Proteome-wide observation of the phenomenon of life on the edge of solubility

Giulia Vecchi, Pietro Sormanni, Benedetta Mannini, Andrea Vandelli, Gian Gaetano Tartaglia, Christopher M. Dobson, F. Ulrich Hartl, and Michele Vendruscolo

*Centre for Misfolding Diseases, Department of Chemistry, University of Cambridge, CB2 1EW Cambridge, United Kingdom; **Gene Function and Evolution, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology (BIST), 08003 Barcelona, Spain; *Department of Neuroscience and Brain Technologies, Istituto Italiano di Tecnologia, 16163 Genoa, Italy; **CRG, BIST, 08003 Barcelona, Spain; †Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain; ‡Department of Biology "Charles Darwin," Sapienza University of Rome, 00185 Rome, Italy; and §Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, 82152 Martinsried, Germany

To function effectively, proteins must avoid aberrant aggregation, and hence they are expected to be expressed at concentrations safely below their solubility limits. By analyzing proteome-wide mass spectrometry data of Caenorhabditis elegans, however, we show that the levels of about three-quarters of the nearly 4,000 proteins analyzed in adult animals are close to their intrinsic solubility limits, indeed exceeding them by about 10% on average. We next asked how aging and functional self-assembly influence these solubility limits. We found that despite the fact that the total quantity of proteins within the cellular environment remains approximately constant during aging, protein aggregation sharply increases between days 6 and 12 of adulthood, after the worms have reproduced, as individual proteins lose their stoichiometric balances and the cellular machinery that maintains solubility undergoes functional decline. These findings reveal that these proteins are highly prone to undergoing concentration-dependent phase separation, which on aging is rationalized in a decrease of their effective solubilities, in particular for proteins associated with translation, growth, reproduction, and the chaperone system.

It is therefore of great importance to understand how the physiological levels of proteins relate to their critical concentrations. It has been suggested that proteins have coevolved with their cellular environment to be sufficiently soluble to enable their expression at the levels needed in cells for their optimal functioning, but with almost no margin of safety against genetic or environmental factors that either decrease their solubilities or increase their cellular concentrations. This concept has been referred to as the life on the edge hypothesis (16). The original suggestion was based on the observation of an anticorrelation between the aggregation rates measured in vitro of a small group of human proteins and the corresponding human mRNA expression levels measured in vivo (16). This anticorrelation was rationalized as being the net result of 2 opposing pressures acting on the amino acid sequences of proteins. The first is the effect of random mutations, which tend on average to increase the aggregation propensity of a protein, and the second is the effect of evolutionary selection of mutations, which tends to select solubilizing mutations to ensure that a protein is soluble and stable enough at the concentration required in the cell for its biological role (16–22). While the action of these 2 contrasting forces may have left proteins at risk for aggregation, evolution has developed a robust protein homeostasis system capable of maintaining the functional balance of the proteome (7, 23–28).

Significance

More than a decade ago, we put forward the “life on the edge of solubility” hypothesis, according to which proteins are expressed in the cellular environment at levels close to their solubility limits. This observation was based on the analysis of a small number of proteins for which solubility and cellular concentration information was available at the time. To confirm this hypothesis we have now taken advantage of recent advances in mass spectrometry that have enabled the proteome-wide analysis of protein concentrations in both the soluble and insoluble forms. We have been able to show in this way that the vast majority of proteins in a model organism are indeed expressed above their solubility limits, and to investigate the consequences of this phenomenon.
vivo, aggregation is thus inhibited by a plethora of molecular chaperones, which assist proteins to remain in their soluble native states (25, 26, 29, 30). Under conditions of cellular stress (31–34) and during aging (35–37), however, the protein homeostasis system becomes progressively impaired and challenged, and eventually fails to prevent aggregation, which in turn places further stress on the cellular environment, promoting yet higher levels of aggregation (7, 23, 25, 27, 38–40).

The characterization of the extent and nature of the connection among protein concentration, protein aggregation, and the protein homeostasis system in a living organism on a global scale is therefore of great importance. Whether the concept that proteins are expressed at their solubility limits could have general validity still remains an open question. The model organism C. elegans is particularly useful for addressing this issue, as it is widely used to study the changes in the protein homeostasis system on aging and stress (35–37, 39, 41, 42). In particular, proteomic studies using mass spectrometry have shown that widespread protein aggregation occurs on aging in this organism (36, 39, 41, 42). Consistent with the hypothesis discussed here (16), proteins expressed at high levels have been found to have a lower aggregation propensity than proteins expressed at low levels (16, 42), a result also more recently observed in Escherichia coli, Saccharomyces cerevisiae, Thermus thermophilus, and human cells (32).

In the present study, we use extensive data on protein abundance in C. elegans derived from mass spectrometry (42) to reveal highly quantitative proteome-wide evidence that the physiological concentrations of proteins are close to their critical levels. The data also show that with age there is a sharp increment in the quantity of aggregated proteins between days 6 and 12 of adulthood, after the worms have reproduced, which is not the result of an increase in the overall protein content in the worms. The proteins most responsible for this proliferation of aggregates are mainly associated with translation, homeostasis, and structural functional classes. Notably, proteins enriched in low-complexity regions and highly prone to liquid–liquid phase separation (43–45) are significantly overrepresented in the deposits that proliferate on aging. These findings indicate that the intracellular proteome is expressed at its solubility limits, with proteins highly prone to undergoing liquid–liquid phase separation driven to aggregate on aging as a result of a decrease in their effective solubilities, rather than an increase in their expression levels.

### Results and Discussion

**Protein Levels Commonly Exceed Their Solubility Limits in Adult C. elegans.** We first sought to examine the evidence at a proteomic level for the hypothesis that proteins are expressed at their critical levels (16). To this end, we analyzed proteome-wide mass spectrometry data for wild-type C. elegans from experiments in which total, soluble (supernatant), and insoluble (pellet) protein abundances were measured in adult worms (42). For the analysis we considered only those proteins detected in at least 2 of 3 replicates (42) and found that in adult wild-type worms (day 12 of adulthood from L4 stage), about 74% (2,792 of 3,775 proteins) are found also in the aggregated (pellet) fraction. This observation implies that about three-quarters of all proteins detected are expressed above their critical concentrations.

Fig. 1. Comparison between the cellular concentrations and the critical concentrations of proteins in adult C. elegans. Density plot of the total abundance (T) and soluble abundance (S), in logarithmic scale, for the 1,163 proteins quantified as at least at their solubility limits (Materials and Methods). Each point is a protein colored from a heat map scale (black to yellow) according to the density of neighboring points, where black indicates an isolated protein, corresponding to a density value close to 0, while yellow indicates a protein that is surrounded by many others in that area, corresponding to a density value close to 1. (Inset) The density values are obtained with a standard Gaussian kernel density estimator and are reported in 3D. The gray bisector line in the scatterplot corresponds to the solubility limit. Protein IDs are indicated for the proteins found to be further from the solubility limit. These proteins are intermediate filaments proteins, collagen, and 2 uncharacterized proteins (Q9NE57, O02141).
To calculate the total mass difference at a given day compared with day 1, we restricted our analysis to those proteins detected and quantified at all times (3,078 proteins; Fig. 2A). We found that no significant change occurs to the total cellular protein mass on aging (Fig. 2A), even though approximately one-third of the proteins in the worms were found to change in abundance by at least 2-fold from day 1 to day 17, either by increasing or decreasing their abundance levels with age (42). This observation indicates that despite the fact that a substantial degree of remodeling occurs on aging in terms of the relative concentrations of individual proteins, the proteome as a whole maintains its total intracellular mass at a specific level.

Using the abundance data for the insoluble fraction, and restricting the analysis to those proteins consistently quantified in the pellet fraction from day 1 to day 17 (965 proteins), we evaluated by means of the procedure described here the change at days 6, 12, and 17 of the insoluble fraction of the cellular content with respect to day 1 of adulthood. We observed in particular a sharp increase in the mass of insoluble proteins occurring between days 6 and 12 of adulthood (Fig. 2B), despite the absence of a corresponding increase in the total mass of protein (Fig. 2A). As a control, we verified that this observed increase of aggregated mass does not correspond to an increase in the total mass of these 965 proteins under scrutiny (Fig. 2C). An increment in the mass of these 965 proteins that form aggregates, which are expressed at or above their solubility limits, could lead to an increase in the total aggregate mass, while the total cellular protein content could in principle be compensated by a corresponding decrease in the abundance of other nonaggregating proteins, to yield the level trend observed in Fig. 2A. We therefore evaluated the total mass relative to day 1 of the 965 proteins forming aggregates, and observed no change on aging (Fig. 2C), a result also found by considering the larger set of all cellular proteins (Fig. 2A). These results show that this set of 965 proteins increases the overall aggregate mass, but without increasing the total abundance. This result is in accordance with previous evidence, where increased aggregation between young and old worms was observed not to be the result of an increase in expression levels (36). Furthermore, these outcomes are conserved when we do not consider in the analysis proteins involved in forming functional filaments (e.g., cytoskeletal proteins; SI Appendix, Fig. S1), indicating that this phenomenon concerns the proteome as a whole. We thus suggest that the effective solubility threshold is lowered on aging. We found no significant change in the total amount of soluble and aggregating protein in aging worms, despite the previously reported change in the composition of the cellular proteome (42). Hence, even if the reshaping of the composition of the proteome does not involve a change in the total quantity of proteins, it causes an overall change in the cellular environment that results in an increase in the fraction of proteins that is in the form of aggregates, with proteins that increase in abundance contributing further to the aggregate load (42). The sum of the contributions of each of the 965 proteins present in the insoluble pellet reveals that the total aggregate load doubles between days 6 and 12 (Fig. 2B), indicating a corresponding decrease of the effective overall protein solubility. We also analyzed whether proteins with longer turnover times could be those more present in the aggregates. By analyzing the results of a recent experiments in which these turnover times were measured (49), we found that this is indeed the case (SI Appendix, Fig. S2), suggesting that proteins are removed less readily in the aggregate states than their soluble forms.

Taken together, these results indicate, therefore, that the mechanism by which proteins exceed their solubility limits with aging is not simply a consequence of an overall increase in their total abundance, but rather results from the reduction of their effective cellular solubility, a result in agreement with the evidence of observed widespread aggregation (36, 42). We also found that this process is not gradual with age, but manifests sharply between days 6 and 12 of adulthood, after the worms have stopped reproducing. This phenomenon could be the result of either an age-dependent loss of regulatory control of the protein homeostasis system or a time-dependent increase in the quantity of the aggregated states of proteins related to their soluble states, as most of the cellular proteins are supersaturated.

**Proteins Involved in Functional Liquid–Liquid Phase Separation Are Particularly Vulnerable to Age-Dependent Aggregation.** Having observed how the cellular concentrations of proteins are linked to their critical concentrations on a global scale and on aging, we next sought to understand how this relationship could specifically affect the proteins involved in phase-separation phenomena inside cells, which are characterized by the formation of membraneless organelles. Such organelles have been described as resulting from functional liquid–liquid phase transitions, characterized by fast diffusion and exchange rates (seconds to minutes) (11, 50). Most of these highly dynamical structures, known as ribonucleoprotein
granules or ribonucleoprotein droplets, have high proteinaceous and RNA or DNA content (43, 51). An increasing body of literature has revealed that the proteins that form membraneless organelles play central roles in neurodegenerative disorders, and in particular amyotrophic lateral sclerosis and frontotemporal dementia (52–56). Hence, it is of major importance to investigate the relationship between critical concentration and physiological concentration in terms of the proteins involved in age-related phase transition phenomena.

It has been shown that a key requirement for a protein to be able to form membraneless organelles and liquid droplets is the presence of regions of low complexity (LC) in the primary sequence (43). To explore how aging affects, on a global scale, the formation of dynamical functional assemblies driven by liquid–liquid phase separation, we next tested whether proteins that can initiate liquid demixing phenomena in cells are more or less vulnerable toward age-dependent aggregation. Since LC regions have been associated with the capacity of forming membraneless organelles, we first compared the fraction of proteins found in aggregates throughout aging (Fig. 2B and C) that have LC regions in their sequence with the total number of proteins that have LC regions in the whole sample (Fig. 2C). We found that 73% of proteins that form deposits from days 1 to 17 have at least one LC region compared with 67% in the total aging proteome (Fig. 3A). Hence, the direct comparison of the proteins forming aggregates with the remaining ones from the proteome result in a relative increase (proportional fraction) of 14.2%. This increase is highly significant, with a P value of about 10^{-8} (Fisher’s exact test; Materials and Methods and Fig. 3A).

As disorder is only one of the properties of proteins that undergo liquid-demixing, we also directly tested whether the proteins forming aggregates throughout aging were intrinsically more prone to phase-separate by using the recently published predictor of granule formation (catGRANULE), which has been used to determine with high accuracy the propensity of proteins to phase-separate based on key physicochemical properties (44, 56). For each protein measured from days 1 to 17 of adulthood (Fig. 2A), both consistently forming aggregates (Fig. 2B and C) or not, making up the total proteome (Fig. 2A), we evaluated the propensity score for granule formation. We found that the proteins forming aggregates with age have a much higher and strongly significant propensity of undergoing liquid phase transitions, as their granule formation scores are consistently globally higher than the total proteome that comprises them (Fig. 3B).

Combined with the analysis of LC regions, these results indicate that the proteins associated with liquid–liquid phase separation are particularly prone to aggregation during aging. We rationalize this finding as being most likely a consequence of their need to be closer to their solubility limits for functional purposes, and hence more vulnerable to an effective solubility decrease on age-related impairment of the protein homeostasis system.

**Proteins Associated with Homeostasis, Translation, and Cellular Structure Are Primarily Responsible for the Age-Dependent Increase of the Mass of Protein Aggregates.** To add up to the physico-chemical characteristics of the age-dependent solubility decrease a functional perspective, we next analyzed the identity of the cellular proteins forming aggregates from days 1 to 12 and of the subset of these proteins that are most responsible for the increase in the aggregate mass on aging. Functional annotation enrichment (Fig. 4) performed with the DAVID software (57) revealed that the set of 965 proteins found within the deposits from days 1 to 12 is enriched in the gene ontology terms belonging to a wide variety of functions, from translation (green bars), to reproduction (pink bars), to cell cycle (orange bars), metabolism (blue bars), and other classes, also including cellular structure (gray bars, Fig. 4). These enriched terms are consistent with functional enrichment that was previously observed in a widespread protein aggregation study in young versus old C. elegans (36). We found in particular that 32 proteins (SI Appendix, Table S1) contribute most to the change in the aggregate mass relative to day 1 (Fig. 2B and Materials and Methods). These proteins are found in 3 major functional classes: molecular chaperones (in particular the small heat-shock proteins Sip-1 and Hsp25 and the heat shock proteins Hsp70 and Hsp90), proteins involved in RNA-binding and translation (ribosomal components and elongation factors), and a series of structural proteins (including intermediate filaments, actin, and tubulin).

Small heat shock proteins have previously been found to coaggregate with, and to drive, the aggregation of a variety of proteins (58–61), a role consistent with previous reports that aggregation in vivo could have a protective function by sequestering potentially toxic protein species (42, 62–65). Ribosomal proteins and proteins...
related to translation have previously been observed to be significantly enriched in aggregate inclusions of older worms compared with younger ones, and to modulate the lifespan of the organism on RNAi knock-down (37). Proteins belonging to this functional class have also been predicted to be at the highest risk for oxidative damage, which is a dominant source of the loss of protein stability and solubility on aging (66). On oxidative stress, several proteins decrease their solubility because of oxidation. We may expect the levels of these oxidated proteins to be reduced and the levels of the molecular chaperones that regulate them to be increased. A recent study (67) using bulk proteomics has provided initial support for both these expectations. The protein contributing most strongly to the aggregate proliferation is Sip-1, a small heat shock protein that becomes active under acidic conditions and is essential for nematode development and reproduction (68). In addition, Sip-1 has been shown to be an important and specific molecular chaperone for RNA binding proteins and cytoskeletal proteins (68, 69).

Interestingly, the molecular chaperones found in the aggregated form in aging C. elegans (Hsp90, Hsp70 and some small hsps) have also been previously shown to be repressed at the transcriptional level throughout aging, being part of a core-chaperone network required to safeguard the aging proteome (70). Moreover, former evidence highlighted that during aging, the induced expression of molecular chaperones through activation of stress responses is reduced because of epigenetic changes on the genome that impair access of transcription factor to hsp consensus sequences (71). Thus, the observed decrease in solubility tackling a variety of processes and function with particular influence on the translation and homeostasis system is likely the consequence of changes that occur at different and multiple levels during aging, causing a resulting environmental change that shifts the critical concentration of the proteins in the cell.

Conclusions

We have shown that three-quarters of the proteins in C. elegans are expressed at levels close to their solubility limits, and indeed most exceed this value slightly, on average by about 10% (Fig. 1). The existence of a solubility edge provides a rationalization of the widespread expression previously observed (36, 42). These results provide quantitative support for the hypothesis that proteins are expressed at concentrations close to their critical values (16). These findings also reveal that the almost 2-fold increase in the levels of aggregated proteins formed in aged worms compared with young animals is not associated with an overall increase in the total protein concentration, which remains approximately constant during aging. Instead, this change is associated with a decrease in the effective solubility of proteins within the worms (Fig. 2), especially of a subgroup of just more than 30 proteins involved in translation and cellular structure and also associated with the protein homeostasis system (Fig. 4 and SI Appendix, Table S1). In particular, proteins involved in the formation of membraneless organelles are particularly vulnerable to this solubility shift, as they tend to be overrepresented in the group of proteins forming aggregates on aging (Fig. 3). We also note that as the concentrations of proteins could be expected to vary significantly in a wide range of cases, including cell types, cell cycle, stress, and disease, the solubility limits that we described at the whole-worm level represent a soft threshold for the possible concentrations that can be observed in individual cells. With continuing advances in proteomics, we can expect data to become available in the near future to quantify exactly how soft this threshold could be.

Overall, therefore, these results indicate that as proteins are expressed at levels close to their solubility limits, the protein homeostasis system should be always active to maintain them in their soluble states. During the course of aging, however, the ability of this quality control system to keep proteins soluble is no longer capable of preventing the proliferation of aggregates, especially for those proteins, involved functionally in liquid–liquid phase separation phenomena, that need to be expressed closely to their critical solubility limits for functional reasons.

Materials and Methods

The calculation of total and soluble normalized protein abundance in adult worms was carried out as described in SI Appendix, Materials and Methods. The calculation of the changes in total and insoluble protein levels on aging was carried out as described in SI Appendix, Materials and Methods. The bioinformatic analysis was carried out as described in SI Appendix, Materials and Methods. Full methods are available in SI Appendix, Materials and Methods.

Data Availability Statement. All data are provided in the main text and SI Appendix.

ACKNOWLEDGMENTS. We acknowledge support from the Centre for Misfolding Diseases.
