ARTICLE

Recent advances in bioanalytical methods to measure proteome stability in cells

Shouxiang Zhang,[a] David W. Greening,[b,c,*] and Yuning Hong [a,*]

Proteome stability constitutes an essential aspect of protein homeostasis (proteostasis). Proteostasis networks maintain proteins and their interactors in a defined conformation for their activity, localisation, and function. However, endogenous or exogenous stressors can perturb proteostasis integrity and deplete folding capacity, generating destabilized folding intermediates and deleterious aggregated species. Over the years, protein unfolding, misfolding and aggregation have been reported to be associated with aging and many diseases such as neurodegenerative diseases, diabetes, cardiac disease and toxicity, and cancers. Therefore, monitoring proteome stability is central to understanding underlying biological processes and mechanisms of disease progression. Herein, we review the recent bioanalytical methods to measure protein stability in cells on a proteome-wide scale.

Introduction

Proteostasis refers to the fine balance of protein biogenesis, folding, trafficking, and degradation, and is thus essential to preserve normal cellular metabolism and integrity of the proteome.[1] In these biological processes, protein stability plays an important role in that proteins require a stable three-dimensional structure for normal functions and interactions within and between cells. However, endogenous or exogenous stimuli can deplete protein folding capacity in cells, causing proteostasis imbalance and dysfunction.[2] Increasingly, more diseases have been shown attributed to protein unfolding and misfolding, ranging from neurodegeneration (e.g. Huntington's, Alzheimer's, and Parkinson's diseases) to cystic fibrosis, obesity, aging, cardiac disease and toxicity, and cancer development and progression.[3-8] Characterizing protein unfolding, misfolding and aggregation would contribute to understandings of cellular adaptive responses, regulatory mechanisms and interconnected pathways, as well as provide insights on future therapeutic development as well as pharmacologic intervention.[9]

Assessing proteome stability (i.e. protein stability in cells on a proteome-wide scale) can be challenging. First, some structural biology methods (e.g. crystallography and cryogenic electron microscopy)[10], although powerful in determining protein structures *in vitro*, require preparation of homogenous purified samples and are thus not suitable for studying protein stability *in situ*. For one thing, destabilized proteins can be highly heterogeneous and dynamic.[11] For another, protein stability is extensively influenced by the cellular environment and molecular interactions, thus studying protein stability *in situ* is beneficial for biological interpretation. Second, in-cell NMR spectroscopy[12] has achieved great success in monitoring conformational changes within the cellular milieu, but requires isotopic labelling of specific proteins. Third, fluorescence protein-tag based methods enable accurate quantitation of folding stability and dynamics of a protein of interest in cells,[13, 14] however, is limited by identification of protein structural change directly on a proteome-wide scale. Therefore, developing new bioanalytical methods to assess proteome stability *in situ* to further understand cellular physiology are highly desirable. In this minireview, we provide an overview of methods that have been reported over the past three years for measuring cellular or subcellular proteome stability.

Proteome stability can be indirectly reflected by behaviours of biosensors, or directly measured in a high-throughput manner (Table 1). Instead of investigating individual proteins, biosensors are applicable to a whole proteome scale, capable of reporting proteome stress or represent the folding status of the proteome. We will mainly describe a nucleic acid-based sensor (Example 1) for quantifying unfolded protein response, and a novel protein-based sensor, i.e. Halo system (Example 2), for monitoring proteostatic stress. Further, protein stability can be converted to other parameters that can be measured and profiled. In this regard, we present proteome solubility profiling (Example 3-5), proteome oxidation profiling (Example 6), limited proteolysis – mass spectrometry (Example 7), and cysteine exposure profiling (Example 8). For each example we will describe both measurement and analysis methods and the biological context and notable limitations of using these methods.

Biosensors

Biosensors here are referred to engineered biological elements for measuring, visualizing and quantifying cellular or subcellular

Table **1** Summary of all methods.

Example	Description	Principle	Method variations	Applications	Limitations
1	A gene signal amplifier platform for quantifying URP gene expression[15]	Proteostasis imbalance triggers URP gene expression	Involving other promoter regions	Monitoring UPR dynamics	Non-high-throughput; not always represent proteome foldedness
2	A fluorogenic protein-based sensor for detecting proteostatic stress[16-18]	Proteostasis imbalance causes AgHalo unfolding or aggregation	Other metastable protein sensors	Detecting trace proteostasis damage and evaluating drug safety	
3	Characterization of proteome thermostability in 13 species[19]	Proteins precipitate at high temperatures	Quantitative measurement of abundance of soluble and insoluble proteins in different organelles	Protein-ligand interaction, drug design, biomarker discovery, etc.	High workload (i.e. requiring preparation of enough sample points for plotting protein denaturation curves); difficult analyses with proteins not following a two-state folding model
4	ATP Interactome in cells[20]	Ligands can interact with proteins and alter their protein melting curves			
5	Profile proteome solubility in response to proteostatic stress[21]	Destabilized proteins transit from supernatant to cell pellet			
6	Global analysis of methionine oxidation using SPROX[22]	Unfolded regions are more prone to oxidation	Other protein footprinting methods, e.g. HRF[23]		Tryptic peptides need to contain methionine residues
7	Cell-wide analysis of protein thermal unfolding using LiP-MS[24]	Unfolded regions are more prone to proteolysis	Other proteolysis-based methods, e.g. DARTS[25], PP[26]		Proteins cannot be resistant to proteolysis
8	Fluorescent tools for measuring unfolded protein load in cells[27-29]	Unfolded regions expose more free cysteines	Incorporation of other ESFs	Fundamental research and disease diagnosis	Cysteines are rare and may not be contained in some proteins/regions

proteome stability. The development and application of biosensors require several criteria. First, they should be bio-orthogonal, namely the introduction and expression of the biosensor should have minimum effects on the experimental system. Specifically, the biosensor should not change the proteostasis capacity, nor involve any interactions that introduce or cause bias. Second, they should specifically respond to proteostasis perturbation by exhibiting altered functionalities that can be easily quantified or visualized, with appropriate levels of detection sensitivity. Third, the biosensor should reflect proteostasis perturbation or the folding status of the proteome.

The following will introduce nucleic acid-based and protein-based sensors, mainly focusing on the development of a novel gene signal amplifier for dynamically reporting unfolded protein response (Example 1), and a fluorogenic metastable protein system for assessing proteostatic stress (Example 2).

Nucleic acid-based sensors

Nucleic acid-based sensors are pathway-focused reporters that can perceive activation or deactivation of promoters involved in the proteostasis network resulting from proteome destabilization. In the design of a typical nucleic acid-based sensor, a reporter gene, such as fluorescent protein (e.g. eGFP) or enzyme (e.g. xylanase[30]), is placed under control of a promoter whose downstream gene is involved in the proteostasis network. At a cellular level, the promoter, serving as a sensor will be activated upon cell stress and result in expression of the reporter protein.

One of the determinants for the specificity, sensitivity and precision of a nucleic acid-based sensor is the selection of the promoter, whose activation or deactivation accounts for the response from the corresponding pathway. Commonly used promoter regions include those in the central organization of proteostasis network, such as molecular chaperones[31] and heat shock factors[32], unfolded protein response (UPR) pathways,[33] and protein degradation pathways[34, 35], universally accepted as critical regulators of the proteostasis network.

UPR is a cellular stress response to accumulation of unfolded/misfolded proteins in the endoplasmic reticulum (ER).[15] UPR functions through activation of three transmembrane ER-resident stress sensors: inositol requiring enzyme 1 (IRE1), protein kinase RNA (PKR)-like ER kinase (PERK) and activating transcription factor 6 (ATF6). Such sensors mediate a complex series of signalling cascades aiming to ameliorate ER stress and restore proteostasis by upregulation of UPR genes involved in ER protein quality control and ER-associated degradation (ERAD), and inhibition of global protein translation, etc.[36] Origel Marmolejo et al. (Example 1) developed a gene signal amplifier platform for quantifying UPR. In the design of this biosensor, the reporter gene GFP is placed under the same transcriptional and translational regulation as three UPR target genes (Fig. 1A), namely DNAJB9 (ERdj4), EIF4EBP1 (EIF4) and HSPA5 (BIP), respectively. This is achieved by integration of a cassette containing an internal ribosome entry site (IRES) and the GFP gene (IRES_GFP) downstream of

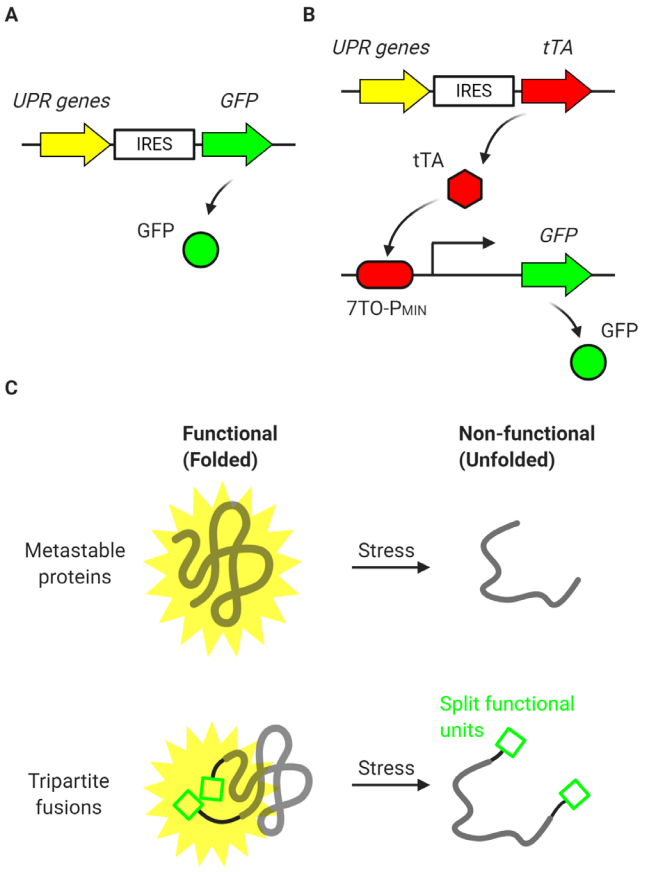

Fig. 1 Schematic illustration of (A, B) nucleic acid-based sensors and (B) protein-based sensors, including metastable proteins or recombinant tripartite fusions.

ERdj4, EIF4 and BIP using CRISPR–Cas9 via homologous direct-repair[37]. This resulted in transcription of a polycistronic mRNA encoding both the UPR target gene and GFP. These three engineered cell lines (i.e. BIP-GFP, ERdj4-GFP and EIF4-GFP) were validated to increase GFP expression following stimulation of pathway-specific transcriptional factors and canonical UPR inducers (e.g. tunicamycin that inhibits *N*-glycosylation of proteins). Furthermore, the same group integrated orthogonal regulatory elements (*vide infra* Perspective, Fig. 1B) into this sensor for enhancing transcriptional and post-translational control of the reporter, to improve signal amplification and dynamic resolution, and achieve sensitive quantitation of UPR genes.

Protein-based sensors

Protein-based sensors are designed to report the deficiency of cellular folding capacity, which can be classified as engineered metastable functional proteins and recombinant tripartite fusions (Fig. 1C), where proteome destabilization can be reflected on the readout, for example, using fluorescence. Gupta et al. reported a metastable firefly luciferase mutant developed by rational design based on X-ray crystal structure,[38] in which proteostasis impairment resulted in reduced luciferase solubility and luminescence activity (Fig. 1C upper). For recombinant tripartite fusions, the strategy is to flank the metastable protein with two half-functional modules. These two modules, one on the N-terminus and the other on the C-

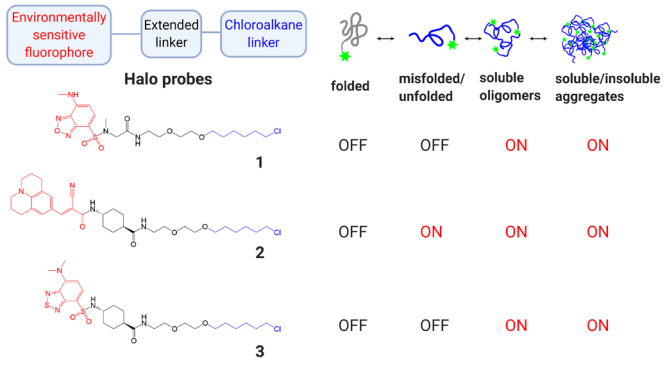

Fig.2 Schematic illustration of HaloTag multicolor fluorogenic biosensors.

terminus, come into close proximity and display functionality only when the metastable protein is in the folded state (Fig. 1C lower). The combination of the two-half modules can include fluorescence resonance energy transfer (FRET) pairs,[39] split domains of an antibiotic resistance markers,[40] for example.

Self-labeling protein tags, which are usually small proteins (<40 kDa) that can be fused with the protein of interest and specifically recognised by a small-molecule probe to form covalent conjugation,[41] have been used to construct biosensors. In this case, through using environmentally sensitive fluorophores (ESFs), the change of the local environment encountered by the sensor protein can be monitored. Liu et al. (Example 2) developed the first fluorogenic self-labeling protein-based sensor for detecting proteome stress in live cells, termed AgHalo system.[16] This system comprised of two essential modules: the Halo protein and the Halo probe (Fig. 2). The Halo protein is a modified bacterial haloalkane dehalogenase designed to specifically, rapidly and covalently bind to a chloroalkane linker under physiological conditions in cells.[42] By directed evolution,[43] Halo was evolved into an aggregation-prone mutant, i.e. AgHalo, that is vulnerable to proteostatic stress and forms different folding intermediates, such as misfolded monomers, soluble and insoluble oligomers, soluble and insoluble aggregates, etc. The Halo probe (Compound **1** in Fig. 2), comprising a chloroalkane linker attached to an ESF that is sensitive to solubility, is used in this work. While non-emission when AgHalo protein is in the monomeric state, the fluorescence of the Halo probe is turned on when soluble oligomers or aggregates are formed.

The advantage of using self-labelling protein tag is the ease of changing the Halo probe to achieve different functionality. By using different ESFs, the change of not only the solubility of the labelled protein but also the polarity and viscosity of the microenvironment can be reported.[44] Liu et al. reported a toolkit for monitoring different aggregated species of AgHalo by incorporating diverse ESFs into Halo probes, and the resultant AgHalo-Halo probe conjugates can serve as an indicator for proteome instability. Halo probe **1**, as mentioned above, is capable to fluoresce when capturing the exposure and stacking of hydrophobic residues in protein aggregates, due to the polarity-dependent fluorogenicity of sulfonyl-benzoxadiazole (SBD).[16] However, the ability of **1** to detect early-stage unfolded AgHalo monomers is limited due to the high background signals. To address this issue, they incorporated 9-(2-carboxy-2-cyanovinyl)julolidine (CCVJ),[17] a molecular rotor that only fluoresces when its excited state is rotationally restricted in a viscous or rigid local environment, into the design of Halo probe **2**. Results showed **2** can detect both misfolded conformations, and soluble and insoluble aggregates. Compared to **1**, probe **2** is more effective and sensitive to detect minor proteome stress that induces misfolding of the AgHalo protein, as well as severe proteome imbalance that involves protein aggregations. They later further synthesized more Halo probe derivatives based on solvatochromic fluorophores[45] and molecular rotors[46]. For example, probe **3**, an optimized benzothiadiazole analogue of probe **1** was shown to be fluorogenic upon aggregation of AgHalo *in vitro* and in live cells with higher photostability and more red-shifted excitation spectrum.[18] The above work expanded the library of protein-based biosensors with improved photophysical properties and biological applicability to assess the integrity of proteostasis networks.

Proteome stability profiling

Protein structural changes induced by external perturbations or internal factors can profoundly influence protein activity and can be assessed by a variety of approaches. This section will describe advances in proteome stability profiling by structural proteomics methods[47-49] and small molecule tools. For structural proteomic approaches, we focus on five representative studies (Examples 3-7) based on measuring protein solubility, oxidation rates and susceptibility to proteolysis. These methods are often coupled with high-throughput protein mass spectrometry,[48] to assess stability of proteins in the proteome. As to small molecule tools, we mainly describe the development of a novel fluorogenic toolkit (Example 8) for capturing cysteine exposure in unfolded proteins.

Proteome solubility profiling

One of the key features of protein destabilization is the solvent exposure of hydrophobic amino acids, causing undesirable intra- and intermolecular interactions that lead to insoluble aggregates and precipitation. Thus, the proportion of soluble proteins correlates to the amount of proteins that remain structured and stabilized.[50] Based on this, one can quantify protein stability by protein denaturation/stabilization curve, where the independent variable, X, is any possible factor that affects protein stability, while the dependent variable, Y, represents the abundance of soluble proteins (e.g. Fig. 3A). In general, X can be temperature, concentrations of stressors/ligands, incubation time of drugs, etc., while Y can be measured and quantified by Western blot, enzyme-linked immunosorbent assay (ELISA), mass-spectrometry proteomics, fluorescence readout, etc. Protein abundance refers to the normalized absolute abundance which can be compared across different treatments, or the relative abundance, calculated by the ratio of abundance in supernatant versus cell lysate.

One common and easily accessible stressor is high-temperature, as heat-induced unfolding often results in fast and apparent protein precipitates. The protein melting curve is then generated by plotting temperature versus soluble protein abundance, termed cellular thermal shift assay (CETSA) first reported by Martinez et al.[51] The experimental procedures include sample treatment with an increasing temperature gradient, isolation of soluble fraction by centrifugation or filtration, and quantitation of protein abundance.[50] In particular, the abundance change of the identified peptide/proteins versus temperature increment can be used to create melting curves of each individual protein and thus to characterize thermostability of the whole proteome (i.e. thermal proteome profiling (TPP)[52]). The melting curve is determined by fitting data points to sigmoidal curves where the melting temperature (T_m) is defined as the temperature at which at least 50% of proteins precipitate. Area under (melting) curve (AUC) is calculated by numerical integration of the melting curve between the lowest and highest measured temperature points. A left or right shift of T_m, and decrease or increase of AUC can infer protein destabilization or stabilization, respectively.

Jarzab et al. (Example 3) compiled a meltome atlas (i.e. proteome melting characteristics) across 13 species covering melting temperatures of 30-90 °C by TPP.[19] This work discovered many parameters that affect thermal proteome stability. For example, comparison of thermal profiles generated by heating cells or lysates showed medium Spearman correlation coefficients[53] (i.e. 0.3 – 0.6), indicating the cellular context has a substantial influence on protein stability. Comparison of amino acid compositions revealed that thermophile *T. thermophilus*, having the most thermostable proteome, contains far fewer polar but more hydrophobic amino acids compared to the mesophile *E. coli*, in accordance with previous observations that hydrophobic interactions form more stable structures. Also, hierarchical clustering of the correlation coefficients between the AUCs of orthologous proteins from different species approximately recovered the known phylogenetic tree, suggesting protein stability is at least partly evolutionarily conserved. In summary, this comprehensive profiling characterized proteome thermostability across different organisms and illustrated the relationship between protein stability and their sequence, function, phylogenetics, etc.

Sridharan et al. (Example 4) quantitated nucleotide triphosphate (NTP)- proteome interaction affinities by profiling thermal stability of proteins (termed Solubility Proteome Profiling) upon treatment of NTPs over a range of concentrations, termed two dimensional (2D)-TPP.[20] On top of parameters measurable by 1D-TPP (e.g. T_m), protein-ligand interaction can be inferred from the relationship between protein solubility and ligand concentrations (i.e. protein stabilization curve). Interestingly, ATP specifically interacts with nucleotide-binding proteins at low concentrations, while at medium and high concentrations ATP exhibits its protective roles in stabilizing protein complexes and modulates the solubility state of a quarter of insoluble proteomes, which reveals a versatile regulatory role of ATP in affecting protein complex stability and solubility.

For less abundant proteins or proteins that form only a small proportion of insoluble aggregates, measurement of abundance changes in the pellet can be more sensitive and accurate. Sui et al. (Example 5) quantified proteome solubility by the abundance difference in the supernatant and cell pellet between control and proteostatic stressor treated cells.[21] Solubility differentials, along with Gene Ontology (GO) and protein-protein interaction analysis, showed no consensus features of metastable subproteomes (i.e. a proteome subset highly expressed and susceptible to aggregation)[54]. Moreover, protein solubility can vary in different treatments, indicating specific functional remodelling of proteome stability in response to different proteostatic stressors.

Proteome oxidation profiling

Destabilized proteins are more vulnerable to oxidation due to increased accessibility and chemical reactivity of amino acids. Based on this assumption, West et al. developed a method for thermodynamic analysis of protein stability, i.e. stability of proteins from rates of oxidation (SPROX).[55] Methionine can be converted to methionine sulfoxide by H_2O_2, during which the oxidation rate depends on its local structural environment. In particular, methionines in unfolded regions are prone to be exposed and oxidized at a faster rate than those buried in folded regions. The oxidized fraction of a protein, which can be acquired by quantitative detection of protein mass-spectrometry, can infer local structural flexibility and solvent exposure, as a proxy of its unfoldedness.

Walker et al. (Example 6) measured thermodynamic folding stabilities of around 10,000 unique domains within approx. 3000 proteins in human cell extracts by SPROX.[22] The protein denaturation curve (Fig. 3A) is generated by monitoring the oxidized fraction of a protein in the presence of increasing concentrations of a chemical denaturant (i.e. guanidine hydrochloride (GdmCl) in this case). This curve can potentially provide some characteristic information that describes protein stability, such as baseline oxidation (i.e. the level of oxidation in the absence of denaturant) that reflects intrinsic susceptibility to oxidation in the native state and $[denaturant]_{1/2}$ (i.e. defined as the minimum denaturant concentration unfolding half of the total protein) that works like T_m for evaluating protein stability. More importantly, for two-state proteins (i.e. proteins where only the fully folded and fully unfolded states are significantly populated but not any partially folded intermediate states), their protein denaturation curve can be interpreted by $\Delta G_{folding}$ (i.e. free energy of folding) by least-squares fitting to a linear extrapolation model of two-state folding.[56] GO term enrichment by $[GdmCl]_{1/2}$ and $\Delta G_{folding}$ measurements identified the most stable subproteome as proteins localized at the lysosomal lumen and extracellular space. These proteins are situated in a relatively extreme environment, such as low pH, and possibly have been evolved to be more resistant to denaturation. Furthermore, decreased baseline oxidation and increased stability of the proteome were observed in the

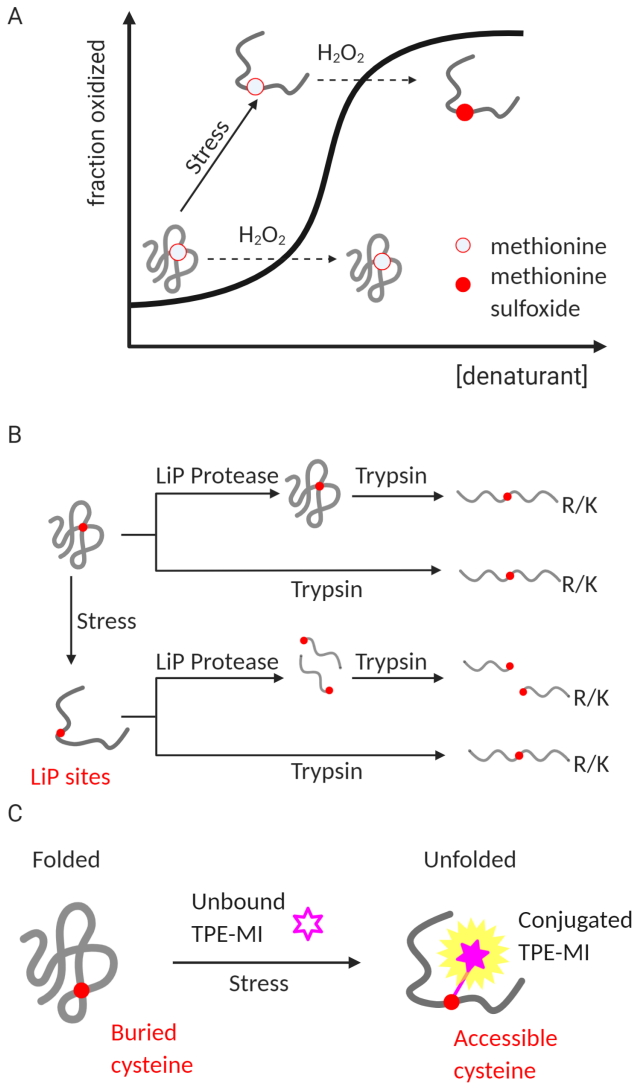

Fig. 3 Schematic illustration of methods of (A) SPROX, (B) Limited proteolysis – mass spectrometry (LiP-MS) and (C) cysteine exposure profiling.

presence of a chemical chaperone, trimethylamine *N*-oxide (TMAO), which is consistent with its reported protective role. This suggests the potential role of using protein oxidation profiling for screening drugs that regulate protein stability.

Limited proteolysis – mass spectrometry

Limited proteolysis (LiP), which uses proteases to cleave the polypeptide chain of a protein under controlled conditions, has successfully been used to probe conformational features of proteins. As structural transitions alter sidechain susceptibility to proteolytic digestion, LiP can generate conformation-specific peptides. Coupled to protein mass spectrometry (MS), LiP-MS was developed and validated to be a power tool to detect structural changes in a complex biological background on a proteome-wide scale by Feng et al.[57] The workflow mainly includes generation of structure-dependent "single digestion" and "double digestion" peptides amenable to MS analysis (Fig. 3B). For "double digestion", the proteome is treated with proteases with a broad spectrum (i.e. LiP protease, such as proteinase K) in nondenaturing conditions, followed by a second digestion with trypsin under denaturing conditions. Comparatively, an aliquot of the same proteome is subjected to trypsinization only for "single digestion". As a result, a fully tryptic peptide, embedding LiP cleavage sites, detected in the "single digestion" sample can be potentially replaced by two half-tryptic sub-peptides in the "double digestion" sample. The disappearance of the fully tryptic peptide and the emergence of half-tryptic sub-peptides can infer increased structural flexibility in local regions. As a proof-of-concept, LiP-MS successfully detected structural transitions of α-Synuclein amyloid maturation and heme dissociation from myoglobin.[57]

It was reported that proteolytic digestion preferably occurs in locally unfolded regions along a protein chain due to enhanced backbone flexibility.[58] Inspired by this, Leuenberger et al. (Example 7) applied LiP-MS in profiling proteome thermal stability in peptide-level resolution.[24] As temperature increases, proteins start to unfold and expose more LiP cleavage sites, thus the abundance of associated fully tryptic peptides decreases accordingly. The intensities of various tryptic peptides from each protein are graphed as a function of temperature and fitted by a sigmoidal model of a two-state denaturation curve. Different from the classical T_m (temperature at which at least 50% of proteins precipitate), T_m in this context is defined as the inflection point of the fitted curve, termed "apparent melting temperature", as it reflects the melting of a protein but equilibrium may not have been reached at temperatures examined. The above approach was first validated by a parallel comparison of thermal unfolding measurement of model proteins with other biophysical methods (i.e. far-ultraviolet circular dichroism measurements) and then used to profile proteome thermostability from organisms with different complexities (i.e. *E. coli, S. cerevisiae, T. thermophilus*, and *H sapiens*).[24]

The proteome thermostability analysis by using LiP-MS has addressed the following questions. First, assuming T_ms of peptides mapping to the same folding domain coincide, hierarchical clustering of T_ms from multiple peptides from the same protein should match the number of domains. By cross-referencing to domain annotation from the Pfam database,[59] seven identified peptides from nicotinamide adenine dinucleotide (reduced form) (NADH)–quinone oxidoreductase 1 (Nqo1) resulted in three different T_m clusters. Peptides from each of these three T_m clusters correctly fell within boundaries of three known domains. On a proteome-wide scale, 70% of all identified peptides from associated Pfam-annotated domains shared the same T_m cluster, indicating this approach can potentially identify domains in proteins. Second, Leuenberger et al. calculated $T_{90\%}$ for each protein, the temperature at which 90% of the protein was denatured and protein function was presumably lost. Considering a previous study in which *E. coli* displayed heat-induced physiological impairment at 47 °C, 83 proteins with $T_{90\%}$ less than 47 °C were further investigated regarding molecular functions and roles in the *E. coli* protein-protein interaction (PPI) network. In a typical PPI network, nodes denote proteins and edges between nodes denote

interactions between proteins. Hubs are nodes with a large number of edges and generally have essential functions for the survival and homeostasis of an organism, while bottlenecks are nodes necessary for interactions between different biological network branches.[60] By calculating the node degree and betweenness using the Cytoscape plug-in CentiScaPe[61], results showed among these 83 proteins losing their functions at 47 °C, 21 were hubs, 22 were bottlenecks, and 17 were both hubs and bottlenecks. These indicated thermal collapse of a cell are attributed to loss-of-function in a small set of key effectors and regulators of distinct yet different biological processes. Third, studies on amino acid composition showed lysine and aspartic acid are significantly enriched in stable proteins (i.e. top 10% of T_ms) and unstable proteins (i.e. bottom 10% of T_ms), respectively. Regarding secondary structure content, β-sheet is enriched in stable proteins while α-helical structure is enriched in unstable proteins, and stable proteins have significant fewer disordered regions compared to unstable proteins. Fourth, thermostability and melting characteristics of intrinsically disordered proteins (IDPs) were under further investigation. In theory, IDPs should not comply with the two-state folding model and thus not exhibit sigmoidal thermal denaturation curve. However, based on proteins previously predicted to be highly disordered,[62] 46 % of them displayed a sigmoidal melting curve. This indicated that nearly half of predicted IDPs underwent two-state folding transitions locally or globally, where they might interact with other molecules or form transient structures. Also, 54% of IDPs with sigmoidal melting curve have GO molecular functions associated with ligand bindings, and were enriched in RNA-binding proteins, ribonucleoprotein complexes and proteins interacting with lipid vesicles. Collectively, this study demonstrated the capability and robustness of LiP-MS to profile proteome thermostability and provided insights on understandings of related regulation mechanisms and biological processes.

Cysteine exposure profiling

Structural changes during protein unfolding can result in variations in solvent accessibility and reactivity of side chains to special chemicals. In particular, cysteines are of low abundance on protein surfaces and classified as the least exposed residue.[63] Protein unfolding thus results in increased cysteine exposure that can be a signature of protein instability. Based on this, the tetraphenylethene maleimide (TPE-MI) series were developed to specifically target unfolded proteins in cells as a novel fluorescent small molecule tool for measuring proteome stability.

Chen et al. (Example 8) first reported a thiol-reactive fluorogen, TPE-MI, for measuring unfolded protein load in cells.[27] TPE-MI is composed of a reactive group, maleimide (MI), for bio-conjugation to thiols via Michael addition reaction under physiological conditions, and a reporter group, tetraphenylethene (TPE), an aggregation-induced emission (AIE)[64] fluorophore that fluoresces upon restriction of intramolecular rotation (RIR)[65] (Fig. 4A). Free TPE-MI is inherently non-fluorescent since MI quenches TPE emission via

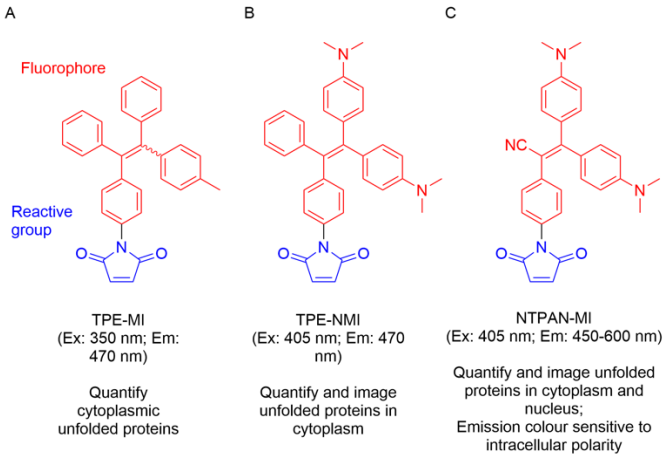

Fig.4 Chemical structures of (A) TPE-MI, (B) TPE-NMI and (C) NTPAN-MI.

the photoinduced electron transfer mechanism in which the low- lying n–π* state between HOMO and LUMO of TPE induced by the MI group. This quenching effect is disrupted by thiol conjugation. On the other hand, the reactivity of TPE-MI is affected by the environment surrounding the cysteine. As TPE is bulky and hydrophobic, cysteines buried in the folded core, or surface exposed in a hydrophilic environment have lower probability to react with TPE-MI to turn on its fluorescence. Collectively, unlocking TPE-MI fluorescence requires both thiol conjugation and sufficient local rigidity. This happens when protein unfolding exposes buried cysteines and nearby hydrophobic amino acid side chains trigger RIR and induce AIE, which renders specific fluorescence turn-on effect upon protein unfolding (Fig. 3C). Results showed TPE-MI displayed much increased fluorescence upon unfolded model proteins but not other biothiol species (e.g. glutathione (GSH)), folded proteins, proteins without cysteines or in the presence of a competitive thiol reactant, *N*-methylmaleimide (NMM). Moreover, TPE-MI is cell-permeable and thus capable to quantify unfolded protein abundance by flow cytometric assay, track proteome unfoldedness by proteomics approach and visualize build-up of unfolded proteins by microscopy in various cell models (e.g. hematopoietic stem cells (HSCs)[66] and Neuro-2a cells[67]) and subcellular proteomes or cell lysates (e.g. mitochondrial fractions isolated from SH-SY5Y cells[68]).

As a follow-up study, Zhang et al. reported TPE-NMI[28] (Fig. 4B), a modified derivative of TPE-MI with improved water miscibility and red-shifted spectral profile, which is compatible with the commonly used 405 nm laser as excitation source. Comparable to the parent compound, TPE-NMI displayed turn-on fluorescence upon protein unfolding in vitro and in cells. Later on, Owyong et al. synthesized another TPE-MI analogue NTPAN-MI[29] (Fig. 4C), which is a solvatochromic fluorogenic probe, for assessing the polarity of the local environment surrounding unfolded proteins in cells. The strategy is to replace one of the TPE core in TPE-MI with a push-pull AIE fluorophore. Such push-pull structure with both electron donor and acceptor experience excited-state charge transfer, which is stabilized by the interaction with the dipoles of solvent, result in emission redshift in polar solvents. Compared with TPE-MI and TPE-NMI which can only react with unfolded proteins in the cytoplasm,

NTPAN-MI has been found to report on unfolded protein load not only in the cytoplasm but also in the nucleus, thus offering the opportunity to quantify nucleus proteostasis capacity separately. By using spectral phasor analysis, polarity surrounding the labelled unfolded proteins in terms of dielectric constant (ε) can be quantified in subcellular resolution. To be specific, unfolded proteins in the ER were mostly in a hydrophobic environment (ε = 22-32) while in the nucleus the environment was more hydrophilic (ε = 36-45). Upon treated with a range of proteostatic stressor, the environment of the labelled unfolded protein in the ER can be more hydrophobic or hydrophilic depending on the specific stressor, whereas in the nucleus, all stressed conditions lead to a more hydrophilic environment, possibly due to the increase of RNA levels in the nucleus upon stress. NTPAN-MI was further applied to reveal proteostasis collapse in cells upon influenza virus infection that disrupts the protein quality control of the host cells.[69] Collectively, this fluorescence toolkit, TPE-MI/TPE-NMI/NTPAN-MI, was used to study proteostasis status, folding capacity and protein quality control machinery by quantitation and visualization of unfolded proteins in cells. The design and improvement of TPE-MI showed the outlook and potential to make use of small molecule tools, especially ESFs, to characterize protein folding intermediate species in cells/cell compartments and investigate cellular responses and regulatory mechanisms to proteotoxic stress.

Perspective

Protein stability on a proteome-wide scale can be interpreted indirectly from biological behaviours of biosensors, or directly quantified by parameters such as protein solubility and side chain accessibility with high-sensitivity methods (e.g. mass-spectrometry). These methods, including their design principle, method variation, application and limitations, are summarized in Table 1. Accurate quantitation enables understandings of latent cellular and biological processes in responses to proteostasis stimuli. Here, we provide an overview of perspectives on the development of bioanalytical methods and chemical probes to measure proteome stability.

Proteome dynamics profiling and accurate control

Protein unfolding is dynamic rather than static. Thus, characterizing and tracing of the whole unfolding process provides a more comprehensive view, but remains challenging. Example 1 described a biosensor that can continually report the UPR gene expression dynamics with high sensitivity and resolution. To achieve this, instead of simply placing GFP under the same transcriptional and translational regulation as UPR target genes (Fig. 1A), a transactivator (i.e. the tetracycline-dependent transactivator, tTA)[70] of GFP was placed downstream of the UPR gene while GFP was placed in an extrachromosomal element (Fig. 1B). This enabled signal amplification and larger dynamic range. The final optimized genetic circuit also included a repressor (the erythromycin-dependent transrepressor, EKRAB)[71] and a post-translational regulator of GFP (i.e. a GFP-specific NanoDeg)[72] for higher sensitivity and dynamic resolution, which relate to genetic engineering and are beyond the scope of this minireview. Collectively, Example 1 demonstrated how cutting-edge technologies contribute to improved performances of biosensors in understanding complicated regulations and alterations in a biological matrix. Example 2 reported a fluorogenic self-labeling protein-based sensor which consists of a Halo protein and a Halo probe. The metastable Halo mutant contributes to the sensitivity of the sensor as only Halo responds to proteostatic stress and changes its folding state can Halo sensor increases its fluorescence signal. One application of AgHalo conjugated with Halo probe **1**, was to detect proteostasis damage prior to observable cytotoxicity and to evaluate drug safety. This provides insights on an unknown biological question in the field of protein homeostasis, "which comes first: protein unfolding (i.e. part) or proteostasis imbalance (i.e. the whole)". To address this issue, a more sensitive biosensor capable to capture trace unfolding event is a prerequisite. Thus, this inspired Zhang's group to construct a set of Halo mutants spanning a wide range of free energy of folding (i.e. $\Delta G_{folding}$) that may be applied in different conditions. Similarly, Wood et al. reported a biosensor set of 15 barnase mutants with $\Delta G_{folding}$ that spans from -25 kJ/mol to 1 kJ/mol to quantify the protein quality control capacity.[73] These improvements expand the sensitivity and applicability of biosensors catering to different experimental requirements and will provide unique tools to understand how perturbation-induced protein structural alterations can influence proteome stability.

Development of small molecule tools

Environmentally sensitive fluorophores (ESFs) have become a key tool for studying protein unfolding and protein aggregations. For example, traditional organic luminophores are mainly composed of planar aromatic rings. They emit efficiently in dilute solutions while being weakened or even totally quenched due to formation of aggregates, which facilitates exciton interactions and nonradiative pathways. This phenomenon is referred to as aggregation-caused quenching (ACQ).[74] Opposite to ACQ, some organic luminophores with freely-rotating groups exhibit higher photoluminescence efficiency in the solid or aggregation state than in solution, termed aggregation-induced emission (AIE).[64, 75] By using the aggregation-disaggregation triggered fluorescence change, these fluorophores would potentially be applied to report protein aggregation or disaggregation dynamic events.[76] Example 2 and Example 8 both demonstrated that further development of ESFs is to expand the toolkit equipped with multi-responsive probes that cover the full fluorescence spectrum for pulse-chase experiment and multiplex imaging as well as display improved sensitivity and selectivity (higher signal-to-noise ratio). For example, to monitor the post-stress recovery of the AgHalo sensor *in situ*, the Halo probe should be highly biocompatible and photostable as it stays in the cells for a long period of time and is excited constantly. Furthermore, organelle-specific

probes would also be desired to facilitate the studies on the cellular component(s) of interest and their role in the proteostasis network. Finally, small amyloid ligands (SALs) that can interact with amyloid aggregate species in a broader sense would be useful for studying how they are formed, transported and degraded in the cells.[77, 78] However, available SALs for targeting amyloid fibrils (e.g. Thioflavin T), prefibrillar protein aggregates (e.g. bis(triphenylphosphonium) tetraphenylethene (TPE-TPP))[79], or oligomeric species (e.g. BD-Oligo)[80] have limited applications in cellular context. Thus, SALs that can be used in cells, especially live cells, remain to be further explored and optimized.

Focusing on cellular context and cellular compartments

X-ray crystallography, NMR and various spectroscopic techniques have been used to determine protein conformations *in vitro*. However, cellular environment (e.g. molecular crowding, interactions with other biomolecules) impacts protein structure and protein stability.[81] A direct comparison between proteome thermostability from Example 7 and data from purified recombinant proteins in ProTherm database[82] showed a poor correlation (R^2 = 0.13). Although Examples 3, 6, and 7 all demonstrated that protein stability is highly correlated with intrinsic properties (e.g. amino acid compositions, secondary structures), interestingly in Example 7, almost half of proteins previously predicted to be highly unstructured unexpectedly display a two-state denaturation curve, indicating they might not be simply disordered but can be locally, globally, or transiently structured for functionality under some circumstance (e.g. binding to ligands or proteins) in the cellular matrix. In Example 3, thermal profiles, generated by heating cells or lysates, mediumly correlate, indicating that the cellular context (i.e., complete membrane and organelle), influence protein thermostability. Furthermore, Example 4 revealed that ATP interacts with different proteins and modulates their solubility and stability at different concentrations. Thus, such findings highlight the importance of assessing protein stability in cells, while characterization of proteins in vitro might be loss of biological relevance.

In Example 8, NTPAN-MI detected the unfolded proteins in the nucleus for the first time but not former TPE-MI versions, indicating cell membrane can prevent small molecules from entering cell compartments and organelles. Biosensors, e.g. developed by similar methods in Example 2, may have limited access to specific locations in cells. Most research studies generally work on total cell extract or cytosolic proteins. However, protein abundance and PPI network can be very different in specific cell compartments. This inspires development of methods/small molecules to enrich/target unfolded proteins of interest in specific organelles.

Solutions of biomacromolecules such as proteins or nucleic acid, can condense into a dense phase that coexists with a dilute phase, termed liquid-liquid phase separation (LLPS).[83] Intracellular LLPS is thought to drive the formation of membraneless compartments in cells that contain proteins, RNA, etc.[84] The exploration of physical and pathological context involving LLPS has illustrated its functional and regulatory roles in signaling pathways and stress adaptation.[85] Thus, development of small molecule tools and biosensors that can report on the formation and dynamics of LLPS, can expand our understanding of biological functions and regulatory mechanisms of these cellular compartments.

Data integration and cross-referencing

In previous examples, proteome stability measurement is always combined and integrated with reference databases or other informatic analysis methods. First, Gene Ontology (GO) provides a formal, unified and systematic representation for the consistent description of gene and gene product attributes across all species,[86] contributed from a large, international group of scientists in the disciplines of biology and computer science. Gene set enrichment analysis[87] can further identify GO terms that are over- or under-represented from a group of proteins of interest (e.g. identified as the most/least stable proteins). Second, cross-referencing to other database or previous studies facilitates exploration of determinants of protein thermostability and its biological relevance. Example 6 mapped methionines from the dataset to different secondary structures using Define Secondary Structure of Proteins (DSSP)[88] so that the baseline oxidation levels for each secondary structure could be compared. Results showed highly oxidizable methionines tend to be enriched in turns, bends, and unstructured coils, while more protected methionines were found to reside in rigid secondary-structures, such as β-strands and α-helices. Besides, solvent-accessible surface area (SASA) of methionines of a particular protein can be calculated by using structures from protein database and molecular visualization programs[89]. Data showed a strong positive correlation between the SASA of a methionine residue and its baseline oxidation. Third, protein stability is tightly regulated by the interactions with small molecule ligands and other bio-macromolecules. Proteins can be stabilized upon binding to cofactors and subunits in a complex, but unfold and co-aggregate following dissociation, unfoldedness or aggregation of their interactors.[90, 91] This interactome aspect focuses on protein complex stability and highlights the importance of protein-protein interactions (PPIs) and omic-based data analyses. For instance, Example 7 demonstrated the functional importance of proteins denatured at detrimental high temperature by assessing their roles and associations in the PPI network. Collectively, data integration and cross-referencing expand understandings of how proteome integrity and stability are well-organized and corelated.

Biological questions relevant to proteome stability

What can we do with proteome stability data? This raises a question why researchers measure proteome stability and what insights and understandings of stress responses and cellular regulatory mechanisms are gained. First, proteome stability measurement characterizes stable/unfolded structures in proteins (i.e. what proteins/structures in the proteome are more/less stable). Different methods draw similar conclusions in some aspect but opposing in different context. In terms of

amino acid composition and secondary structure, Example 3 using TPP, Example 6 using SPROX and Example 7 using LiP-MS all supported that hydrophobic interactions and rigid secondary-structures are more stable and enriched in stable proteins. Examples 6 and 7 both showed unstable proteins have significant more disordered regions than stable proteins. However, Example 5 concluded that there is no consensus feature of metastable subproteomes, while Example 6 demonstrated that the most stable subproteome are proteins localized at the lysosomal lumen and extracellular space, and the least unstable proteins include ATP-binding and ribosomal proteins. These measurements were conducted under various experimental settings and can be very heterogenous and incomparable, which require proper translation and explanation. Second, protein denaturation measurement reveals how proteostasis deteriorates and collapses. Example 7 using LiP-MS identified proteins losing their functions at a temperature previously reported to impair cellular physiology. Based on their roles in the PPI network, results showed heat-induced breakdowns of cell integrity originated from denaturation of a specific class of proteins, i.e. a small set of key effectors and regulators of biological processes, but not from simultaneous, generalized or random loss of proteins. Third, Examples 3 and 7 shed light on the evolutionary conservation of protein stability. For example, Example 7 identified orthologous and paralogous genes from different species used eggNOG database[92], and discovered that recently diverged paralogs with above 60% sequence identity had more similar thermostability than random protein pairs. Collectively, proteome stability profiling aids in understanding complex and inter-regulatory processes.

Applications in measuring proteome stability

Methods for measuring proteome stability have diverse applications in other fields – from biology, biochemistry, chemistry, systems biology, medicine and pharmaceutical development. Since small molecules can alter the protein melting/denaturation curves by protein-ligand interactions, TPP can be used for preclinical drug development, monitoring drug-target engagement,[51] and discovering biomarkers for drug efficacy and toxicity[52]. One major application of TPP revealed drug efficacy by the identification of targets and off-targets.[52] In this work, identified interactors of a kinase inhibitor with a known spectrum of targets, staurosporine, included not only kinases but also non-kinases. Also, drug treatment shifted melting curves of the regulatory subunits of kinase complexes and downstream effectors. This provides a robust and comprehensive way for a proteome-wide assessment of drug-target engagement and explanation of potential adverse side effects.

Limitations and method variations

Table 1 summarizes limitations of biosensors and proteome stability profiling methods. Nucleic acid sensors only respond to a specific pathway, and protein-based sensors aim to report the folding status of cellular or subcellular proteome. Therefore, these biosensors only apply to comparative measurements with proper controls and under given experimental conditions. Other concerns include the potential immunogenicity and cytotoxicity induced by reporter proteins,[93] resistance of proteasomal degradation of overexpressed reporter proteins,[94, 95] ect. Regarding proteome stability profiling methods, methionine oxidation is not a benign modification and may impact localized folding stability of a second methionine located elsewhere in the same protein.[22] Regional unfolding may be detected by LiP-MS, but not TPP if it does not cause protein precipitation.

There are alternative methods and method variations based on Examples 1 – 8. For example, SPROX belongs to the protein footprinting method, which also include hydroxyl-radical footprinting (HRF)[23], etc. Drug Affinity Responsive Target Stability (DARTS)[25] and Pulse Proteolysis (PP)[26] are also proteolysis-based methods. These approaches complement each other and can be further combined for specific applications and improved characterizations.

Conclusion

In this minireview, we have introduced two main classes of bioanalytical methods for assessing proteome stability, i.e. employing novel biosensors and proteome stability profiling methods. Biosensors, including nucleic acid-based and protein-based biosensors, are designed to report proteostasis impairment or represent the folding status of the proteome. Proteome stability profiling methods quantitate a certain biophysical parameter that reflects protein stability for all proteins in the proteome, allowing evaluation of effects of molecules and cellular perturbations on protein stability in different disease-relevant cell types, to study regulation of aggregation-susceptible proteins and organelles. . These bioanalytical methods align different strategies to detect, identify and monitor regulators on a proteome-level, and incorporate genetic and protein engineering, systems biology and synthetic chemistry. They contribute to characterization of various unfolded protein species and further understanding of cellular physiology and protein structural changes influencing proteome stability . Eventually these tools and methodologies can facilitate the process of drug development strategy as well as pharmacologic therapy.[96]

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This work was supported by grants to Y.H. (Australian Research Council DE170100058, Rebecca L. Cooper Medical Research Foundation PG2018043, National Health and Medical Research Council 1161803 and Australia-China Science and Research Fund-Joint Research Centre on Personal Health Technologies) and to D.G. (National Health and Medical Research Council 1057741 and 1139489, and Helen Amelia Hains Fellowship).

References

1. J. Labbadia and R. I. Morimoto, *Annu. Rev. Biochem.*, 2015, **84**, 435-464.
2. S. Kaushik and A. M. Cuervo, *Nat. Med.*, 2015, **21**, 1406-1415.
3. F. Chiti and C. M. Dobson, *Annu. Rev. Biochem.*, 2017, **86**, 27-68.
4. A. Kurtishi, B. Rosen, K. S. Patil, G. W. Alves and S. G. Møller, *Mol. Neurobiol.*, 2019, **56**, 3676-3689.
5. M. S. Hipp, S. H. Park and F. U. Hartl, *Trends Cell Biol.*, 2014, **24**, 506-514.
6. C. L. Klaips, G. G. Jayaraj and F. U. Hartl, *J Cell Biol*, 2018, **217**, 51-63.
7. J. S. Pattison and J. Robbins, *Autophagy*, 2008, **4**, 821-823.
8. F. Del Monte and G. Agnetti, *Proteomics Clin. Appl.*, 2014, **8**, 534-542.
9. W. E. Balch, R. I. Morimoto, A. Dillin and J. W. Kelly, *Science*, 2008, **319**, 916-919.
10. H. W. Wang and J. W. Wang, *Protein Sci.*, 2017, **26**, 32-39.
11. M. S. Hipp, P. Kasturi and F. U. Hartl, *Nat Rev Mol Cell Biol*, 2019, **20**, 421-435.
12. D. Sakakibara, A. Sasaki, T. Ikeya, J. Hamatsu, T. Hanashima, M. Mishima, M. Yoshimasu, N. Hayashi, T. Mikawa, M. Wälchli, B. O. Smith, M. Shirakawa, P. Güntert and Y. Ito, *Nature*, 2009, **458**, 102-105.
13. S. Ebbinghaus, A. Dhar, J. D. McDonald and M. Gruebele, *Nature Methods*, 2010, **7**, 319-323.
14. R. Feng, M. Gruebele and C. M. Davis, *Nature Communications*, 2019, **10**, 1179.
15. C. A. Origel Marmolejo, B. Bachhav, S. D. Patibandla, A. L. Yang and L. Segatori, *Nat. Chem. Biol.*, 2020, **16**, 520-528.
16. Y. Liu, M. Fares, N. P. Dunham, Z. Gao, K. Miao, X. Jiang, S. S. Bollinger, A. K. Boal and X. Zhang, *Angew. Chem. Int. Ed. Engl.*, 2017, **56**, 8672-8676.
17. M. Fares, Y. Li, Y. Liu, K. Miao, Z. Gao, Y. Zhai and X. Zhang, *Bioconjug. Chem.*, 2018, **29**, 215-224.
18. Y. Liu, K. Miao, Y. Li, M. Fares, S. Chen and X. Zhang, *Biochemistry*, 2018, **57**, 4663-4674.
19. A. Jarzab, N. Kurzawa, T. Hopf, M. Moerch, J. Zecha, N. Leijten, Y. Bian, E. Musiol, M. Maschberger, G. Stoehr, I. Becher, C. Daly, P. Samaras, J. Mergner, B. Spanier, A. Angelov, T. Werner, M. Bantscheff, M. Wilhelm, M. Klingenspor, S. Lemeer, W. Liebl, H. Hahne, M. M. Savitski and B. Kuster, *Nature Methods*, 2020, **17**, 495-503.
20. S. Sridharan, N. Kurzawa, T. Werner, I. Günthner, D. Helm, W. Huber, M. Bantscheff and M. M. Savitski, *Nature Communications*, 2019, **10**, 1155.
21. X. Sui, D. E. V. Pires, A. R. Ormsby, D. Cox, S. Nie, G. Vecchi, M. Vendruscolo, D. B. Ascher, G. E. Reid and D. M. Hatters, *Proceedings of the National Academy of Sciences*, 2020, **117**, 2422-2431.
22. E. J. Walker, J. Q. Bettinger, K. A. Welle, J. R. Hryhorenko and S. Ghaemmaghami, *Proceedings of the National Academy of Sciences*, 2019, **116**, 6081-6090.
23. J. Kiselar and M. R. Chance, *Annual Review of Biophysics*, 2018, **47**, 315-333.
24. P. Leuenberger, S. Ganscha, A. Kahraman, V. Cappelletti, P. J. Boersema, C. von Mering, M. Claassen and P. Picotti, *Science*, 2017, **355**, eaai7825.
25. B. Lomenick, R. Hao, N. Jonai, R. M. Chin, M. Aghajan, S. Warburton, J. Wang, R. P. Wu, F. Gomez, J. A. Loo, J. A. Wohlschlegel, T. M. Vondriska, J. Pelletier, H. R. Herschman, J. Clardy, C. F. Clarke and J. Huang, *Proceedings of the National Academy of Sciences*, 2009, **106**, 21984-21989.
26. C. Park and S. Marqusee, *Nature Methods*, 2005, **2**, 207-212.
27. M. Z. Chen, N. S. Moily, J. L. Bridgford, R. J. Wood, M. Radwan, T. A. Smith, Z. Song, B. Z. Tang, L. Tilley, X. Xu, G. E. Reid, M. A. Pouladi, Y. Hong and D. M. Hatters, *Nat Commun*, 2017, **8**, 474.
28. S. Zhang, M. Liu, L. Y. F. Tan, Q. Hong, Z. L. Pow, T. C. Owyong, S. Ding, W. W. H. Wong and Y. Hong, *Chemistry – An Asian Journal*, 2019, **14**, 904-909.
29. T. C. Owyong, P. Subedi, J. Deng, E. Hinde, J. J. Paxman, J. M. White, W. Chen, B. Heras, W. W. H. Wong and Y. Hong, *Angew. Chem. Int. Ed. Engl.*, 2019, DOI: 10.1002/anie.201914263.
30. G. Cedras, H. Kroukamp, W. H. Van Zyl and R. Den Haan, *Biotechnol. Appl. Biochem.*, 2020, **67**, 82-94.
31. Y. E. Kim, M. S. Hipp, A. Bracher, M. Hayer-Hartl and F. U. Hartl, *Annu. Rev. Biochem.*, 2013, **82**, 323-355.
32. M. Akerfelt, R. I. Morimoto and L. Sistonen, *Nature reviews. Molecular cell biology*, 2010, **11**, 545-555.
33. P. Walter and D. Ron, *Science*, 2011, **334**, 1081-1086.
34. M. Galves, R. Rathi, G. Prag and A. Ashkenazi, *Trends Biochem. Sci.*, 2019, **44**, 872-884.
35. I. Dikic, *Annu. Rev. Biochem.*, 2017, **86**, 193-224.
36. G. E. Karagöz, D. Acosta-Alvear and P. Walter, *Cold Spring Harb. Perspect. Biol.*, 2019, **11**.
37. F. A. Ran, P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott and F. Zhang, *Nat. Protoc.*, 2013, **8**, 2281-2308.
38. R. Gupta, P. Kasturi, A. Bracher, C. Loew, M. Zheng, A. Villella, D. Garza, F. U. Hartl and S. Raychaudhuri, *Nat Methods*, 2011, **8**, 879-884.
39. B. T. Bajar, E. S. Wang, S. Zhang, M. Z. Lin and J. Chu, *Sensors (Basel, Switzerland)*, 2016, **16**, 1488.
40. V. Sachsenhauser, X. Deng, H.-h. Kim, M. Jankovic and J. C. A. Bardwell, *ACS Chem. Biol.*, 2020, DOI: 10.1021/acschembio.0c00083.
41. C. A. Hoelzel and X. Zhang, *Chembiochem*, 2020, **21**, 1935-1946.
42. G. V. Los, L. P. Encell, M. G. McDougall, D. D. Hartzell, N. Karassina, C. Zimprich, M. G. Wood, R. Learish, R. F. Ohana, M. Urh, D. Simpson, J. Mendez, K. Zimmerman, P. Otto, G. Vidugiris, J. Zhu, A. Darzins, D. H. Klaubert, R. F. Bulleit and K. V. Wood, *ACS Chem. Biol.*, 2008, **3**, 373-382.
43. F. H. Arnold, *Angew. Chem. Int. Ed. Engl.*, 2018, **57**, 4143-4148.
44. A. S. Klymchenko, *Acc. Chem. Res.*, 2017, **50**, 366-375.
45. A. S. Klymchenko, *Acc. Chem. Res.*, 2017, **50**, 366-375.
46. W. L. Goh, M. Y. Lee, T. L. Joseph, S. T. Quah, C. J. Brown, C. Verma, S. Brenner, F. J. Ghadessy and Y. N. Teo, *J. Am. Chem. Soc.*, 2014, **136**, 6159-6162.
47. U. Kaur, H. Meng, F. Lui, R. Ma, R. N. Ogburn, J. H. R. Johnson, M. C. Fitzgerald and L. M. Jones, *J. Proteome Res.*, 2018, **17**, 3614-3627.
48. P. Grandi and M. Bantscheff, *Drug Discovery Today: Technologies*, 2019, **31**, 99-108.
49. N. de Souza and P. Picotti, *Curr. Opin. Struct. Biol.*, 2020, **60**, 57-65.
50. D. M. Molina and P. Nordlund, *Annu. Rev. Pharmacol. Toxicol.*, 2016, **56**, 141-161.

51. D. M. Molina, R. Jafari, M. Ignatushchenko, T. Seki, E. A. Larsson, C. Dan, L. Sreekumar, Y. Cao, P. Nordlund and xe, *Science*, 2013, **341**, 84-87.
52. M. M. Savitski, F. B. M. Reinhard, H. Franken, T. Werner, M. F. Savitski, D. Eberhard, D. M. Molina, R. Jafari, R. B. Dovega, S. Klaeger, B. Kuster, P. Nordlund, M. Bantscheff and G. Drewes, *Science*, 2014, **346**, 1255784.
53. C. Spearman, *The American Journal of Psychology*, 1904, **15**, 72-101.
54. R. Kundra, P. Ciryam, R. I. Morimoto, C. M. Dobson and M. Vendruscolo, *Proceedings of the National Academy of Sciences*, 2017, **114**, E5703-E5711.
55. G. M. West, L. Tang and M. C. Fitzgerald, *Anal. Chem.*, 2008, **80**, 4175-4185.
56. C. N. Pace, in *Methods Enzymol.*, Academic Press, 1986, vol. 131, pp. 266-280.
57. Y. Feng, G. De Franceschi, A. Kahraman, M. Soste, A. Melnik, P. J. Boersema, P. P. de Laureto, Y. Nikolaev, A. P. Oliveira and P. Picotti, *Nat. Biotechnol.*, 2014, **32**, 1036-1044.
58. A. Fontana, P. P. de Laureto, B. Spolaore, E. Frare, P. Picotti and M. Zambonin, *Acta Biochim. Pol.*, 2004, **51**, 299-321.
59. S. El-Gebali, J. Mistry, A. Bateman, S. R. Eddy, A. Luciani, S. C. Potter, M. Qureshi, L. J. Richardson, G. A. Salazar, A. Smart, E. L L. Sonnhammer, L. Hirsh, L. Paladin, D. Piovesan, S. C E. Tosatto and R. D. Finn, *Nucleic Acids Res.*, 2018, **47**, D427-D432.
60. H. Yu, P. M. Kim, E. Sprecher, V. Trifonov and M. Gerstein, *PLoS Comput. Biol.*, 2007, **3**, e59.
61. G. Scardoni, M. Petterlini and C. Laudanna, *Bioinformatics*, 2009, **25**, 2857-2859.
62. J. Gsponer, M. E. Futschik, S. A. Teichmann and M. M. Babu, *Science*, 2008, **322**, 1365-1368.
63. S. M. Marino and V. N. Gladyshev, *J. Mol. Biol.*, 2010, **404**, 902-916.
64. Y. Hong, J. W. Y. Lam and B. Z. Tang, *Chem. Soc. Rev.*, 2011, **40**, 5361-5388.
65. N. L. Leung, N. Xie, W. Yuan, Y. Liu, Q. Wu, Q. Peng, Q. Miao, J. W. Lam and B. Z. Tang, *Chemistry (Easton)*, 2014, **20**, 15349-15353.
66. L. Hidalgo San Jose, M. J. Sunshine, C. H. Dillingham, B. A. Chua, M. Kruta, Y. Hong, D. M. Hatters and R. A. J. Signer, *Cell Rep.*, 2020, **30**, 69-80.e66.
67. S. Parakh, S. Shadfar, E. R. Perri, A. M. G. Ragagnin, C. V. Piattoni, M. B. Fogolín, K. C. Yuan, H. Shahheydari, E. K. Don, C. J. Thomas, Y. Hong, M. A. Comini, A. S. Laird, D. M. Spencer and J. D. Atkin, *iScience*, 2020, **23**, 101097.
68. D. Hu, X. Sun, X. Liao, X. Zhang, S. Zarabi, A. Schimmer, Y. Hong, C. Ford, Y. Luo and X. Qi, *Acta Neuropathol.*, 2019, **137**, 939-960.
69. M. Marques, B. Ramos, A. R. Soares and D. Ribeiro, *Cells*, 2019, **8**.
70. M. Gossen and H. Bujard, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 5547-5551.
71. W. Weber, C. Fux, M. Daoud-El Baba, B. Keller, C. C. Weber, B. P. Kramer, C. Heinzen, D. Aubel, J. E. Bailey and M. Fussenegger, *Nat. Biotechnol.*, 2002, **20**, 901-907.
72. W. Zhao, L. Pferdehirt and L. Segatori, *ACS Synthetic Biology*, 2018, **7**, 540-552.
73. R. J. Wood, A. R. Ormsby, M. Radwan, D. Cox, A. Sharma, T. Vöpel, S. Ebbinghaus, M. Oliveberg, G. E. Reid, A. Dickson and D. M. Hatters, *Nature Communications*, 2018, **9**, 287.
74. G. Chen, W. Li, T. Zhou, Q. Peng, D. Zhai, H. Li, W. Z. Yuan, Y. Zhang and B. Z. Tang, *Advanced Materials*, 2015, **27**, 4496-4501.
75. H. Wang, E. Zhao, J. W. Y. Lam and B. Z. Tang, *Materials Today*, 2015, **18**, 365-377.
76. J. Mei, Y. Huang and H. Tian, *ACS applied materials & interfaces*, 2018, **10**, 12217-12261.
77. P. Verwilst, H. S. Kim, S. Kim, C. Kang and J. S. Kim, *Chem. Soc. Rev.*, 2018, **47**, 2249-2265.
78. X. L. Bu, P. P. N. Rao and Y. J. Wang, *Mol. Neurobiol.*, 2016, **53**, 3565-3575.
79. M. Kumar, Y. Hong, D. C. Thorn, H. Ecroyd and J. A. Carver, *Anal. Chem.*, 2017, **89**, 9322-9329.
80. C. L. Teoh, D. Su, S. Sahu, S.-W. Yun, E. Drummond, F. Prelli, S. Lim, S. Cho, S. Ham, T. Wisniewski and Y.-T. Chang, *J. Am. Chem. Soc.*, 2015, **137**, 13503-13509.
81. H.-X. Zhou, *FEBS Lett.*, 2013, **587**, 1053-1061.
82. K. A. Bava, M. M. Gromiha, H. Uedaira, K. Kitajima and A. Sarai, *Nucleic Acids Res.*, 2004, **32**, D120-D121.
83. S. Alberti, A. Gladfelter and T. Mittag, *Cell*, 2019, **176**, 419-434.
84. Y. Shin and C. P. Brangwynne, *Science*, 2017, **357**, eaaf4382.
85. C. Rabouille and S. Alberti, *Curr. Opin. Cell Biol.*, 2017, **47**, 34-42.
86. C. Gene Ontology, *Nucleic Acids Res.*, 2008, **36**, D440-D444.
87. J. Reimand, R. Isserlin, V. Voisin, M. Kucera, C. Tannus-Lopes, A. Rostamianfar, L. Wadi, M. Meyer, J. Wong, C. Xu, D. Merico and G. D. Bader, *Nat. Protoc.*, 2019, **14**, 482-517.
88. W. Kabsch and C. Sander, *Biopolymers*, 1983, **22**, 2577-2637.
89. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33-38, 27-38.
90. M. Iljina, A. J. Dear, G. A. Garcia, S. De, L. Tosatto, P. Flagmeier, D. R. Whiten, T. C. T. Michaels, D. Frenkel, C. M. Dobson, T. P. J. Knowles and D. Klenerman, *ACS Nano*, 2018, **12**, 10855-10866.
91. C. S. H. Tan, K. D. Go, X. Bisteau, L. Dai, C. H. Yong, N. Prabhu, M. B. Ozturk, Y. T. Lim, L. Sreekumar, J. Lengqvist, V. Tergaonkar, P. Kaldis, R. M. Sobota and P. Nordlund, *Science*, 2018, **359**, 1170-1177.
92. L. J. Jensen, P. Julien, M. Kuhn, C. von Mering, J. Muller, T. Doerks and P. Bork, *Nucleic Acids Res.*, 2008, **36**, D250-254.
93. A. M. Ansari, A. K. Ahmed, A. E. Matsangos, F. Lay, L. J. Born, G. Marti, J. W. Harmon and Z. Sun, *Stem Cell Rev Rep*, 2016, **12**, 553-559.
94. A. Khmelinskii, M. Meurer, C. T. Ho, B. Besenbeck, J. Füller, M. K. Lemberg, B. Bukau, A. Mogk and M. Knop, *Mol. Biol. Cell*, 2016, **27**, 360-370.
95. E. A. Specht, E. Braselmann and A. E. Palmer, *Annu. Rev. Physiol.*, 2017, **79**, 93-117.
96. R. M. Sebastian and M. D. Shoulders, *Annu. Rev. Biochem.*, 2020, **89**, 529-555.