitin–proteasome system[72] and the autophagy-lysosome pathway.[73] It is useful to remember, however, that our efforts to solve the protein folding problem were similarly overmatched in its early days. Perhaps the most important contributions that can be made at this point to solving the proteome folding problem are (1) continuing to elucidate chaperone mechanisms; (2) assessing protein folding biophysics on proteomic scales,[122,123,125] which will help refine computational molecular physics models for sequence-based estimates of protein folding, misfolding, and aggregation energetics; and (3) assessing the effects of chaperones and co-translational folding on folding outcomes on proteomic scales, which will help to define the interactions of proteostasis network components with their substrates. Once these pieces of the puzzle begin to fall into place, we expect rapid progress to be made on revealing the secrets of the proteome folding problem.

CRediT authorship contribution statement

Evan T. Powers: Conceptualization, Writing – original draft, Writing – review & editing. **Lila M. Gierasch:** Conceptualization, Writing – review & editing.

References

1. Kendrew, J.C., Bodo, G., Dintzis, H.M., Parrish, R.G., Wyckoff, H., Phillips, D.C., (1958). A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, **181**, 662–666.
2. Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R., Phillips, D.C., et al., (1960). Structure of myoglobin: a three-dimensional Fourier synthesis at 2 A. resolution. *Nature*, **185**, 422–427.
3. Anfinsen, C.B., Haber, E., Sela, M., White Jr., F.H., (1961). The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. USA*, **47**, 1309–1314.
4. Haber, E., Anfinsen, C.B., (1962). Side-chain interactions governing the pairing of half-cystine residues in ribonuclease. *J. Biol. Chem.*, **237**, 1839–1844.
5. Sela, M., White Jr., F.H., Anfinsen, C.B., (1957). Reductive cleavage of disulfide bridges in ribonuclease. *Science*, **125**, 691–692.
6. White Jr., F.H., (1961). Regeneration of native secondary and tertiary structures by air oxidation of reduced ribonuclease. *J. Biol. Chem.*, **236**, 1353–1360.
7. Dill, K.A., MacCallum, J.L., (2012). The protein-folding problem, 50 years on. *Science*, **338**, 1042–1046.
8. Dill, K.A., Ozkan, S.B., Shell, M.S., Weikl, T.R., (2008). The protein folding problem. *Annu. Rev. Biophys.*, **37**, 289–316.
9. Brini, E., Simmerling, C., Dill, K., (2020). Protein storytelling through physics. *Science*, **370**
10. Lupas, A.N., Pereira, J., Alva, V., Merino, F., Coles, M., Hartmann, M.D., (2021). The breakthrough in protein structure prediction. *Biochem. J.*, **478**, 1885–1890.

11. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al., (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*,.

12. Tunyasuvunakool, K., Adler, J., Wu, Z., Green, T., Zielinski, M., Zidek, A., et al., (2021). Highly accurate protein structure prediction for the human proteome. *Nature*,.

13. Anson, M.L., (1945). Protein denaturation and the properties of protein groups. *Adv. Protein Chem.*, **2**, 361–386.

14. Gershenson, A., Gierasch, L.M., Pastore, A., Radford, S. E., (2014). Energy landscapes of functional proteins are inherently risky. *Nature Chem. Biol.*, **10**, 884–891.

15. Clark, P.L., (2004). Protein folding in the cell: reshaping the folding funnel. *Trends Biochem. Sci.*, **29**, 527–534.

16. Braselmann, E., Chaney, J.L., Clark, P.L., (2013). Folding the proteome. *Trends Biochem. Sci.*, **38**, 337–344.

17. Balch, W.E., Morimoto, R.I., Dillin, A., Kelly, J.W., (2008). Adapting proteostasis for disease intervention. *Science*, **319**, 916–919.

18. Knowles, T.P., Vendruscolo, M., Dobson, C.M., (2014). The amyloid state and its association with protein misfolding diseases. *Nature Rev. Mol. Cell. Biol.*, **15**, 384–396.

19. Powers, E.T., Morimoto, R.I., Dillin, A., Kelly, J.W., Balch, W.E., (2009). Biological and chemical approaches to diseases of proteostasis deficiency. *Annu. Rev. Biochem.*, **78**, 959–991.

20. Winklhofer, K.F., Tatzelt, J., Haass, C., (2008). The two faces of protein misfolding: gain- and loss-of-function in neurodegenerative diseases. *EMBO J.*, **27**, 336–349.

21. Powers, E.T., Balch, W.E., (2013). Diversity in the origins of proteostasis networks–a driver for protein function in evolution. *Nature Rev. Mol. Cell Biol.*, **14**, 237–248.

22. Kim, Y.E., Hipp, M.S., Bracher, A., Hayer-Hartl, M., Hartl, F.U., (2013). Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.*, **82**, 323–355.

23. Hartl, F.U., Hayer-Hartl, M., (2002). Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science*, **295**, 1852–1858.

24. Frydman, J., (2001). Folding of newly translated proteins in vivo: the role of molecular chaperones. *Annu. Rev. Biochem.*, **70**, 603–647.

25. Babinchak, W.M., Surewicz, W.K., (2020). Liquid-liquid phase separation and its mechanistic role in pathological protein aggregation. *J. Mol. Biol.*, **432**, 1910–1925.

26. Edun, D.N., Flanagan, M.R., Serrano, A.L., (2021). Does liquid-liquid phase separation drive peptide folding? *Chem. Sci.*, **12**, 2474–2479.

27. Choi, J.M., Holehouse, A.S., Pappu, R.V., (2020). Physical principles underlying the complex biology of intracellular phase transitions. *Annu. Rev. Biophys.*, **49**, 107–133.

28. Banani, S.F., Lee, H.O., Hyman, A.A., Rosen, M.K., (2017). Biomolecular condensates: organizers of cellular biochemistry. *Nature Rev. Mol. Cell Biol.*, **18**, 285–298.

29. Bentley, E.P., Frey, B.B., Deniz, A.A., (2019). Physical chemistry of cellular liquid-phase separation. *Chemistry*, **25**, 5600–5610.

30. Auton, M., Rosgen, J., Sinev, M., Holthauzen, L.M., Bolen, D.W., (2011). Osmolyte effects on protein stability and solubility: a balancing act between backbone and side-chains. *Biophys. Chem.*, **159**, 90–99.

31. Davis-Searles, P.R., Saunders, A.J., Erie, D.A., Winzor, D.J., Pielak, G.J., (2001). Interpreting the effects of small uncharged solutes on protein-folding equilibria. *Annu. Rev. Biophys. Biomol. Struct.*, **30**, 271–306.

32. Ferreon, A.C., Moosa, M.M., Gambin, Y., Deniz, A.A., (2012). Counteracting chemical chaperone effects on the single-molecule alpha-synuclein structural landscape. *Proc. Natl. Acad. Sci. USA*, **109**, 17826–17831.

33. Zhou, H.X., Rivas, G., Minton, A.P., (2008). Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.*, **37**, 375–397.

34. Miklos, A.C., Sarkar, M., Wang, Y., Pielak, G.J., (2011). Protein crowding tunes protein stability. *J. Am. Chem. Soc.*, **133**, 7116–7120.

35. Schlesinger, A.P., Wang, Y., Tadeo, X., Millet, O., Pielak, G.J., (2011). Macromolecular crowding fails to fold a globular protein in cells. *J. Am. Chem. Soc.*, **133**, 8082–8085.

36. Epstein, C.J., Anfinsen, C.B., Young, D.M., Goldberger, R.F., (1962). A study of factors influencing rate extent of enzymic reactivation during reoxidation of reduced ribonuclease. *Arch. Biochem. Biophys.*,, 223.

37. Goldberger, R.F., Epstein, C.J., Anfinsen, C.B., (1963). Acceleration of reactivation of reduced bovine pancreatic ribonuclease by a microsomal system from rat liver. *J. Biol. Chem.*, **238**, 628–635.

38. Kelly, J.W., (1996). Alternative conformations of amyloidogenic proteins govern their behavior. *Curr. Opin. Struct. Biol.*, **6**, 11–17.

39. Chi, E.Y., Krishnan, S., Randolph, T.W., Carpenter, J.F., (2003). Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharm. Res.*, **20**, 1325–1336.

40. Pappu, R.V., Wang, X., Vitalis, A., Crick, S.L., (2008). A polymer physics perspective on driving forces and mechanisms for protein aggregation. *Arch. Biochem. Biophys.*, **469**, 132–141.

41. Michaels, T.C.T., Saric, A., Habchi, J., Chia, S., Meisl, G., Vendruscolo, M., et al., (2018). Chemical kinetics for bridging molecular mechanisms and macroscopic measurements of amyloid fibril formation. *Annu. Rev. Phys. Chem.*, **69**, 273–298.

42. Eisenberg, D.S., Sawaya, M.R., (2017). Structural studies of amyloid proteins at the molecular level. *Annu. Rev. Biochem.*, **86**, 69–95.

43. Lu, J.X., Qiang, W., Yau, W.M., Schwieters, C.D., Meredith, S.C., Tycko, R., (2013). Molecular structure of beta-amyloid fibrils in Alzheimer's disease brain tissue. *Cell*, **154**, 1257–1268.

44. Walti, M.A., Ravotti, F., Arai, H., Glabe, C.G., Wall, J.S., Bockmann, A., et al., (2016). Atomic-resolution structure of a disease-relevant Abeta(1–42) amyloid fibril. *Proc. Natl. Acad. Sci. USA*, **113**, E4976–E4984.

45. Rodriguez, J.A., Ivanova, M.I., Sawaya, M.R., Cascio, D., Reyes, F.E., Shi, D., et al., (2015). Structure of the toxic core of alpha-synuclein from invisible crystals. *Nature*, **525**, 486–490.

46. Sawaya, M.R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S.A., Apostol, M.I., et al., (2007). Atomic

structures of amyloid cross-beta spines reveal varied steric zippers. *Nature*, **447**, 453–457.

47. Fitzpatrick, A.W.P., Falcon, B., He, S., Murzin, A.G., Murshudov, G., Garringer, H.J., et al., (2017). Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature*, **547**, 185–190.

48. Schmidt, M., Wiese, S., Adak, V., Engler, J., Agarwal, S., Fritz, G., et al., (2019). Cryo-EM structure of a transthyretin-derived amyloid fibril from a patient with hereditary ATTR amyloidosis. *Nature Commun.*, **10**, 5008.

49. Banerjee, P.R., Deniz, A.A., (2014). Shedding light on protein folding landscapes by single-molecule fluorescence. *Chem. Soc. Rev.*, **43**, 1172–1188.

50. Borgia, M.B., Borgia, A., Best, R.B., Steward, A., Nettels, D., Wunderlich, B., et al., (2011). Single-molecule fluorescence reveals sequence-specific misfolding in multidomain proteins. *Nature*, **474**, 662–665.

51. Yu, H., Dee, D.R., Liu, X., Brigley, A.M., Sosova, I., Woodside, M.T., (2015). Protein misfolding occurs by slow diffusion across multiple barriers in a rough energy landscape. *Proc. Natl. Acad. Sci. USA*, **112**, 8308–8313.

52. Rhoades, E., Gussakovsky, E., Haran, G., (2003). Watching proteins fold one molecule at a time. *Proc. Natl. Acad. Sci. USA*, **100**, 3197–3202.

53. Bowman, M.A., Riback, J.A., Rodriguez, A., Guo, H., Li, J., Sosnick, T.R., et al., (2020). Properties of protein unfolded states suggest broad selection for expanded conformational ensembles. *Proc. Natl. Acad. Sci. USA*, **117**, 23356–23364.

54. Riback, J.A., Bowman, M.A., Zmyslowski, A.M., Knoverek, C.R., Jumper, J.M., Hinshaw, J.R., et al., (2017). Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science*, **358**, 238–241.

55. Alderson, T.R., Kay, L.E., (2020). Unveiling invisible protein states with NMR spectroscopy. *Curr. Opin. Struct. Biol.*, **60**, 39–49.

56. Whitley, M.J., Xi, Z., Bartko, J.C., Jensen, M.R., Blackledge, M., Gronenborn, A.M., (2017). A combined NMR and SAXS analysis of the partially folded cataract-associated V75D gammaD-Crystallin. *Biophys. J.*, **112**, 1135–1146.

57. Srivastava, K.R., Lapidus, L.J., (2017). Prion protein dynamics before aggregation. *Proc. Natl. Acad. Sci. USA*, **114**, 3572–3577.

58. Gidalevitz, T., Ben-Zvi, A., Ho, K.H., Brignull, H.R., Morimoto, R.I., (2006). Progressive disruption of cellular protein folding in models of polyglutamine diseases. *Science*, **311**, 1471–1474.

59. Olzscha, H., Schermann, S.M., Woerner, A.C., Pinkert, S., Hecht, M.H., Tartaglia, G.G., et al., (2011). Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell*, **144**, 67–78.

60. Park, S.H., Kukushkin, Y., Gupta, R., Chen, T., Konagai, A., Hipp, M.S., et al., (2013). PolyQ proteins interfere with nuclear degradation of cytosolic proteins by sequestering the Sis1p chaperone. *Cell*, **154**, 134–145.

61. Yu, A., Shibata, Y., Shah, B., Calamini, B., Lo, D.C., Morimoto, R.I., (2014). Protein aggregation can inhibit clathrin-mediated endocytosis by chaperone competition. *Proc. Natl. Acad. Sci. USA*, **111**, E1481–E1490.

62. De Lorenzo, F., Goldberger, R.F., Steers Jr., E., Givol, D., Anfinsen, B., (1966). Purification and properties of an enzyme from beef liver which catalyzes sulfhydryl-disulfide interchange in proteins. *J. Biol. Chem.*, **241**, 1562–1567.

63. Fuchs, S., De Lorenzo, F., Anfinsen, C.B., (1967). Studies on the mechanism of the enzymic catalysis of disulfide interchange in proteins. *J. Biol. Chem.*, **242**, 398–402.

64. Venetianer, P., Straub, F.B., (1963). The enzymic reactivation of reduced ribonuclease. *Biochim. Biophys. Acta*, **67**, 166–168.

65. Venetianer, P., Straub, F.B., (1964). The mechanism of action of the ribonuclease-reactivating enzyme. *Biochim. Biophys. Acta*, **89**, 189–190.

66. Mayer, M.P., Gierasch, L.M., (2019). Recent advances in the structural and mechanistic aspects of Hsp70 molecular chaperones. *J. Biol. Chem.*, **294**, 2085–2097.

67. Horwich, A.L., Fenton, W.A., (2009). Chaperonin-mediated protein folding: using a central cavity to kinetically assist polypeptide chain folding. *Q. Rev. Biophys.*, **42**, 83–116.

68. Schopf, F.H., Biebl, M.M., Buchner, J., (2017). The HSP90 chaperone machinery. *Nature Rev. Mol. Cell Biol.*, **18**, 345–360.

69. Carra, S., Alberti, S., Arrigo, P.A., Benesch, J.L., Benjamin, I.J., Boelens, W., et al., (2017). The growing world of small heat shock proteins: from structure to functions. *Cell Stress Chaperones*, **22**, 601–611.

70. Doyle, S.M., Genest, O., Wickner, S., (2013). Protein rescue from aggregates by powerful molecular chaperone machines. *Nature Rev. Mol. Cell Bio*, **14**, 617–629.

71. Mahmoud, S.A., Chien, P., (2018). Regulated proteolysis in bacteria. *Annu. Rev. Biochem.*, **87**, 677–696.

72. Amm, I., Sommer, T., Wolf, D.H., (2014). Protein quality control and elimination of protein waste: the role of the ubiquitin-proteasome system. *Biochim. Biophys. Acta*, **1843**, 182–196.

73. Bento, C.F., Renna, M., Ghislat, G., Puri, C., Ashkenazi, A., Vicinanza, M., et al., (2016). Mammalian autophagy: how does it work? *Annu. Rev. Biochem.*, **85**, 685–713.

74. Gray, M.J., Wholey, W.Y., Wagner, N.O., Cremers, C.M., Mueller-Schickert, A., Hock, N.T., et al., (2014). Polyphosphate is a primordial chaperone. *Mol. Cell*, **53**, 689–699.

75. Hebert, D.N., Lamriben, L., Powers, E.T., Kelly, J.W., (2014). The intrinsic and extrinsic effects of N-linked glycans on glycoproteostasis. *Nature Chem. Biol.*, **10**, 902–910.

76. Evans, M.S., Sander, I.M., Clark, P.L., (2008). Cotranslational folding promotes beta-helix formation and avoids aggregation in vivo. *J. Mol. Biol.*, **383**, 683–692.

77. Nilsson, O.B., Nickson, A.A., Hollins, J.J., Wickles, S., Steward, A., Beckmann, R., et al., (2017). Cotranslational folding of spectrin domains via partially structured states. *Nature Struct. Mol. Biol.*, **24**, 221–225.

78. Stein, K.C., Frydman, J., (2019). The stop-and-go traffic regulating protein biogenesis: how translation kinetics controls proteostasis. *J. Biol. Chem.*, **294**, 2076–2084.

79. Walsh, I.M., Bowman, M.A., Soto Santarriaga, I.F., Rodriguez, A., Clark, P.L., (2020). Synonymous codon substitutions perturb cotranslational protein folding in vivo and impair cell fitness. *Proc. Natl. Acad. Sci. USA*, **117**, 3528–3534.

80. Matsusaki, M., Kanemura, S., Kinoshita, M., Lee, Y.H., Inaba, K., Okumura, M., (2020). The protein disulfide

isomerase family: from proteostasis to pathogenesis. *Biochim. Biophys. Acta Gen Subj.*, **1864**, 129338

81. Gothel, S.F., Marahiel, M.A., (1999). Peptidyl-prolyl cis-trans isomerases, a superfamily of ubiquitous folding catalysts. *Cell Mol. Life Sci.*, **55**, 423–436.

82. Plate, L., Wiseman, R.L., (2017). Regulating secretory proteostasis through the unfolded protein response: from function to therapy. *Trends Cell Biol.*, **27**, 722–737.

83. Akerfelt, M., Morimoto, R.I., Sistonen, L., (2010). Heat shock factors: integrators of cell stress, development and lifespan. *Nature Rev. Mol. Cell Biol.*, **11**, 545–555.

84. Pakos-Zebrucka, K., Koryga, I., Mnich, K., Ljujic, M., Samali, A., Gorman, A.M., (2016). The integrated stress response. *EMBO Rep.*, **17**, 1374–1395.

85. Rudiger, S., Germeroth, L., Schneider-Mergener, J., Bukau, B., (1997). Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. *EMBO J.*, **16**, 1501–1507.

86. Rudiger, S., Schneider-Mergener, J., Bukau, B., (2001). Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone. *EMBO J.*, **20**, 1042–1050.

87. Van Durme, J., Maurer-Stroh, S., Gallardo, R., Wilkinson, H., Rousseau, F., Schymkowitz, J., (2009). Accurate prediction of DnaK-peptide binding via homology modelling and experimental data. *PLoS Comput. Biol.*, **5**, e1000475

88. Calloni, G., Chen, T., Schermann, S.M., Chang, H.C., Genevaux, P., Agostini, F., et al., (2012). DnaK functions as a central hub in the E. coli chaperone network. *Cell Rep.*, **1**, 251–264.

89. Kirschke, E., Goswami, D., Southworth, D., Griffin, P.R., Agard, D.A., (2014). Glucocorticoid receptor function regulated by coordinated action of the Hsp90 and Hsp70 chaperone cycles. *Cell*, **157**, 1685–1697.

90. Wegele, H., Wandinger, S.K., Schmid, A.B., Reinstein, J., Buchner, J., (2006). Substrate transfer from the chaperone Hsp70 to Hsp90. *J. Mol. Biol.*, **356**, 802–811.

91. Hipp, M.S., Park, S.H., Hartl, F.U., (2014). Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends Cell Biol.*, **24**, 506–514.

92. Germain, D.P., Fan, J.Q., (2009). Pharmacological chaperone therapy by active-site-specific chaperones in Fabry disease: in vitro and preclinical studies. *Int. J. Clin. Pharmacol. Ther.*, **47** (Suppl 1), S111–S117.

93. Parenti, G., Andria, G., Valenzano, K.J., (2015). Pharmacological chaperone therapy: preclinical development, clinical translation, and prospects for the treatment of lysosomal storage disorders. *Mol. Ther.*, **23**, 1138–1148.

94. Sawkar, A.R., Cheng, W.C., Beutler, E., Wong, C.H., Balch, W.E., Kelly, J.W., (2002). Chemical chaperones increase the cellular activity of N370S beta -glucosidase: a therapeutic strategy for Gaucher disease. *Proc. Natl. Acad. Sci. USA*, **99**, 15428–15433.

95. Powers, E.T., Kelly, J.W., (2021). From uncovering the mechanism of transthyretin aggregation to the drug tafamidis for ameliorating neurodegeneration and cardiomyopathy. In: Smith, R.A., Kaspar, B.K., Svendsen, C.N. (Eds.), *Neurotherapeutics in the Era of Translational Medicine,*. Academic Press, London, United Kingdom, pp. 65–103.

96. Eisele, Y.S., Monteiro, C., Fearns, C., Encalada, S.E., Wiseman, R.L., Powers, E.T., et al., (2015). Targeting protein aggregation for the treatment of degenerative diseases. *Nature Rev. Drug Discov.*, **14**, 759–780.

97. Calamini, B., Morimoto, R.I., (2012). Protein homeostasis as a therapeutic target for diseases of protein conformation. *Curr. Top. Med. Chem.*, **12**, 2623–2640.

98. Calamini, B., Silva, M.C., Madoux, F., Hutt, D.M., Khanna, S., Chalfant, M.A., et al., (2012). Small-molecule proteostasis regulators for protein conformational diseases. *Nature Chem. Biol.*, **8**, 185–196.

99. Grandjean, J.M.D., Wiseman, R.L., (2020). Small molecule strategies to harness the unfolded protein response: where do we go from here? *J. Biol. Chem.*, **295**, 15692–15711.

100. Sidrauski, C., McGeachy, A.M., Ingolia, N.T., Walter, P., (2015). The small molecule ISRIB reverses the effects of eIF2alpha phosphorylation on translation and stress granule assembly. *Elife*, **4**

101. Wang, A.M., Miyata, Y., Klinedinst, S., Peng, H.M., Chua, J.P., Komiyama, T., et al., (2013). Activation of Hsp70 reduces neurotoxicity by promoting polyglutamine protein degradation. *Nature Chem. Biol.*, **9**, 112–118.

102. Aoki, K., Motojima, F., Taguchi, H., Yomo, T., Yoshida, M., (2000). GroEL binds artificial proteins with random sequences. *J. Biol. Chem.*, **275**, 13755–13758.

103. Clerico, E.M., Tilitsky, J.M., Meng, W., Gierasch, L.M., (2015). How hsp70 molecular machines interact with their substrates to mediate diverse physiological functions. *J. Mol. Biol.*, **427**, 1575–1588.

104. Landry, S.J., Gierasch, L.M., (1991). The chaperonin GroEL binds a polypeptide in an alpha-helical conformation. *Biochemistry*, **30**, 7359–7362.

105. Wang, Z., Feng, H., Landry, S.J., Maxwell, J., Gierasch, L.M., (1999). Basis of substrate binding by the chaperonin GroEL. *Biochemistry*, **38**, 12537–12546.

106. Tartaglia, G.G., Dobson, C.M., Hartl, F.U., Vendruscolo, M., (2010). Physicochemical determinants of chaperone requirements. *J. Mol. Biol.*, **400**, 579–588.

107. Langer, T., Lu, C., Echols, H., Flanagan, J., Hayer, M.K., Hartl, F.U., (1992). Successive action of DnaK, DnaJ and GroEL along the pathway of chaperone-mediated protein folding. *Nature*, **356**, 683–689.

108. Pratt, W.B., Toft, D.O., (2003). Regulation of signaling protein function and trafficking by the hsp90/hsp70-based chaperone machinery. *Exp. Biol. Med. (Maywood)*, **228**, 111–133.

109. Wiseman, R.L., Powers, E.T., Buxbaum, J.N., Kelly, J.W., Balch, W.E., (2007). An adaptable standard for protein export from the endoplasmic reticulum. *Cell*, **131**, 809–821.

110. Powers, E.T., Powers, D.L., Gierasch, L.M., (2012). FoldEco: a model for proteostasis in E. coli. *Cell Rep.*, **1**, 265–276.

111. Pobre, K.F.R., Powers, D.L., Ghosh, K., Gierasch, L.M., Powers, E.T., (2019). Kinetic versus thermodynamic control of mutational effects on protein homeostasis: a perspective from computational modeling and experiment. *Protein Sci.*, **28**, 1324–1339.

112. Bershtein, S., Mu, W., Serohijos, A.W., Zhou, J., Shakhnovich, E.I., (2013). Protein quality control acts on folding intermediates to shape the effects of mutations on organismal fitness. *Mol. Cell*, **49**, 133–144.

113. Dickson, A., Brooks 3rd., C.L., (2013). Quantifying chaperone-mediated transitions in the proteostasis network of E. coli. *PLoS Comput Biol.*, **9**, e1003324.

114. Royle, K., Kontoravdi, C., (2013). A systems biology approach to optimising hosts for industrial protein production. *Biotechnol. Letters*, **35**, 1961–1969.

115. Hall, D., (2020). On the nature of the optimal form of the holdase-type chaperone stress response. *FEBS Letters*, **594**, 43–66.

116. Santra, M., Dill, K.A., de Graff, A.M.R., (2018). How do chaperones protect a cell's proteins from oxidative damage? *Cell Syst.*, **6** 743-51 e3.

117. Santra, M., Dill, K.A., de Graff, A.M.R., (2019). Proteostasis collapse is a driver of cell aging and death. *Proc. Natl. Acad. Sci. USA*, **116**, 22173–22178.

118. de Graff, A.M., Mosedale, D.E., Sharp, T., Dill, K.A., Grainger, D.J., (2020). Proteostasis is adaptive: balancing chaperone holdases against foldases. *PLoS Comput. Biol.*, **16**, e1008460

119. Chen, K., Gao, Y., Mih, N., O'Brien, E.J., Yang, L., Palsson, B.O., (2017). Thermosensitivity of growth is determined by chaperone-mediated proteome reallocation. *Proc. Natl. Acad. Sci. USA*, **114**, 11548–11553.

120. Santra, M., Farrell, D.W., Dill, K.A., (2017). Bacterial proteostasis balances energy and chaperone utilization efficiently. *Proc. Natl. Acad. Sci. USA*, **114**, E2654–E2661.

121. Hingorani, K.S., Metcalf, M.C., Deming, D.T., Garman, S.C., Powers, E.T., Gierasch, L.M., (2017). Ligand-promoted protein folding by biased kinetic partitioning. *Nature Chem. Biol.*, **13**, 369–371.

122. Leuenberger, P., Ganscha, S., Kahraman, A., Cappelletti, V., Boersema, P.J., von Mering, C., et al., (2017). Cell-wide analysis of protein thermal unfolding reveals determinants of thermostability. *Science*, **355**

123. Mateus, A., Bobonis, J., Kurzawa, N., Stein, F., Helm, D., Hevler, J., et al., (2018). Thermal proteome profiling in bacteria: probing protein state in vivo. *Mol. Syst. Biol.*, **14**, e8242

124. Zou, T., Williams, N., Ozkan, S.B., Ghosh, K., (2014). Proteome folding kinetics is limited by protein halflife. *PLoS One*, **9**, e112701

125. Niwa, T., Ying, B.W., Saito, K., Jin, W., Takada, S., Ueda, T., et al., (2009). Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins. *Proc. Natl. Acad. Sci. USA*, **106**, 4201–4206.

126. Niwa, T., Kanamori, T., Ueda, T., Taguchi, H., (2012). Global analysis of chaperone effects using a reconstituted cell-free translation system. *Proc. Natl. Acad. Sci. USA*, **109**, 8937–8942.

127. Cho, Y., Zhang, X., Pobre, K.F., Liu, Y., Powers, D.L., Kelly, J.W., et al., (2015). Individual and collective contributions of chaperoning and degradation to protein homeostasis in E. coli. *Cell Rep.*, **11**, 321–333.