Update
TRENDS in Biochemical Sciences Vol.xxx No.x
Full text provided by www.sciencedirect.com
ScienceDirect
Letters

Life on the edge: a link between gene expression levels and aggregation rates of human proteins

Gian Gaetano Tartaglia, Sebastian Pechmann, Christopher M. Dobson and Michele Vendruscolo

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

We have found that expression levels of human genes *in vivo* are remarkably anti-correlated with the aggregation rates of the corresponding proteins measured *in vitro* by experiment. This result indicates that human proteins have evolved to resist aggregation and to function efficiently, but with almost no margin of safety to respond to genetic and environmental factors that decrease their solubility or increase their concentration *in vivo*. We speculate that this result provides a compelling reason for the existence of disorders that are associated with protein aggregation, such as Alzheimer's and Parkinson's diseases [1].

Evidence for this conclusion is presented in Figure 1, where the experimental *in vitro* aggregation rates of a set of peptides and proteins are plotted against the *in vivo* expression levels of their respective genes, as assessed using DNA microarray technology [2]. The plot includes all the experimental data we could extract from the literature (Table 1), and the correlation coefficient between expression levels and aggregation rates is an astonishing 0.97.

The existence of such a strong degree of anti-correlation for this set of peptides and proteins indicates that the aggregation propensities of the proteins needed by the cell are precisely tuned by mutation and evolutionary selection to levels that enable them to be functional at the concentrations required for optimally efficient performance. It also indicates that protein molecules have co-evolved with their cellular environments to be sufficiently soluble for their biological roles, but not more so. Hence, aggregation can result from even minor changes in the chemistry (e.g. as a result of oxidative stress) and in the regulation (e.g. as a result of changes associated with ageing) of otherwise harmless proteins. Indeed, when proteins are expressed at higher levels than those found naturally, it is unlikely that their aggregation can be avoided, except for relatively short lengths of time.

A particularly interesting test of the evolutionary link between expression levels and aggregation propensities is the existence of amyloid structures that are functional rather than being disease-associated. In humans, the best characterized example of such a 'functional amyloid' (see

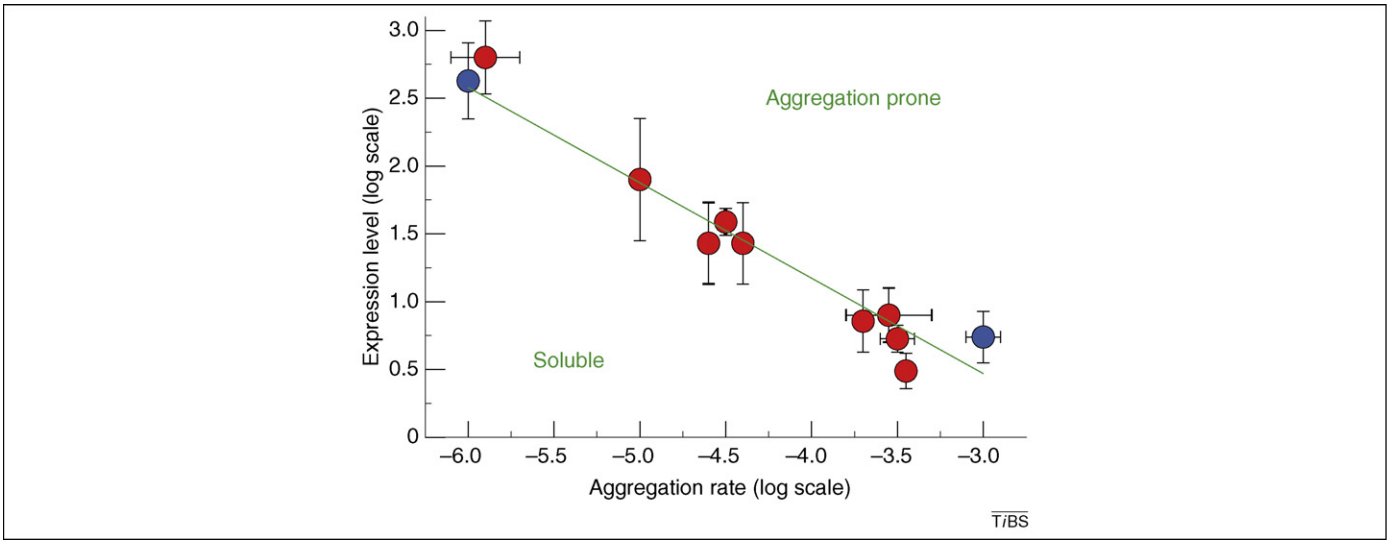

Figure 1. Correlation between expression levels and the measured aggregation rates of the corresponding proteins. The expression levels are estimates taken from measurements of the cellular mRNA concentrations [2]. All the data obtained from a comprehensive search in the databases of human expression levels and of the amyloid aggregation literature are included provided that the aggregation rates are measured at pH values between 4.0 and 8.0 (see Table 1). The standard deviations of the rates are reported only in four cases [7] because these values are not available or difficult to extract from the published data. The two proteins not involved in any known disease (acylphosphatase and glucagon) are represented by blue circles [1]. 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 [2] were normalized by median scaling and between the samples by quantile normalization [17]. The expression levels were averaged over all the tissues in which a particular gene is expressed.

Corresponding author: Vendruscolo, M. (mv245@cam.ac.uk).
Available online xxxxxx.

www.sciencedirect.com 0968-0004/$ – see front matter doi:10.1016/j.tibs.2007.03.005

Please cite this article in press as: Tartaglia, G.G. et al., Life on the edge: a link between gene expression levels and aggregation rates of human proteins, Trends Biochem. Sci. (2007), doi:10.1016/j.tibs.2007.03.005

Table 1. Conditions used to monitor the aggregation rates for the peptide and proteins considered in this work

Name	EL[a]	Rate	pH	IS	Conc.	Refs
Acylphosphatase	0.7	−3.0	5.5	43.0	0.04	[7]
Prion106–126	0.5	−3.5	5.0	1.2	0.3	[8]
Calcitonin	0.9	−3.5	7.4	25.0	1.5	[9]
Tau protein	0.8	−3.6	7.6	50.0	0.004	[10]
Acetylcholinesterase$_{586–599}$	1.4	−3.7	7.0	7.7	0.2	[11]
Aβ_{40} peptide	1.6	−4.6	7.4	81.0	0.03	[12]
Aβ_{42} peptide	1.4	−4.3	7.4	81.0	0.01	[12]
Amylin	1.8	−4.5	5.0	0.06	0.001	[13]
Transthyretin	2.1	−5.2	4.4	130.0	0.01	[14]
α-synuclein	2.6	−5.9	7.4	2.0	0.1	[15]
Glucagon	2.8	−6.0	7.0	10.0	0.8	[16]
Pmel17	2.7	−0.1	7.4	100.0	0.01	[4]

[a]EL: normalized expression levels [2,17]. Rate: logarithm in base 10 of the aggregation rates measured in seconds. The ionic strength (IS) and the protein concentration (Conc.) are in mM units. See Ref. [1] for further information about the proteins and their links (or absence thereof) with disease.

Fowler *et al.* [3] in this issue) is Pmel17, the presence of which in melanosomes has been linked to the production and storage of melanin [4]. The aggregation rate of Pmel17 (see Table 1) is much higher than that predicted on the basis of the plot in Figure 1, implying that its assembly into its functional amyloid state *in vivo* is highly favourable.

If the link that we have identified for the group of peptides and proteins for which data are available proves true for proteins in general, an important question concerns its origin. Our own view is that this intimate relationship is the net result of the opposing effects on protein sequence of an evolutionary pressure to remain soluble at the concentrations needed by the cell and of random mutational processes that tend to increase their aggregation propensity.

The existence of such a relationship provides dramatic evidence for the view that the intrinsic propensity of proteins to revert to the amyloid state, rather than the presence of pathogenic sequence motifs that encode this type of structure, is the cause of amyloid diseases [5]. In particular, it rationalizes observations that even small deviations either in the expression levels or in the aggregation rates of proteins can have dramatic pathological consequences [6].

In summary, the correlation described here indicates that proteins have co-evolved with the quality-control mechanisms present in the cellular environment to have a low enough aggregation propensity to enable an organism to function optimally. There is, however, almost no scope for dealing with any situation in which these levels rise further or whereby the aggregation rates are increased, for example, by chemical modifications, genetic mutations or impaired regulatory processes. In the context of protein solubility, therefore, we are constantly living our lives at the edge of a molecular precipice.

References

1 Chiti, F. and Dobson, C.M. (2006) Protein misfolding, functional amyloid and human disease. *Annu. Rev. Biochem.* 75, 333–366
2 Su, A.I. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci. U. S. A.* 101, 6062–6067
3 Fowler, D.M. *et al.* Functional amyloid – from bacteria to humans. *Trends Biochem. Sci.* doi:10.1016/j.tibs.2007.03.003.
4 Fowler, D.M. *et al.* (2006) Functional amyloid formation within mammalian tissue. *PLoS Biol.* 4, 100–107
5 Dobson, C.M. (1999) Protein misfolding, evolution and disease. *Trends Biochem. Sci.* 24, 329–332
6 Lansbury, P.T. and Lashuel, H.A. (2006) A century old debate on protein aggregation and neurodegeneration enters the clinic. *Nature* 443, 774–779
7 Chiti, F. *et al.* (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. *Nature* 424, 805–808
8 Salmona, M. *et al.* (1999) Molecular determinants of the physicochemical properties of a critical prion protein region comprising residues 106–126. *Biochem. J.* 342, 207–214
9 Fowler, S.B. *et al.* (2005) Rational design of aggregation-resistant bioactive peptides: reengineering human calcitonin. *Proc. Natl. Acad. Sci. U. S. A.* 102, 10105–10110
10 Gamblin, T.C. *et al.* (2003) Tau polymerization: role of the amino terminus. *Biochemistry* 42, 2252–2257
11 Cottingham, M.G. *et al.* (2002) Amyloid fibril formation by a synthetic peptide from a region of human acetylcholinesterase that is homologous to the Alzheimer's amyloid-β peptide. *Biochemistry* 41, 13539–13547
12 Fezoui, Y. and Teplow, D.B. (2002) Kinetic studies of amyloid β-protein fibril assembly – differential effects of α-helix stabilization. *J. Biol. Chem.* 277, 36948–36954
13 Goldsbury, C.S. *et al.* (2000) Studies on the *in vitro* assembly of Aβ1–40: implications for the search for Aβ fibril formation inhibitors. *J. Struct. Biol.* 130, 217–231
14 Hammarstrom, P. *et al.* (2002) Sequence-dependent denaturation energetics: a major determinant in amyloid disease diversity. *Proc. Natl. Acad. Sci. U. S. A.* 99, 16427–16432
15 Conway, K.A. *et al.* (2000) Acceleration of oligomerization, not fibrillization, is a shared property of both α-synuclein mutations linked to early-onset Parkinson's disease: implications for pathogenesis and therapy. *Proc. Natl. Acad. Sci. U. S. A.* 97, 571–576
16 Beaven, G.H. *et al.* (1969) Formation and structure of gels and fibrils from glucagons. *Eur. J. Biochem.* 11, 37–42
17 Bolstad, B.M. *et al.* (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185–193

Please cite this article in press as: Tartaglia, G.G. et al., Life on the edge: a link between gene expression levels and aggregation rates of human proteins, Trends Biochem. Sci. (2007), doi:10.1016/j.tibs.2007.03.005