Proteome Metastability in Health, Aging, and Disease

Prajwal Ciryam[1,2], Gian G Tartaglia[3], Richard I Morimoto[2], Christopher M Dobson[1], Edward P. O'Brien[4], Michele Vendruscolo[1]
[1]Department of Chemistry, University of Cambridge. Cambridge, UK. [2]Department of Molecular Biosciences, Northwestern University, Evanston, USA.
[3]Centre for Genomic Regulation, Barcelona, Spain. [4]Department of Chemistry, Pennsylvania State University, State College, USA.

Abstract

Maintaining protein solubility is fundamental to proteostasis, as the formation of diverse aggregated species is associated with a variety of cytotoxic events and disorders, including Alzheimer's and Parkinson's diseases. Increasing evidence indicates that protein aggregation can be widespread in living systems, as many different proteins aggregate upon cellular stress. The origins of this proteomic metastability, however, remain unclear, as do the reasons only certain proteins aggregate in vivo. We have applied simple models of **cotranslational folding** and **protein supersaturation** to quantify metastability at a proteome scale. We find that many proteins can shift from cotranslational to posttranslational folding because of translation kinetics, a source of metastability for nascent chains. Further, we show that the proteins most vulnerable to aggregation are those whose cellular concentrations are high relative to their intrinsic solubilities. These supersaturated proteins constitute a metastable sub-proteome involved in forming pathological assemblies in stress and ageing. We find that such proteins are overrepresented in the biochemical processes associated with neurodegenerative disorders, helping to rationalise their specific cellular pathologies. We anticipate that this type of analysis can provide a generally applicable basis for tracking the instability of proteomes in ageing, stress, and disease.

Identifying risk points for metastability across the proteome

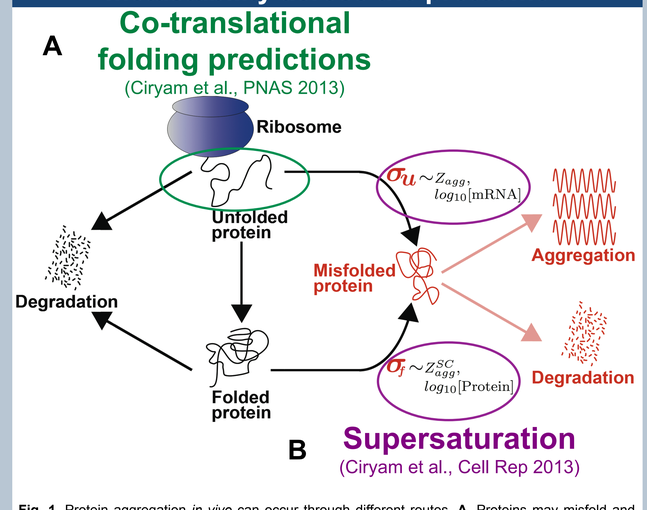

Fig. 1. Protein aggregation *in vivo* can occur through different routes. **A.** Proteins may misfold and aggregate as they emerge from the ribosome or when they unfold at least transiently from the native state. **B.** The proteins most at risk of aggregation are those whose concentration is high with respect to their solubility.

Distinguishing kinetics and thermodynamics in cotranslational folding on the ribosome

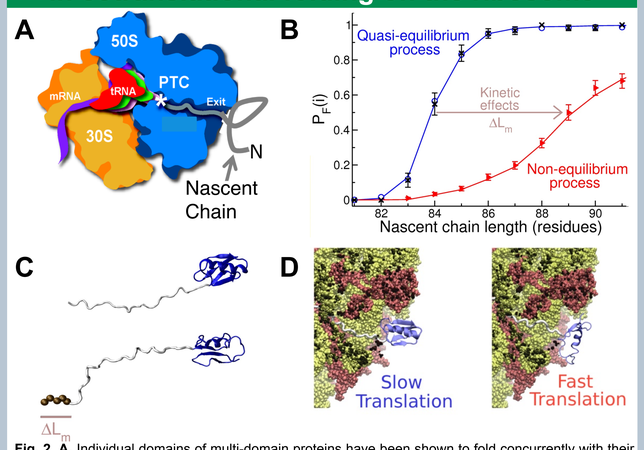

Fig. 2. A. Individual domains of multi-domain proteins have been shown to fold concurrently with their synthesis on the ribosome[1-4]. Because nascent chain synthesis is time-dependent and irreversible, co-translational (CoT) folding occurs out-of-equilibrium, triggering a chain-length delay in domain folding. **B.** CoT folding can be considered a quasi-equilibrium process if amino acid synthesis is arrested at each nascent chain length to permit the partially-synthesized chain to reach equilibrium, and is otherwise a non-equilibrium process. The nascent chain length at which a domain achieves a given probability of CoT folding, P_F, may be delayed in the continuous translation, non-equilibrium scenario, as illustrated in **B.** and **C.** This distinction can arise based on the translation rate, as seen in **D.**

A race between time scales for cotranslational folding

$$P_F(i) = \sum_{j=1}^{i} \frac{\tau_{F,j}^{-1}}{\tau_{A,j+1}^{-1}} \prod_{k=j}^{i} \frac{\tau_{A,k+1}^{-1}}{\tau_{A,k+1}^{-1} + \tau_{F,k}^{-1} + \tau_{D,k}^{-1}}$$

Folding

Unfolding

Amino acid addition

Fig. 3. For a ribosome that exhibits potential dwell times at each codon, the ensemble averaged domain folding probability $P_F(i)$ as a function of the nascent chain length i is given by the equation above, where $\tau_{F,i}$, $\tau_{U,i}$, $\tau_{A,i+1}$ are, respectively, the mean times of **domain folding**, **domain unfolding**, and **amino acid addition** at nascent chain length i. In this equation[5], $P_F(i)$ is the probability that the domain is folded immediately prior to adding the next amino acid.

A systems approach to cotranslational folding

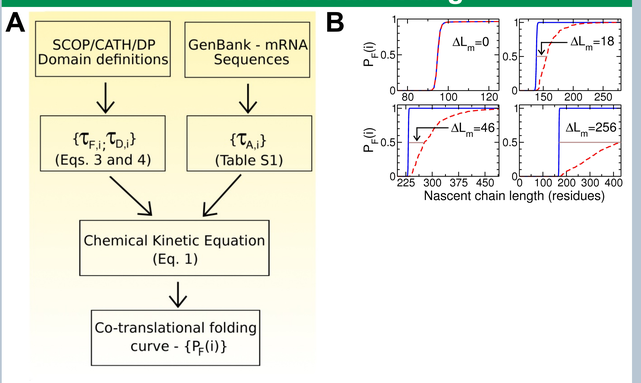

Fig. 4. A. Illustration of the work flow of this systems approach. **B.** Examples of CoT folding curves calculated for four different protein domains in E. coli at in vivo (red) and infinitely slow (blue) translation rates. The domains correspond to ASNC ECOLI, domain 1 (upper left); 3MG2 ECOLI, domain 1 (upper right); ILVC ECOLI, domain 1 (lower left); and ENO, domain 1 (lower right). Note that for ASNC ECOLI, domain 1, the red and blue lines are superimposed.

Kinetic effects can drive major delays in cotranslational folding

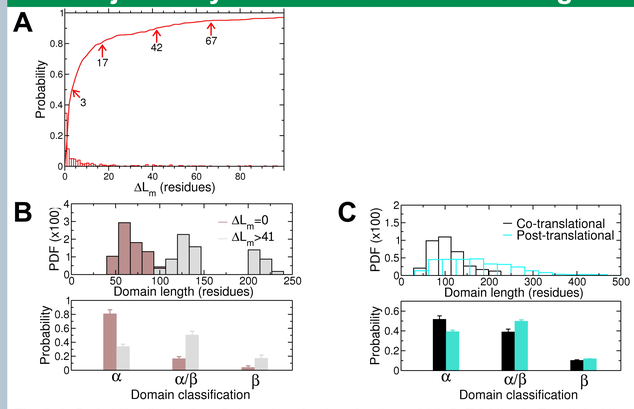

Fig. 5. A. Probability distribution of ΔL_m values for domains that exhibit CoT folding in E. coli doubling every 150 min. at 37°C (n = 422). The cumulative distribution function (CDF) is shown as a solid red line. The arrows and numbers indicate (from left to right) the ΔLm values at which the CDF equals 0.5, 0.8, 0.9, and 0.95, respectively. **B.** Structural characterization of domains that fold cotranslationally and posttranslationally in E. coli cells that are dividing every 150 min. at 37°C. (Upper) Probability density function (PDF) vs. domain length. (Lower) Probability of different domain classifications in terms of mostly α (α), mostly β (β), or mixed α/β secondary structure. **C.** As in **B.**, except the data are from protein domains that fold cotranslationally with ΔL_m = 0 and those that fold with ΔLm values greater than 41 residues. The noncontiguous distribution for the ΔLm > 41 distribution arises from the small number of domains used in its construction (n = 41 data points).

Concentration, aggregation propensity, and the concept of supersaturation

Supersaturation: The state in which **more solute** is dissolved in a solvent than is possible under normal conditions

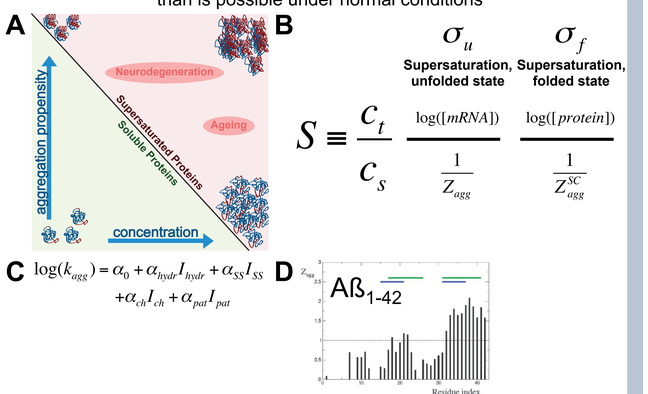

Fig. 6. A. The risk of proteins to be supersaturated increases with their concentration, aggregation propensity, or a combination of the two. **B.** Drawing from work on sickle hemoglobin[6], we define the unfolded supersaturation score, σ_u, using microarray expression and the unfolded Zyggregator score as proxies for unfolded protein concentration and solubility of the unfolded sequence, respectively. We define the folded supersaturation score, σ_f, using normalized spectral abundance factors from MS and the structurally-corrected Zyggregator score as proxies for folded protein concentration and solubility of the folded sequence, respectively. The unfolded Zyggregator score[7] calculated as described in **C.** predicts aggregation-prone regions in amyloid proteins such as **D.** the $A\beta_{1-42}$ peptide.

Widespread aggregation, neurodegenerative diseases, and supersaturated proteins

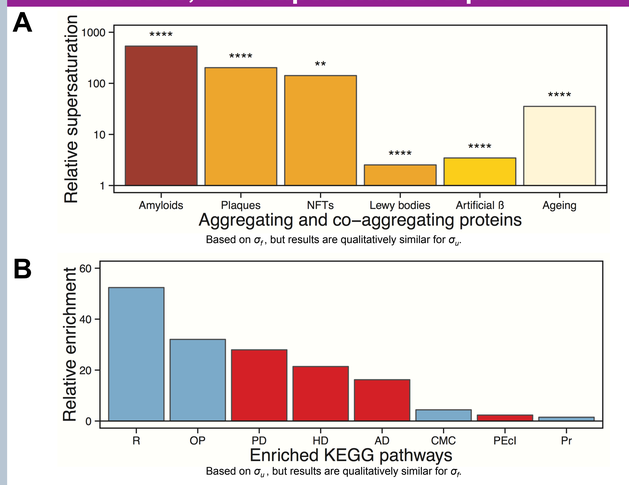

Fig. 7. A. Summary of the results for the different classes of aggregating proteins analyzed in this work. Results are given in terms of the increases in the supersaturation scores (σ_f) over the average value for the whole proteome or for an experimental lysate ("fold change"). **B.** List of the KEGG pathways identified here as significantly enriched (Bonferroni-corrected p values) in proteins at or above the 95th percentile of supersaturation (σ_u).

Supersaturated proteins in the Alzheimer's KEGG pathway

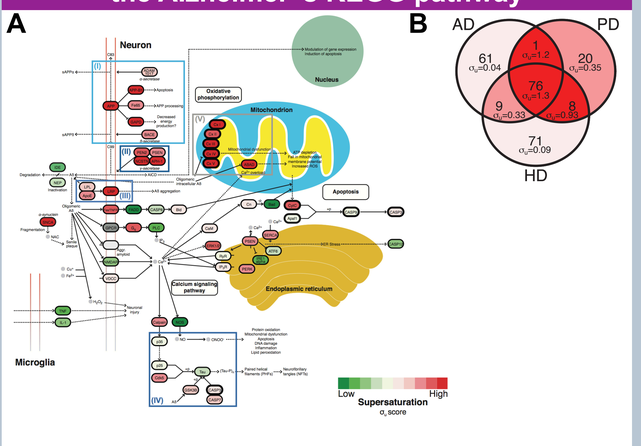

Fig. 8. A. The KEGG pathway for Alzheimer's disease is curated from the literature and includes processes proximal to the cause of the disease such as (I) APP, (II) presenilin, (III) ApoE, and (IV) tau processing, as well as processes more distal to the disease such as (V) oxidative phosphorylation. **B.** Comparison of the σ_u scores for the proteins in the Alzheimer's, Parkinson's, and Huntington's pathways. Colors are assigned based on the division of the σu scores into deciles from low (green) to high (red).

Disruption of proximal and distal Alzheimer's disease processes

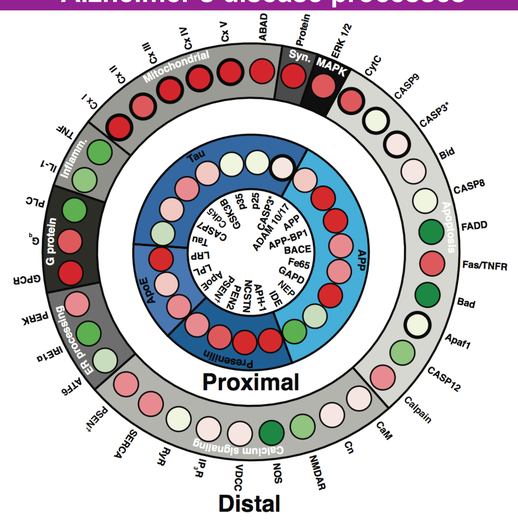

Fig. 9. Most proteins in Alzheimer's pathways proximal to the cause of disease are supersaturated, while select distal pathways, especially those associated with the mitochondria, are also disproportionately composed of supersaturated proteins.

Transcriptional regulation in Alzheimer's disease

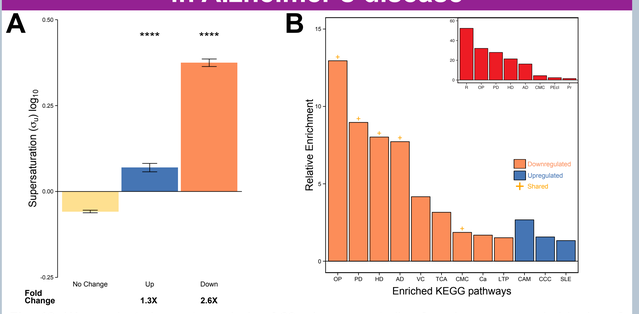

Fig. 10. We conducted a meta-analysis of 20 microarray studies from human autopsied brains of control and Alzheimer's tissue. **A.** Supersaturation scores (Mean ± SEM) (where the median for the full data set of 19,554 proteins was normalized to 0) for proteins corresponding to genes in expression categories No Change (N=16,747), Up (N=1,322), or Down (N=1,485). ****: Holm-Bonferroni-corrected Wilcoxon/Mann-Whitney-U P-values < .0001. Supersaturation scores are calculated using unfolded aggregation propensities calculated as previously reported and the $\log_{10}$ of the geometric mean of expression levels from normal tissue. **B.** Nine pathways are significantly enriched in genes downregulated in Alzheimer's Disease and three pathways are significantly enriched in genes upregulated in Alzheimer's Disease. Five of the downregulated pathways are shared with those previously found to enriched in the most supersaturated proteins (see inset). Inset: Pathways enriched in the 5% most supersaturated proteins in the unfolded state. R: Ribosome; OP: Oxidative Phosphorylation; PD: Parkinson's Disease; HD: Huntington's Disease; AD: Alzheimer's Disease; CMC: Cardiac Muscle Contraction; PEcl: Pathogenic *E. coli* Infection; Pr: Proteasome; VC: *Vibrio cholerae* Infection; TCA: Tricarboxylic Acid Cycle; Ca: Calcium Signaling; LTP: Long Term Potentiation; CAM: Cell Adhesion Molecules; CCC: Complement and Coagulation Cascade; SLE: Systepic Lupus Erythematosus.

Conclusions

- About **1/3** of *E. coli* cytosolic proteins arepredicted to exhibit **CoT folding**
- The majority of *E. coli* CoT-folding domains exhibit **kinetic delays**
- For about **20%** of cytosolic proteins, this delay can **switch** proteins from co- to posttranslational folding
- **Large ß-strand-rich domains** are more sensitive to kinetic delays than small α-helical domains
- Our systems approach offers a framework to analyze other organisms, re-engineer transcriptomes, and understand risk factors for **metastability**
- Our supersaturation predictions reconcile **many observations in the literature** related to **widespread aggregation**
- Proteins associated with Alzheimer's, Parkinson's, and Huntington's are disproportionately supersaturated, and the proteins these pathways share are particularly supersaturated, suggesting a **core metastable proteome**
- Supersaturated proteins are **preferentially regulated in Alzheimer's**, with **downregulation** of disease-associated supersaturated pathways

1. Nicola AV, Chen W, & Helenius A (1999) Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. *Nat Cell Biol* 1(6):341-345.
2. Netzer WJ & Hartl FU (1997) Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* 388(6640):343-349.
3. Agashe VR, *et al.* (2004) Function of trigger factor and DnaK in multidomain protein folding: Increase in yield at the expense of folding speed. *Cell* 117(2):199-209.
4. Elcock AH (2006) Molecular Simulations of cotranslational protein folding: Fragment stabilities, folding cooperativity, and trapping in the ribosome. *Plos Comput Biol* 2(7):824-841.
5. O'Brien EP, Vendruscolo M, & Dobson CM (2012) Prediction of variable translation rate effects on cotranslational protein folding. *Nature Communications* 3:868.
6. J. Hofrichter, P.D. Ross, and W.A. Eaton. Supersaturation in sickle cell hemoglobin solutions. *Proc Natl Acad Sci USA*, 73(9):3035–3039, 1976.
7. Tartaglia GG, et al. The Zyggregator method for predicting protein aggregation propensities. *Chem Soc Rev* (2008), 37: 1395-1401.

Funding: *Fulbright Scholarship, St John´s College Benefactors Scholarship, and Northwestern Medical Scientist Training Program Grant to* ***PC****, Wellcome Trust Grant MAAG868 RG 61114 to* ***CMD*** *and* ***MV.***