Cell Reports
Resource

Cell PRESS
Open ACCESS

Cellular Strategies for Regulating Functional and Nonfunctional Protein Aggregation

Jörg Gsponer[1,*] and M. Madan Babu[2,*]
[1]Centre for High-Throughput Biology, Department of Biochemistry and Molecular Biology, University of British Columbia, East Mall, Vancouver V6T 1Z4, Canada
[2]MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK
*Correspondence: gsponer@chibi.ubc.ca (J.G.), madanm@mrc-lmb.cam.ac.uk (M.M.B.)
http://dx.doi.org/10.1016/j.celrep.2012.09.036

SUMMARY

Growing evidence suggests that aggregation-prone proteins are both harmful and functional for a cell. How do cellular systems balance the detrimental and beneficial effect of protein aggregation? We reveal that aggregation-prone proteins are subject to differential transcriptional, translational, and degradation control compared to nonaggregation-prone proteins, which leads to their decreased synthesis, low abundance, and high turnover. Genetic modulators that enhance the aggregation phenotype are enriched in genes that influence expression homeostasis. Moreover, genes encoding aggregation-prone proteins are more likely to be harmful when overexpressed. The trends are evolutionarily conserved and suggest a strategy whereby cellular mechanisms specifically modulate the availability of aggregation-prone proteins to (1) keep concentrations below the critical ones required for aggregation and (2) shift the equilibrium between the monomeric and oligomeric/aggregate form, as explained by Le Chatelier's principle. This strategy may prevent formation of undesirable aggregates and keep functional assemblies/aggregates under control.

INTRODUCTION

The process of protein aggregation has been linked to several human pathologies, such as Alzheimer's and Parkinson's disease (Chiti and Dobson, 2006). While the potentially harmful effects of protein aggregation have been well established by several studies, it is less often emphasized that protein aggregation can also have beneficial effects to cellular systems. A number of recent studies have shown that several human physiological processes depend on protein aggregation or even fibril formation (Fowler et al., 2007; Reijns et al., 2008; Salazar et al., 2010). Remarkably, the dynamic formation of a variety of cellular bodies, such as stress granules and processing bodies, has been shown to depend on protein aggregation (Balagopal and Parker, 2009). For instance, assembly of stress granules is mediated by aggregation of a glutamine-rich domain in the RNA-binding proteins TIA-1 (Gilks et al., 2004) and Pum (Salazar et al., 2010). Similarly, glutamine/asparagine (Q/N)-rich segments have been shown to be essential for the formation of processing bodies. Although it is unlikely that all aggregates formed in these cellular bodies have a fibrillar character, it is certain that the aggregation propensity of proteins has been exploited to mediate the formation of these assemblies (Fiumara et al., 2010; Salazar et al., 2010). Nonetheless, recent studies have shown that certain protein interactions (for example, hdm2-arf) indeed involve formation of amyloid-like structures (Sivakolundu et al., 2008) and that several peptide and protein hormones are stored in an amyloid-like conformation within cells (Maji et al., 2009).

The observations that extant genomes contain a significant proportion of proteins with the potential to form aggregates and that stretches of aggregation-prone regions are evolutionarily conserved (see Extended Results; Figure S1) suggest that, though potentially harmful, such regions might be structurally and functionally important (Goldschmidt et al., 2010; Linding et al., 2004; Monsellier et al., 2008). For instance, they may be part of the essential hydrophobic core of globular proteins (Linding et al., 2004) or may form patches that mediate protein interactions (Masino et al., 2011; Pechmann et al., 2009). Taken together, these considerations raise the following fundamental questions: (1) how do cells minimize the likelihood of spontaneous aggregation of proteins containing aggregation-prone regions? (2) How are functional aggregates kept under control? The fact that protein aggregation can have harmful effects suggests that "nonfunctional" aggregation should be avoided and "functional" aggregation has to be highly regulated. Indeed, for individual cases of functional aggregates, control mechanisms that regulate the aggregation process have been identified (Fowler et al., 2007). However, very little is known about the regulation of the majority of proteins that are known to form aggregates in a cell or that contain evolutionarily conserved aggregation-prone segments. We hypothesized that cellular systems could have evolved regulatory mechanisms to keep protein aggregation under control by ensuring that the levels of these proteins are low and that they are turned over rapidly. In this work, we present evidence that supports this hypothesis, define a framework for protein aggregation regulation, and discuss its implications.

RESULTS

Identification of Aggregation and Nonaggregation Prone Proteins

Protein aggregates that have been linked to human disease and those found in several functional complexes are primarily

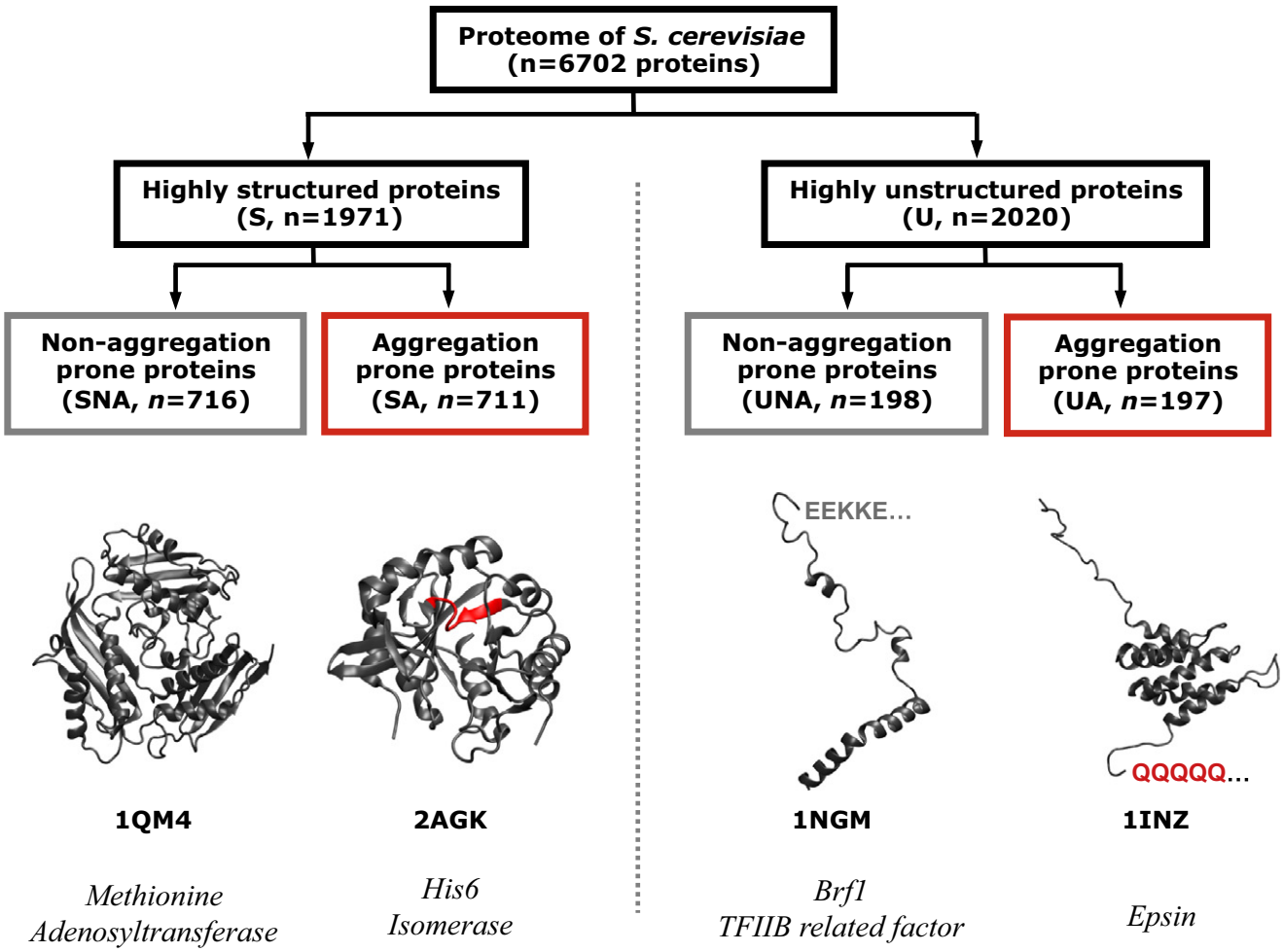

Figure 1. Identification of Structured and Unstructured Nonaggregation-Prone Proteins

The *S. cerevisiae* proteome was grouped into four categories: highly structured protein without aggregation prone elements (SNA), highly structured proteins with aggregation prone elements (SA), highly unstructured proteins with nonaggregating K/E-rich stretches (UNA), and highly unstructured proteins with aggregation prone Q/N-rich stretches (UA). PDB codes are provided for the structures as the four-letter code.

See also Figure S1 and Table S4.

beta-sheet aggregates (Chiti and Dobson, 2006; Fowler et al., 2007; Maji et al., 2009). Though the morphologies (e.g., fibrillar or amorphous) of aggregates may differ, their formation depends on the propensity to form beta-sheets (Rousseau et al., 2006a). We therefore aimed to identify proteins that are likely to form beta-sheet aggregates, irrespective of the morphology of the resulting aggregate. Increased beta-sheet aggregation potential of proteins is associated with the presence of aggregation-prone elements, such as hydrophobic and Q/N-rich stretches (Krobitsch and Lindquist, 2000; Michelitsch and Weissman, 2000). The former are predominantly found buried within folded domains and may need to be exposed to form aggregates, while the latter are often part of unstructured segments and do not have the requirement to unfold to form aggregates (Chen et al., 2001; Linding et al., 2004). Therefore, we first distinguished the proteins in *S. cerevisiae* that are highly structured (S) or highly unstructured (U) (Gsponer et al., 2008) in order to identify the aggregation-prone proteins in this proteome (see Extended Experimental Procedures).

We then identified aggregation-prone, structured proteins by detecting stretches of hydrophobic amino acids using the TANGO algorithm (Fernandez-Escamilla et al., 2004). As proteins with low aggregation likelihood, we identified those highly structured proteins that lack any stretch of consecutive hydrophobic residues. In the highly unstructured proteins, we detected Q/N-rich and lysine/glutamate (K/E)-rich regions using the algorithm described by Michelitsch and Weissman (2000), as their presence in unstructured proteins has been associated with increased and decreased aggregation likelihood, respectively (Lawrence et al., 2007; Santner et al., 2012) (see Extended Results).

In this manner, we divided the *S. cerevisiae* proteome into four groups (Figure 1) and investigated whether the aggregation-prone proteins are regulated differently from the nonaggregation-prone proteins by integrating this structural information

with different genome-scale data sets that describe most of the regulatory steps that influence protein synthesis or degradation (Tables 1 and S1). Additional sequence and structure features, such as the thermodynamic stability of a protein, its folding/unfolding pathway, and involvement in physical interactions with other proteins in a cell, will affect the manifestation of the aggregation-prone elements described above and, ultimately, the likelihood of the protein to aggregate in vivo. While the importance of these features has been investigated using individual proteins (Masino et al., 2011; Münch and Bertolotti, 2010), they are not trivial to assess on a genomic scale.

Transcripts Encoding Aggregation-Prone Proteins Are Present in Low Levels due to Slower Transcription Rate

A comparison of transcript levels revealed that messenger RNAs (mRNA) encoding aggregation-prone proteins are expressed at lower levels than transcripts encoding nonaggregation-prone proteins (Figure 2B), which is consistent with recent reports (de Groot and Ventura, 2010; Tartaglia et al., 2007, 2009). This difference in transcript levels appears to primarily result from a differential rate of transcription (Figure 2A), because we did not observe a statistically significant difference in transcript half-lives between the groups (Table S2A). An analysis of histone modification data did not reveal a consistent difference in the promoter or the open reading frame (ORF) between genes that encode aggregation-prone or nonaggregation-prone proteins, suggesting that the differential rate of transcription cannot be explained due to global differences in histone modification alone. However, we do observe statistically significant differences for a few histone modification marks that are associated with transcription within the ORF region between the two groups of unstructured proteins (Table S2C).

Aggregation-Prone Proteins Are Present in Low Abundance and for a Short Time

A comparison of protein levels showed that the intracellular concentration of proteins that are likely to form aggregates is significantly lower than that of nonaggregation-prone proteins (Figure 2E). Reduced protein abundance could be the result of low transcript abundance, increased protein turnover, or decreased protein synthesis, due to tight translational regulation. An analysis of differences in protein half-lives showed that aggregation-prone proteins have a shorter half-life than the nonaggregation-prone proteins, suggesting that a rapid turnover of such proteins could contribute to their limited availability in a cell (Figure 2F). Reduced protein synthesis due to translational regulation may be mediated by (1) protein-mRNA interaction, (2) complex 5′ untranslated regions (UTRs), (3) decreased ribosomal density on transcripts, and (4) restricted choice of codon usage. We therefore systematically investigated each of these regulatory steps.

RNA-Binding Proteins Preferentially Bind mRNA of Aggregation-Prone Proteins

An investigation of protein-mRNA interaction data for 45 RNA-binding proteins in yeast revealed that several RNA-binding proteins show a significant enrichment for interactions with transcripts that encode aggregation-prone proteins (Table S2B). Interestingly, the translation initiation repressor protein KHD1p binds to more than 65% (P_{UA}: 2×10^{-9}; Fisher's test) of all transcripts encoding aggregation-prone, Q/N-rich, unstructured proteins, but only 19% of the nonaggregation-prone, unstructured proteins. In addition, KHD1p shows enrichment for binding transcripts that encode aggregation-prone, structured proteins (P_{SA}: 2×10^{-4}; Fisher's exact test). These observations suggest that translational regulation via protein-mRNA interaction is an important factor that might influence the availability of some aggregation-prone proteins.

mRNA of Aggregation-Prone Proteins Have Complex 5′UTR and RNA Structure

An analysis of the 5′UTR regions of transcripts showed that mRNA encoding aggregation-prone proteins tend to have much larger 5′UTR sequences (Figure 2C). The longer 5′UTR sequences form energetically more favorable secondary structures when compared to mRNA encoding non-aggregation-prone proteins (Figure S2A). Consistent with these finding, an analysis of the secondary structure profile of transcripts of *S. cerevisiae* that was recently measured using a high-throughput experimental approach (Table 1) revealed more secondary structure in the transcripts of aggregation-prone proteins when compared to nonaggregation-prone proteins (Figure S2B). Moreover, transcripts of aggregation-prone, unstructured proteins contain G-quadruplex-forming sequences more often than nonaggregation-prone proteins ($P_{SA\text{-}SNA}$: 0.1; $P_{UA\text{-}UNA}$: 1×10^{-4}; Fisher's exact test). Such folded structures might contribute to the observed reduced protein levels by regulating translational initiation (Kudla et al., 2009; Kumari et al., 2007). These findings suggest that translation initiation is likely to be differently regulated for mRNA encoding aggregation-prone proteins.

mRNA Encoding Aggregation-Prone Proteins Have Lower Translation Efficiency

In order to investigate the role of codon bias, we compared the transfer RNA (tRNA) adaptation index (tAi) of genes encoding aggregation-prone and nonaggregation-prone proteins. The tAi is based on the copy number of each of the tRNAs in a given genome and can be used to establish translational selection as well as to score objectively translational efficiency (dos Reis et al., 2004; Gingold and Pilpel, 2011). We find that transcripts encoding aggregation-prone proteins have a significantly lower tAi compared to those encoding nonaggregation-prone proteins (Figure 2D). It has been noted recently that translationally optimal codons are associated with buried residues in proteins, irrespective of their expression level, possibly to minimize protein misfolding (Drummond and Wilke, 2008; Lee et al., 2010). We also analyzed recently published ribosome-profiling data to further investigate this difference in translational efficiency. Indeed, aggregation-prone, structured proteins are less efficiently translated than nonaggregation-prone ones (Figure S3; Extended Results). An analysis of the polysome profiling data revealed that mRNA encoding aggregation-prone, unstructured proteins have a lower density of ribosomes per transcript (Table S2A). These observations collectively suggest that transcripts that encode aggregation-prone proteins are globally less efficiently

Table 1. Compendium of Data Sets Used in Our Study

Type of Information and Citation [PubMed ID]	Description of the Method Used to Obtain the Data
Histone modifications O'conner and Wyrick [17485428]	Database of published ChIP-microarray experiments that gives the relative enrichment of each histone modification at selected promoter regions or ORFs.
Transcriptional rate Holstege et al. [9845373] and Schwanhausser et al. [21593866]	Transcriptional rates for yeast grown in YPD were calculated by the authors based on the transcript abundances and mRNA half-lives. These were in turn determined by obtaining and comparing transcript levels of the wild-type and the temperature-sensitive RNA polymerase *rpb1-1* mutant strains using an Affymetrix microarray. For mouse cells, transcription rate was computed from experimentally obtained transcript steady state levels and turnover rates through next generation sequencing of mRNA.
Transcript abundance Holstege et al. [9845373], Vogel et al. [20739923], Lackner et al. [17434133], and Schwanhausser et al. [21593866]	Transcript abundances for the yeast and human cells were determined using high-density oligonucleotide arrays. For mouse cells, next generation sequencing was used to quantify transcript levels.
Transcript half-life Wang et al. [11972065], Lackner et al. [17434133], Yang et al. [12902380], and Schwanhausser et al. [21593866]	Transcript half-lives were determined by measuring transcript levels over several minutes after inhibiting transcription. This was estimated using (1) the temperature-sensitive RNA polymerase *rpb1-1* mutant *S. cerevisiae* strain, (2) 300 μg/ml 1,10-phenanthroline to block transcription in *S. pombe*, and (3) actinomycin D for human cell lines. For mouse cells, mRNA was labeled using 4-thiouridine. mRNA abundance was monitored over time using next generation sequencing to obtain turnover rates of transcripts.
RBP-bound transcripts Hogan et al. [18959479]	To identify RNAs associated with RNA-binding proteins (RBPs), (TAP)-tagged proteins were affinity purified from whole-cell extracts of cultures grown to mid-log phase in rich medium. RNA was extracted from the extracts, reverse transcribed, and then hybridized to DNA microarrays.
Transcript 5′ UTR length Nagalakshmi et al. [18451266] and Vogel et al. [20739923]	To map transcribed regions of the yeast genome, polyadenylate [poly(A)] RNA was isolated from yeast cells grown in rich media and used to generate double-stranded complementary DNA (cDNA) by reverse transcription. The double-stranded cDNA was fragmented and subjected to high-throughput Illumina sequencing, in which 35 base pairs of sequence were determined from one end of each fragment and mapped back onto the genome.
Transcript secondary structure and G-quadruplexes Kertesz et al. [20811459] and Capra et al. [20676380]	Parallel analysis of RNA structure: To identify RNA secondary structure location, in vitro-folded RNAs were first treated with different structure-specific enzymes, fragmented, and then determined by deep sequencing. G-quadruplexes were computationally identified using G4 DNA motif pattern and the loop length threshold approach.
Translational efficiency/codon bias Man and Pilpel [17277776] and Ingloia et al. [19213877]	The tRNA adaptation index (tAi) is determined by calculating a weight for each of the sense codons, derived from the copy number of all tRNA types that recognize it (including wobble interactions). For a given coding sequence, the tAi value is the geometric mean of the weights of all its sense codons. The tAi of a coding sequence ranges theoretically from 0 to 1 (0.2–0.7 for *S. cerevisiae* genome), with high values corresponding to high levels of translational efficiency. Ingolia et al. performed a ribosome foot-printing experiment to enrich for protected parts of the mRNA and subsequently performed a next generation sequencing experiment to obtain nucleotide resolution ribosome occupancy data in yeast.
Translational rate Arava et al. [12660367], Lackner et al. [17434133], and Schwanhausser et al. [21593866]	To obtain a profile of ribosome association for the yeast transcriptome, which is an indicator for translational rate, the authors fractionated polysomes using velocity sedimentation. Following this, a quantitative microarray analysis of several fractions across the gradient was used to estimate the translational status of each mRNA. Translation rates for mouse proteins were computed from experimentally obtained steady state levels and turnover rates of proteins using SILAC and mass spectrometry.

Table 1. *Continued*

Type of Information and Citation [PubMed ID]	Description of the Method Used to Obtain the Data
Protein abundance Ghaemmaghami et al. [14562106], Newman et al. [16699522], Matsuyama et al. [16823372], Vogel et al. [20739923], and Schwanhausser et al. [21593866]	Estimates of the endogenous protein expression levels during log-phase were obtained by tagging every protein with TAP-tag and/or green fluorescent protein and measuring the intensity for *S. cerevisiae* and *S. pombe*. Liquid chromatography-tandem mass spectrometry (LC-MS/MS) was used to obtain the abundance of proteins in the medulloblastoma Daoy cell line for human and NIH 3T3 cells for mouse.
Protein half-life Belle et al. [16916930] and Schwanhausser et al. [21593866]	Protein half-lives were determined by first inhibiting protein synthesis via the addition of cyclohexamide and by monitoring the abundance of each TAP-tagged protein in the yeast genome as a function of time. For the mouse cells, SILAC labeling of proteins followed by LC-MS/MS over time was employed to obtain protein half-life.
Overexpression phenotypes Sopko et al. [16455487]	To examine gene overexpression effects, an ordered array of 5,280 yeast strains was constructed, each conditionally overexpressing a unique yeast gene, covering 85% of all yeast genes. To catalog the spectrum of genes that affect cellular fitness when overexpressed, the array was transferred to a medium containing galactose, and each strain was examined for defects in colony formation.
Genetic screen for aggregation Willingham et al. [14657499] and Nollen et al. [15084750]	For *S. cerevisiae*, a gene deletion set (YGDS) of 4,850 viable mutant haploid strains was used to identify genes that enhance toxicity of a mutant huntingtin fragment or alpha-synuclein. In *C. elegans*, genome-wide RNA interference was used to identify genes that, when suppressed, resulted in the premature appearance of protein aggregates.

See also Table S1.

translated when compared to those encoding nonaggregation-prone proteins.

Synthesis and Degradation of Aggregation-Prone Proteins Is Tightly Regulated

While transcript and protein abundance may be related quantities, they are not strictly correlated, and recent genome-scale studies show extensive evidence for posttranscriptional regulation (Vogel, 2011). Thus it is important to study them independently and identify the influence of the different steps that affect protein abundance. Therefore, we performed a comprehensive statistical analysis of the major contributors in the gene expression process that affect protein abundance for the (non-) aggregation-prone proteins through a detailed partial least square regression (PLSR) analysis (see Extended Experimental Procedures and Extended Results; Tables S2D and S2E). The results of the PLSR calculations and the reported findings consistently suggest that (1) the cellular regulation of the aggregation-prone proteins is different compared to the nonaggregation-prone proteins and (2) a combination of reduced transcript abundance, rapid protein turnover, and translational regulation contributes to the low availability of aggregation-prone proteins.

The Observed Trends Are Evolutionary Conserved

We then assessed whether the tight regulation of aggregation-prone proteins is likely to be an evolutionarily conserved mechanism. To this end, we analyzed several published data sets (Table 1) for *Schizosaccharomyces pombe* and *Homo sapiens* and found similar trends to those observed for *S. cerevisiae* for the available data (Table 2). We also analyzed a recently published data set for *Mus musculus* and found that aggregation-prone and nonaggregation-prone proteins are regulated significantly differently at the protein level (Tables 2 and S3G). Overall these results suggest that the tight regulation of aggregation-prone proteins may be an evolutionarily conserved strategy.

Control Calculations for Alternative Explanations and Confounding Factors

We observed that the trends were not a result of differences in protein length and intrinsic disorder in the respective groups (Table S3A). Elimination of membrane proteins in the structured group, which have stretches of hydrophobic amino acids, does not affect the observed differences (Table S3B). While we observed that the four groups of proteins defined here are enriched to occur in different subcellular compartments (Table S3A), the observed differences are not primarily a consequence of their location within a cell (see Extended Results). As unstructured proteins are more tightly regulated than structured proteins due to their involvement in regulatory and signaling roles (Babu et al., 2011; Gsponer et al., 2008), we investigated only the subset of unstructured proteins that were not associated with regulatory or signaling roles and found similar trends (Table S3C).

Although TANGO has been benchmarked on very different aggregating and nonaggregating peptides, our selection of aggregation-prone, structured proteins may contain an unknown bias toward those with low abundance. Therefore, we used an additional predictor, PASTA (Trovato et al., 2007), to identify aggregation-prone, structured proteins and found consistent

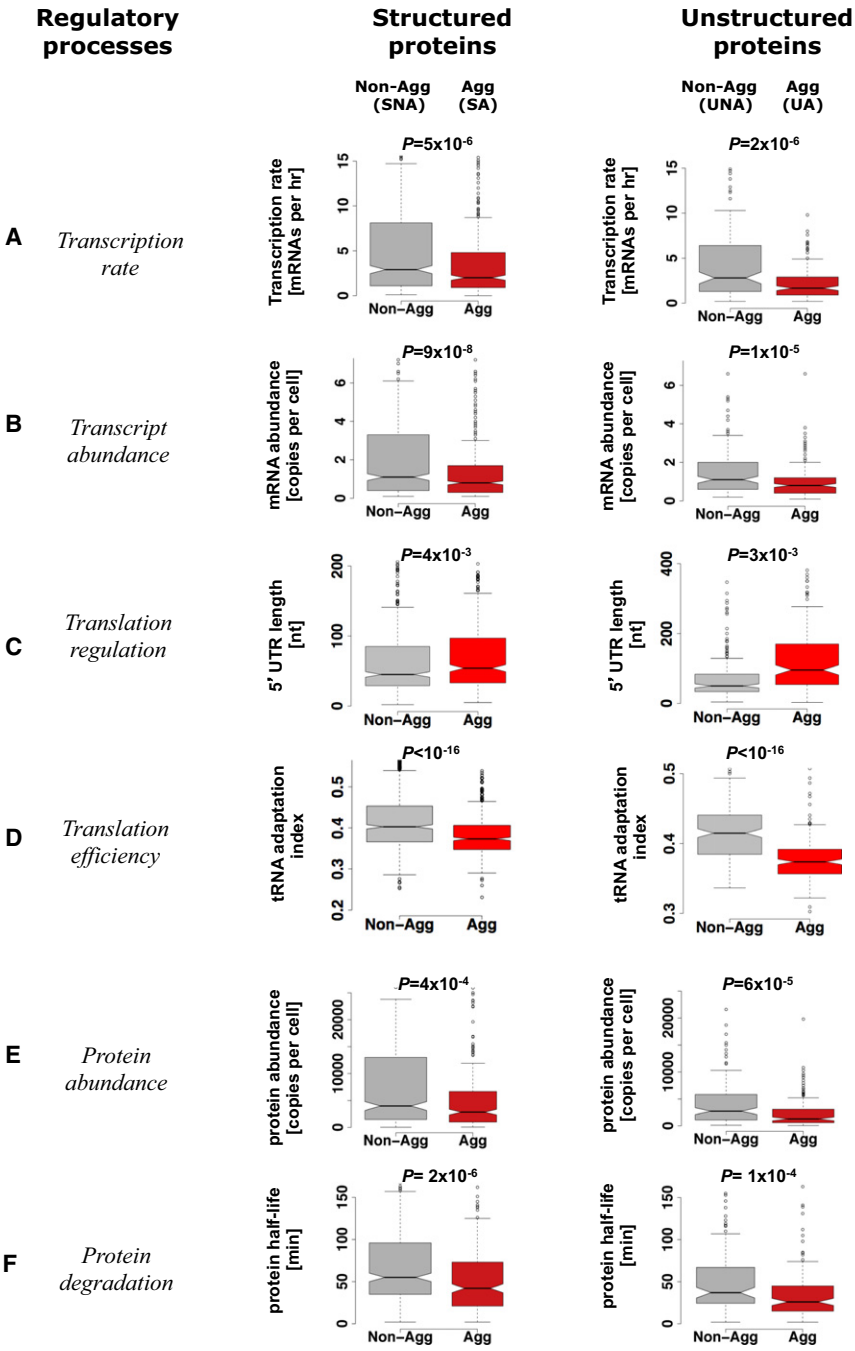

Figure 2. Box-Plot of the Distribution of Values for Various Regulatory and Cellular Properties

This is shown for the different groups of proteins that are nonaggregation-prone (structured SNA, unstructured UNA, gray boxes) and aggregation-prone (structured SA, unstructured UA, red boxes) in *S. cerevisiae*. Box-plot identifies the middle 50% of the data, the median, and the extreme points. (A–F) (A) Transcription rate, (B) transcript abundance, (C) translational regulation, (D) translational efficiency, (E) protein abundance, and (F) protein degradation. p values were computed using the Wilcoxon test.

See also Tables 1 and S2, and Figures S2 and S3.

trends (Table S3D). Similarly, we selected a recently identified list of prionogenic proteins (Alberti et al., 2009) in *S. cerevisiae* for the aggregation-prone, unstructured proteins and found that they have the expected low abundances and short half-lives (Table S3E). The four groups of proteins that we selected allowed for a clear distinction between aggregation-prone and nonaggregation-prone proteins as well as the necessary distinction between structured and unstructured proteins (see above). However, the four groups cover only part of the *S. cerevisiae* proteome, and a more continuous classification with respect to aggregation likelihood is required to assess the regulation of all yeast proteins. To address this, we used the algorithm AGGRESCAN (Conchillo-Solé et al., 2007) that identifies hydrophobic and nonhydrophobic aggregation-prone segments in structured and unstructured proteins. Analysis using the alternative classification scheme confirmed all the trends that we identified (Table S3F). The differences reported here are consistently determined to be statistically significant by two independent statistical tests: the Wilcoxon rank-sum test and the Kolmogorov–Smirnov test (Table S3H).

Modulators Enhancing Aggregation Phenotype Are Enriched in Genes Influencing Expression Homeostasis

The reported observations lead to the following predictions: (1) If the regulation at multiple stages is crucial for minimizing protein aggregation in a cell, then overexpression of the aggregation-prone proteins should be more often detrimental than overexpression of nonaggregation-prone proteins. Indeed, an investigation of the overexpression phenotype data (Sopko et al., 2006) revealed that aggregation-prone proteins are twice as often lethal when overexpressed compared to nonaggregation-prone proteins (Figure 3A; Extended Results). (2) In addition to genes that influence protein folding, multiple genetic loci that participate in gene expression homeostasis, such as RNA-binding proteins, should modulate protein aggregation in vivo. To test this hypothesis, we investigated published screens in yeast (Willingham et al., 2003) and *C. elegans* (Nollen et al., 2004) that systematically identified genetic backgrounds in which the phenotype due to the expression of an aggregation-prone protein is enhanced. We found, in both organisms, that alterations in genetic background that enhance

Table 2. Comparison of Aggregation-Prone and Non-Aggregation-Prone Proteins in *S. pombe*, *M. musculus*, and *H. sapiens*

Cellular Quantity	SNA		SA	
S. pombe				
	$\tilde{x}$	n	$\tilde{x}$	n
Transcript abundance [signal intensity]	2318 ± 264	676	**1,867 ± 193** (3×10^{-5})	488
Transcript half-life [% with long half-life]	77%	246	**47%** (3×10^{-11})	189
Translational efficiency [tAi]	0.40 ± 0.00	633	**0.38 ± 0.00** (3×10^{-11})	451
Protein abundance [signal intensity]	0.50 ± 0.09	621	**0.17 ± 0.04** ($<10^{-16}$)	289
H. sapiens				
Transcript abundance [arbitrary unit]	2,530 ± 1,005	107	**345 ± 330** (3×10^{-7})	89
Translation regulation [5′UTR length in nt]	85 ± 15	107	113 ± 30 (4×10^{-2})	89
Protein abundance [molecules/cell]	1,776 ± 1,100	107	**505 ± 247** (3×10^{-5})	89
M. musculus				
Transcriptional rate [mRNAs/cell*h]	2.0 ± 0.2	505	1.6 ± 0.2 (0.3)	359
Transcript abundance [copies/cell]	21 ± 1	497	20 ± 2 (0.9)	357
Transcript half-life [h]	11 ± 1	542	11 ± 1 (0.02)	392
Translational rate [proteins/mRNA*h]	63 ± 6	486	**39 ± 5** (3×10^{-8})	346
Protein abundance [proteins/cell]	36 k ± 7 k	570	**13 k ± 5 k** (6×10^{-9})	408
Protein half-life [h]	79 ± 7	570	**56 ± 6** (5×10^{-7})	408

Median values and their confidence intervals (C.I. $= 1.58 \times IQR/\sqrt{n}$, where IQR is the inter-quartile range and *n* the group sample size) are reported for highly structured proteins without aggregation-prone elements (SNA) and highly structured proteins with aggregation-prone elements (SA). There are only a low number of Q/N-enriched proteins in the data set for *S. pombe* and few proteins with Q/N-enriched domains in the data set available for *H. sapiens*. Therefore, no statistically significant analysis on the U group was possible. Results for the U group in *M. musculus* can be seen in Table S3G. "*n*" is the number of data points. Statistical significance was calculated with the Wilcoxon test and Fischer's exact test. Statistically significant differences are highlighted in bold. See also Table S3.

the aggregation phenotype were enriched for multiple genes that directly or indirectly influence transcript and protein availability (Figure 3B). Thus our findings support the emerging view that, in addition to mutations in the proteins themselves, mutations which affect their expression level or the genes that influence the expression level of an aggregation-prone protein may contribute to the disease phenotype involving protein aggregation (Powers et al., 2009).

DISCUSSION

We have provided a study that analyses the control of aggregation-prone proteins at multiple levels of gene expression regulation in evolutionarily diverse organisms within a single framework. Previous studies on individual data sets have indicated that aggregation-prone proteins may be regulated differently than non-aggregation-prone ones, but our findings reveal that *S. cerevisiae* keeps aggregation-prone proteins at low abundance by combining several strategies at nearly every regulatory level: from the initiation of transcription up to degradation of proteins. The differential regulation of aggregation-prone and nonaggregation-prone proteins seems to be evolutionarily conserved, as several of the trends are also found in *S. pombe*, *H. sapiens*, and *M. musculus*. This conservation of a differential regulation of aggregation-prone proteins in multiple organisms may underline the significance and generality of our findings. Considering the growing evidence for the importance of aggregation in various physiological processes, the conserved differential control of aggregation-prone proteins may be part of a general regulatory framework that not only minimizes unwanted/potentially harmful aggregation, but also keeps functional aggregation in check (see below).

The results presented here were possible only because of a conceptual framework, in which we distinguish between structured and unstructured aggregation-prone proteins. The trends we report would be missed by grouping them together. The fact that we find similar regulatory differences between aggregation-prone and nonaggregation-prone structured and unstructured proteins is, nevertheless, quite intriguing, particularly considering that hydrophobic stretches in structured proteins and Q/N-enriched stretches in unstructured proteins have different functions and are likely to have potentially different pathological consequences. These observations emphasize the importance of future studies to investigate the detailed molecular mechanisms that underlie the regulation of aggregation-prone proteins. In this direction, our analyses provide interesting pointers for regulatory mechanisms that may be of particular interest. For instance, we find that the 5′UTR structure of transcripts encoding aggregation-prone proteins is more complex and longer than that of nonaggregation-prone ones and that specific RNA-binding proteins target a large fraction of the transcripts that encode aggregation-prone proteins. It is likely that the fine-tuning of the different regulatory steps that keep protein concentrations low is different for individual aggregation-prone proteins or that even additional, specific control mechanisms are in place (Fowler et al., 2007). We wish to note that, while the general outcome from this study is biologically meaningful, as we observe consistent differences across many

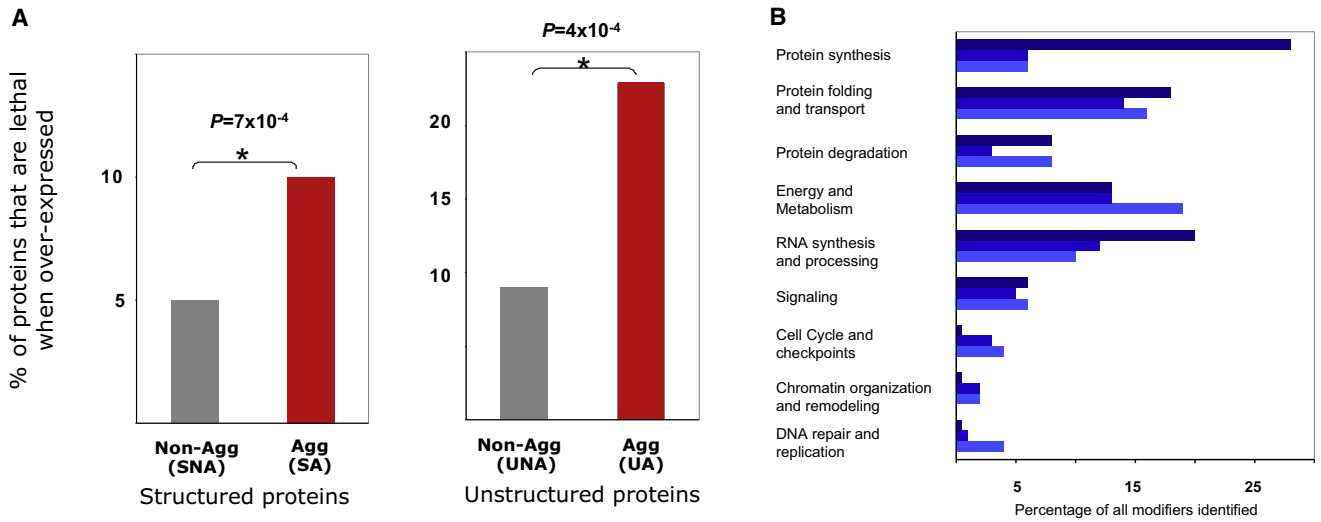

Figure 3. Overexpression Toxicity and Genetic Modulators of Protein Aggregation

(A) Overexpression toxicity phenotype. P values obtained from Fisher's test.

(B) Distribution of the functional categories of the genes, which when deleted or downregulated result in enhanced lethality upon expression of an aggregation-prone protein (expression of huntingtin fragment in *C. elegans* [dark blue], α-synuclein [blue], and huntingtin [light blue] in *S. cerevisiae*). The genes were grouped according to the GO annotation of the proteins they encode. The GO annotations influencing transcript and protein availability are terms in the figure other than DNA repair and replication, cell cycle and checkpoints, signaling, energy, and metabolism. All these terms influencing transcript and protein availability are significantly enriched ($p < 4 \times 10^{-2}$) for the *C. elegans* data set, except for chromatin organization and remodeling, for which the number of genes was small. For the yeast data sets, only the term protein synthesis was enriched significantly ($p < 10^{-3}$). However, p values for this data set have to be interpreted with care, as the number of identified genes in the screen is small.

tests, the reader should be aware that, for specific individual comparisons, a significant p value does not always mean that the difference will be biologically significant.

Below, we discuss how these findings fit in with the current understanding of avoidance of protein aggregation and regulating functional aggregates.

Evolution by Negative Design Minimizes Nonfunctional Protein Aggregation

During the course of evolution, when an aggregation promoting mutation is introduced in a gene, two main scenarios can be envisioned: either the mutated protein with increased aggregation likelihood provides a fitness advantage or it does not. If the mutated protein does not provide a fitness advantage, but forms nonfunctional or toxic aggregates, individuals harboring such sequences are likely to be eliminated from the population, resulting in selection for sequences that are less likely to form aggregates. Reports that support this outcome have been accumulated in recent years (Geiler-Samerotte et al., 2011; Morell et al., 2011). Accordingly, sequence motifs that are highly aggregation prone are significantly underrepresented (Broome and Hecht, 2000; Patki et al., 2006). These findings have been interpreted as strong indicators for avoidance of aggregation as a major evolutionary driving force in the design of protein sequences (Rousseau et al., 2006b) (Figure 4; Extended Discussion).

Proteostasis-Chaperone Network Minimizes Undesirable Protein Aggregation

If the mutated protein provides a fitness advantage, individuals that are able to prevent nonfunctional or undesirable aggregation of the mutated protein will be selected for in a population. Evidence for this outcome has also been presented in the literature. Despite the underrepresentation of sequence motifs that are highly aggregation-prone, extant genomes still contain a significant proportion of proteins that have aggregation-prone stretches (Linding et al., 2004; Monsellier et al., 2008) (see Figure S1; Table S4A). This suggests that cells minimize the harmful effects of aggregation-prone proteins. Accordingly, a significant body of work has shown that a substantial part of any organism's proteome (the proteostasis network) is dedicated to minimizing nonfunctional aggregation by ensuring protein folding, solubility, and removal of aggregates by specific cellular mechanisms (De Baets et al., 2011; Gidalevitz et al., 2010; Glover and Lindquist, 1998; Hishiya and Takayama, 2008; Powers et al., 2009; Roth and Balch, 2011) (Figure 4; Extended Discussion).

Tight Regulation as a Means to Control Functional and Toxic Aggregates by Le Chatelier's Principle

While the proteostasis-chaperone network may ensure folding of aggregation-prone proteins, too much buffering may be harmful for a cell. For instance, increased chaperone activity can minimize aggregation (Kitamura et al., 2006), but too high levels of chaperones may lead to tumorigenesis (Dai et al., 2007). These considerations, together with the fact that protein aggregation can be functional, raises the fundamental question: how do cells balance the benefits and risks of protein aggregation?

We suggest that the observed regulation of aggregation-prone proteins facilitates prevention of the formation of undesirable aggregates and may keep functional assemblies/aggregates

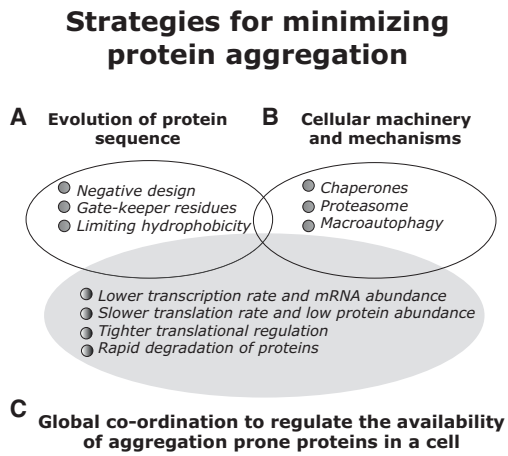

Figure 4. Evolutionary and Cellular Strategies to Deal with Proteins that Are Likely to Aggregate

Two major solutions to explain how cells are robust to the presence of aggregation-prone proteins have been proposed.

(A) The sequences of such proteins have been subjected to rigorous selection such that they do not easily form aggregates (i.e., avoidance of aggregation is a major evolutionary driving force in the design of protein sequences).

(B) Cellular systems have evolved machineries and mechanisms to either avoid or efficiently clear aggregates.

(C) The results from our integrated analysis provide insights into a third possible solution to this problem. Cellular systems have evolved regulatory strategies that control the availability of aggregation-prone proteins and their encoding transcripts such that they are present for short periods, low quantities, and in precise amounts in a cell. This strategy may ensure that the abundance of aggregation-prone proteins is below the critical concentration. See also Figure 5 and Figure S4.

under control, as explained by Le Chatelier's principle (Figure 5). This principle states that if any change (e.g., altered concentration due to rapid turnover) is imposed on a system that is in equilibrium (e.g., soluble versus oligomeric or fibrillar), then the system tends to adjust to a new equilibrium, counteracting the change. Since (1) aggregate formation is a nucleation-dependent process that relies on the amount (i.e., critical concentration) of aggregation-prone proteins in the cell (Harper and Lansbury, 1997; Serio et al., 2000) and (2) individual subunits can recycle in and out of fibrillar or amorphous aggregates in a concentration-dependent manner or by active mechanisms (Carulla et al., 2005; Colby et al., 2006; Kim et al., 2002; Stenoien et al., 2002), control of the abundance of aggregation-prone proteins may (1) ensure that levels are lower than the critical concentration and (2) permit shifting of the dynamic equilibrium between the soluble monomeric form and the aggregate form, as explained by Le Chatelier's principle (Figures 5 and S4; Extended Discussion).

The critical concentration above which fibrillar aggregation takes place is largely determined by the amino-acid composition and the environment of the protein. These properties are evolvable as long as protein function is not compromised during sequence changes. However, if sequence evolution is constrained due to functional reasons, evolution of a tight control over the availability of aggregation-prone proteins can keep their effective intracellular concentration below their critical concentration and minimize the chances of aggregation, even when the chaperone-system or clearance mechanisms fail due to stress or functional overload (i.e., upon failure of the chaperone proteostasis network) (Bence et al., 2001; Satyal et al., 2000). Hence, proteins that are kept at low concentrations can "accept" mutations that provide functional advantages, but equally increase the aggregation propensity. In other words, proteins that are kept at low concentrations for functional purposes can, by evolutionary drift, end up with a higher aggregation propensity without incurring significant negative selection. This concept is consistent with the observation of Drummond et al. (2005) that highly abundant proteins evolve slowly. Thus the tight regulation of aggregation-prone proteins offers a distinct solution to the problem of minimizing protein aggregation as compared to evolution by negative design and the proteostasis network. Supporting this hypothesis, it has been shown that (1) an iron response element (IRE) in the 5′UTR of α-synuclein and the amyloid precursor protein (APP) ensures a tight translational control, which when disrupted results in increased protein abundance and leads to protein aggregation (Avramovich-Tirosh et al., 2008; Friedlich et al., 2007; Rogers et al., 2002), and (2) increased expression of full-length TIA-1 that contains a Q/N-enriched region induces stress granule formation, and overexpression of the Q/N-rich region of TIA-1 alone forms cytoplasmic microaggregates that sequester endogenous TIA-1 (Gilks et al., 2004).

Importantly, if protein aggregation, higher-order oligomerization, or even amyloid-like fibrillation is required for functional reasons, local (super) saturation, which is an increase in the effective local concentration above the critical one, can be mediated. This can be achieved, for example, by specific regulatory proteins that "scaffold" aggregation-prone proteins in a particular location or by confinement of the protein within cellular compartments (Brangwynne et al., 2009; Decker et al., 2007; Fowler et al., 2006; Harper and Lansbury, 1997; Hu et al., 2009; Li et al., 2012; van Ham et al., 2010). Release of the aggregation-prone proteins from the confinement will reverse the aggregation process and likely resolve soluble aggregates and partially insoluble aggregates, as explained by Le Chatelier's principle. This may be further enhanced by active mechanisms, such as the disaggresome (Bieschke et al., 2009) (Figure 5; Extended Discussion).

Support for this concept is provided by exciting recent reports. (1) Intrinsically disordered low complexity regions (LCR) of several RNA-binding proteins have been shown to enable the formation of reversible granule-like macromolecular assemblies by promoting amyloid-like interactions. Though the regulation of the formation of the highly dynamic aggregates is not yet understood, it has been proposed that RNA may act as a scaffold, allowing the LCRs to reach high local concentrations that are necessary for aggregation to occur (Han et al., 2012; Kato et al., 2012). (2) Under normal conditions, the yeast protein Lsb2p is expressed at low levels, but under stress conditions, its expression increases. Importantly, as it is a scaffold for Sup35p, the increase in Lsb2p concentration promotes local accumulation of soluble Sup35 and its conversion into the fibrillar state (Chernova et al., 2011). Taken together, these recent findings suggest that a phenomenon similar to Le Chatelier's principle is exploited for functional and reversible assembly/aggregation.

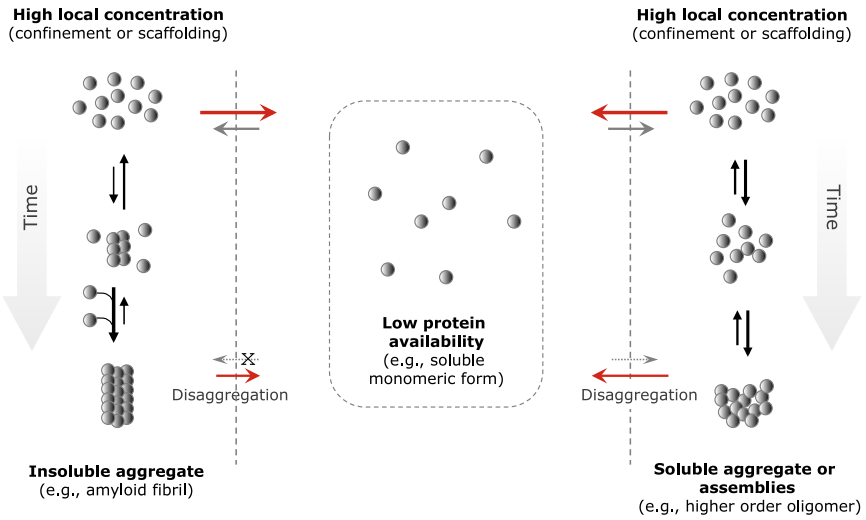

Figure 5. Proposed Model for the Avoidance of Undesirable Aggregation and the Control of Functional Aggregation as Explained by Le Chatelier's Principle

The differential regulation of aggregation-prone protein results in low intracellular abundance of these proteins. We suggest that it is below the critical concentration, thereby preventing undesired aggregation (center panel). Local increase in monomer concentration above the critical one (i.e., supersaturation) via confinement or a scaffolding protein can facilitate formation of aggregates. This local supersaturation would start the aggregation process and, depending on the time span of supersaturation, allow the formation of oligomers or even fibrils (left and right panels, respectively). Release from the confinement/scaffold (gray dashed vertical lines) before an aggregation nucleus has formed will cause dissociation of all soluble oligomeric assemblies/aggregates (bottom right). As a result, the equilibrium will shift significantly toward the soluble monomeric form (red horizontal arrow). Even in the case where insoluble aggregates (e.g., fibrillar structures) have been formed (bottom left), their formation may be partially reversed (small red arrow) by the low abundance of the soluble monomeric form achieved by tight regulation, because it has been shown that monomers can recycle in and out of the fibrillar aggregate in a concentration-dependent manner or disaggregate by active mechanisms.

While it is clear that Le Chatelier's principle is only valid for closed thermodynamic systems and cells are not "closed" in the strict sense, it is certain that changes in local concentration of proteins in a cell can affect the equilibrium between monomeric and multimeric states of proteins.

Implications and Significance of the Proposed Strategy to Regulate Aggregation-Prone Proteins

The observation that the availability of aggregation-prone proteins is likely to be tuned by several factors at multiple levels during transcription, translation, and degradation complements and underscores the emerging view that the pathophysiology of aggregation-related diseases is multifactorial in nature. In addition, the process of aging is also likely to alter the mechanisms that help maintain protein homeostasis and thereby increases the risk for aggregation-related diseases (Gidalevitz et al., 2010; Powers et al., 2009; Roth and Balch, 2011). In addition to this aspect, our findings have important implications:

(1) Mutations that disrupt the tight control of aggregation-prone proteins, thereby altering the dynamic equilibrium between the soluble monomeric and aggregated states, are likely to be a common mechanism that may underlie several expression level–dependent disease phenotypes that involve protein aggregation. Cross-seeding by other aggregation-prone proteins may also affect such equilibrium (Sandefur and Schnell, 2011). Indeed, it has been shown that huntingtin aggregates cross-seed TIA-1 and is likely to repress the physiological function of TIA-1 due to loss of function caused by sequestration of the functional protein into aggregates (Furukawa et al., 2009).

(2) Our findings suggest potential candidate loci (for example, 5′ and 3′UTR regions of aggregation-prone proteins, RNA-binding proteins, E3 ubiquitin ligases, etc.), which are likely to manifest as low-frequency mutant alleles with partial penetrance, that should be prioritized for further detailed investigation in whole-genome association studies aimed at identifying causal mutations of aggregation diseases. Indeed, a recent work has identified a single nucleotide mutation in the 5′UTR of puratrophin-1 to be associated with autosomal-dominant cerebellar ataxia, a group of heterogeneous neurodegenerative disorders (Ishikawa et al., 2005).

(3) The observations provide a framework for identifying intracellular pathways that regulate levels of specific aggregation-prone proteins (e.g., the presence of IRE in 5′UTR of α-synuclein and APP). Therefore, it could provide avenues and enumerate strategies for tailoring drugs that modulate expression of specific aggregation-prone proteins rather than develop generic drugs that may disrupt both harmful and functional aggregates indiscriminately.

Finally, the reported strategy to control protein aggregates provides robustness to cellular systems by minimizing the potentially harmful effects of aggregation-prone proteins and, at the same time, permits their vital contribution to the functioning of a cell.

EXPERIMENTAL PROCEDURES

See Extended Experimental Procedures for more details.

Identification of Highly Structured and Unstructured Group of Proteins

The proteome of *S. cerevisiae* was divided into proteins that are highly structured and those that are highly unstructured. The prediction of intrinsic disorder was carried out using Disopred2 (Ward et al., 2004). We then calculated the fraction of the sequence that was predicted to be unstructured. Depending on this fraction, we classified each protein as highly structured (S) (1,971 proteins with 0%–10% of all residues unstructured), highly unstructured (U) (2,020 proteins with 30%–100% unstructured residues), and not fitting within each group (2,711 proteins).

Identification of Aggregation-Prone and Nonaggregation-Prone Proteins in Each Structural Class

To identify aggregation-prone proteins among the highly structured and highly unstructured proteins, we used TANGO (Fernandez-Escamilla et al., 2004) for the structured proteins and the algorithm described by Michelitsch and Weissman (2000) for the unstructured proteins. We divided the group of highly structured proteins into those that are highly aggregation prone (SAG; 711 proteins with more than seven consecutive residues identified as aggregation-prone by TANGO) and those with very low aggregation likelihood (SNA; 716 proteins with no residues identified as aggregation-prone by TANGO; for details, see Extended Experimental Procedures). We identified aggregation-prone stretches in the highly unstructured proteins by searching for segments that contain large Q- or N-enriched segments. We used the algorithm described by Michelitsch and Weissman (2000) and identified 197 aggregation-prone, unstructured proteins (UAG) that contained 25 glutamines or asparagines in a segment of 80 residues. As unstructured proteins with low aggregation likelihood (UNA), we identified 198 proteins that contain 30 lysines or glutamic acids in a segment of 80 residues (see Extended Experimental Procedures). Large numbers of charged amino acids, such as K and E, have been shown to prevent aggregation, even under conditions that normally cause proteins to aggregate, such as heating or chemical denaturation (Lawrence et al., 2007).

Data Set and Statistical Analysis

The complete proteome sequences of *S. cerevisiae*, *S. pombe*, human, and mouse were obtained from the National Center for Biotechnology Information (NCBI). Information on mRNA abundance, transcriptional rate, transcript half-life, transcripts bound to RNA-binding proteins, protein abundance, protein half-life, translational rate, overexpression phenotype, and data on genetic background that enhanced protein aggregation were all obtained from published literature (see Table 1). All statistical analyses to estimate significance (Wilcoxon, Kolmogorov–Smirnov, and Fisher's exact test) and the partial least square regression (PLSR) analysis were carried out using the R statistical analysis package.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Results, and Extended Discussion, Extended Experimental Procedures, four figures, and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2012.09.036.

LICENSING INFORMATION

ACKNOWLEDGMENTS

This work was supported by the Medical Research Council (U105185859). We thank A. Bertolotti, E. Levy, J. Scott-Brown, M. Buljan, M. Garcia-Alai, M. Goedert, R. van der Lee, S. Michnick, S. Radford, S. Teichmann, and S. Sarkar for providing helpful comments. We apologize for, because of lack of space, not citing the work that describes the datasets and other references. M.M.B. acknowledges Darwin College, Schlumberger Ltd, Trinity College, HFSP (RGY0073/2010), ERASysBio+ (GRAPPLE; BBSRC), and the EMBO YI Programme for support. J.G. is funded by PrioNet Canada and the University of British Columbia.

Received: March 15, 2012
Revised: July 23, 2012
Accepted: September 27, 2012
Published: November 15, 2012

REFERENCES

Alberti, S., Halfmann, R., King, O., Kapila, A., and Lindquist, S. (2009). A systematic survey identifies prions and illuminates sequence features of prionogenic proteins. Cell *137*, 146–158.

Avramovich-Tirosh, Y., Amit, T., Bar-Am, O., Weinreb, O., and Youdim, M.B. (2008). Physiological and pathological aspects of Abeta in iron homeostasis via 5'UTR in the APP mRNA and the therapeutic use of iron-chelators. BMC Neurosci. *9*(*Suppl 2*), S2.

Babu, M.M., van der Lee, R., de Groot, N.S., and Gsponer, J. (2011). Intrinsically disordered proteins: regulation and disease. Curr. Opin. Struct. Biol. *21*, 432–440.

Balagopal, V., and Parker, R. (2009). Polysomes, P bodies and stress granules: states and fates of eukaryotic mRNAs. Curr. Opin. Cell Biol. *21*, 403–408.

Bence, N.F., Sampat, R.M., and Kopito, R.R. (2001). Impairment of the ubiquitin-proteasome system by protein aggregation. Science *292*, 1552–1555.

Bieschke, J., Cohen, E., Murray, A., Dillin, A., and Kelly, J.W. (2009). A kinetic assessment of the C. elegans amyloid disaggregation activity enables uncoupling of disassembly and proteolysis. Protein Sci. *18*, 2231–2241.

Brangwynne, C.P., Eckmann, C.R., Courson, D.S., Rybarska, A., Hoege, C., Gharakhani, J., Jülicher, F., and Hyman, A.A. (2009). Germline P granules are liquid droplets that localize by controlled dissolution/condensation. Science *324*, 1729–1732.

Broome, B.M., and Hecht, M.H. (2000). Nature disfavors sequences of alternating polar and non-polar amino acids: implications for amyloidogenesis. J. Mol. Biol. *296*, 961–968.

Carulla, N., Caddy, G.L., Hall, D.R., Zurdo, J., Gairí, M., Feliz, M., Giralt, E., Robinson, C.V., and Dobson, C.M. (2005). Molecular recycling within amyloid fibrils. Nature *436*, 554–558.

Chen, S., Berthelier, V., Yang, W., and Wetzel, R. (2001). Polyglutamine aggregation behavior in vitro supports a recruitment mechanism of cytotoxicity. J. Mol. Biol. *311*, 173–182.

Chernova, T.A., Romanyuk, A.V., Karpova, T.S., Shanks, J.R., Ali, M., Moffatt, N., Howie, R.L., O'Dell, A., McNally, J.G., Liebman, S.W., et al. (2011). Prion induction by the short-lived, stress-induced protein Lsb2 is regulated by ubiquitination and association with the actin cytoskeleton. Mol. Cell *43*, 242–252.

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

Colby, D.W., Cassady, J.P., Lin, G.C., Ingram, V.M., and Wittrup, K.D. (2006). Stochastic kinetics of intracellular huntingtin aggregate formation. Nat. Chem. Biol. *2*, 319–323.

Conchillo-Solé, O., de Groot, N.S., Avilés, F.X., Vendrell, J., Daura, X., and Ventura, S. (2007). AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC Bioinformatics *8*, 65.

Dai, C., Whitesell, L., Rogers, A.B., and Lindquist, S. (2007). Heat shock factor 1 is a powerful multifaceted modifier of carcinogenesis. Cell *130*, 1005–1018.

De Baets, G., Reumers, J., Delgado Blanco, J., Dopazo, J., Schymkowitz, J., and Rousseau, F. (2011). An evolutionary trade-off between protein turnover rate and protein aggregation favors a higher aggregation propensity in fast degrading proteins. PLoS Comput. Biol. *7*, e1002090.

de Groot, N.S., and Ventura, S. (2010). Protein aggregation profile of the bacterial cytosol. PLoS ONE *5*, e9383.

Decker, C.J., Teixeira, D., and Parker, R. (2007). Edc3p and a glutamine/asparagine-rich domain of Lsm4p function in processing body assembly in Saccharomyces cerevisiae. J. Cell Biol. *179*, 437–449.

dos Reis, M., Savva, R., and Wernisch, L. (2004). Solving the riddle of codon usage preferences: a test for translational selection. Nucleic Acids Res. *32*, 5036–5044.

Drummond, D.A., and Wilke, C.O. (2008). Mistranslation-induced protein misfolding as a dominant constraint on coding-sequence evolution. Cell *134*, 341–352.

Drummond, D.A., Bloom, J.D., Adami, C., Wilke, C.O., and Arnold, F.H. (2005). Why highly expressed proteins evolve slowly. Proc. Natl. Acad. Sci. USA *102*, 14338–14343.

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.

Fiumara, F., Fioriti, L., Kandel, E.R., and Hendrickson, W.A. (2010). Essential role of coiled coils for aggregation and activity of Q/N-rich prions and PolyQ proteins. Cell *143*, 1121–1135.

Fowler, D.M., Koulov, A.V., Alory-Jost, C., Marks, M.S., Balch, W.E., and Kelly, J.W. (2006). Functional amyloid formation within mammalian tissue. PLoS Biol. *4*, e6.

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.

Friedlich, A.L., Tanzi, R.E., and Rogers, J.T. (2007). The 5′-untranslated region of Parkinson's disease alpha-synuclein messengerRNA contains a predicted iron responsive element. Mol. Psychiatry *12*, 222–223.

Furukawa, Y., Kaneko, K., Matsumoto, G., Kurosawa, M., and Nukina, N. (2009). Cross-seeding fibrillation of Q/N-rich proteins offers new pathomechanism of polyglutamine diseases. J. Neurosci. *29*, 5153–5162.

Geiler-Samerotte, K.A., Dion, M.F., Budnik, B.A., Wang, S.M., Hartl, D.L., and Drummond, D.A. (2011). Misfolded proteins impose a dosage-dependent fitness cost and trigger a cytosolic unfolded protein response in yeast. Proc. Natl. Acad. Sci. USA *108*, 680–685.

Gidalevitz, T., Kikis, E.A., and Morimoto, R.I. (2010). A cellular perspective on conformational disease: the role of genetic background and proteostasis networks. Curr. Opin. Struct. Biol. *20*, 23–32.

Gilks, N., Kedersha, N., Ayodele, M., Shen, L., Stoecklin, G., Dember, L.M., and Anderson, P. (2004). Stress granule assembly is mediated by prion-like aggregation of TIA-1. Mol. Biol. Cell *15*, 5383–5398.

Gingold, H., and Pilpel, Y. (2011). Determinants of translation efficiency and accuracy. Mol. Syst. Biol. *7*, 481.

Glover, J.R., and Lindquist, S. (1998). Hsp104, Hsp70, and Hsp40: a novel chaperone system that rescues previously aggregated proteins. Cell *94*, 73–82.

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., Futschik, M.E., Teichmann, S.A., and Babu, M.M. (2008). Tight regulation of unstructured proteins: from transcript synthesis to protein degradation. Science *322*, 1365–1368.

Han, T.W., Kato, M., Xie, S., Wu, L.C., Mirzaei, H., Pei, J., Chen, M., Xie, Y., Allen, J., Xiao, G., and McKnight, S.L. (2012). Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. Cell *149*, 768–779.

Harper, J.D., and Lansbury, P.T., Jr. (1997). Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. Annu. Rev. Biochem. *66*, 385–407.

Hishiya, A., and Takayama, S. (2008). Molecular chaperones as regulators of cell death. Oncogene *27*, 6489–6506.

Hu, X., Crick, S.L., Bu, G., Frieden, C., Pappu, R.V., and Lee, J.M. (2009). Amyloid seeds formed by cellular uptake, concentration, and aggregation of the amyloid-beta peptide. Proc. Natl. Acad. Sci. USA *106*, 20324–20329.

Ishikawa, K., Toru, S., Tsunemi, T., Li, M., Kobayashi, K., Yokota, T., Amino, T., Owada, K., Fujigasaki, H., Sakamoto, M., et al. (2005). An autosomal dominant cerebellar ataxia linked to chromosome 16q22.1 is associated with a single-nucleotide substitution in the 5′ untranslated region of the gene encoding a protein with spectrin repeat and Rho guanine-nucleotide exchange-factor domains. Am. J. Hum. Genet. *77*, 280–296.

Kato, M., Han, T.W., Xie, S., Shi, K., Du, X., Wu, L.C., Mirzaei, H., Goldsmith, E.J., Longgood, J., Pei, J., et al. (2012). Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. Cell *149*, 753–767.

Kim, S., Nollen, E.A., Kitagawa, K., Bindokas, V.P., and Morimoto, R.I. (2002). Polyglutamine protein aggregates are dynamic. Nat. Cell Biol. *4*, 826–831.

Kitamura, A., Kubota, H., Pack, C.G., Matsumoto, G., Hirayama, S., Takahashi, Y., Kimura, H., Kinjo, M., Morimoto, R.I., and Nagata, K. (2006). Cytosolic chaperonin prevents polyglutamine toxicity with altering the aggregation state. Nat. Cell Biol. *8*, 1163–1170.

Krobitsch, S., and Lindquist, S. (2000). Aggregation of huntingtin in yeast varies with the length of the polyglutamine expansion and the expression of chaperone proteins. Proc. Natl. Acad. Sci. USA *97*, 1589–1594.

Kudla, G., Murray, A.W., Tollervey, D., and Plotkin, J.B. (2009). Coding-sequence determinants of gene expression in Escherichia coli. Science *324*, 255–258.

Kumari, S., Bugaut, A., Huppert, J.L., and Balasubramanian, S. (2007). An RNA G-quadruplex in the 5′ UTR of the NRAS proto-oncogene modulates translation. Nat. Chem. Biol. *3*, 218–221.

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

Lee, Y., Zhou, T., Tartaglia, G.G., Vendruscolo, M., and Wilke, C.O. (2010). Translationally optimal codons associate with aggregation-prone sites in proteins. Proteomics *10*, 4163–4171.

Li, P., Banjade, S., Cheng, H.C., Kim, S., Chen, B., Guo, L., Llaguno, M., Hollingsworth, J.V., King, D.S., Banani, S.F., et al. (2012). Phase transitions in the assembly of multivalent signalling proteins. Nature *483*, 336–340.

Linding, R., Schymkowitz, J., Rousseau, F., Diella, F., and Serrano, L. (2004). A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. J. Mol. Biol. *342*, 345–353.

Maji, S.K., Perrin, M.H., Sawaya, M.R., Jessberger, S., Vadodaria, K., Rissman, R.A., Singru, P.S., Nilsson, K.P., Simon, R., Schubert, D., et al. (2009). Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. Science *325*, 328–332.

Masino, L., Nicastro, G., Calder, L., Vendruscolo, M., and Pastore, A. (2011). Functional interactions as a survival strategy against abnormal aggregation. FASEB J. *25*, 45–54.

Michelitsch, M.D., and Weissman, J.S. (2000). A census of glutamine/asparagine-rich regions: implications for their conserved function and the prediction of novel prions. Proc. Natl. Acad. Sci. USA *97*, 11910–11915.

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

Morell, M., de Groot, N.S., Vendrell, J., Avilés, F.X., and Ventura, S. (2011). Linking amyloid protein aggregation and yeast survival. Mol. Biosyst. *7*, 1121–1128.

Münch, C., and Bertolotti, A. (2010). Exposure of hydrophobic surfaces initiates aggregation of diverse ALS-causing superoxide dismutase-1 mutants. J. Mol. Biol. *399*, 512–525.

Nollen, E.A., Garcia, S.M., van Haaften, G., Kim, S., Chavez, A., Morimoto, R.I., and Plasterk, R.H. (2004). Genome-wide RNA interference screen identifies previously undescribed regulators of polyglutamine aggregation. Proc. Natl. Acad. Sci. USA *101*, 6403–6408.

Patki, A.U., Hausrath, A.C., and Cordes, M.H. (2006). High polar content of long buried blocks of sequence in protein domains suggests selection against amyloidogenic non-polar sequences. J. Mol. Biol. *362*, 800–809.

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.

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

Reijns, M.A., Alexander, R.D., Spiller, M.P., and Beggs, J.D. (2008). A role for Q/N-rich aggregation-prone regions in P-body localization. J. Cell Sci. *121*, 2463–2472.

Rogers, J.T., Randall, J.D., Cahill, C.M., Eder, P.S., Huang, X., Gunshin, H., Leiter, L., McPhee, J., Sarang, S.S., Utsuki, T., et al. (2002). An iron-responsive element type II in the 5′-untranslated region of the Alzheimer's amyloid precursor protein transcript. J. Biol. Chem. *277*, 45518–45528.

Roth, D.M., and Balch, W.E. (2011). Modeling general proteostasis: proteome balance in health and disease. Curr. Opin. Cell Biol. *23*, 126–134.

Rousseau, F., Schymkowitz, J., and Serrano, L. (2006a). Protein aggregation and amyloidosis: confusion of the kinds? Curr. Opin. Struct. Biol. *16*, 118–126.

Rousseau, F., Serrano, L., and Schymkowitz, J.W. (2006b). How evolutionary pressure against protein aggregation shaped chaperone specificity. J. Mol. Biol. *355*, 1037–1047.

Salazar, A.M., Silverman, E.J., Menon, K.P., and Zinn, K. (2010). Regulation of synaptic Pumilio function by an aggregation-prone domain. J. Neurosci. *30*, 515–522.

Sandefur, C.I., and Schnell, S. (2011). A model of threshold behavior reveals rescue mechanisms of bystander proteins in conformational diseases. Biophys. J. *100*, 1864–1873.

Santner, A.A., Croy, C.H., Vasanwala, F.H., Uversky, V.N., Van, Y.Y., and Dunker, A.K. (2012). Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression. Biochemistry *51*, 7250–7262.

Satyal, S.H., Schmidt, E., Kitagawa, K., Sondheimer, N., Lindquist, S., Kramer, J.M., and Morimoto, R.I. (2000). Polyglutamine aggregates alter protein folding homeostasis in Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA *97*, 5750–5755.

Serio, T.R., Cashikar, A.G., Kowal, A.S., Sawicki, G.J., Moslehi, J.J., Serpell, L., Arnsdorf, M.F., and Lindquist, S.L. (2000). Nucleated conformational conversion and the replication of conformational information by a prion determinant. Science *289*, 1317–1321.

Sivakolundu, S.G., Nourse, A., Moshiach, S., Bothner, B., Ashley, C., Satumba, J., Lahti, J., and Kriwacki, R.W. (2008). Intrinsically unstructured domains of Arf and Hdm2 form bimolecular oligomeric structures in vitro and in vivo. J. Mol. Biol. *384*, 240–254.

Sopko, R., Huang, D., 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.

Stenoien, D.L., Mielke, M., and Mancini, M.A. (2002). Intranuclear ataxin1 inclusions contain both fast- and slow-exchanging components. Nat. Cell Biol. *4*, 806–810.

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., Pechmann, S., Dobson, C.M., and Vendruscolo, M. (2009). A relationship between mRNA expression levels and protein solubility in E. coli. J. Mol. Biol. *388*, 381–389.

Trovato, A., Seno, F., and Tosatto, S.C. (2007). The PASTA server for protein aggregation prediction. Protein Eng. Des. Sel. *20*, 521–523.

van Ham, T.J., Holmberg, M.A., van der Goot, A.T., Teuling, E., Garcia-Arencibia, M., Kim, H.E., Du, D., Thijssen, K.L., Wiersma, M., Burggraaff, R., et al. (2010). Identification of MOAG-4/SERF as a regulator of age-related proteotoxicity. Cell *142*, 601–612.

Vogel, C. (2011). Translation's coming of age. Mol. Syst. Biol. *7*, 498.

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J. Mol. Biol. *337*, 635–645.

Willingham, S., Outeiro, T.F., DeVit, M.J., Lindquist, S.L., and Muchowski, P.J. (2003). Yeast genes that enhance the toxicity of a mutant huntingtin fragment or alpha-synuclein. Science *302*, 1769–1772.