An Evolutionary Trade-Off between Protein Turnover Rate and Protein Aggregation Favors a Higher Aggregation Propensity in Fast Degrading Proteins

Greet De Baets[1,2], Joke Reumers[1,2], Javier Delgado Blanco[1,2], Joaquin Dopazo[3,4,5], Joost Schymkowitz[1,2]*, Frederic Rousseau[1,2]*

1 VIB Switch Laboratory, VIB, Brussels, Belgium, 2 Vrije Universiteit Brussel, Brussels, Belgium, 3 Department of Bioinformatics and Genomics, Centro de Investigación Príncipe Felipe (CIPF), Valencia, Spain, 4 CIBER de Enfermedades Raras (CIBERER), Valencia, Spain, 5 Functional Genomics Node (INB) at CIPF, Valencia, Spain

Abstract

We previously showed the existence of selective pressure against protein aggregation by the enrichment of aggregation-opposing 'gatekeeper' residues at strategic places along the sequence of proteins. Here we analyzed the relationship between protein lifetime and protein aggregation by combining experimentally determined turnover rates, expression data, structural data and chaperone interaction data on a set of more than 500 proteins. We find that selective pressure on protein sequences against aggregation is not homogeneous but that short-living proteins on average have a higher aggregation propensity and fewer chaperone interactions than long-living proteins. We also find that short-living proteins are more often associated to deposition diseases. These findings suggest that the efficient degradation of high-turnover proteins is sufficient to preclude aggregation, but also that factors that inhibit proteasomal activity, such as physiological ageing, will primarily affect the aggregation of short-living proteins.

Introduction

Biological networks are fine-tuned to respond to narrow changes in protein concentration. The ability of a cell to maintain metabolic and signal transduction fluxes is therefore highly dependent on a tight regulation of its proteostatic network [1]. The capacity of the protein quality control system to regulate protein folding and degradation erodes with age, resulting in increased protein aggregation and aggregation-associated diseases [2,3]. Which proteins first fall prey to misfolding is most likely a stochastic process that is modulated by both tissue-specific expression levels and environmental factors [4]. However, sensitivity to protein aggregation is also determined by intrinsic protein parameters such as the efficiency of the folding process [5], thermodynamic stability [6,7], the aggregation propensity of the protein sequence [8,9] and its ability to be recognized by the protein quality control system [10]. We previously showed that evolutionary forces shape protein sequences in order to minimize their aggregation propensity, by strategically placing aggregation-opposing gatekeeper residues along the sequence [11,12]. Although this insight has been confirmed by independent studies [13,14,15,16], the extent to which selective pressures mould protein sequences is most likely not uniform, but determined by the biological context in which the protein functions [17]. For instance, it has been shown that proteins with high expression

levels on average have a lower aggregation propensity than proteins with lower expression levels [18]. We reasoned that proteins with high turnover rate and thus short lifetime will have, on average, lower risk of misfolding than long-living proteins. Their respective sequences should therefore also experience different selective pressures against protein aggregation. Such evolutionary pressure might have resulted in different affinities towards molecular chaperones and different implications towards aggregation-related diseases.

In order to determine the relationship between protein lifetime and protein aggregation we here combine experimental lifetime measured for 611 proteins [19] with the corresponding gene expression data in 532 healthy individuals. We also correlated experimental chaperone interaction data and structural information of these proteins to their aggregation propensity using TANGO [20], an algorithm that accurately predicts the intrinsic aggregation propensity of protein sequences. This analysis resulted in two major observations: i) short-living proteins on average are predicted to have longer and more severe aggregating regions than long-living proteins, and ii) the evolutionary enrichment of aggregation breaking gatekeeper residues is less pronounced in short-living proteins, suggesting that they experience milder selective pressure to minimize aggregation. Further, we also found significantly less interactions between short-living proteins and molecular chaperones in the IntAct database [21]. Our results

Author Summary

In order to carry out their biological function, proteins need to fold into well-defined three-dimensional structures. Protein aggregation is a process whereby proteins misfold into inactive and often toxic higher order structures, which is implied in about 30 human diseases such as Alzheimer's disease, Parkinson's disease and systemic amyloidosis. In earlier work it has been shown that although protein aggregation is an intrinsic property of polypeptide chains that cannot be entirely avoided, evolution has optimized protein sequences to minimize the risk of aggregation in a proteome. Here we show that this pressure is not uniform, but that proteins with a short lifetime have on average a higher aggregation propensity than long-living proteins. In addition, we show that high turnover proteins also make fewer interactions with chaperones. Taken together, these observations suggest that under normal physiological conditions the aggregation propensity of short-lived proteins does not represent a significant treat for the biochemistry of the cell. Presumably the strong dependence of these proteins on proteasomal degradation is sufficient to preclude the accumulation of aggregates. As proteasomal activity declines with age this would also explain why we observe a higher association of high turnover proteins with age-dependent aggregation-related diseases.

suggest that under normal circumstances, protein aggregation of short-living proteins is not problematic, and thus there is little evolutionary pressure to reduce the intrinsic aggregation propensity or optimize chaperone interaction. This would turn such proteins into the Achilles' heel of the proteome in conditions where proteasomal function is significantly reduced, such as is reported for normal human ageing [22,23,24,25]. In support of this hypothesis, we found that all but one of the proteins with experimentally determined turnover rates that are involved in a protein deposition disease belong to the fastest turnover rate group.

Materials and Methods

Scope and limitation of protein aggregation prediction

The current study focuses on short-stretch mediated protein aggregation, where specific segments of a polypeptide chain assemble into an intermolecular beta-sheet and thus nucleate aggregation. Since current knowledge in the field suggests that the short-stretch mediated protein aggregation covers the majority of disease-associated protein deposition, and no reliable prediction methods exist for alternative protein aggregation mechanisms, we feel justified to ignore alternative aggregation mechanisms such as 3D domain swapping and native protein aggregation. Like all current protein aggregation prediction algorithms, TANGO calculates intrinsic aggregation propensity of an input polypeptide sequence and returns short stretches predicted to have a high propensity to nucleate protein aggregation through the formation of intermolecular beta-sheets. These regions constitute the intrinsic aggregation propensity of the sequence in the absence of globular structure. Since these aggregation prone regions are nearly always part of the hydrophobic core when the protein resides in its native conformation, the aggregating stretches identified computationally need to become exposed by (partial) unfolding of the protein before they can actually nucleate protein aggregation. So, although three dimensional relationships that existed in the folded state are no longer relevant during assembly into an intermolecular beta-sheet,

they are highly relevant to determine if a particular region is likely to become exposed in the first place. In order to estimate the likelihood that a given short polypeptide segment may become exposed by (partial) protein unfolding, we employ the FoldX force field, which calculates the contribution of each amino acid to the thermodynamic stability of the three dimensional structure of the protein, thus allowing to determine if an aggregation prone region is in a stable or less stable part of the structure.

Datasets

Protein selection. Trans-membrane (TM) and extracellular proteins in the experimental dataset were excluded from the analysis. As hydrophobic trans-membrane regions of the TM proteins are not under selective pressure against aggregation they should not be considered for the analysis of the relation between protein lifetime and aggregation tendency. Since this study analyses the relation between proteasomal degradation and aggregation, extracellular proteins that are degraded by lysosomes are also deleted. We selected these proteins using the keywords "Membrane" (KW-472) and "Extracellular matrix" (KW-0272). This resulted in a dataset of 191 short-living (PSI ≤ 2) and 420 long-living (PSI ≥ 5) proteins.

Lifetime of proteins. Yen et al. developed a global stability analysis, a high throughput approach for proteome-scale protein-turnover analysis, resulting in a protein stability index (PSI) for 8000 human proteins [19]. PSI scores ranges from 1 to 7, with higher value indicating higher protein stability. Using a low and high cut-off value to eliminate proteins with intermediate lifetime, the dataset is split in two groups of short (PSI ≤ 2) versus long-living (PSI ≥ 5) proteins.

Determination of aggregating sequences and flanking gatekeeper residues

The statistical mechanics algorithm TANGO [20] was used to determine the aggregation-prone regions in the human proteins. This resulted in an aggregation propensity (0–100%) for each residue, whereby an aggregating segment is defined as a continuous stretch of at least five consecutive residues, each with a TANGO score higher than 5%. The five positions before and after aggregation-prone regions are considered as "gatekeeping flanks", with each P, R, K, E or D counting as gatekeepers [17]. No distinction was made between gatekeepers at the N or C terminus of the aggregating stretch.

Gene expression analysis

Our dataset was composed of 532 HG-U133_Plus_2 type microarray experiments extracted from GEO (Gene Expression Omnibus) [26]. Queries were carried out using GEOmetadb module from R [27]. The dataset is composed of cancer healthy control samples only. HG-U133_Plus_2 microarrays contains probe sets of 54675 human genes per chip. All 532 chips were preprocessed in one single block using robust multichip average (RMA). RMA processing consists of three steps: background adjustment, quantile normalization and finally summarization. A list of common housekeeping genes (EIF4G2, RPL9, SFR9, GUK1, H3F3A, RHOA, ACTB) was used to confirm that the expression levels remain constant for the whole dataset. The dataset was divided into two subsets according to long-living and short-living proteins. Conversion of Affymetrix to Uniprot identifiers was done using Babelomics4 id converter [28,29].

FoldX modeling

Structures were selected according to the following criteria: (1) 100% sequence identity with the sequence of interest, (2) crystal
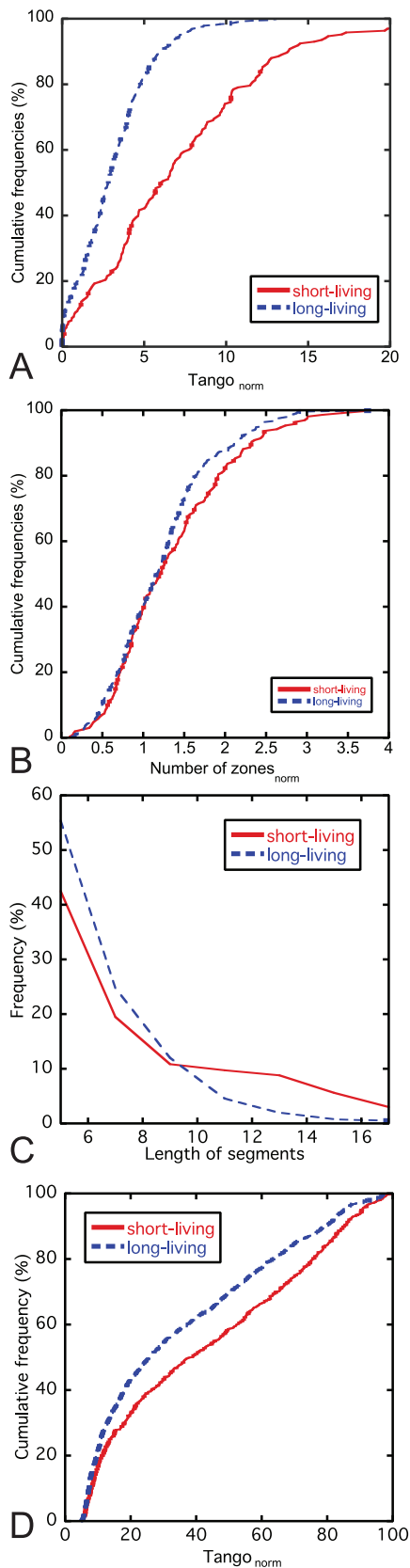
Figure 1. Short-living proteins display stronger aggregation propensity than long-living proteins. (A) Cumulative frequency of the length normalized TANGO scores for the short-living (PSI ≤ 2) and long-living proteins (PSI ≥ 5). The occurrence of stronger aggregating sequences (higher TANGO score) is higher in short-living proteins. (B) Cumulative frequency of the number of aggregating segments for the short-living (PSI ≤ 2) and long-living proteins (PSI ≥ 5). (C) Frequency of the length of the aggregating segments in short-living (PSI ≤ 2) and long-living proteins (PSI ≥ 5). (D) Cumulative frequency of the aggregation propensity of each aggregating segments for short-living (PSI ≤ 2) and long-living proteins (PSI ≥ 5).
doi:10.1371/journal.pcbi.1002090.g001

structure, (3) resolution at least 3 Ä. All modeling was performed using the FoldX 2.8 force field and tool suite [30,31]. All structures were repaired using the RepairPDB command and homology models were constructed using the BuildModel command. The stability of the aggregation nucleating regions was extracted using the SequenceDetail command.

Statistics

Comparison of the distributions for each parameter tested in the analysis of short versus long-living proteins was performed using Mann-Whitney and Kolmogorov-Smirnov tests.

Results

Determination of the aggregation propensity

Yen et al. developed a global stability analysis, a high throughput approach for proteome-scale protein-turnover analysis, resulting in a protein stability index (PSI) for 8000 human proteins [19]. PSI scores ranges from 1 to 7, with higher values indicating higher biological protein stability and thus slower protein turnover. To simplify the analysis, we used a low and a high cut-off value to eliminate proteins with intermediate lifetime, so that the data were split in two groups of short (PSI ≤ 2) versus long-living (PSI ≥ 5) proteins (Text S1). A number of characteristics of the aggregation propensity of these 611 proteins were determined using the TANGO algorithm [20]: *i)* the average aggregation propensity of the protein (total TANGO score normalized by protein length), *ii)* the number of aggregating segments in the protein, *iii)* the length of aggregating segments, and *iv)* the aggregation propensity of each aggregating segment. The correlation with the experimentally determined biological lifetime of the protein was tested for each individual parameter and significant differences were found (Text S1): Short-living proteins display a higher average aggregation propensity (Figure 1A), which is not caused by an increase in the average number of aggregating segments (Figure 1B), but by an significant increase in their length (Figure 1C) and aggregation propensity (Figure 1D). As previous studies have shown that long proteins on average have less effective aggregation-promoting regions than shorter proteins [32] and the average length of short and long-living proteins is respectively 263 and 357 amino acids, the aforementioned observations could also be due to the longer mean length of long-living proteins. In order to exclude this possibility, we repeated the analysis after the exclusion of proteins longer than 300 amino acids, and found that the difference in aggregation tendency between the two lifetime categories remains significant (p<0.001), showing that the observed difference in aggregation tendency is linked to the disparity in lifetime, and is independent of the difference in mean length of the proteins. This conclusion is confirmed by plotting the average aggregation tendency in function of the protein length for each lifetime category (Figure 2A). In view of the idea introduced by Vendrusculo and co-workers that protein expression levels are tuned to the solubility limit of the protein [18], we need to exclude that the difference in aggregation load in our data is simply due to a lower expression
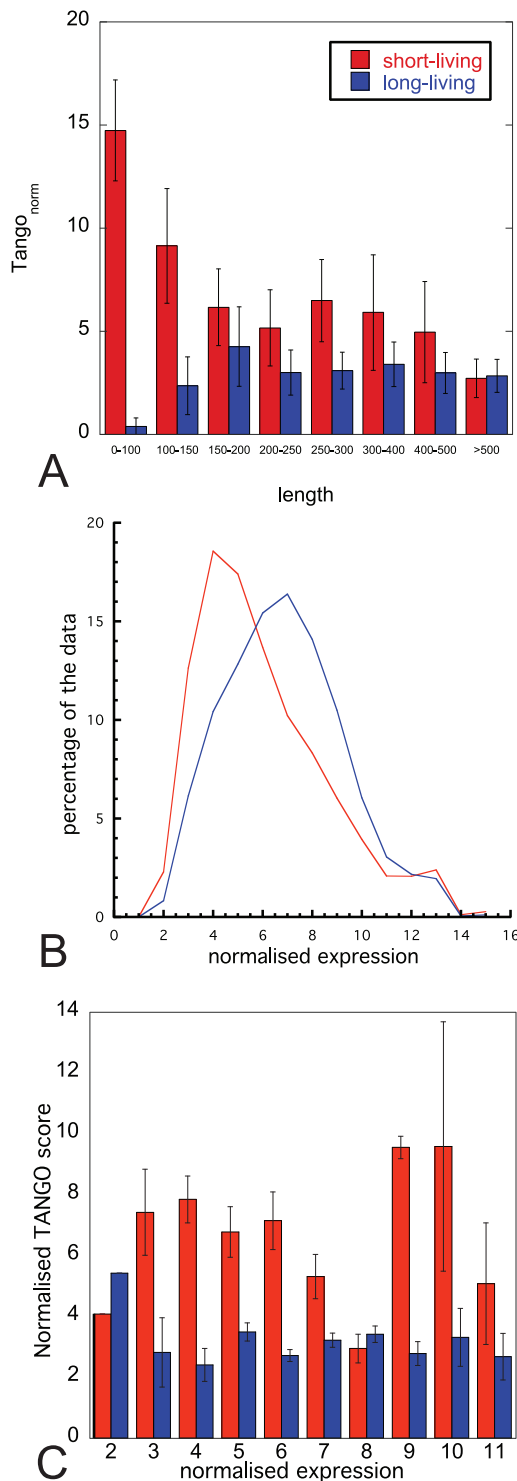
Figure 2. The effect of other factors that are known to modulate aggregation propensity. (A) Average aggregation propensity in function of the protein length for short-living (PSI ≤ 2) and long-living (PSI ≥ 5) proteins. (B) Density plot of the normalized expression level of the proteins from our short-living and long-living groups recorded from microarray data from 532 healthy human individuals. (C) Difference in aggregation load between fast and slow turnover proteins separated by expression level.
doi:10.1371/journal.pcbi.1002090.g002

level for the fast turnover proteins. To address this, we employed publically available microarray data from the Gene Expression Omnibus (GEO) [33], corresponding to 532 healthy individuals from 62 studies to compare expression levels of the proteins in our lifetime dataset. The density plot of the normalized expression levels for all proteins from the short lifetime and long lifetime groups reveals indeed a different composition of both groups in terms of expression levels (Figure 2B). However, when we plot the length normalized aggregation score of the short and long-living proteins grouped per expression level (Figure 2C), we see that the expression level is not the determining factor in the difference in aggregation propensity between fast and slow turnover proteins. These results suggest that proteins with a short biological lifetime undergo less evolutionary pressure to minimize the burden of aggregation.

The influence of thermodynamic stability of the protein

An alternative explanation for the lower sensitivity of fast turnover proteins to the evolutionary pressure against protein aggregation could be that these proteins possess native structures with inherently superior thermodynamic stability to those of proteins from the long lifetime group. Given the significant structural coverage of our dataset, i.e. there are high resolution crystallographic structures available for 127 proteins in our dataset of 611 (Text S1), we can address this question using a modeling approach. To do so we employed the FoldX force field [31] to calculate the thermodynamic stability of the aggregation nucleating regions predicted by TANGO in the corresponding crystal structures. We then plotted the average thermodynamic stability of the aggregating nucleating regions per bin of aggregation propensity according to TANGO (Figure 3A). In this plot, we observe a clear correlation between the aggregation propensity of a polypeptide stretch and thermodynamic stability of the same region in the context of its native three-dimensional structure, so that sequences with the highest aggregation propensity form the most stable parts of the protein structure under native conditions, which is in accordance with previous observations [5]. Importantly, Figure 3A reveals no significant differences between proteins with a long or a short lifetime, showing that the difference in aggregation propensity between these groups is not due to fundamental differences in protein architecture or thermodynamic stability.

Occurrence of gatekeeper residues to oppose protein aggregation

It has been well established that evolutionary pressure against protein aggregation has resulted in the enrichment at the flanks of aggregation prone segments of *gatekeeper* residues, a term used to indicate amino acids that counteract aggregation [12,15,34]. This disruption of the aggregation prone stretches is achieved by a) the repulsive effect of charge (arginine, aspartate, glutamate), b) the entropic penalty for burial (arginine and lysine) or c) incompatibility with beta-structure conformation (proline) [34]. We analyzed the frequency of occurrence of gatekeeper residues in our short- and long-living protein datasets and found that the frequency of occurrence of gatekeeper residues shows a small but significant reduction in short-living proteins (Figure 3B), which indicates that the introduction of gatekeepers as an evolutionary mechanism, to minimize aggregation is less pronounced in this set. This is consistent with the observation of longer aggregating stretches since they are less frequently interrupted by aggregation breaking residues, resulting also in a higher aggregation propensity of the stretches.
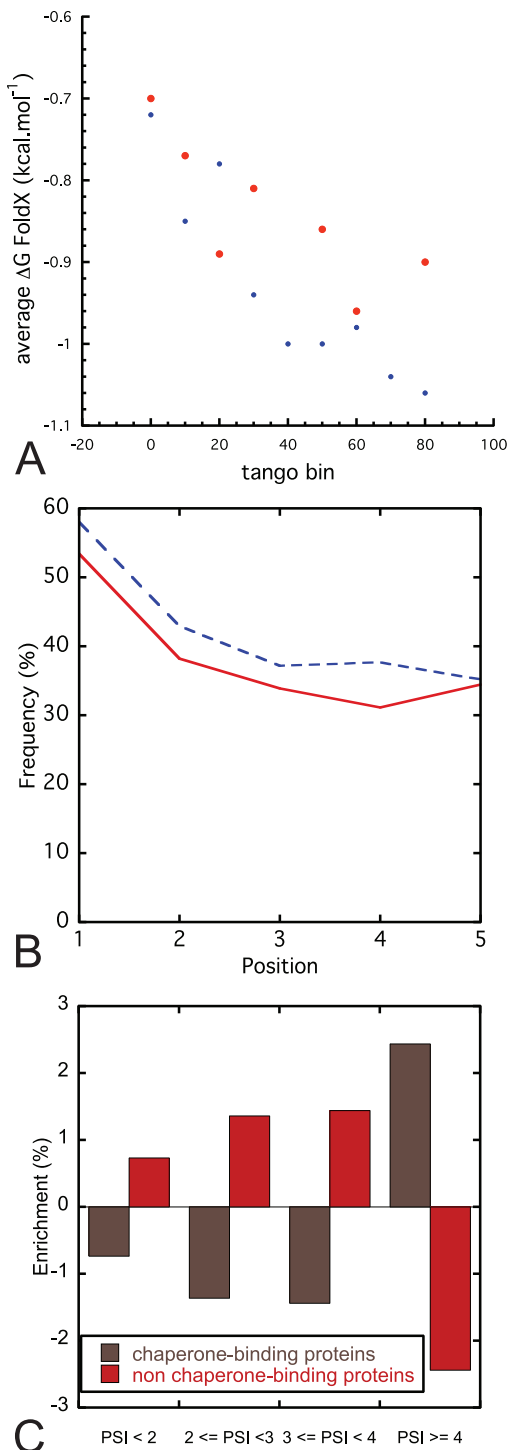
Figure 3. The effect of cellular and evolutionary mechanisms that counteract protein aggregation. (A) A plot of the average thermodynamic stability of aggregation nucleating regions in the context of a native folded protein in function of the aggregation propensity shows that the most strongly aggregating segments are on average buried in the most stable regions of the protein and that this trend is similar for the long and short-living proteins in our set. The correlation values on the raw unaveraged data are 0.43 and 0.30 for short and long-living protein respectively, which rise to 0.79 and 0.91 in the bin-average plot shown here. (B) Frequency of the gatekeepers (residues P, R, K, D, E) for the short-living (PSI ≤ 2) and long-living proteins (PSI ≥ 5). (C) Enrichment of the chaperone-binding and non-chaperone-binding proteins for the different lifetime categories.
doi:10.1371/journal.pcbi.1002090.g003

## Relation between protein stability and chaperone binding

A major component of the protein quality control system that evolved in all forms of cellular life to deal with the unavoidable burden of protein misfolding and aggregation is formed by the diverse families of molecular chaperones, which are a class of proteins that assist other proteins in (re)folding and disaggregation and eventually shuttle substrates to the degradation machinery [35]. In order to address the question if protein turnover rates influence the requirement of chaperone assistance of a protein, we searched the protein interaction database IntAct [21] (release March 19, 2010) for experimentally recorded interactions between proteins from our dataset and an extensive list of known human molecular chaperones (listed in Text S1). A total of 237 chaperone-binding proteins were identified, but experimentally determined protein stability was available for only 114 proteins. Based on Yen *et al.*, we divided this set of proteins into four categories according to their PSI turnover scores: short half-life (PSI < 2), medium half-life (2 ≤ PSI < 3), long half-life (3 ≤ PSI < 4) and extra-long half-life (PSI ≥ 4) [19]. For each category we calculated the enrichment of chaperone-binding proteins, where enrichment is defined as $PSI_N/PSI_T - SUM_N/SUM_T$. PSI_x is the number of proteins in a given set x, belonging to a given PSI category and SUM_x the total number of proteins in a given set x. X points to the total set (T) or the (non-) chaperone-binding proteins (N). Comparison of the chaperone enrichment in short-living versus long-living proteins shows that in our limited dataset, proteins that interact with molecular chaperones are significantly enriched in the group of long-living proteins (Figure 3C). Given we observed no fundamental differences in the thermodynamic stability or protein architecture between these groups (see FoldX analysis above), this suggests that short-living proteins on average require less chaperone intervention than long-living proteins, consistent with the notion that their fast degradation rate is sufficient to protect against misfolding and aggregation.

Relation between disease-associated mutations and lifetime

We investigated which of the proteins in our dataset are involved in a human disease associated with protein deposition and found 16 proteins with known PSI score (Text S1). Interestingly, all but one of these proteins belong to the category of short (PSI < 2) or medium (2 ≤ PSI < 3) half-life. Although this analysis is not exhaustive, the data does suggest that the lack of evolutionary pressure to reduce aggregation in short-living proteins can backfire in circumstances were their turnover is altered.

Discussion

Protein aggregation is triggered by short polypeptide stretches within a protein sequence that assemble into intermolecular beta-sheets when they become exposed to the solvent [8,36,37] (Figure 4). These aggregation nucleating regions can be predicted with good accuracy with biocomputational tools [20,38,39,40,41,42,43,44,45, 46,47,48,49,50,51] and earlier work has shown that their occurrence is an inevitable consequence of the structural requirements of protein structure [52]. Globular protein architecture requires the tertiary packing of hydrophobic secondary structure elements to form a stable hydrophobic core. Unfortunately, these physicochemical parameters are also associated to a high probability for self-assembly of such secondary structure elements into β-aggregates [53,54]. Indeed, less than 10% of globular protein domains are devoid of aggregation propensity [12]. As a consequence of these
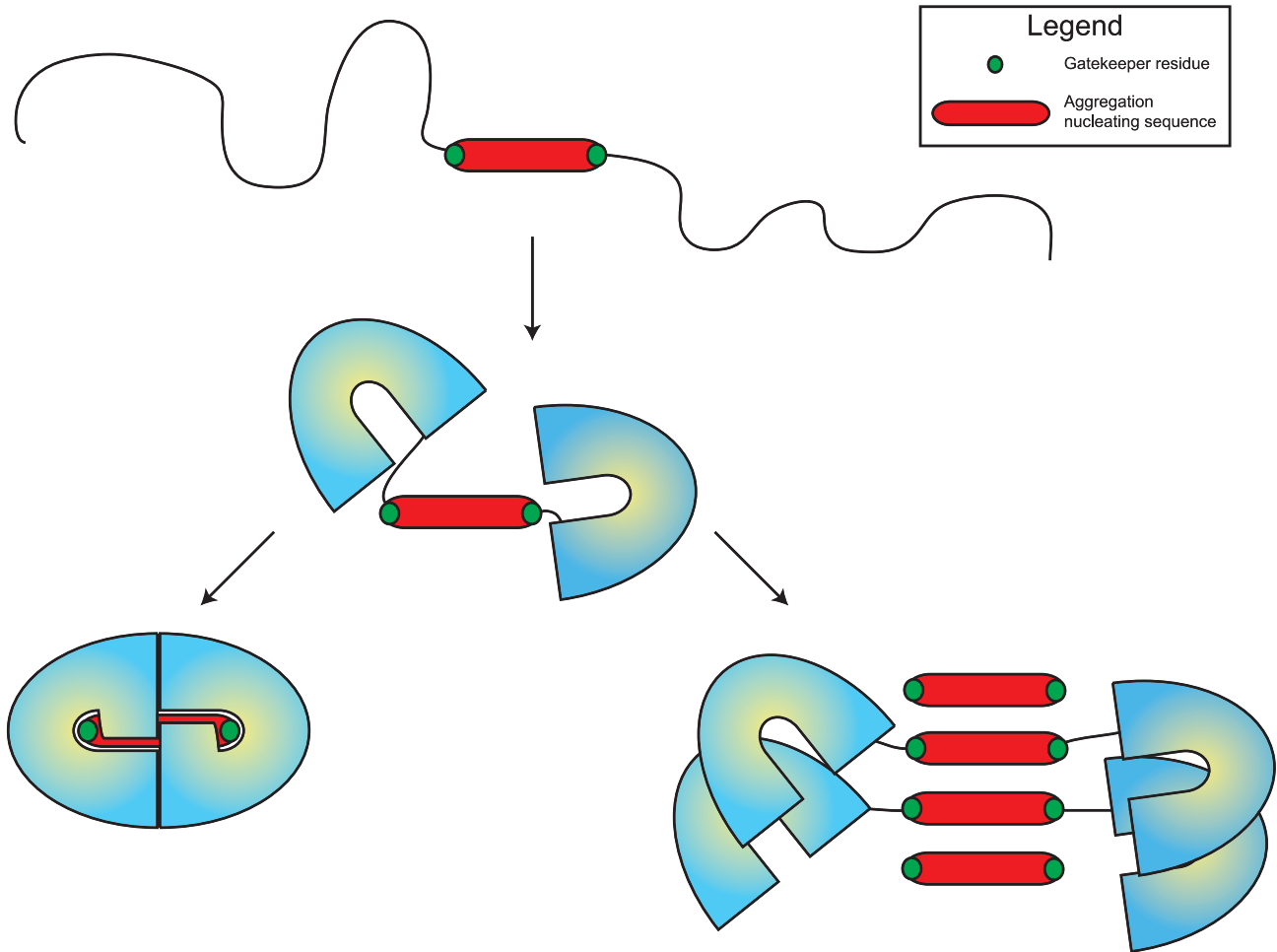
Figure 4. Schematic representation of the short-stretch hypothesis of protein aggregation. In the unfolded state (top) and partially folded intermediates, the protein exposes an aggregation nucleating stretch, that becomes buried upon folding into the globular native structure (bottom left). In a competing reaction, aggregation-prone stretches may align into an intermolecular β-sheet, effectively nucleating the formation of a protein aggregate (bottom right). The gatekeeper residues indicated in green reduce the rate of the aggregation reaction by interfering with the beta-sheet structure through steric hindrance and charge repulsion.
doi:10.1371/journal.pcbi.1002090.g004

overlapping but opposing forces that govern protein folding and aggregation, protein folding is generally a very inefficient process [55,56]. Moreover, aggregation is detrimental for the cell as misfolded proteins are inactive [57] and can acquire toxic gain-of-function [58]. Protein homeostasis is therefore tightly regulated by the protein quality control machinery of the cell.

Given the high burden of protein aggregation on the proteome, and even if aggregation propensity cannot be avoided altogether, selective pressure to minimize the aggregation propensity of protein sequences is still to be expected. Indeed, it was found that aggregation-opposing residues are enriched at specific sites along the sequence of proteins [12,59]. These so-called aggregation-gatekeepers residues, consisting of prolines and charged amino acids, are systematically found at the flanks of aggregation-prone sequences stretches within proteins. Due to their β-breaking nature or charge they efficiently lower the aggregation propensity of hydrophobic stretches while at the same time preserving hydrophobic cores by their peripheral placement (Figure 4). Removal of gatekeepers increases aggregation and as a result gatekeeper mutations are three times more frequent in human disease mutants than in human polymorphisms [17,60].

Selective pressure against aggregation is not homogeneous. We previously showed that enrichment of gatekeeper residues is more pronounced at the flanks of strongly aggregating sequences [12] and it was also shown that aggregation propensity inversely correlates with gene expression [18]. In this study we employed the TANGO aggregation prediction tool [20] to compare the aggregation characteristics of proteins taken from the extremes of the protein lifetime distribution from the large scale data by Yen *et al* [19]. We observe a significantly higher aggregation propensity in proteins with a short lifetime than in proteins with a long lifetime. Analysis of gene expression data in 532 healthy individuals excluded the possibility that the observed difference in aggregation propensity arises from differences in gene expression levels between short-living and long-living proteins. Additionally the FoldX [31] analysis of the structures from both groups of proteins clearly show that this is not a result from a superior thermodynamic stability of short lifetime proteins, but rather from a genuinely higher aggregation propensity of their protein sequence. The higher aggregation propensity of short-living proteins does not originate from a higher number of aggregating regions, but rather from the higher average length and

aggregation propensity of these regions, which can be traced back to a reduction in the amount of aggregation breaking gatekeeper residues. Hence, the reduced placement of gatekeepers in short-living proteins and the resulting higher average aggregation propensity, is evidence for the fact that proteins with a fast turnover rate experience less selective pressure to minimize aggregation than proteins with a longer biological lifetime.

Moreover, a search of the IntAct database [21] revealed that there are significantly more recorded chaperone interactions for long-living proteins than short-living proteins. So, not only do short-living proteins experience milder selective pressure against aggregation, but at the same time they also interact less frequently with molecular chaperones or at least form less stable interactions of the type that can be recorded by current experimental techniques. Taken together, these data strongly suggest that the misfolding of short-living proteins is generally not affecting the fitness of the cell, as presumably the strong dependence of these proteins on proteasomal degradation suffices to avoid the accumulation of protein aggregates.

On the other hand, it is known that the efficiency of the proteasomal system erodes as a result of physiological ageing [61,62,63]. Under these changing conditions, proteins with a higher aggregation propensity and lacking sufficient affinity for chaperones would form the Achilles' heel of the proteome and be among the most susceptible to aggregate. In this respect it is interesting to see that some of the fast turnover proteins from the dataset are indeed associated with human diseases with a protein deposition phenotype.

Supporting Information

Text S1 Supplementary data. Table 1. Comparison between the aggregation parameters for short-living and long-living proteins. The analysed population is the group of short-living protein, the reference population are the long-living proteins. ++ and − indicate that the population has a distribution significantly (p<0.001) shifted to respectively higher or lower values than the reference population in the performed statistical test, idem for + and − where p<0.01. **Table 2. Lifetime data for disease-associated proteins.** We show the lifetime values of the proteins from the Yen dataset [19] on protein lifetime that are associated with protein deposition diseases. **Table 3. Overview of the protein set used.** From the Yen dataset [19] on protein lifetime, we here show the lifetime values for the 611 proteins that fall in the extreme categories (longest and shortest lifetimes respectively). Where high resolution structural information is available in the Protein Structure Databank (PDB) (http://www.pdb.org) [64] we indicate the PDBID. **Table 4. Overview of the chaperone set.** This table contains the chaperones used in the IntAct [21] interaction study, represented by their accession number, entry name and UniProt comment. (DOC)

Author Contributions

Conceived and designed the experiments: JS FR. Performed the experiments: GDB JR JDB. Analyzed the data: GDB JR JDB JS FR. Contributed reagents/materials/analysis tools: JD. Wrote the paper: GDB JR JS FR.

References

1. Powers ET, Morimoto RI, Dillin A, Kelly JW, Balch WE (2009) Biological and chemical approaches to diseases of proteostasis deficiency. Annu Rev Biochem 78: 959–991.
2. Pechmann S, Levy ED, Tartaglia GG, Vendruscolo M (2009) Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. Proc Natl Acad Sci U S A 106: 10159–10164.
3. Ben-Zvi A, Miller EA, Morimoto RI (2009) Collapse of proteostasis represents an early molecular event in Caenorhabditis elegans aging. Proc Natl Acad Sci U S A 106: 14914–14919.
4. Balch WE, Morimoto RI, Dillin A, Kelly JW (2008) Adapting proteostasis for disease intervention. Science 319: 916–919.
5. Tartaglia GG, Vendruscolo M (2008) Proteome-level interplay between folding and aggregation propensities of proteins. J Mol Biol 402: 919–928.
6. Masino L, Nicastro G, Calder L, Vendruscolo M, Pastore A (2011) Functional interactions as a survival strategy against abnormal aggregation. FASEB J 25: 45–54.
7. Luheshi LM, Crowther DC, Dobson CM (2008) Protein misfolding and disease: from the test tube to the organism. Curr Opin Chem Biol 12: 25–31.
8. Esteras-Chopo A, Serrano L, de la Paz ML (2005) The amyloid stretch hypothesis: Recruiting proteins toward the dark side. Proc Natl Acad Sci U S A 102: 16672–16677.
9. Rousseau F, Schymkowitz J, Serrano L (2006) Protein aggregation and amyloidosis: confusion of the kinds? Curr Opin Struct Biol 16: 1–9.
10. Prahlad V, Morimoto RI (2009) Integrating the stress response: lessons for neurodegenerative diseases from C. elegans. Trends Cell Biol 19: 52–61.
11. Reumers J, Maurer-Stroh S, Schymkowitz J, Rousseau F (2009) Protein sequences encode safeguards against aggregation. Hum Mutat 30: 431–437.
12. Rousseau F, Serrano L, Schymkowitz JW (2006) How evolutionary pressure against protein aggregation shaped chaperone specificity. J Mol Biol 355: 1037–1047.
13. Monsellier E, Ramazzotti M, Taddei N, Chiti F (2008) Aggregation Propensity of the Human Proteome. Plos Comput Biol 4.
14. Monsellier E, Ramazzotti M, de Laureto PP, Tartaglia GG, Taddei N, et al. (2007) The distribution of residues in a polypeptide sequence is a determinant of aggregation optimized by evolution. Biophys J 93: 4382–4391.

15. Monsellier E, Chiti F (2007) Prevention of amyloid-like aggregation as a driving force of protein evolution. EMBO Rep 8: 737–742.
16. de Groot NS, Ventura S (2010) Protein aggregation profile of the bacterial cytosol. PLoS One 5: e9383.
17. Reumers J, Rousseau F, Schymkowitz J (2009) Multiple evolutionary mechanisms reduce protein aggregation. Open Biol 2: 176–184.
18. Tartaglia GG, Pechmann S, Dobson CM, 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.
19. Yen HC, Xu Q, Chou DM, Zhao Z, Elledge SJ (2008) Global protein stability profiling in mammalian cells. Science 322: 918–923.
20. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nat Biotechnol 22: 1302–1306.
21. Aranda B, Achuthan P, Alam-Faruque Y, Armean I, Bridge A, et al. (2010) The IntAct molecular interaction database in 2010. Nucleic Acids Res 38: D525–531.
22. Li Y, Wang YS, Shen XF, Hui YN, Han J, et al. (2008) Alterations of activity and intracellular distribution of the 20S proteasome in ageing retinal pigment epithelial cells. Exp Gerontol 43: 1114–1122.
23. Bregegere F, Milner Y, Friguet B (2006) The ubiquitin-proteasome system at the crossroads of stress-response and ageing pathways: a handle for skin care? Ageing Res Rev 5: 60–90.
24. Carrard G, Dieu M, Raes M, Toussaint O, Friguet B (2003) Impact of ageing on proteasome structure and function in human lymphocytes. Int J Biochem Cell Biol 35: 728–739.
25. Stolzing A, Grune T (2001) The proteasome and its function in the ageing process. Clin Exp Dermatol 26: 566–572.
26. Barrett T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, et al. (2007) NCBI GEO: mining tens of millions of expression profiles–database and tools update. Nucleic Acids Res 35: D760–765.
27. Zhu Y, Davis S, Stephens R, Meltzer PS, Chen Y (2008) GEOmetadb: powerful alternative search engine for the Gene Expression Omnibus. Bioinformatics 24: 2798–2800.
28. Medina I, Carbonell J, Pulido L, Madeira SC, Goetz S, et al. (2010) Babelomics: an integrative platform for the analysis of transcriptomics, proteomics and

genomic data with advanced functional profiling. Nucleic Acids Res 38 Suppl: W210–213.

29. Al-Shahrour F, Carbonell J, Minguez P, Goetz S, Conesa A, et al. (2008) Babelomics: advanced functional profiling of transcriptomics, proteomics and genomics experiments. Nucleic Acids Res 36: W341–346.

30. Schymkowitz JW, Rousseau F, Martins IC, Ferkinghoff-Borg J, Stricher F, et al. (2005) Prediction of water and metal binding sites and their affinities by using the Fold-X force field. Proc Natl Acad Sci U S A 102: 10147–10152.

31. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, et al. (2005) The FoldX web server: an online force field. Nucleic Acids Res 33: W382–388.

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

33. Wheeler DL, Barrett T, others (2008) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 36: D13–21.

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

35. McClellan AJ, Tam S, Kaganovich D, Frydman J (2005) Protein quality control: chaperones culling corrupt conformations. Nat Cell Biol 7: 736–741.

36. Teng PK, Eisenberg D (2009) Short protein segments can drive a non-fibrillizing protein into the amyloid state. Protein Eng Des Sel 22: 531–536.

37. Ventura S, Zurdo J, Narayanan S, Parreno M, Mangues R, et al. (2004) Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. Proc Natl Acad Sci U S A 101: 7258–7263.

38. Maurer-Stroh S, Debulpaep M, Kuemmerer N, de la Paz ML, Martins IC, et al. (2010) Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. Nat Methods 7: 237–242.

39. Trovato A, Seno F, Tosatto SCE (2007) The PASTA server for protein aggregation prediction. Protein Eng Des Sel 20: 521–523.

40. Tartaglia GG, Cavalli A, Pellarin R, Caflisch A (2005) Prediction of aggregation rate and aggregation-prone segments in polypeptide sequences. Protein Sci 14: 2723–2734.

41. Caflisch A (2006) Computational models for the prediction of polypeptide aggregation propensity. Curr Opin Chem Biol 10: 437–444.

42. Tartaglia GG, Vendruscolo M (2008) The Zyggregator method for predicting protein aggregation propensities. Chem Soc Rev 37: 1395–1401.

43. Conchillo-Sole O, de Groot NS, Aviles FX, Vendrell J, Daura X, et al. (2007) AGGRESCAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. BMC Bioinformatics 8: 65.

44. Zibaee S, Makin OS, Goedert M, Serpell LC (2007) A simple algorithm locates beta-strands in the amyloid fibril core of alpha-synuclein, Abeta, and tau using the amino acid sequence alone. Protein Sci 16: 906–918.

45. Bryan AW, Jr., Menke M, Cowen LJ, Lindquist SL, Berger B (2009) BETASCAN: probable beta-amyloids identified by pairwise probabilistic analysis. PLoS Comput Biol 5: e1000333.

46. Rojas Quijano FA, Morrow D, Wise BM, Brancia FL, Goux WJ (2006) Prediction of nucleating sequences from amyloidogenic propensities of tau-related peptides. Biochemistry 45: 4638–4652.

47. Saiki M, Konakahara T, Morii H (2006) Interaction-based evaluation of the propensity for amyloid formation with cross-beta structure. Biochem Biophys Res Commun 343: 1262–1271.

48. Thompson MJ, Sievers SA, Karanicolas J, Ivanova MI, Baker D, et al. (2006) The 3D profile method for identifying fibril-forming segments of proteins. Proc Natl Acad Sci U S A 103: 4074–4078.

49. Goldschmidt L, Teng PK, Riek R, Eisenberg D (2010) Identifying the amylome, proteins capable of forming amyloid-like fibrils. Proc Natl Acad Sci U S A 107: 3487–3492.

50. Galzitskaya OV, Garbuzynskiy SO, Lobanov MY (2006) Prediction of amyloidogenic and disordered regions in protein chains. PLoS Comput Biol 2: e177.

51. Yoon S, Welsh WJ (2004) Detecting hidden sequence propensity for amyloid fibril formation. Protein Sci 13: 2149–2160.

52. Linding R, Schymkowitz J, Rousseau F, Diella F, 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.

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

54. Chiti F, Taddei N, Baroni F, Capanni C, Stefani M, et al. (2002) Kinetic partitioning of protein folding and aggregation. Nat Struct Biol 9: 137–143.

55. Schubert U, Anton LC, Gibbs J, Norbury CC, Yewdell JW, et al. (2000) Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. Nature 404: 770–774.

56. Kaganovich D, Kopito R, Frydman J (2008) Misfolded proteins partition between two distinct quality control compartments. Nature 454: 1088–1095.

57. Rajan RS, Kopito RR (2005) Suppression of wild-type rhodopsin maturation by mutants linked to autosomal dominant retinitis pigmentosa. J Biol Chem 280: 1284–1291.

58. Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, et al. (2002) Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. Nature 416: 507–511.

59. Otzen DE, Kristensen O, Oliveberg M (2000) Designed protein tetramer zipped together with a hydrophobic Alzheimer homology: a structural clue to amyloid assembly. Proc Natl Acad Sci U S A 97: 9907–9912.

60. Reumers J, Maurer-Stroh S, Schymkowitz J, Rousseau F (2009) Protein Sequences Encode Safeguards Against Aggregation. Human Mutation 30: 431–437.

61. Tonoki A, Kuranaga E, Tomioka T, Hamazaki J, Murata S, et al. (2009) Genetic evidence linking age-dependent attenuation of the 26S proteasome with the aging process. Mol Cell Biol 29: 1095–1106.

62. Hwang JS, Chang I, Kim S (2007) Age-associated decrease in proteasome content and activities in human dermal fibroblasts: restoration of normal level of proteasome subunits reduces aging markers in fibroblasts from elderly persons. J Gerontol A Biol Sci Med Sci 62: 490–499.

63. Proctor CJ, Tsirigotis M, Gray DA (2007) An in silico model of the ubiquitin-proteasome system that incorporates normal homeostasis and age-related decline. BMC Syst Biol 1: 17.

64. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, et al. (2000) The Protein Data Bank. Nucleic Acids Res 28: 235–242.