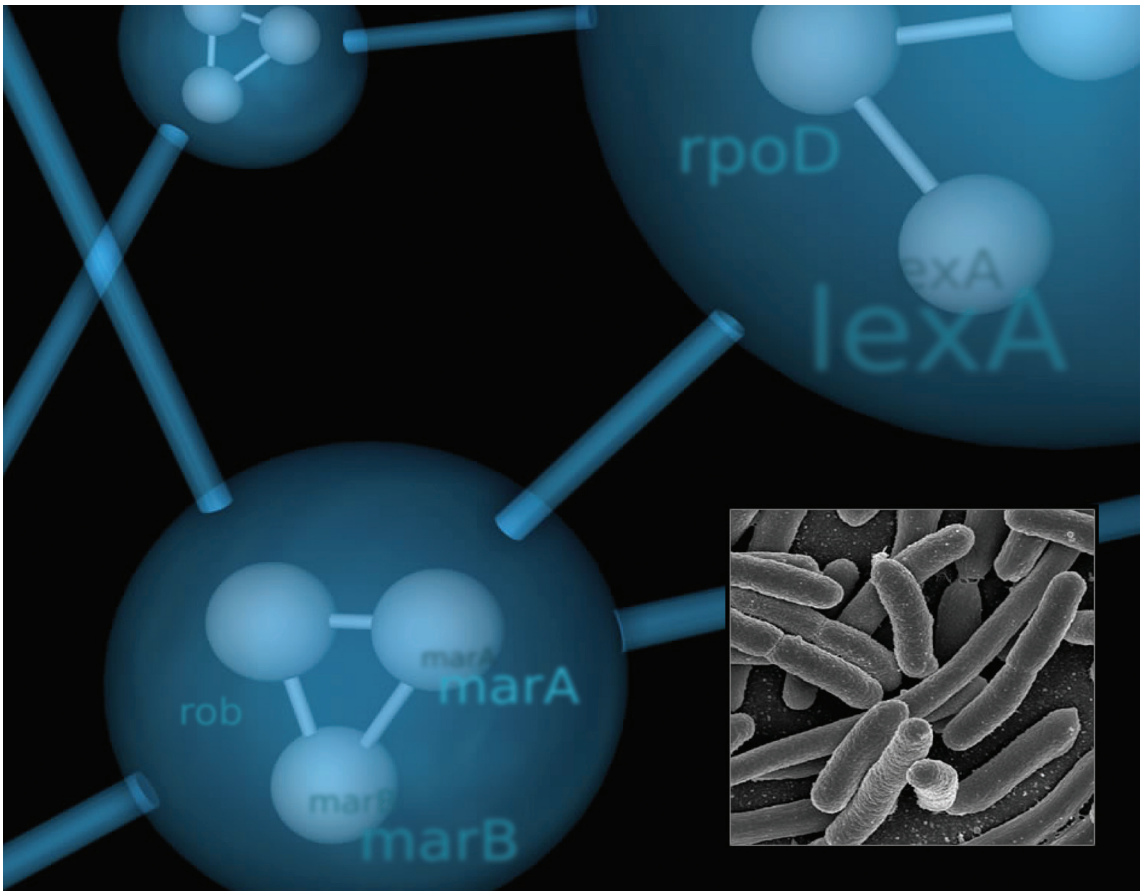
Molecular BioSystems

This article was published as part of the

Computational and Systems Biology themed issue

Please take a look at the full table of contents to access the other papers in this issue.



Correlation between mRNA expression levels and protein aggregation propensities in subcellular localisations†

Gian Gaetano Tartaglia and Michele Vendruscolo

Received 1st July 2009, Accepted 10th August 2009
First published as an Advance Article on the web 3rd September 2009
DOI: 10.1039/b913099n

We investigate the relationship between mRNA expression levels and protein aggregation propensities at the proteomic level, and find that these quantities exhibit a significant correlation when they are averaged across subcellular localisations. In order to investigate this phenomenon, we study the dependence of mRNA expression levels and protein aggregation propensities on the volume of the corresponding subcellular localisations, finding that proteins tend to be increasingly more abundant and more soluble with decreasing volumes of their subcellular localisations. These results indicate that the maintenance of protein solubility plays an even greater role than previously thought in sustaining protein homeostasis.

Introduction

We recently presented initial evidence that the maximum levels of mRNA expression are highly correlated with the aggregation rates of the corresponding proteins.[1] We suggested that this correlation arises from a balance between random mutagenesis processes that increase the tendency of proteins to aggregate[2] and the need to maintain their solubility at the concentrations required for their optimal function in the cell.[1] In this view, protein sequences have co-evolved with the cellular environment and quality control mechanisms to resist aggregation, but only to the minimal levels required for cell viability.[1,3]

Our initial analysis, which was focused on a relatively small set of human proteins for which both expression levels and aggregation rates were measured experimentally under near-physiological conditions, prompted a series of questions about the possible implications of our findings.[3,4] We thus investigate here the relationship between expression levels and aggregation propensities at a proteomic level. Our results indicate that mRNA expression levels and protein aggregation propensities are highly correlated if proteins are grouped according to their subcellular localisations. In order to understand these results, we begin to study how the properties of the individual subcellular localisations are related to the abundance and solubility of the proteins that are present in them. We found a significant correlation of the volumes of the subcellular localisations with the aggregation propensities of the corresponding proteins, as well as with the respective mRNA expression levels. Taken together, these results indicate that proteins tend to be simultaneously highly abundant and highly soluble in small subcellular localisations.

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge, UK CB2 1EW
† This article is part of a *Molecular BioSystems* themed issue on Computational and Systems Biology.

Correlation between mRNA expression levels and protein aggregation propensities in different subcellular localisations

It is well known that different cell compartments exhibit different mRNA expression levels[5] and different aggregation propensities.[6,7] By bringing these observations together with our previous analysis,[1] we observe here a strong anticorrelation between mRNA expression levels and the aggregation propensities of the corresponding proteins, when the latter are grouped according to their subcellular localisations (Fig. 1). Since, despite the presence of significant fluctuations induced by variable turnover levels, mRNA expression levels tend to be on average proportional to protein abundances,[8] these results also suggest that average protein abundances and average protein aggregation propensities should also be correlated at the subcellular localisation level.

Consistently with previous results,[6,7] we found ribosomal proteins to be highly aggregation-resistant, while cytosolic and endoplasmic reticulum proteins show a higher tendency to form aggregates. These results can be explained, at least in part, by considering that ribosomal proteins are highly basic in order to bind to negatively charged ribosomal RNA, which involves unfavorable contributions to their aggregation propensities.[9,10] Although in this study we excluded membrane proteins of the rough and smooth endoplasmic reticulum and of the Golgi apparatus, we observed that proteins associated with the endoplasmic reticulum have an elevated hydrophobic content and exhibit high aggregation propensities.[9,10] We also observed that by excluding nuclear proteins, the correlation in Fig. 1 improves from −93% to −97%, which indicates that these proteins tend to have lower aggregation propensities and lower expression levels than expected. These findings can be interpreted by considering that proteins involved in transcription constitute about 42% of the nuclear proteins in our database and that they are on average as abundant as the remaining proteins in the nucleus (2.41 on the scale reported in Fig. 1), but are characterized by a lower aggregation

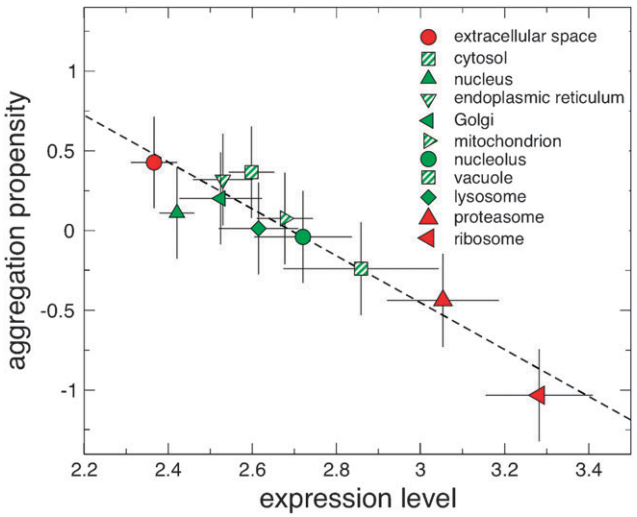


Fig. 1 The relationship between expression levels and aggregation propensities in different subcellular localisations in a human cell. Aggregation propensities of proteins present in 11 different subcellular localisations are plotted against the expression levels of the corresponding mRNAs; the coefficient of correlation is −93%. In each case, the average and the variance are reported for both mRNA expression levels and protein aggregation propensities; aggregation propensities and mRNA expression levels are averaged within each subcellular localisation and over the 79 tissues that we analysed. Aggregation propensities were calculated using the Zyggregator algorithm;[3,8] a pH of 4.5 was used for these calculations for proteins in lysosomes and vacuoles. Gene Ontology (GO) annotations were used for identifying protein localisations.[22]

propensity (−0.26 on the scale reported in Fig. 1), thus suggesting that solubility in transcription is particularly important to maintain cell viability.[1,3,11] By contrast, we found that transport and signal transduction proteins, which are especially abundant in the cytosol and in the extracellular space, have a significantly high aggregation propensity, in agreement with previous results.[6] We also analysed the proteins in the *E. coli* cytosol and found that they are more aggregation prone than the proteins in the human cytosol (0.88 and 0.38, respectively, on the scale reported in Fig. 1), which suggests that higher aggregation propensities are expected in organisms with a simpler cellular organisation.[12]

Correlation between mRNA expression levels and protein aggregation propensities in different tissues

We found that although mRNA expression levels vary within and among tissues, their values tend to be correlated (Fig. 2). In our analysis we used 79 different human tissues; data were taken from the gene atlas of the human transcriptome,[13] and scaled using median scaling and quantile normalization.[14] We also observed strong correlations between the mRNA expression levels and predicted aggregation propensities of the corresponding proteins in specific tissues when these properties are averaged over different subcellular localisations (Fig. 3a). When the mRNA expression levels and the aggregation propensities of the corresponding proteins are not averaged in this way, but across tissues, the overall relationship between them is much less significant (Fig. 3b).

Proteins in small subcellular localisations are highly concentrated and highly soluble

The correlation that we reported in Fig. 1 is likely to depend on the specific environmental characteristics of the different subcellular localisations. We started investigating these effects by considering the relationship between the volumes of the subcellular localisations, the levels of expression and the aggregation propensities. We found that the mRNA expression levels corresponding to proteins resident in small subcellular localisations are high (Fig. 4a) and the respective aggregation propensities are low (Fig. 4b). These results suggest that the maintenance of solubility is a particularly important requirement

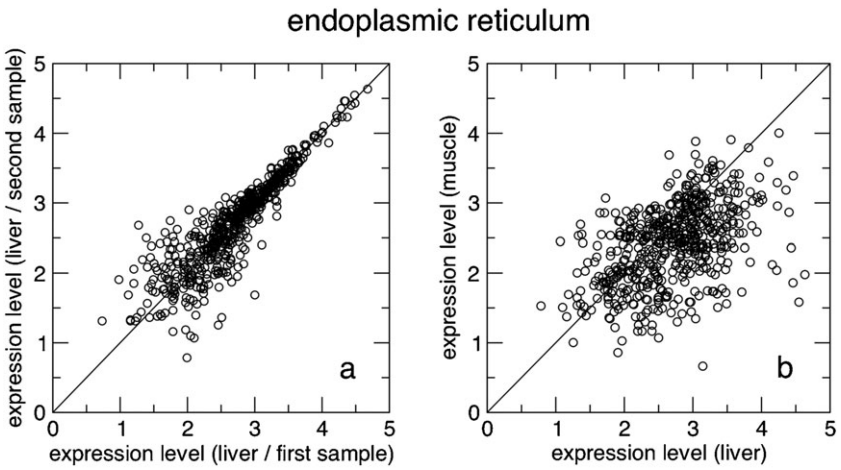


Fig. 2 Examples of correlations between mRNA expression levels in different human tissues. We observe correlations between measurements of mRNA expression levels from the same tissue (a) and, although at a weaker level, between samples of mRNA expression levels taken from different tissues (b). The coefficient of correlation between liver samples is 90% (a), while the coefficient of correlation between liver and muscle samples is 50% (b).

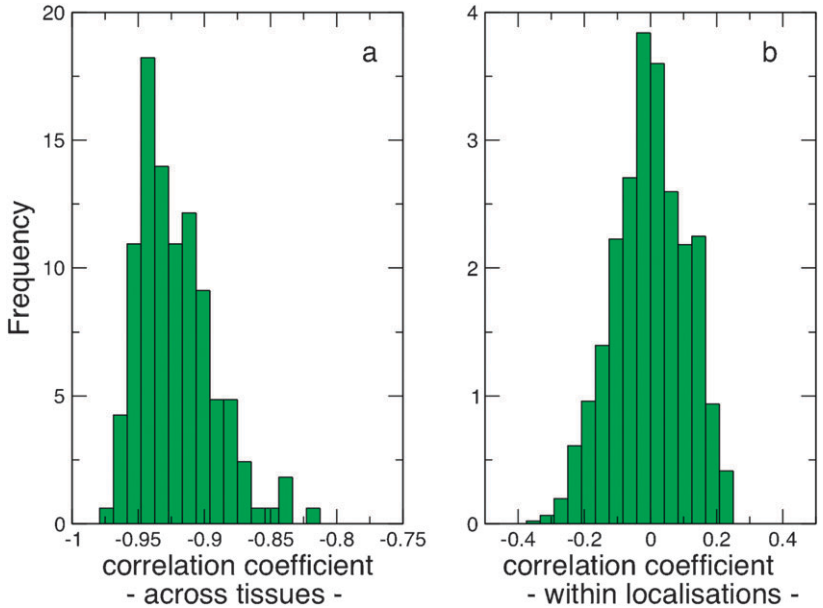


Fig. 3 Correlations between protein aggregation propensities and mRNA expression levels for tissues and subcellular localisations. (a) The correlations between aggregation propensities and mRNA expression levels are calculated for 79 different tissues averaging over 11 different subcellular localisations; significant anticorrelations are found in this case. The lowest coefficients of correlation were found for liver, amygdala and cervical ganglion (−83%), while the lymph node, skin, and uterus showed the highest anticorrelations (97%). (b) We did not find significant correlations between aggregation propensities and mRNA expression levels calculated for the different subcellular localisations when averaging over the different tissues; the highest coefficient of correlation was observed for the Golgi apparatus (26%), while the lowest was observed for the nucleolus (−34%).

to enable biochemical reactions in the cell to take place by avoiding dysfunctional associations when confining the participating molecules in small volumes. We also note that effects caused by molecular crowding can also further influence the properties of proteins confined in small subcellular localisations.[15,16]

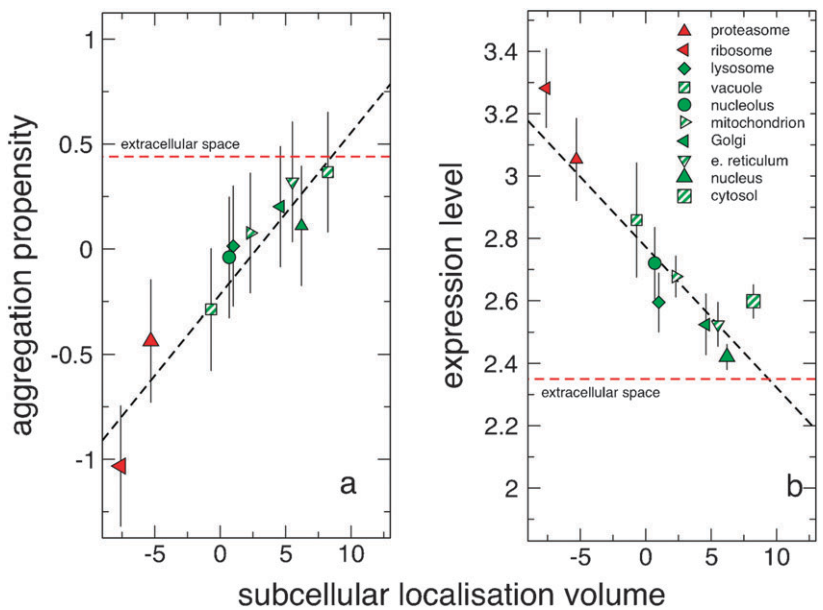


Fig. 4 Relationships between mRNA expression levels, protein aggregation propensities and volumes of the corresponding subcellular localisations. The coefficient of correlation between protein aggregation propensities and volumes is 88% (a) and between mRNA expression levels and volumes is −87% (b); volumes are calculated in μm^3 (Table 1) and reported on a logarithmic scale, and protein aggregation propensities and mRNA expression levels in the extracellular space are not associated with specific volumes (dashed red line). These results indicate that proteins primarily resident in larger subcellular localisations, such as the nucleus or the endoplasmic reticulum, tend to have high aggregation propensities and low concentrations. On the contrary, proteins confined in small subcellular localisations, such as ribosomes, vacuoles or lysosomes, tend to have low aggregation propensities and high concentrations.

Table 1 Volumes of the subcellular localisations and numbers of genes used in this study. Volumes were calculated using data available from http://www.rkm.com.au/CELL/animalcell.html and http://porpax.bio.miami.edu/~cmallery/150/cells/organelle.htm#two, and through a literature search.[17–23] Genes were taken from the atlas of the human transcriptome[13] and their corresponding protein sequences were downloaded from http://www.ensembl.org/

Subcellular Localisation	Expression Levels (ln scale)	Aggregation Propensity	Volume (μm^3)	Number of Genes
Extracellular space	2.35	0.44	—	1515
Cytosol	2.59	0.38	3750	1319
Nucleus	2.41	0.10	500	3178
E. reticulum	2.53	0.33	250	650
Golgi apparatus	2.52	0.23	100	153
Mitochondrion	2.67	0.11	10	729
Nucleolus	2.72	0.01	2	114
Lysosome	2.59	0.05	2	159
Vacuole	2.85	−0.23	0.5	12
Proteasome	3.05	−0.38	0.005	25
Ribosome	2.28	−0.95	0.0005	168

Conclusions

We have reported a relationship between mRNA expression levels and the aggregation propensities of the corresponding proteins when their values are averaged across subcellular localisations, thus supporting the suggestion that the abundance and the solubility of proteins are correlated.[1,3] We have also shown that proteins tend to be concentrated and soluble to a level inversely proportional to the volume of their subcellular localisations. These results suggest that the organisation of a cell into compartments makes biochemical processes more efficient by concentrating the molecules that carry them out, but only by simultaneously ensuring that their solubility is kept at levels at which aggregation is avoided under normal circumstances. In terms of their solubility, therefore, proteins are maintained at the edge of aggregation.

Methods

We investigated the correlation between the mRNA expression levels and the predicted aggregation propensities of the corresponding proteins in a dataset that comprise proteins in different subcellular localisations.

The aggregation propensity of a protein i is calculated as

$$\xi_i = \frac{z_i}{L_i} + p\frac{z_i^p}{P_i} + n\frac{z_i^n}{N_i} \qquad (1)$$

where z_i represents the overall aggregation score,[9,10] z_i^p is the score for regions characterized by $z_i \geq 0$ and z_i^n is the score for regions characterized by $z_i < 0$. The variable L_i is the length of the polypeptide chain, P_i is the number of regions with positive aggregation scores, and N_i is the number of regions with negative aggregation scores. The parameters p and n were chosen so that the expected value of ξ_i is 0 and its variance is 1 over the entire database used in this study.

The expression levels are estimates taken from measurements of the cellular mRNA concentrations.[13] To generate a homogeneous set of variables, the expression levels of each of the 158 human samples (79 different tissues) in the database used here[13] were normalized by median scaling and between the samples by quantile normalization.[14]

References

1 G. G. Tartaglia, S. Pechmann, C. M. Dobson and M. Vendruscolo, Life on the edge: A link between gene expression levels and aggregation rates of human proteins, *Trends Biochem. Sci.,* 2007 **32**, 204–206.
2 M. A. DePristo, D. M. Weinreich and D. L. Hartl, Missense meanderings in sequence space: A biophysical view of protein evolution, *Nat. Rev. Genet*., 2005, **6**, 678–687.
3 M. Vendruscolo and G. G. Tartaglia, Towards quantitative predictions in cell biology using chemical properties of proteins, *Mol. BioSyst*., 2008, **4**, 1170–1175.
4 G. G. Tartaglia, S. Pechmann, C. M. Dobson and M. Vendruscolo, A relationship between mRNA expression levels and protein solubility in *E. Coli*, *J. Mol. Biol*., 2009, **388**, 381–389.
5 A. Drawid, R. Jansen and M. Gerstein, Genome-wide analysis relating expression level with protein subcellular localization, *Trends Genet*., 2000, **16**, 426–430.
6 G. G. Tartaglia and A. Caflisch, Computational analysis of the S. Cerevisiae proteome reveals the function and cellular localization of the least and most amyloidogenic proteins, *Proteins: Struct., Funct., Bioinf*., 2007, **68**, 273–278.
7 E. Monsellier, M. Ramazzotti, N. Taddei and F. Chiti, Aggregation propensity of the human proteome, *PLoS Comput. Biol*., 2008, **4**, e1000199.
8 P. Lu, C. Vogel, R. Wang, X. Yao and E. M. Marcotte, Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation, *Nat. Biotechnol*., 2007, **25**, 117–124.
9 G. G. Tartaglia, A. P. Pawar, S. Campioni, C. M. Dobson, F. Chiti and M. Vendruscolo, Prediction of aggregation-prone regions in structured proteins, *J. Mol. Biol*., 2008, **380**, 425–436.
10 G. G. Tartaglia and M. Vendruscolo, The Zyggregator method for predicting protein aggregation propensities, *Chem. Soc. Rev*., 2008, **37**, 1395–1401.
11 A. von Mikecz, Xenobiotic-induced autoimmunity and protein aggregation diseases share a common subnuclear pathology, *Autoimmun. Rev*., 2005, **4**, 214–218.
12 G. G. Tartaglia, R. Pellarin, A. Cavalli and A. Caflisch, Organism complexity anti-correlates with proteomic beta-aggregation propensity, *Protein Sci*., 2005, **14**, 2735–2740.
13 A. I. Su, T. Wiltshire, S. Batalov, H. Lapp, K. A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, M. P. Cooke, J. R. Walker and J. B. Hogenesch, A gene atlas of the mouse and human protein-encoding transcriptomes, *Proc. Natl. Acad. Sci. U. S. A*., 2004, **101**, 6062–6067.
14 B. M. Bolstad, R. A. Irizarry, M. Astrand and T. P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias, *Bioinformatics*, 2003, **19**, 185–193.
15 A. P. Minton, The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media, *J. Biol. Chem*., 2001, **276**, 10577–10580.
16 R. J. Ellis, Macromolecular crowding: Obvious but underappreciated, *Trends Biochem. Sci*., 2001, **26**, 597–604.
17 B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, *Molecular Biology of the Cell*, Garland Science Publishing, New York, 2002.
18 J. M. Sturgess and F. A. Delaigle, Morphometry of Golgi apparatus in developing liver, *J. Cell Biol*., 1972, **55**, 524–530.
19 I. E. Scheffler, *Mitochondria*, Wiley & Sons, New York, 1999.
20 P. Hozak, P. R. Cook, C. Schofer, W. Mosgoller and F. Wachtler, Site of transcription of ribosomal-rna and intranucleolar structure inhela-cells, *J. Cell Sci*., 1994, **107**, 639–648.
21 W. Kuhnel, *Color Atlas of Cytology, Histology and Microscopic Anatomy*, Georg Thieme Verlag, New York, 2003.
22 A. Tinazli, J. Piehler, M. Beuttler, R. Guckenberger and R. Tampe, Native protein nanolithography that can write, read and erase, *Nat. Nanotechnol*., 2007, **2**, 220–225.
23 A. T. Dinh, C. Pangarkar and S. Mitragotri, Understand discrete nanoscale transport, *Chem. Eng. Prog*., 2008, **104**, 62–68.