Cell Reports
Article

Widespread Aggregation and Neurodegenerative Diseases Are Associated with Supersaturated Proteins

Prajwal Ciryam,[1,2] Gian Gaetano Tartaglia,[3] Richard I. Morimoto,[2,*] Christopher M. Dobson,[1,*] and Michele Vendruscolo[1,*]
[1]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK
[2]Department of Biochemistry, Molecular Biology and Cell Biology, Rice Institute for Biomedical Research, Northwestern University, Evanston, IL 60208-3500, USA
[3]Centre for Genomic Regulation (CRG) and Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain
*Correspondence: r-morimoto@northwestern.edu (R.I.M.), cmd44@cam.ac.uk (C.M.D.), mv245@cam.ac.uk (M.V.)
http://dx.doi.org/10.1016/j.celrep.2013.09.043

SUMMARY

The maintenance of protein solubility is a fundamental aspect of cellular homeostasis because protein aggregation is associated with a wide variety of human diseases. Numerous proteins unrelated in sequence and structure, however, can misfold and aggregate, and widespread aggregation can occur in living systems under stress or aging. A crucial question in this context is why only certain proteins appear to aggregate readily in vivo, whereas others do not. We identify here the proteins most vulnerable to aggregation as those whose cellular concentrations are high relative to their solubilities. We find that these supersaturated proteins represent a metastable subproteome involved in pathological aggregation during stress and aging and are overrepresented in biochemical processes associated with neurodegenerative disorders. Consequently, such cellular processes become dysfunctional when the ability to keep intrinsically supersaturated proteins soluble is compromised. Thus, the simultaneous analysis of abundance and solubility can rationalize the diverse cellular pathologies linked to neurodegenerative diseases and aging.

INTRODUCTION

Neurodegenerative disorders are increasingly prevalent in our society and represent a very significant challenge to healthcare systems (Balch et al., 2008; Dobson, 2003). A number of explanations of the fundamental origins of these diseases have been proposed, including mitochondrial dysfunction, disruptions of the endoplasmic reticulum and membrane trafficking, effects on protein folding and clearance, and the activation of inflammatory responses (Balch et al., 2008; Dobson, 2003; Querfurth and LaFerla, 2010; Selkoe, 2011). One common feature associated with these conditions, however, is the aggregation of certain peptides and proteins, which generates a cascade of pathological events, including the secondary aggregation of various other proteins and the consequent failure of protein homeostasis to preserve normal biological function (Balch et al., 2008; Dobson, 2003; Gidalevitz et al., 2006; Selkoe, 2011).

Given the evidence that protein aggregation is a widespread phenomenon (Chapman et al., 2006; David et al., 2010; Gidalevitz et al., 2006; Koga et al., 2011; Koplin et al., 2010; Liao et al., 2004; Narayanaswamy et al., 2009; Olzscha et al., 2011; Reis-Rodrigues et al., 2012; Wang et al., 2005; Xia et al., 2008), two key questions are why some proteins, but not others, aggregate in vivo and generate pathological states, and whether or not the identities of these proteins differ substantially between diseases. If particular proteins aggregate in response to specific stresses, different sets of aggregated proteins will appear under each condition. Alternatively, the various sets of aggregating proteins may correspond to a fraction of the proteome with distinctive characteristics that increase the risk of aggregation under many kinds of stress. The latter possibility is consistent with observations that aggregation-prone proteins share general physicochemical features (Chiti et al., 2003; Fernandez-Escamilla et al., 2004; Olzscha et al., 2011; Tartaglia et al., 2008).

Our aim in this work has been to answer a fundamental question about widespread protein aggregation: why certain proteins aggregate in stress, aging, or disease, whereas others do not. To address this problem, we have sought to establish a proteome-wide method of identifying the proteins that are vulnerable to aggregation in vivo. Using this method, we have identified a number of proteins that are expressed at levels that are high relative to their solubilities. These proteins are supersaturated because their concentrations exceed their solubility levels. Early evidence that supersaturation predisposes proteins to aggregate was provided by the finding that the rate and extent of aggregation of hemoglobin S, which is associated with sickle cell anemia, are strongly concentration dependent (Hofrichter et al., 1976). More recently, this idea has been used to compare the aggregation and crystallization behavior of proteins (Yoshimura et al., 2012). Here, we have extended the concept of supersaturation to the proteome level by considering both the unfolded and folded states that can be populated by individual proteins, as well as their association into complexes. Thus, for

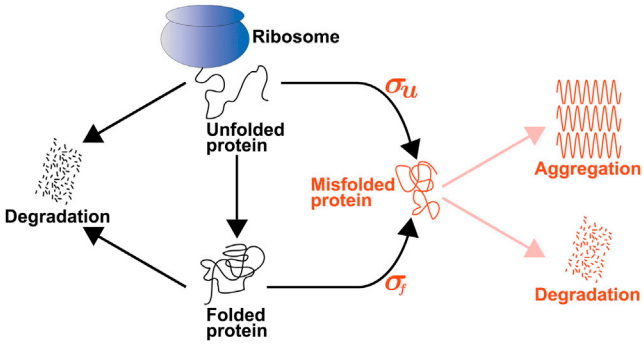

Figure 1. Protein Aggregation In Vivo Can Occur through Different Routes

Proteins may misfold and aggregate as they emerge from the ribosome or when they unfold at least transiently from the native state. The proteins most at risk for aggregation are those whose concentrations are high with respect to their solubilities. We quantify this risk by defining the supersaturation σ_u and σ_f scores. The supersaturation score σ_u, which measures the tendency of proteins to aggregate from the unfolded state, is based on the Zyggregator score (Tartaglia et al., 2008) (Z_{agg}) and mRNA expression levels. The supersaturation score σ_f, which measures the tendency of proteins to aggregate from the folded state, is based on the structurally corrected Zyggregator score (Tartaglia et al., 2008) (Z_{agg}^{SC}) and protein concentration. See also Tables S1 and S3.

example, an intrinsically aggregation-prone protein is not necessarily dangerous unless it is expressed at a relatively high concentration. Similarly, a highly concentrated protein may not be at risk of losing its solubility unless its intrinsic propensity to aggregate is relatively high.

By predicting supersaturation from estimated protein concentrations and aggregation propensities at a proteome scale, we are able to rationalize a variety of phenomena associated with aggregation and misfolding diseases. We find through our analysis of the human and *C. elegans* proteomes that those proteins known to interact with aggregates or to aggregate upon aging are highly supersaturated and that the cellular processes known to be associated with neurodegenerative diseases are at risk of disruption because they involve an exceptionally large number of supersaturated proteins. These results show how the initial appearance of protein aggregates in the presence of other vulnerable proteins can precipitate a series of uncontrolled aggregation events with severe pathological consequences and that proteins in a supersaturated state comprise the subproteome most at risk of misfolding and aggregation under conditions of stress. These proteins and the biochemical pathways to which they belong may be the first to suffer from an impairment of protein homeostasis and, therefore, represent the underlying basis for the cellular damage caused by diseases of misfolding, including neurodegenerative conditions such as Alzheimer's, Huntington's, and Parkinson's diseases.

RESULTS

Prediction of Protein Supersaturation from Concentration and Aggregation Propensity

In order to identify those proteins most at risk of misfolding and aggregating in vivo, we calculated their level of supersaturation using a score that we define in terms of the concentrations of proteins relative to their aggregation propensities (see Experimental Procedures; Supplemental Experimental Procedures). The cellular concentrations of proteins with high supersaturation scores are more likely to exceed their critical values under varying conditions, leading these proteins to become insoluble. We here used the aggregation propensities of proteins as estimates of their solubility because experimental measurements of critical concentrations of proteins in vivo are extremely difficult to carry out at the proteome level. To evaluate the risk of proteins to aggregate from their unfolded or native states, we define the parameters σ_u and σ_f as the supersaturation scores, respectively (Figure 1). The risk of aggregation is different in these two states because in the folded state the most aggregation-prone regions tend to be buried in the core of the structure, and thus they are prevented from forming intermolecular interactions (Tartaglia et al., 2008). The critical concentrations of proteins in their unfolded states thus are generally expected to be lower than in their folded states, hence the necessity of introducing the σ_u and σ_f scores separately. Because the largest pool of unfolded proteins corresponds to newly synthesized proteins, whose concentrations can be estimated from the corresponding mRNA concentrations, we used the logarithmic average of scores derived from microarray analysis of over 70 types of human tissue or of the nematode *C. elegans* at a range of ages to represent levels of newly synthesized proteins (Golden et al., 2008; Su et al., 2004). For folded proteins, in order to define the σ_f score, we used the logarithm of the normalized spectral abundance factors (NSAFs) derived from mass spectrometry (Schrimpf et al., 2012).

We estimated the propensity of proteins to aggregate from the unfolded state using the Z_{agg} score calculated with the Zyggregator method (Tartaglia et al., 2008), which is based on the analysis of the physicochemical properties of amino acid sequences (Chiti et al., 2003). The Zyggregator method employs algorithms that have been parameterized to reproduce the aggregation behavior of a set of known amyloidogenic proteins and has been validated in a series of studies in which it has been shown to lead to accurate predictions of aggregation rates both in vitro and in vivo (Belli et al., 2011; Luheshi et al., 2007; Roodveldt et al., 2012; Tartaglia et al., 2008). For proteins that aggregate from the native state, we used an aggregation propensity score that accounts for the protective effects of the folded structure, which is defined by assigning corrections to the aggregation propensities of individual residues on the basis of the extent of the structural fluctuations that they experience in the folded state and that lead to their temporary exposure to the solvent (Tartaglia et al., 2008). This structurally corrected aggregation propensity score (Z_{agg}^{SC}) has been shown to correlate well with protein solubility determined from an in-vitro-reconstituted bacterial ribosome system (Agostini et al., 2012). We then summed the logarithms of the concentration and aggregation propensity values (see Experimental Procedures; Supplemental Experimental Procedures) to construct a human database of σ_u scores for 16,263 proteins and σ_f scores for 6,155 proteins, and a *C. elegans* database of σ_u scores for 16,623 proteins and σ_f scores for 10,149 proteins (Table S2). The sizes of our databases were limited primarily by the availability of expression and

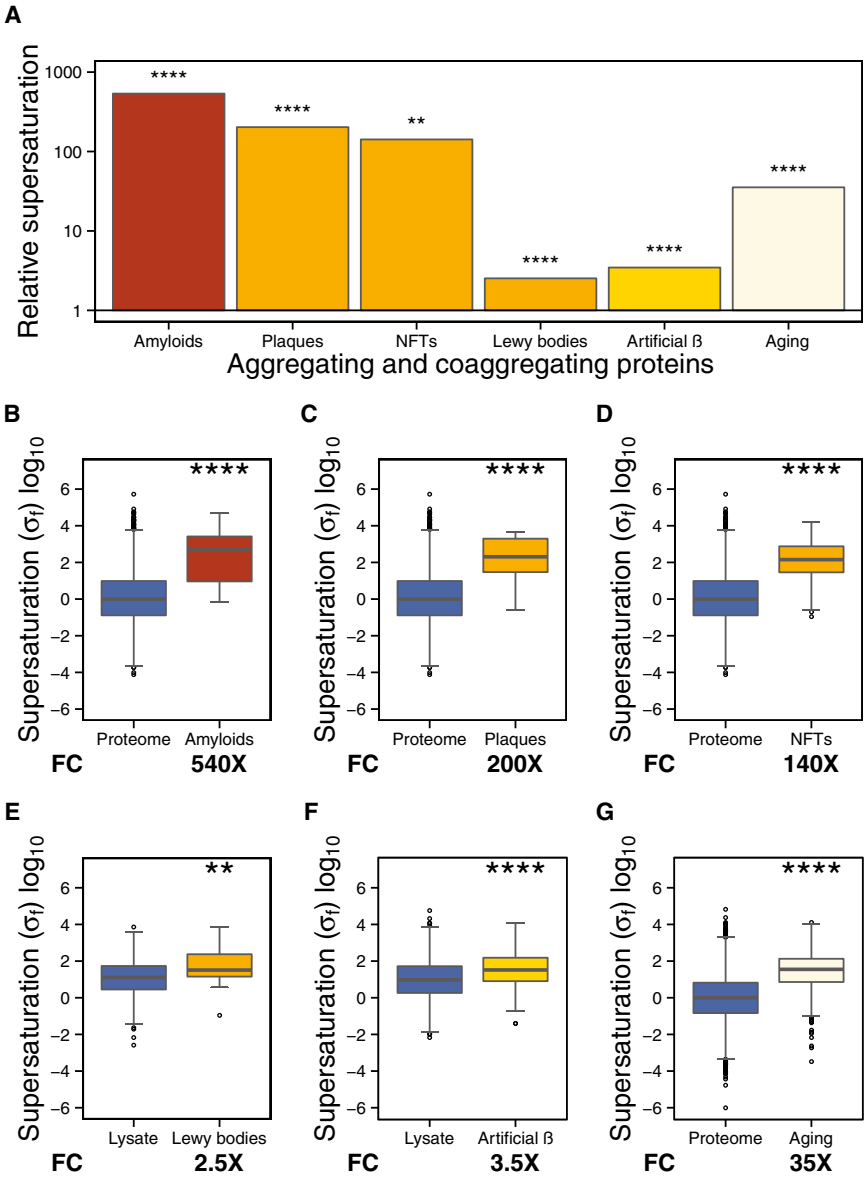

Figure 2. Widespread Protein Aggregation Is Associated with High Supersaturation Scores

(A–G) Summary of the results for the different classes of aggregating proteins analyzed in this work. Results are given in terms of the increases in the supersaturation scores over the average value for the whole proteome or for an experimental lysate ("fold change" [FC]). We compared the supersaturation scores σ_f for (B) the whole proteome and the human "amyloid proteins" in UniProt (UniProt Consortium, 2012), (C) the whole proteome and proteins that coprecipitate in amyloid plaques (Liao et al., 2004), (D) the whole proteome and proteins that coprecipitate in neurofibrillary tangles (NFTs) (Wang et al., 2005), (E) the whole lysate and proteins that coprecipitate in Lewy bodies (Xia et al., 2008), (F) the whole lysate and proteins that coprecipitate in artificial β sheet protein aggregates (Olzscha et al., 2011), and (G) the whole proteome and proteins found to aggregate in *C. elegans* during aging (David et al., 2010; Reis-Rodrigues et al., 2012). Boxplots extend from the lower to the upper quartiles, with the internal lines referring to the median values. Whiskers range from the lowest to highest value data points within 150% of the interquartile ranges. The statistical significance was assessed by the Wilcoxon/Mann-Whitney U test with Bonferroni-corrected p values (*$p < 0.05$, **$p < 0.01$, and ****$p < 0.0001$). See also Figures S1–S3, S5, and S6.

abundance data that could be unambiguously mapped to specific protein identifiers.

Supersaturation Scores Rationalize Widespread Protein Aggregation under Stress

Because a wide range of proteins is known to form fibrillar assemblies within the cell (Chapman et al., 2006; Chiti and Dobson, 2006; David et al., 2010; Fowler et al., 2007; Gidalevitz et al., 2006; Haass and Selkoe, 2007; Hartl et al., 2011; Koplin et al., 2010; Liao et al., 2004; Olzscha et al., 2011; Reis-Rodrigues et al., 2012; Wang et al., 2005; Xia et al., 2008), we investigated the relationship between the phenomena of supersaturation and aggregation. For instance, the function of actin, a highly abundant protein also identified as being highly supersaturated in our calculations (Table S2), relies on its ability to assemble reversibly into filaments, and it has been suggested that given the typical cytosolic concentration of actin, it would always remain in a polymerized form were it not for the presence of specific regulatory mechanisms (Pollard et al., 2000).

More generally, we find that the "amyloid proteins," as annotated in the UniProt database (UniProt Consortium, 2012), have elevated supersaturation scores (Figures 2A, 2B, S1A, S1B, S1H, and S1I). For these proteins, the σ_f score is more than 500-fold greater than the median value over the proteome (540×, $p = 1.9 \times 10^{-5}$), indicating that these proteins are on average at greater risk for aggregation upon accumulation in the cell (Figures 2A and 2B) than when they are in the process of being synthesized (Figures S1A and S1B). This conclusion is consistent with the appearance of such proteins as the predominant constituents of either intracellular inclusions or extracellular deposits in a variety of diseases (Balch et al., 2008; Dobson, 2003).

Given that high supersaturation scores correspond to an increased risk of proteins becoming insoluble, we investigated whether such scores can rationalize additional aspects of protein aggregation, including coaggregation with large, insoluble deposits associated with disease (Liao et al., 2004; Wang et al., 2005; Xia et al., 2008) and artificial β sheet protein (Olzscha et al., 2011) aggregates. Our results indicate that the σ_f score can identify proteins found to incorporate into amyloid plaques

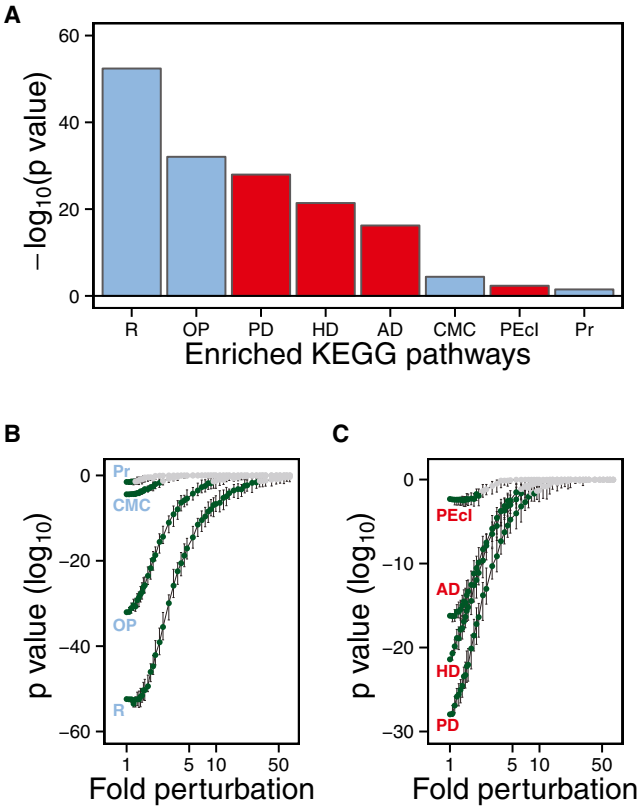

Figure 3. Biochemical Processes Associated with Neurodegenerative Diseases Are Highly Enriched in Supersaturated Proteins

(A) List of the KEGG pathways (Kanehisa et al., 2010) identified here as significantly enriched (Bonferroni-corrected p values) in proteins at or above the 95th percentile of supersaturation (σ_u): R, ribosome; OP, oxidative phosphorylation; PD, Parkinson's disease; HD, Huntington's disease; AD, Alzheimer's disease; CMC, cardiac muscle contraction; PecI, pathogenic *E. coli* infection; Pr, proteasome. Physiological and pathological pathways are shown in blue and red, respectively.

(B and C) Test of the robustness of the significance of the enrichment of the KEGG pathways according to their supersaturation scores. Gaussian noise was introduced 100 independent times into the proteome scores at 50 different levels and plotted (1×, no noise) for (B) physiological pathways, which are robust up to 28× (R), 8.8× (OP), 2.3× (CMC), and 1.2× (Pr), and (C) pathological pathways, which are robust up to 8.2× (PD), 6.2× (HD), 5.5× (AD), and 2.1× (PEcI). Error bars indicate interquartile ranges; green points indicate error levels below the p = 0.05 significance (red dashed line) by the Wilcoxon/Mann-Whitney U test.

See also Figures S4 and S6.

(Liao et al., 2004) (200×, $p = 1.7 \times 10^{-7}$, Figures 2A and 2C), neurofibrillary tangles (Wang et al., 2005) (140×, $p = 1.2 \times 10^{-18}$, Figures 2A and 2D), and Lewy bodies (Xia et al., 2008) (2.5×, $p = 2.9 \times 10^{-3}$, Figures 2A and 2E) isolated from autopsy samples of patients with neurodegenerative disease. We find that proteins that coaggregate with artificial β sheet proteins in human cell cultures (Olzscha et al., 2011) are characterized by increased values of the σ_f score (3.5×, $p = 1.7 \times 10^{-6}$, Figures 2A and 2F), as well. Such proteins also have elevated σ_u scores (1.4×, $p = 1.4 \times 10^{-2}$, Figures S1A and S2F), consistent with observations from pulse-chase experiments that both newly synthesized and preexisting proteins interact with aggregates (Olzscha et al., 2011). The σ_f (Figure 2) and σ_u (Figure S1) scores are broadly consistent for all of these sets, with proteins that coaggregate with large inclusions tending to be relatively highly supersaturated in both the folded and unfolded states.

We then turned our attention to the analysis of proteins that are likely to become susceptible to aggregation when the control of protein homeostasis declines, as in aging (David et al., 2010; Reis-Rodrigues et al., 2012). We observe that the σ_u scores (0.61×, $p = 2.0 \times 10^{-85}$, Figures S1A and S1G) of proteins that aggregate upon aging in *C. elegans* (David et al., 2010; Reis-Rodrigues et al., 2012) are lower than those of the proteome as a whole; by contrast, the σ_f scores are much higher for these proteins (35×, $p = 1.9 \times 10^{-158}$, Figures 2A and 2G). The low values of the σ_u scores can be attributed to the fact that global gene expression in *C. elegans* declines with age (Golden et al., 2008), thus reducing the levels of newly synthesized proteins that may aggregate from their unfolded states. Instead, proteins that aggregate during aging have low expression levels but relatively high concentrations and tend to be metastable by virtue of their high σ_f scores.

Because both the estimates of concentration and the predictions of aggregation propensity are subject to considerable potential errors, to test the robustness of our results against these errors, we introduced Gaussian noise into the calculations of the σ_f and σ_u scores, finding that the results are indeed stable against high levels of noise (in many cases, more than 50-fold error, Figures S2 and S3).

Biochemical Pathways Associated with Neurodegenerative Diseases Are Enriched in Supersaturated Proteins

Because supersaturation scores are able to explain diverse phenomena related directly to aggregation, we wondered whether they might help identify cellular processes particularly sensitive to stress. Therefore, we asked whether particular biochemical pathways are at high risk for disruption by virtue of the supersaturation levels of their constituent proteins. For our analysis, we considered the pathways that are listed in the Kyoto Encyclopedia of Genes and Genomes (KEGG), which are based on manually curated sets of proteins involved in cellular processes or proposed to be involved in disease on the basis of reports in the literature. We used the DAVID software (Huang et al., 2009) to determine if any of the KEGG pathways (Kanehisa et al., 2010) have a large number of human proteins with high σ_u or σ_f scores. We carried out an unbiased search of the roughly 200 KEGG pathways and found only 8 pathways significantly enriched in the proteins with σ_u scores at or above the 95th percentile. Strikingly, all these pathways involve proteins that either form well-defined functional complexes or are related to diseases that involve pathological protein complexes. The three disease pathways that are most dramatically enriched in supersaturated proteins are those of Alzheimer's ($p = 6.0 \times 10^{-17}$), Parkinson's ($p = 1.1 \times 10^{-28}$), and Huntington's ($p = 4.0 \times 10^{-22}$) diseases (Figure 3A). A fourth pathway, involved in pathogenic *E. coli* infection, which is closely associated with the cytoskeleton, is also enriched but to a much less significant

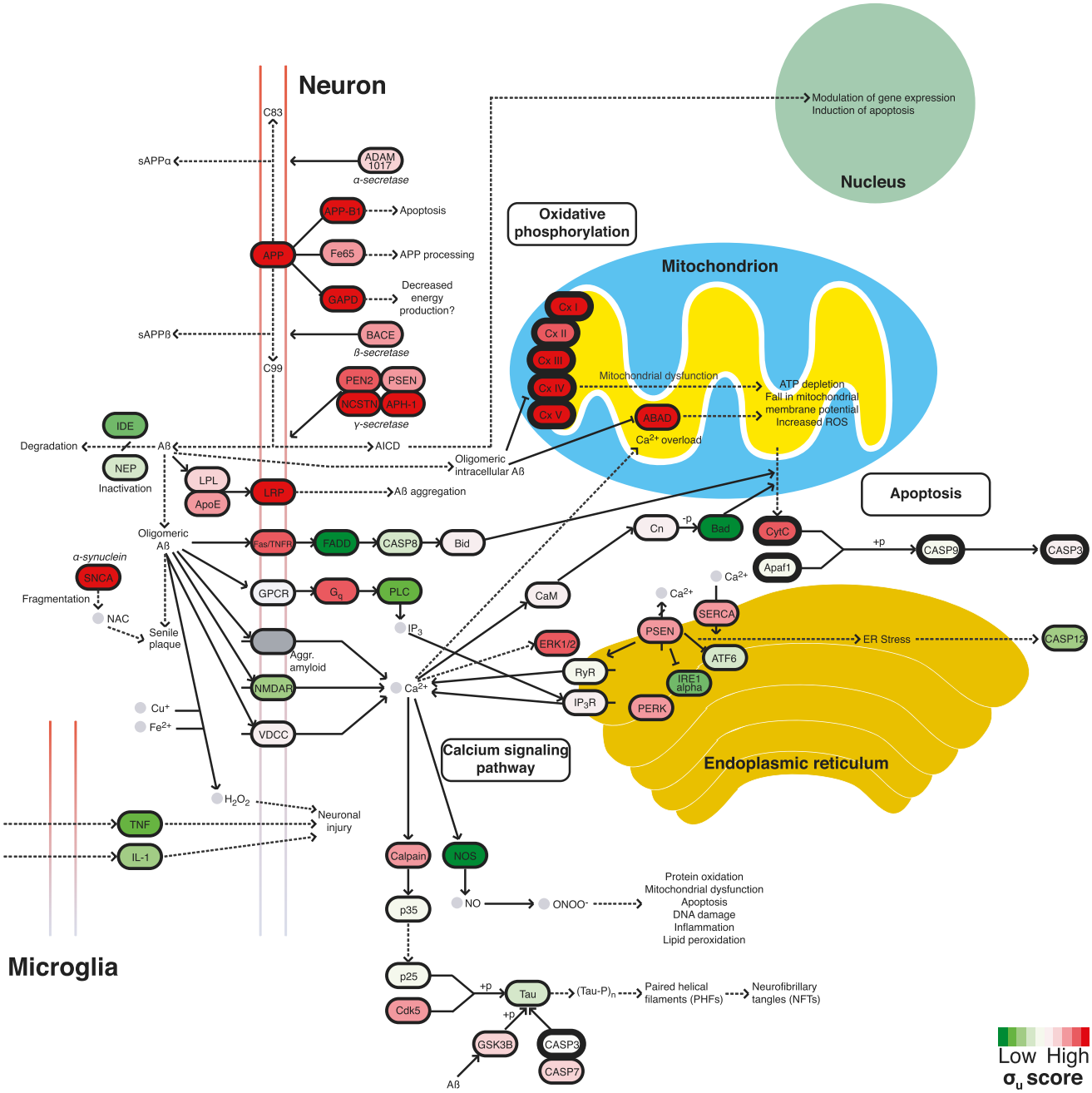

Figure 4. Supersaturated Proteins Are Overrepresented in the KEGG Alzheimer's Disease Pathway

The KEGG pathway for Alzheimer's disease (Kanehisa et al., 2010) is curated from the literature and includes many proteins not directly associated with the disruption of the homeostasis of the Aβ peptide. Those proteins that are part of the overlap (see Figure 5A) of the Alzheimer's, Parkinson's, and Huntington's disease pathways are shown in bold. Approximately 65% of proteins in the Alzheimer's pathway have high σ_u scores. Colors are assigned based on the division of the σ_u database into deciles from low (green) to high (red).

level ($p = 4.6 \times 10^{-3}$). KEGG pathways are broadly constructed, with those related to neurodegenerative diseases including not only proteins known to aggregate but also the proteins that process these aggregates and the cellular systems that become impaired as a result of aggregation (Kanehisa et al., 2010). Despite this disparate collection of proteins, a staggering two-thirds of proteins in the KEGG Alzheimer's disease pathway have supersaturation scores above the median value for the human proteome (Figure 4).

In addition to these disease pathways, we also find that the KEGG pathways associated with the assembly of the ribosome ($p = 4.0 \times 10^{-53}$) and the proteasome ($p = 3.2 \times 10^{-2}$), and

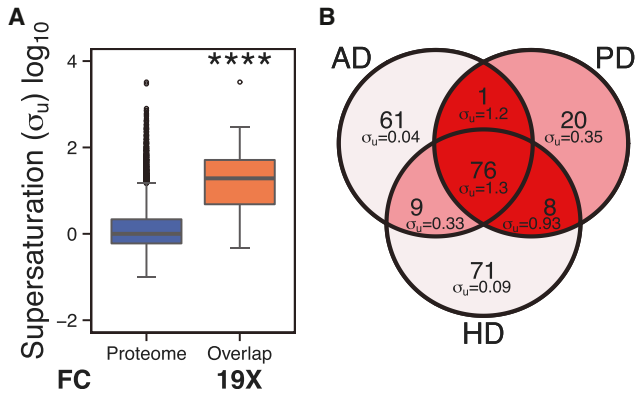

Figure 5. Supersaturated Proteins Are Common to Different Neurodegenerative Pathways

(A) Comparison of the σ_u scores for the proteome and the set of 76 proteins common among the Alzheimer's, Parkinson's, and Huntington's KEGG pathways (Kanehisa et al., 2010); this set of proteins is denoted as "overlap." ****$p < 0.0001$.

(B) Comparison of the σ_u scores for the proteins in the Alzheimer's, Parkinson's, and Huntington's pathways. Colors are assigned based on the division of the σ_u scores into deciles from low to high (color code as in Figure 4).

See also Figures S2–S6.

with the processes of oxidative phosphorylation ($p = 8.8 \times 10^{-33}$) and cardiac muscle contraction ($p = 3.9 \times 10^{-5}$), are enriched in supersaturated proteins (Figure 3A). These results, particularly those associated with the ribosome and oxidative phosphorylation, are highly robust against errors in the calculation of σ_u scores (Figure 3B). All of these pathways involve the assembly of major cellular macromolecular complexes, the components of which must contain interactive surfaces that tend to be aggregation prone (Pechmann et al., 2009). In agreement with this finding, proteins involved in such assemblies, especially the ribosome, have been observed consistently in widespread aggregation studies (David et al., 2010; Koplin et al., 2010; Reis-Rodrigues et al., 2012). Significantly, none of the pathways identified by using supersaturation scores is identifiable from aggregation propensities alone.

These results suggest that widespread aggregation under conditions of stress is defined not only by the specific nature of the stress itself but also by the level of supersaturation of the proteins that aggregate. If this is the case, some proteins should have characteristics that render them susceptible to aggregation under a variety of conditions. To investigate this possibility, we determined the overlap among the Alzheimer's, Parkinson's, and Huntington's disease KEGG pathways. The 76 proteins common to the pathways represented by these diseases have on average a much larger σ_u score (19×, $p = 4.2 \times 10^{-31}$) than that of the proteome as a whole (Figure 5A), or indeed, that of the remaining proteins in the three individual pathways (Figure 5B).

Of the eight pathways enriched in proteins with high σ_u scores, five, including those for Alzheimer's, Parkinson's, and Huntington's diseases, are also enriched in proteins with high σ_f scores, although in the latter case, the data are less statistically significant (Figure S4). The fact that the σ_u scores so strongly single out pathways involved in neurodegeneration suggests that the proteins that misfold and aggregate during the course of these diseases aggregate primarily from unstructured or partially unstructured states. In particular, because many proteins in the KEGG neurodegenerative disease pathways are membrane proteins, they are unlikely to aggregate from their folded state in the membrane environment. They are, however, likely to be at risk during folding or upon removal from the membrane for degradation (MacGurn et al., 2012; Notterpek et al., 1999; Skach, 2009), especially if protein homeostasis dysfunction impairs membrane trafficking, as has been reported by Cooper et al. (2006). Consistent with the view that membrane proteins may have a somewhat elevated risk of aggregation in the unfolded state is that the median σ_u score is modestly elevated for proteins with "membrane" Gene Ontology (Ashburner et al., 2000) tag (1.2×, $p = 1.7 \times 10^{-20}$), whereas the σ_f score is low for such proteins (0.47×, 5.9×10^{-52}). This risk, therefore, is small compared to that of most sets of aggregating proteins in Figure 2. In order to test whether or not the observation that the Alzheimer's, Parkinson's, and Huntington's disease pathways have elevated numbers of supersaturated proteins is simply a consequence of their richness in membrane proteins, we determined whether or not the sets of proteins at or above the 95th percentile of σ_u or σ_f scores are significantly enriched in membrane proteins compared to the proteome at large. Our results indicate that the most supersaturated proteins are not enriched in this way (σ_u : $p = 0.16$, σ_f : $p = 1.0$). However, we do find that the proteins most supersaturated in the unfolded state are enriched in those associated with the Gene Ontology tag "organelle membrane" (σ_u : $p = 6.9 \times 10^{-19}$, σ_f : $p = 1.0$). This result is consistent with recent evidence that membrane proteins in the mitochondrial respiratory chain can misfold when the Parkinson's-associated gene PINK1 is mutated (Pimenta de Castro et al., 2012). In fact, predominant among the overlapping proteins in the Alzheimer's, Parkinson's, and Huntington's disease pathways are the proteins that are members of the respiratory chain.

Because coaggregating proteins are more highly supersaturated in the folded state than in the unfolded state, the ability of the $\boldsymbol{\sigma_u}$ score, in addition to the $\boldsymbol{\sigma_f}$ score, to identify disease pathways suggests at least three possibilities. First, it may be that many proteins involved with disease aggregate independently instead of associating with inclusions, as suggested by the tendency of proteins toward homomeric aggregation (Matsumoto et al., 2006; Rajan et al., 2001; Wright et al., 2005). Second, proteins involved with disease may be degraded if they fail to fold into their native conformation (Goldberg, 2003). Third, because mass spectrometry experiments designed to identify aggregating or coaggregating proteins tend to ignore membrane proteins because of the difficulty of distinguishing them from truly insoluble ones, these experiments may be missing an important component of the disease process. Despite the need for such details to be resolved in the future, our results suggest that supersaturation in the folded state is reporting on significant functional properties of proteins in the cell, as well as on potentially pathological processes. We find, for example, that in addition to proteins that aggregate (Figure 2), those proteins that form functional complexes tend to have high $\boldsymbol{\sigma_f}$ scores (complexes [Licata et al., 2012]: 6.2×, $p = 2.3 \times 10^{-58}$; nuclear complexes

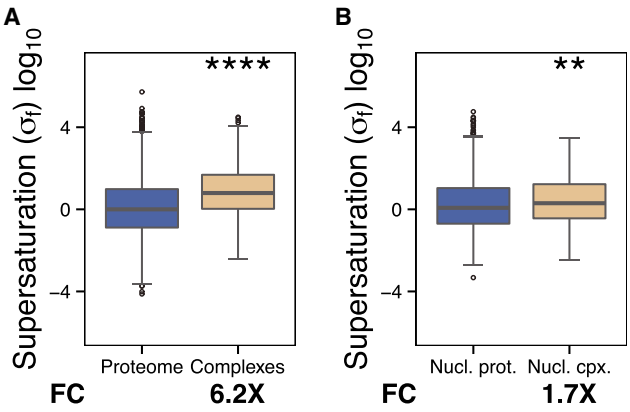

Figure 6. Proteins that Form Complexes Tend to Be More Supersaturated than the Proteome as a Whole

(A and B) We compared the σ_f scores for (A) the proteome and proteins involved in complexes and (B) the nuclear proteome (Nucl. prot.) and proteins involved in nuclear complexes (Nucl. cpx.). Boxplots extend from the lower to the upper quartiles, with the internal lines referring to the median values. Whiskers range from the lowest to highest value data points within 150% of the interquartile ranges. The statistical significance was assessed by the Wilcoxon/Mann-Whitney U test with Bonferroni-corrected p values (**$p < 0.01$ and ****$p < 0.0001$). See also Figures S2 and S5.

[Luc and Tempst, 2004]: 1.7×, $p = 1.2 \times 10^{-3}$, Figure 6). These results are consistent with evidence that the features that mediate normal protein interactions are similar to those that promote aggregation (Pechmann et al., 2009).

To test our results for the σ_u scores, we used another predictor of aggregation from the unfolded state, the TANGO algorithm (Fernandez-Escamilla et al., 2004). A supersaturation score based on this algorithm (σ_{u_T}) (Table S2) produces similar results to those obtained with the σ_u score (Figure S1). Furthermore, to test the possibility that the σ_u and σ_f scores may give too much weight to the expression levels relative to the aggregation propensities, we increased the exponential weight of aggregation propensities in the σ_u and σ_f score. At the highest reweighting that we tested (Figures S5 and S6), supersaturation scores are more strongly correlated with aggregation propensity (human Z_{agg}, σ_u : 0.88; human Z^{SC}_{agg}, σ_f : 0.88; worm Z_{agg}, σ_u : 0.99; and worm Z^{SC}_{agg}, σ_f : 0.96) than concentration score (human mRNA expression, σ_u : 0.35; human protein abundance,σ_f : 0.53; worm mRNA expression, σ_u : 0.10; and worm protein abundance,σ_f : −0.015). We find that most of our results are robust over a wide range of weights for the aggregation propensity (Figures S5 and S6).

Nevertheless, we observed that concentration is a strong predictor of the variety of aggregation-related phenomena that we have analyzed in this work because an important component of the predictive power of σ_f is attributable to this property. Indeed, the widespread aggregation data sets that we considered tend to exhibit protein abundance levels more elevated than corresponding mRNA levels, and whereas they also tend to be more elevated in Z^{SC}_{agg} than Z_{agg}, the difference is much smaller (Figure S1). In addition, the ratio of relative σ_f -to-σ_u values is positively correlated with the ratio of relative protein abundance-to-mRNA levels but negatively correlated with the ratio of relative Z^{SC}_{agg} -to-Z_{agg} levels (Figure S1). The relevance of the concentration levels in rationalizing widespread aggregation measurements is a further indication that the concepts of solubility and supersaturation are key in understanding these data.

Moreover, because both the procedures adopted for pathway construction in the KEGG database (Kanehisa et al., 2010) and the identification of aggregating proteins using mass spectrometry (Olzscha et al., 2011) typically consider only the more abundant proteins, the use of the supersaturation score helps to identify additional proteins that aggregate in disease, particularly those with low concentrations and high aggregation propensities. For example, pathways related to cell surface receptors, including olfactory transduction ($p = 4.0 \times 10^{-90}$) and neuroactive ligand-receptor interaction ($p = 4.9 \times 10^{-5}$), are enriched among those supersaturated proteins that are present at low concentrations.

DISCUSSION

Previous studies that have investigated the causes of proteome-wide aggregation have considered the intrinsic aggregation propensities of proteins (Goldschmidt et al., 2010; Monsellier et al., 2008; Tartaglia et al., 2005; Tartaglia and Vendruscolo, 2010). It has thus been suggested that a diverse collection of proteins can form aggregates, although the presence of molecular chaperones and clearance processes largely prevents proteome-wide aggregation under stress-free conditions (Dobson, 2003; Goldschmidt et al., 2010; Lindquist and Kelly, 2011; Monsellier et al., 2008; Olzscha et al., 2011). It has also been shown that destabilizing mutations may drive soluble proteins toward aggregation and that a genetic background predisposed to such defects can exacerbate this problem (Gidalevitz et al., 2006; Luheshi et al., 2007). More generally, because protein aggregation in the cellular environment has potentially devastating effects, the expression of aggregation-prone proteins is generally maintained at low levels (Tartaglia et al., 2007) and tightly regulated (Gsponer and Babu, 2012). It has been observed, however, that proteins are only just soluble at the levels at which they are expressed in the cell (Tartaglia et al., 2007) and because these trends are conserved in the yeast, mouse, and human proteomes, it has been suggested that monomeric and aggregated forms of proteins are in an effective homeostatic state (Gsponer and Babu, 2012). Similarly, it was recently shown that highly abundant proteins have fewer aggregation-prone surfaces, an observation consistent with their low aggregation propensities (Levy et al., 2012; Tartaglia et al., 2007). Given these evolutionary constraints, it may be surprising that in vivo aggregation is such a common phenomenon under stress (Chapman et al., 2006; Gidalevitz et al., 2006; Koplin et al., 2010; Liao et al., 2004; Narayanaswamy et al., 2009; Olzscha et al., 2011; Wang et al., 2005; Xia et al., 2008) or aging (David et al., 2010; Koga et al., 2011; Reis-Rodrigues et al., 2012). Our results on protein supersaturation provide an explanation for these observations because they indicate that not only the intrinsic propensities of proteins to aggregate but also their cellular concentrations are key factors that distinguish aggregation-prone proteins from those whose homeostasis is more robust (Tables 1 and S2). Although it has

Table 1. Summary of Widespread Aggregation Data sets Analyzed in this Work

Data Set	References	Species	Original Number	Number of UniProt IDs	# σ_u	# σ_f	# σ_{u_T}
Amyloids	UniProt Consortium, 2012	human	27	27	20	13	20
Plaques (coaggregators)	Liao et al., 2004	human	26	26	26	18	26
NFTs (coaggregators)	Wang et al., 2005	human	72	88	75	52	75
Lewy bodies (coaggregators)	Xia et al., 2008	human	36	34	33	22	33
Lewy bodies (detected)	Xia et al., 2008	human	707	557	538	380	538
Artificial β (coaggregators)	Olzscha et al., 2011	human	133	151	141	97	140
Artificial β (lysate)	Olzscha et al., 2011	human	3,055	3,032	2,778	1,858	2,758
Aging aggregators	David et al., 2010	worm	720	719	517	577	512
Aging aggregators	Reis-Rodrigues et al., 2012	worm	203	203	160	178	159
Complexes	Licata et al., 2012	human	1,729	1,637	–	941	–
Nuclear complexes	Luc and Tempst, 2004	human	604	630	–	319	–
Nuclear proteome	Ashburner et al., 2000; Huang et al., 2009	human	–	1,942	–	1,942	–
Membrane proteins	Ashburner et al., 2000; Huang et al., 2009	human	–	–	6,167	2,285	6,140

"Data Set" is a description of the data used. "References" indicate the references for the data. "Species" indicates the species to which the data refer. "Original Number" is the number of data points listed in the original published set. "Number of UniProt IDs" is the number of data points listed after conversion to UniProt ID. For human proteins, only the reviewed UniProt IDs are counted. No. σ_u, no. σ_f, and no. σ_{u_T} are the number of those proteins in the Number of UniProt IDs column, the number for which the given score is available.
See also Table S1.

been observed that widespread aggregation occurs upon overexpression, which raises the supersaturation levels (Gsponer and Babu, 2012; Narayanaswamy et al., 2009; Sopko et al., 2006), our results indicate that a substantial fraction of the proteome is intrinsically supersaturated and therefore requires the constant aid of quality control mechanisms such as molecular chaperones to remain soluble.

The example of serum albumin illustrates some of the strategies adopted by supersaturated proteins to avoid aggregation, as well as their limitations. Albumin, which is exceptionally abundant, is classified in our analysis as supersaturated (Table S2). This protein, which is ubiquitous in the serum, is usually considered to be very soluble, and yet it has been observed to aggregate in vitro (Costantino et al., 1995; Maruyama et al., 2001) and to form toxic aggregates in the synovial fluid of patients with rheumatoid arthritis (Oates et al., 2006). These findings reflect two important aspects that regulate the behavior of supersaturated proteins: the volume that they occupy, and the interactions that they form. Although albumin is highly abundant in the serum, its concentration is still relatively low owing to the large volume of the serum itself. By contrast, when confined in the synovial fluid, which has a smaller volume, the concentration of albumin may become substantially higher. In addition, albumin forms numerous complexes with proteins and other molecules, which may protect it from aggregation, as has been observed for the ribosomal proteins (David et al., 2010; Koplin et al., 2010; Reis-Rodrigues et al., 2012). Future studies should account for the volume occupied by proteins and the complexes they form in order to increase the accuracy of supersaturation predictions.

Although abundant proteins have evolved to be more soluble than those that are of low abundance (Tartaglia et al., 2007), some proteins are expressed at such high concentrations that it may be impossible for them to achieve the necessary solubility with the constraints of functionality and the need for a stable hydrophobic core. We have indeed found that abundance itself is also a good predictor of widespread aggregation in vivo. This result indicates that highly abundant proteins are intrinsically more at risk of aggregation than low-abundant proteins (Table S3), which in turn suggests that highly abundant proteins must be maintained at high solubility levels by the protein homeostasis system. These proteins are therefore more susceptible to aggregation in processes that impair protein homeostasis, such as stress, aging, or disease. The strong predictive power of abundance underscores the importance of solubility in this phenomenon.

Supersaturated proteins are kinetically, but not thermodynamically, stable in their soluble states (Baldwin et al., 2011; Gazit, 2002; Yoshimura et al., 2012) and are, thus, likely to be highly dependent on the systems that control protein homeostasis in order to remain soluble. The instability of supersaturated proteins is thus expected to arise from the failure of cellular systems that contribute to keeping them soluble. The disruption of this machinery under stress and in disease conditions leads to the aggregation of such proteins by shifting the protein homeostasis boundary for their solubility (Hutt et al., 2009; Taylor and Dillin, 2011). The present study suggests that a widespread failure to maintain proteins in their soluble functional states underlies the diverse and complex pathophysiology of neurodegenerative diseases. The sensitivity of neurons to protein aggregation, therefore, may arise from their high dependence on classes of proteins, such as those identified here, that are inherently and unavoidably at risk. Thus, the initial aggregation of the primary amyloid proteins, such as Aβ and α-synuclein, may trigger an aggregation cascade that disrupts cellular pathways involving these supersaturated proteins.

In conclusion, we have shown that the presence of a large number of supersaturated proteins in the human proteome rationalizes a wide variety of aggregation phenomena associated with aging and disease. The finding that such proteins are overrepresented in a broad range of biochemical processes linked to neurodegenerative diseases reveals why such processes are particularly vulnerable to the appearance of aggregated species or other factors that compromise proteins homeostasis. We anticipate that the type of analysis that we have described can provide a general and widely applicable basis for tracking the instability of proteomes under specific circumstances. By exploiting recent advances in techniques for proteomic analysis, it may soon become possible to use supersaturation measures to assess quantitatively the vulnerability of the human proteome to aggregation and the risk of neurodegenerative disease in individuals over the course of their lives.

EXPERIMENTAL PROCEDURES

The concentration and aggregation propensity of a given protein were combined to produce the supersaturation score $\sigma = C + Z$, where C is the logarithm of the concentration derived from the mRNA expression or protein abundance levels (see Tables S1 and S2), and Z is the Zyggregator score. The σ scores were then recentered such that the median of each database was zero and used to analyze proteins associated with widespread aggregation (Chapman et al., 2006; David et al., 2010; Gidalevitz et al., 2006; Koplin et al., 2010; Liao et al., 2004; Olzscha et al., 2011; Reis-Rodrigues et al., 2012; Wang et al., 2005; Xia et al., 2008), by comparing them to a control lysate when available, or to the full proteome database otherwise. Similar procedures were followed for protein complexes (Licata et al., 2012; Luc and Tempst, 2004). A summary of data sets used in this study is provided in Tables 1 and S2. DAVID (Gidalevitz et al., 2006) was used to find KEGG (Kanehisa et al., 2010) pathways enriched in proteins at or above the 95th percentile of each σ score, with the full database set as background. Error tests were performed by introducing Gaussian noise into the full σ_U and σ_f databases in 100 independent trials or by changing the weighting of aggregation propensity in the uncentered scores, and then recomputing Mann-Whitney U p values and median fold changes each time.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.09.043.

ACKNOWLEDGMENTS

P.C. was supported by grants from the US-UK Fulbright Commission and St. John's College, University of Cambridge. R.I.M. was supported by grants from the National Institutes of Health (NIGMS, NIA, and NINDS), the Ellison Medical Foundation, and the Daniel F. and Ada L. Rice Foundation. C.M.D. and M.V. were supported by grants from the Wellcome Trust and the UK Biotechnology and Biological Sciences Research Council. P.C. thanks Francis J. DiTraglia for valuable discussions regarding robustness testing.

Received: May 6, 2013
Revised: August 12, 2013
Accepted: September 27, 2013
Published: October 31, 2013

REFERENCES

Agostini, F., Vendruscolo, M., and Tartaglia, G.G. (2012). Sequence-based prediction of protein solubility. J. Mol. Biol. *421*, 237–241.

Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., et al.; The Gene Ontology Consortium. (2000). Gene ontology: tool for the unification of biology. Nat. Genet. *25*, 25–29.

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

Baldwin, A.J., Knowles, T.P.J., Tartaglia, G.G., Fitzpatrick, A.W., Devlin, G.L., Shammas, S.L., Waudby, C.A., Mossuto, M.F., Meehan, S., Gras, S.L., et al. (2011). Metastability of native proteins and the phenomenon of amyloid formation. J. Am. Chem. Soc. *133*, 14160–14163.

Belli, M., Ramazzotti, M., and Chiti, F. (2011). Prediction of amyloid aggregation in vivo. EMBO Rep. *12*, 657–663.

Chapman, E., Farr, G.W., Usaite, R., Furtak, K., Fenton, W.A., Chaudhuri, T.K., Hondorp, E.R., Matthews, R.G., Wolf, S.G., Yates, J.R., et al. (2006). Global aggregation of newly translated proteins in an *Escherichia coli* strain deficient of the chaperonin GroEL. Proc. Natl. Acad. Sci. USA *103*, 15800–15805.

Chiti, F., and Dobson, C.M. (2006). Protein misfolding, functional amyloid, and human disease. Annu. Rev. Biochem. *75*, 333–366.

Chiti, F., Stefani, M., Taddei, N., Ramponi, G., and Dobson, C.M. (2003). Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature *424*, 805–808.

Cooper, A.A., Gitler, A.D., Cashikar, A., Haynes, C.M., Hill, K.J., Bhullar, B., Liu, K.N., Xu, K.X., Strathearn, K.E., Liu, F., et al. (2006). Alpha-synuclein blocks ER-Golgi traffic and Rab1 rescues neuron loss in Parkinson's models. Science *313*, 324–328.

Costantino, H.R., Langer, R., and Klibanov, A.M. (1995). Aggregation of a lyophilized pharmaceutical protein, recombinant human albumin: effect of moisture and stabilization by excipients. Biotechnology (N. Y.) *13*, 493–496.

David, D.C., Ollikainen, N., Trinidad, J.C., Cary, M.P., Burlingame, A.L., and Kenyon, C. (2010). Widespread protein aggregation as an inherent part of aging in *C. elegans*. PLoS Biol. *8*, e1000450.

Dobson, C.M. (2003). Protein folding and misfolding. Nature *426*, 884–890.

Fernandez-Escamilla, A.M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004). Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat. Biotechnol. *22*, 1302–1306.

Fowler, D.M., Koulov, A.V., Balch, W.E., and Kelly, J.W. (2007). Functional amyloid—from bacteria to humans. Trends Biochem. Sci. *32*, 217–224.

Gazit, E. (2002). The "Correctly Folded" state of proteins: is it a metastable state? Angew. Chem. Int. Ed. Engl. *41*, 257–259.

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

Goldberg, A.L. (2003). Protein degradation and protection against misfolded or damaged proteins. Nature *426*, 895–899.

Golden, T.R., Hubbard, A., Dando, C., Herren, M.A., and Melov, S. (2008). Age-related behaviors have distinct transcriptional profiles in *Caenorhabditis elegans*. Aging Cell *7*, 850–865.

Goldschmidt, L., Teng, P.K., Riek, R., and Eisenberg, D. (2010). Identifying the amylome, proteins capable of forming amyloid-like fibrils. Proc. Natl. Acad. Sci. USA *107*, 3487–3492.

Gsponer, J., and Babu, M.M. (2012). Cellular strategies for regulating functional and nonfunctional protein aggregation. Cell Rep. *2*, 1425–1437.

Haass, C., and Selkoe, D.J. (2007). Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. Nat. Rev. Mol. Cell Biol. *8*, 101–112.

Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. Nature *475*, 324–332.

Hofrichter, J., Ross, P.D., and Eaton, W.A. (1976). Supersaturation in sickle cell hemoglobin solutions. Proc. Natl. Acad. Sci. USA *73*, 3035–3039.

Huang, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. *4*, 44–57.

Hutt, D.M., Powers, E.T., and Balch, W.E. (2009). The proteostasis boundary in misfolding diseases of membrane traffic. FEBS Lett. *583*, 2639–2646.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., and Hirakawa, M. (2010). KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res. *38*(Database issue), D355–D360.

Koga, H., Kaushik, S., and Cuervo, A.M. (2011). Protein homeostasis and aging: the importance of exquisite quality control. Ageing Res. Rev. *10*, 205–215.

Koplin, A., Preissler, S., Ilina, Y., Koch, M., Scior, A., Erhardt, M., and Deuerling, E. (2010). A dual function for chaperones SSB-RAC and the NAC nascent polypeptide-associated complex on ribosomes. J. Cell Biol. *189*, 57–68.

Levy, E.D., De, S., and Teichmann, S.A. (2012). Cellular crowding imposes global constraints on the chemistry and evolution of proteomes. Proc. Natl. Acad. Sci. USA *109*, 20461–20466.

Liao, L., Cheng, D., Wang, J., Duong, D.M., Losik, T.G., Gearing, M., Rees, H.D., Lah, J.J., Levey, A.I., and Peng, J. (2004). Proteomic characterization of postmortem amyloid plaques isolated by laser capture microdissection. J. Biol. Chem. *279*, 37061–37068.

Licata, L., Briganti, L., Peluso, D., Perfetto, L., Iannuccelli, M., Galeota, E., Sacco, F., Palma, A., Nardozza, A.P., Santonico, E., et al. (2012). MINT, the molecular interaction database: 2012 update. Nucleic Acids Res. *40*(Database issue), D857–D861.

Lindquist, S.L., and Kelly, J.W. (2011). Chemical and biological approaches for adapting proteostasis to ameliorate protein misfolding and aggregation diseases: progress and prognosis. Cold Spring Harb. Perspect. Biol. *3*, a004507.

Luc, P.V., and Tempst, P. (2004). PINdb: a database of nuclear protein complexes from human and yeast. Bioinformatics *20*, 1413–1415.

Luheshi, L.M., Tartaglia, G.G., Brorsson, A.C., Pawar, A.P., Watson, I.E., Chiti, F., Vendruscolo, M., Lomas, D.A., Dobson, C.M., and Crowther, D.C. (2007). Systematic in vivo analysis of the intrinsic determinants of amyloid Beta pathogenicity. PLoS Biol. *5*, e290.

MacGurn, J.A., Hsu, P.C., and Emr, S.D. (2012). Ubiquitin and membrane protein turnover: from cradle to grave. Annu. Rev. Biochem. *81*, 231–259.

Maruyama, T., Katoh, S., Nakajima, M., and Nabetani, H. (2001). Mechanism of bovine serum albumin aggregation during ultrafiltration. Biotechnol. Bioeng. *75*, 233–238.

Matsumoto, G., Kim, S., and Morimoto, R.I. (2006). Huntingtin and mutant SOD1 form aggregate structures with distinct molecular properties in human cells. J. Biol. Chem. *281*, 4477–4485.

Monsellier, E., Ramazzotti, M., Taddei, N., and Chiti, F. (2008). Aggregation propensity of the human proteome. PLoS Comp. Biol. *4*, e1000199.

Narayanaswamy, R., Levy, M., Tsechansky, M., Stovall, G.M., O'Connell, J.D., Mirrielees, J., Ellington, A.D., and Marcotte, E.M. (2009). Widespread reorganization of metabolic enzymes into reversible assemblies upon nutrient starvation. Proc. Natl. Acad. Sci. USA *106*, 10147–10152.

Notterpek, L., Ryan, M.C., Tobler, A.R., and Shooter, E.M. (1999). PMP22 accumulation in aggresomes: implications for CMT1A pathology. Neurobiol. Dis. *6*, 450–460.

Oates, K.M.N., Krause, W.E., Jones, R.L., and Colby, R.H. (2006). Rheopexy of synovial fluid and protein aggregation. J. R. Soc. Interface *3*, 167–174.

Olzscha, H., Schermann, S.M., Woerner, A.C., Pinkert, S., Hecht, M.H., Tartaglia, G.G., Vendruscolo, M., Hayer-Hartl, M., Hartl, F.U., and Vabulas, R.M. (2011). Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. Cell *144*, 67–78.

Pechmann, S., Levy, E.D., Tartaglia, G.G., and Vendruscolo, M. (2009). Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. Proc. Natl. Acad. Sci. USA *106*, 10159–10164.

Pimenta de Castro, I., Costa, A.C., Lam, D., Tufi, R., Fedele, V., Moisoi, N., Dinsdale, D., Deas, E., Loh, S.H., and Martins, L.M. (2012). Genetic analysis of mitochondrial protein misfolding in *Drosophila melanogaster*. Cell Death Differ. *19*, 1308–1316.

Pollard, T.D., Blanchoin, L., and Mullins, R.D. (2000). Molecular mechanisms controlling actin filament dynamics in nonmuscle cells. Annu. Rev. Biophys. Biomol. Struct. *29*, 545–576.

Querfurth, H.W., and LaFerla, F.M. (2010). Alzheimer's disease. N. Engl. J. Med. *362*, 329–344.

Rajan, R.S., Illing, M.E., Bence, N.F., and Kopito, R.R. (2001). Specificity in intracellular protein aggregation and inclusion body formation. Proc. Natl. Acad. Sci. USA *98*, 13060–13065.

Reis-Rodrigues, P., Czerwieniec, G., Peters, T.W., Evani, U.S., Alavez, S., Gaman, E.A., Vantipalli, M., Mooney, S.D., Gibson, B.W., Lithgow, G.J., and Hughes, R.E. (2012). Proteomic analysis of age-dependent changes in protein solubility identifies genes that modulate lifespan. Aging Cell *11*, 120–127.

Roodveldt, C., Andersson, A., De Genst, E.J., Labrador-Garrido, A., Buell, A.K., Dobson, C.M., Tartaglia, G.G., and Vendruscolo, M. (2012). A rationally designed six-residue swap generates comparability in the aggregation behavior of α-synuclein and β-synuclein. Biochemistry *51*, 8771–8778.

Schrimpf, S.P., von Mering, C., Bendixen, E., Heazlewood, J.L., Bumann, D., Omenn, G., and Hengartner, M.O. (2012). The initiative on Model Organism Proteomes (iMOP) Session September 6, 2011, Geneva, Switzerland. Proteomics *12*, 346–350.

Selkoe, D.J. (2011). Alzheimer's disease. Cold Spring Harb. Perspect. Biol. *3*, a004457.

Skach, W.R. (2009). Cellular mechanisms of membrane protein folding. Nat. Struct. Mol. Biol. *16*, 606–612.

Sopko, R., Huang, D.Q., Preston, N., Chua, G., Papp, B., Kafadar, K., Snyder, M., Oliver, S.G., Cyert, M., Hughes, T.R., et al. (2006). Mapping pathways and phenotypes by systematic gene overexpression. Mol. Cell *21*, 319–330.

Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G., et al. (2004). A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. USA *101*, 6062–6067.

Tartaglia, G.G., and Vendruscolo, M. (2010). Proteome-level interplay between folding and aggregation propensities of proteins. J. Mol. Biol. *402*, 919–928.

Tartaglia, G.G., Pellarin, R., Cavalli, A., and Caflisch, A. (2005). Organism complexity anti-correlates with proteomic beta-aggregation propensity. Protein Sci. *14*, 2735–2740.

Tartaglia, G.G., Pechmann, S., Dobson, C.M., and Vendruscolo, M. (2007). Life on the edge: a link between gene expression levels and aggregation rates of human proteins. Trends Biochem. Sci. *32*, 204–206.

Tartaglia, G.G., Pawar, A.P., Campioni, S., Dobson, C.M., Chiti, F., and Vendruscolo, M. (2008). Prediction of aggregation-prone regions in structured proteins. J. Mol. Biol. *380*, 425–436.

Taylor, R.C., and Dillin, A. (2011). Aging as an event of proteostasis collapse. Cold Spring Harb. Perspect. Biol. *3*, a004440.

UniProt Consortium. (2012). Reorganizing the protein space at the universal protein resource (UniProt). Nucleic Acids Res. *40*(Database issue), D71–D75.

Wang, Q., Woltjer, R.L., Cimino, P.J., Pan, C., Montine, K.S., Zhang, J., and Montine, T.J. (2005). Proteomic analysis of neurofibrillary tangles in Alzheimer disease identifies GAPDH as a detergent-insoluble paired helical filament tau binding protein. FASEB J. *19*, 869–871.

Wright, C.F., Teichmann, S.A., Clarke, J., and Dobson, C.M. (2005). The importance of sequence diversity in the aggregation and evolution of proteins. Nature *438*, 878–881.

Xia, Q., Liao, L., Cheng, D., Duong, D.M., Gearing, M., Lah, J.J., Levey, A.I., and Peng, J. (2008). Proteomic identification of novel proteins associated with Lewy bodies. Front. Biosci. *13*, 3850–3856.

Yoshimura, Y., Lin, Y.X., Yagi, H., Lee, Y.H., Kitayama, H., Sakurai, K., So, M., Ogi, H., Naiki, H., and Goto, Y. (2012). Distinguishing crystal-like amyloid fibrils and glass-like amorphous aggregates from their kinetics of formation. Proc. Natl. Acad. Sci. USA *109*, 14446–14451.