Derivation of a solubility condition for proteins from an analysis of the competition between folding and aggregation

Sebastian Pechmann†* and Michele Vendruscolo*

Received 19th April 2010, Accepted 13th August 2010
DOI: 10.1039/c005160h

Failure in maintaining protein solubility *in vivo* impairs protein homeostasis and results in protein misfolding and aggregation, which are often associated with severe neurodegenerative and systemic disorders that include Alzheimer's and Parkinson's diseases and type II diabetes. In this work we formulate a model of the competition between folding and aggregation, and derive a condition on the solubility of proteins in terms of the stability of their folded states, their aggregation propensities and their degradation rates. From our model, the bistability between folding and aggregation emerges as an intrinsic aspect of protein homeostasis. The analysis of the conditions that determine such a bistability provides a rationalization of the recently observed relationship between the cellular abundance and the aggregation propensity of proteins. We then discuss how the solubility condition that we derive can help rationalise the correlation that has been reported between evolutionary rates and expression levels or proteins, as well as *in vivo* protein solubility and expression level measurements, and recently elucidated trends of proteome evolution.

Introduction

The myriad biochemical reactions that take place in a cell rely on its ability to maintain proteins in a soluble state through a variety of different processes, which are often collectively referred to as "protein homeostasis".[1–3] Failure in protein homeostasis is linked to several severe protein misfolding conditions that include Alzheimer's and Parkinson's diseases.[4,5]

Among the mechanisms that prevent protein aggregation, a particularly important role is played by the stability of the native fold, which helps protect aggregation-prone regions of the amino acid sequences from becoming available for forming dysregulated inter-molecular interactions.[6–9] In this sense, protein folding and aggregation are in close competition,[5,7,10] and tight quality control mechanisms are present in the cell to guarantee that proteins fold correctly. The presence of abnormal protein aggregates is usually the result of failed clearance,[11] and by the failure of other control mechanisms, such as the heat shock response[12] or the unfolded protein response.[13] Since these safety mechanisms require a significant amount of energy, however, one might expect the amino acid sequences of proteins to have evolved to minimize their dependence on them, at least under normal conditions. Indeed, cellular protein levels appear to have been tuned by evolution for optimal functioning;[14,15] in turn such levels provide also a good predictor of the rate of protein sequence evolution,[16–18] which is itself particularly constrained by the need to avoid protein aggregation and maintain solubility.[19]

In this work we formulate a model of the competition between protein folding and aggregation in the cell, and derive a relationship between protein solubility, aggregation and degradation. According to this solubility condition, the amount of soluble proteins is proportional to the stability and degradation capacity of proteins, and inversely proportional to their propensity to aggregate. These results complement those obtained by Rieger *et al.*[20] by showing that the presence of bistability in protein homeostasis is a general feature of the competition between protein folding and aggregation, which is not limited to chaperone-assisted folding. As a major result of our analysis we show that the solubility condition provides a rationalization of the recently observed relationship between the cellular abundance and the aggregation propensity of proteins.

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK. E-mail: mv245@cam.ac.uk
† Present address: Department of Biology, Stanford University. E-mail: pechmann@stanford.edu

Results

Bistability between the folded and aggregated states

In this work we study the competition between protein folding and aggregation within a model that describes the inter-conversion of proteins between the folded, unfolded, and aggregated states (Fig. 1). We denote by F the number of protein molecules in the folded state, by A the number in the aggregated state, and by U the number in all the states that can give rise to aggregation, which include folding intermediates, as well as partially unstructured, unstructured and misfolded states.[20] In this sense, the near-native conformations that can sometimes give rise to aggregation[9] are considered as a part of the unfolded state, rather than the folded state.

Our analysis is based on a master equation approach, which has been used to describe various types of biological processes, including biochemical,[21] and metabolic[22] networks, protein aggregation,[23] and protein homeostasis.[2,20] The master equation that we consider here describes the fluxes between F, U and A, including the clearance of aggregated proteins,

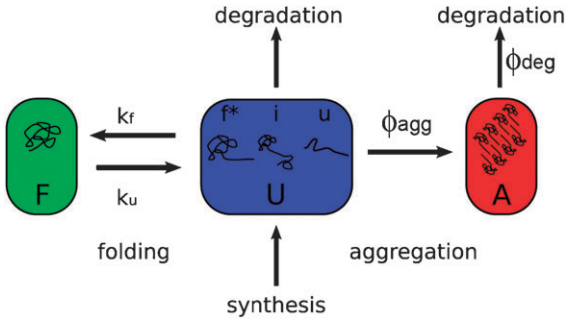

Fig. 1 Model of the competition between folding and aggregation. F, U and A denote, respectively, the number of proteins in the folded, unfolded and aggregated states. The unfolded state U, as defined here, comprises all aggregation-prone species, which include locally unfolded proteins (f*), folding or misfolding intermediates (i), and unfolded proteins (u). Synthesis and degradation of proteins are assumed to be at steady state, and aggregated proteins are degraded at rate ϕ_d. The disaggregation reaction, which is assumed to be slow, is absorbed into an effective aggregation rate ϕ_a.

which give a description of the competition between protein folding and aggregation in the cell:

$$\begin{aligned} \frac{dF}{dt} &= \phi_f - \phi_u = k_f U - k_u F \\ \frac{dA}{dt} &= \phi_a - \phi_d = \Lambda A U - -\phi_d \\ P_{tot} &= F + U + A \end{aligned} \quad (1)$$

The third equation expresses the requirement that the total number of proteins P_{tot} in the cell must be the sum of all the proteins in the folded, unfolded and aggregated states. We assume this protein abundance to be at the steady state level under non-stress conditions; this level is considered to be stable otherwise it would be immediately subject to selection.[24] Thus, the production and degradation of new proteins in the unfolded state can be considered constant and is omitted from the model. The first equation states that the net flux to the folded state is given by the difference of the fluxes between the folded and unfolded states, with constant folding and unfolding rates k_f and k_u. The second equation specifies that A depends on the fluxes of aggregation, ϕ_a, and degradation, ϕ_d. In the following we make specific assumptions about the form of ϕ_a and ϕ_d. Since the aggregation kinetics can be approximated as exhibiting exponential growth over time,[23] we assume that the aggregation rate increases linearly with the number of already aggregated proteins, so that the aggregation flux can be expressed as $\phi_a = \Lambda A U$; where we introduced the aggregation propensity factor Λ, and we absorbed the rate of disaggregation into the aggregation rate. In order to define the degradation flux ϕ_d, we considered that protein degradation refers to a series of complex pathways that eliminate from the cell unfolded, misfolded and aggregated proteins with high specificity.[25] We thus considered three cases: unlimited first order degradation with the rate constant d (I), first order degradation up to a limited capacity Q, then zeroth order degradation (II), and impaired degradation beyond the capacity (III), where $d_0 < d$

$$\begin{aligned} \phi_d^{I} &= dA \\ \phi_d^{II} &= \begin{cases} dA & \text{if } A < Q \\ dQ & \text{if } A \geq Q \end{cases} \\ \phi_d^{III} &= \begin{cases} dA & \text{if } A < Q \\ d_0 Q & \text{if } A \geq Q \end{cases} \end{aligned} \quad (2)$$

From the master equation (eqn (1)), we derive the steady state conditions by setting the 'nullcline conditions' $dF/dt = dA/dt = 0$. Examples of steady state solution for different degradation fluxes (eqn (2)) are illustrated in Fig. 2. For unlimited first order clearance of aggregated proteins, ϕ_d^I, only one stable steady state is possible (I) (Fig. 2B). Since, however, the cell has only limited capacity to remove aberrant aggregates,[2,26,27] we consider in addition two other models of protein degradation that exhibit first order kinetics within the capacity of the degradation machinery,[28] and zeroth order kinetics if the number of proteins to be degraded exceeds the capacity[29] (II). High concentrations of aggregated proteins can furthermore directly disrupt and impair the homeostasis and degradation system[30,31] (III). For a limited degradation capacity, we found that the system can be bistable (Fig. 2D and F).

Bistability and homeostasis

In our description, proteins are bistable as they can be found either in a soluble functional state or in an insoluble aggregated one, which are separated by an unstable state (Fig. 3). This description implies the presence of a threshold behaviour, which has been suggested to increase robustness[32,33] and has been used to explain the occurrence of aggregation in chaperone-assisted folding both *in vitro* and *in vivo*.[20] Threshold phenomena are crucial for evolutionary robustness by favoring strong purifying selection, which is necessary for preventing protein aggregation, and have also been consistently observed in protein evolution.[34,35] In Fig. 3, filled circles denote stable and open circles unstable steady states at the intersections of the $dF/dt = dA/dt = 0$ lines for protein folding (straight solid line) and aggregation (other solid line). The three different possible scenarios for biologically relevant parameters are: (A) one stable steady state at high concentration of folded and low concentration of aggregated proteins, (B) two stable steady states, and (C) one stable steady state at low concentration of folded and high concentration of aggregated proteins. Selective evolutionary pressures acting on protein stability and aggregation propensity and against the random evolutionary drift towards less soluble proteins are necessary to make sure only scenario (A) is realized and aggregation avoided.

Derivation of the solubility condition

By starting from the basic observation that in order to avoid uncontrolled aggregation proteins should remain soluble, we derive a condition on the cellular abundance of proteins by requiring that the master equation (eqn (1)) should have only

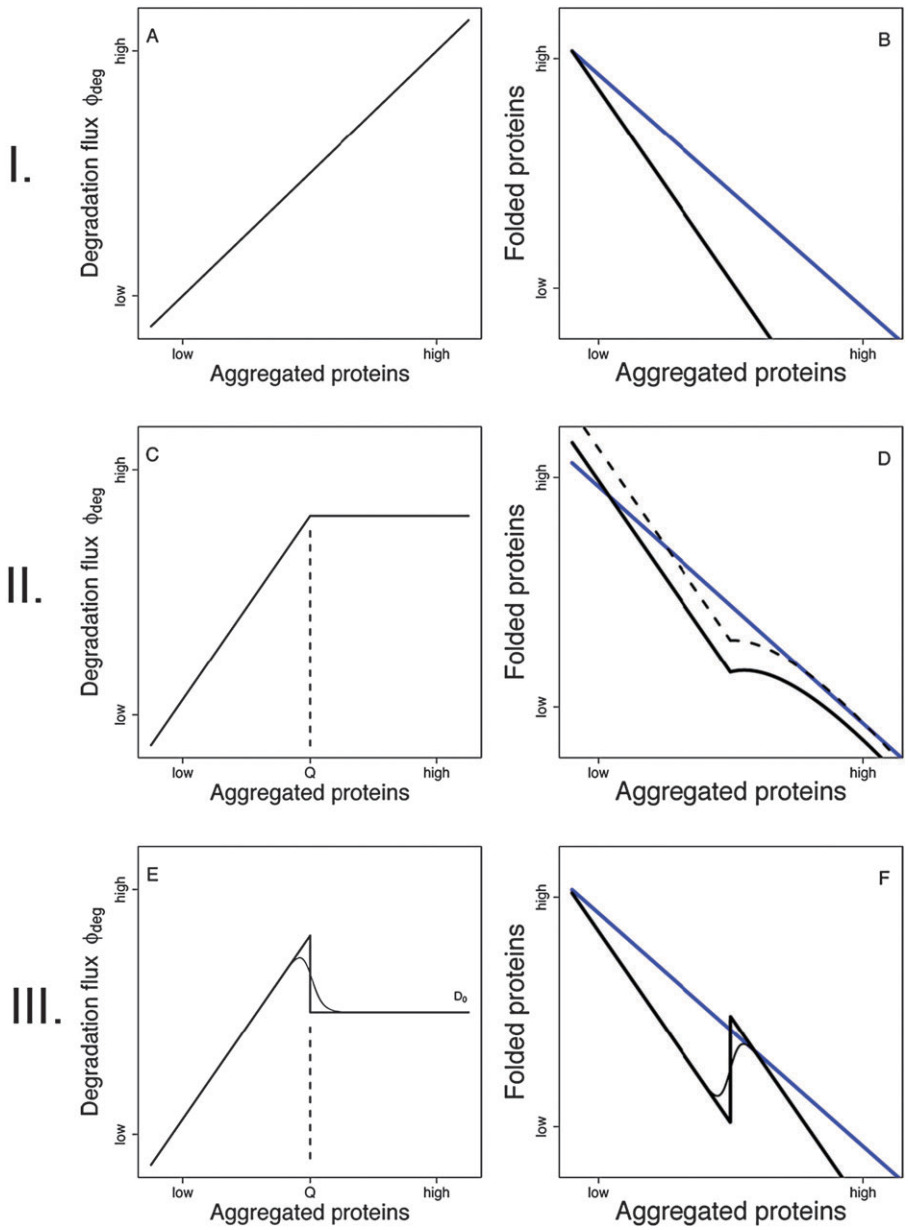

Fig. 2 Examples of steady state solutions (B, D, F) for, respectively, the three types of degradation fluxes that we considered, ϕ_d^I (A), ϕ_d^{II} (C), and ϕ_d^{III} (E), for folding (straight solid line) and aggregation (other solid line). A limited degradation capacity (II, III) can give rise to bistability; the absence (solid line) or presence (dashed line) of bistability depends on the system parameters (D). The use of Hill functions allows a smoother transition to be modeled across the degradation capacity (F, thin line).

the solution in which there is a high concentration of folded and a low concentration of aggregated proteins (see Methods)

$$P < \frac{Sd}{\Lambda} \qquad (3)$$

where $P = P_{tot} - Q$ is the cellular abundance of proteins that is not immediately covered by the degradation capacity Q. According to this relationship, P is proportional to protein stability, $S = 1 - k_f/k_u$ (where k_f and k_u are the folding and unfolding rates), and degradation rate, d, and inversely proportional to the propensity to aggregate Λ. We found that this relationship, which in this work we call 'the solubility condition', is independent of the choice of the specific degradation function (see Methods). The solubility condition links the cellular abundance of proteins P to their stability S, degradation rate d, and aggregation propensity Λ (see Methods), as it states that the number P of proteins that are not directly covered by the degradation machinery must be smaller than the fraction d of the ratio of stability and aggregation propensity. If the solubility condition is fulfilled, then no uncontrolled aggregation takes place and protein solubility and homeostasis are maintained.

Bifurcation diagrams for ϕ_a^{II} and ϕ_a^{II} illustrate the steady states as a function of the aggregation propensity (Fig. 4). If degradation is impaired beyond capacity (Fig. 4B), the folded and the aggregated states are more separated than for

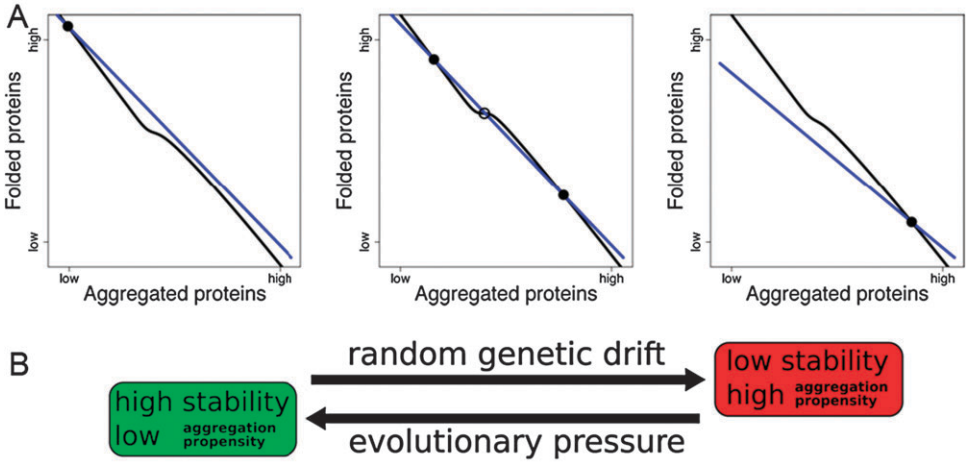

Fig. 3 (A) Phase plane analysis reveals a general possibility for bistability in the competition between protein folding and aggregation. Filled circles denote stable and open circles unstable steady states at the intersections of the nullclines for protein folding (straight solid line) and aggregation (other solid line). The three different possible scenarios for biologically relevant parameters are: (left panel) one stable steady state at high concentration of folded and low concentration of aggregated proteins, (right panel) one stable steady state at low concentration of folded and high concentration of aggregated proteins, and (central panel) both stable steady states are present. Evolutionary pressure acting on protein stability and against aggregation propensity and the random evolutionary drift make sure that the scenario in the left panel is realized and aggregation avoided.

non-impaired degradation beyond capacity (Fig. 4A), which increases robustness. In both cases, several proteins would have to aggregate before the system becomes bistable. This result is in agreement with the finding of a distinct degradation capacity of refoldable aggregates,[36] and the observation of aggresomes.[37] The time evolution of the number of aggregated proteins is simulated numerically, and for parameters that satisfy the solubility condition the system always settles in a folded steady state at low numbers of aggregated proteins. However, if the number of aggregated proteins exceeds the degradation capacity, then homeostasis fails (Fig. 4C).

To explore whether the sharp cut-offs in the degradation rates of type II and III play a significant role, we explored the use of Hill functions to provide an alternative representation of the transition from first to zeroth order degradation (see Methods). A systematic study of the critical aggregation propensity as a function of the Hill coefficient n reveals that the solubility condition appears to depend in a continuous and strictly monotonically decreasing function on the Hill coefficient, and thus can be accounted for by a constant (Fig. 4D).

Stochasticity reduces protein solubility

Stochasticity, which is often an important feature of biochemical reactions in living systems, is caused primarily by the fact that sometimes very small numbers of molecules of a certain type are available in a cell. Fluctuations have been found to play prominent roles in many processes vital for cell function and development, for example in gene expression,[38] signaling[39,40] and differentiation.[41] Although noise can be deleterious for example by disrupting signaling pathways or perturbing a system from an optimal steady state,[42] it can also increase adaptability to changing environments[43] or phenotypic variety of a population,[44] and as such be beneficial.

In the context of the present discussion we consider the observation that noise can induce bistability.[39] To study the effect of noise on protein homeostasis, we carried out stochastic simulations with the Gillespie algorithm[45] (see Methods). In these simulations, we used a two-component noise generator that explicitly models transcription and translation[42] to mimic protein synthesis and degradation as the dominant source of variability in cellular protein levels. We used the Fano factor $\eta = \sigma^2/P$, defined as the ratio of the variance over the mean of the protein abundance P, to describe the noise strength. For the same parameters, protein abundance noise increases the likelihood that protein homeostasis fails, and thus on average decreases protein solubility (Fig. 5A). To compensate for this effect, evolutionary selection has to either down-regulate the cellular abundance or increase the fitness S/Λ, so that the likelihood of failed homeostasis remains low. A major factor determining the susceptibility to noise is the strong non-reversibility of protein aggregation. If the fluctuations are sufficiently large to perturb the system out of the folded steady state and exhaust the degradation capacity, then aggregation will take place.

Proteasomal degradation is subject to fluctuations itself because of the varying degree of availability of proteasomes,[46] while autophagy, which also play an important role in the degradation of protein aggregates,[47] encapsulates and sequesters whole aggregates at once, making it more robust against noise. Variations in the levels of proteasome-constituent proteins also contribute to overall noise.[48] Stochastic simulations that model independent sources of noise for protein levels and degradation capacity reveal that uncorrelated noise that impairs the degradation capacity can dramatically amplify the burden that noise in protein abundance imposes on solubility. Indeed, all simulations resulted in failed homeostasis with additional degradation capacity noise (Fig. 5B). Proteasomal proteins are among the proteins with the smallest fluctuations in their cellular concentrations,[49] a situation that may originate from the need of eliminating the amplification of noise in protein levels, which would lead to failed protein homeostasis and loss of solubility.

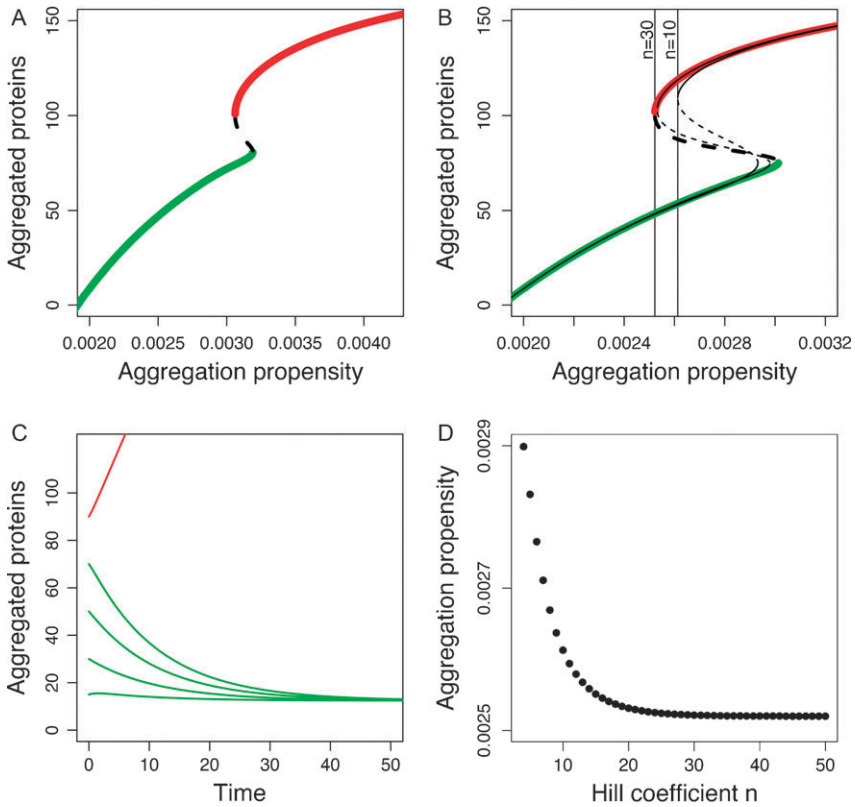

Fig. 4 Bifurcation analysis and numerical verification of the solubility condition. (A, B) Bifurcation diagrams for ϕ_d^{II} (A) and ϕ_d^{III} (B), indicating the steady states as a function of the bifurcation parameter aggregation propensity. Stable steady states are indicated by solid lines, unstable states by dashed lines; folded steady states are found on the left of the unstable states and aggregated steady states on the right. Alterations of Hill coefficients have only little effect (B). (C) The numerical integration of the differential equations provides the time evolution of the system. For initial conditions within the degradation capacity, the system settles on a steady state characterized by a low number of aggregated proteins as long as the solubility condition is met (family of four curves). If the degradation capacity is exceeded, then uncontrolled aggregation takes place (other line). (D) Aggregation propensity at the bifurcation point, just avoiding the bistable regime, as a function of the Hill coefficient n.

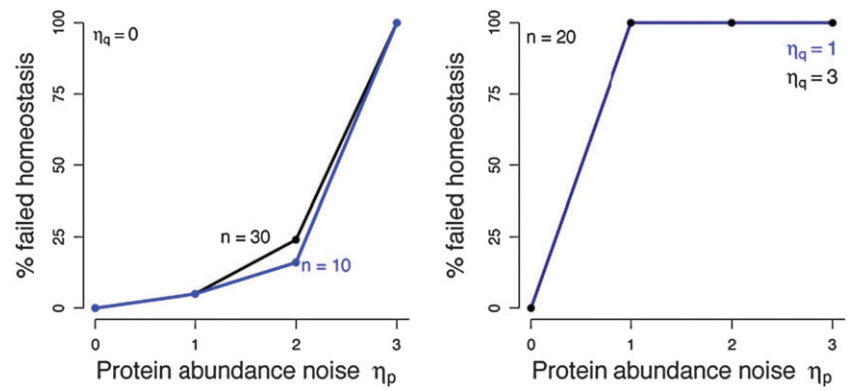

Fig. 5 The presence of noise increases the pressure on protein homeostasis and effectively reduces protein solubility. We report the percentage of stochastic simulations that yielded high numbers of aggregated proteins as a function of noise. (A) An increase in the protein abundance noise (η_p) enhances the likelihood of aggregation and failed homeostasis. (B) Uncorrelated noise that limits the capacity of the degradation machinery (η_q) amplifies protein abundance noise and dramatically promotes aggregation.

Consequences of the solubility condition

According to the solubility condition (eqn (3)), the soluble cellular abundance of a protein depends on its stability and aggregation propensity, as well as the degradation and quality control capacity. To test the validity of this relationship, and establish whether aggregation propensity and stability can be used to better understand the determinants of *in vivo* protein solubility, we discuss the application to experimental observations both at the levels of single proteins and of proteomes.

(1) Correlation between protein abundance and aggregation rates. The solubility condition (eqn (3)) enables a rationalization of the recently proposed 'life on the edge' hypothesis.[50] According

to this condition, highly expressed proteins should be particularly stable and have a low aggregation propensity in order to remain soluble and avoid aggregation. Indeed, the solubility condition, rewritten in logarithmic form, becomes

$$\log(P) < -\log(\Lambda) + \log(Sd) \qquad (4)$$

This relationship is consistent with the strong anti-correlation that has been observed between protein expression levels and aggregation rates on a proteome-wide scale.[50,51] In fact, because protein copy numbers in the cell can vary between a very few proteins up to more than 10^6, an anticorrelation between expression levels and aggregation rates is a particularly effective way to guarantee protein homeostasis and control balanced fluxes to and from the misfolded and aggregated states. In addition, recent experimental studies have reported that indeed more highly expressed proteins are more stable.[52,53]

(2) Solubility of maltose-binding protein mutants. The maltose-binding protein (MBP) of *E. coli* is a component of the transport system for maltose sugars. The experimentally measured *in vivo* solubilities of MBP and of three of its mutants were extracted from the literature.[54,55] Relative stabilities $\Delta\Delta G$ were predicted from the amino acid sequence[56] and are in accordance with experimental findings.[55] The solubility of the mutants correlates with the relative protein stability upon mutation: the higher the loss in stability through the mutation, the higher the loss in solubility. Additionally, the destabilizing mutations increase the aggregation propensity (Fig. 6A) and reduce the overall protein solubility. Intrinsic aggregation propensities were predicted with the Zyggregator method.[7] Two high quantitative linear correlations between solubility and stability (correlation coefficient 0.99) and solubility and aggregation propensity (correlation coefficient −0.86) support the hypothesis that protein solubility in the cell can be rationalized by the solubility condition, under the assumption that the mutations do not disrupt specific degradation signals and thus alter degradation pathways and capacity.

(3) HIV integrase heterologous expression. It is well established that in order to enhance the yields in the heterologous expression of HIV integrase in *E. coli*, an effective strategy is to optimize the solubility of the protein by mutating exposed surface residues.[57] By using the available data, we found that for this protein and its mutants the solubility correlates with the predicted intrinsic aggregation propensity[7] (Fig. 6B), and that the mutations do not change the stability significantly (Fig. 6B). Mutant PDB structures (starting from PDB 1BIZ, chain A) were optimized with MODELLER,[58] and stability differences $\Delta\Delta G$ were calculated with FoldX.[59] The small changes in stability are in agreement with the fact that fully exposed surface residue mutations have generally only a small effect on the stability.[60] As predicted by the solubility condition, we found a correlation between solubility and aggregation propensity together with an approximately constant stability of the HIV integrase mutants.

Discussion

In this work we have derived a solubility condition (eqn (3)), which defines protein solubility in the cell in terms of the competition between the folding, aggregation and degradation processes. This solubility condition provides a conceptual framework that provides insight into the following observations.

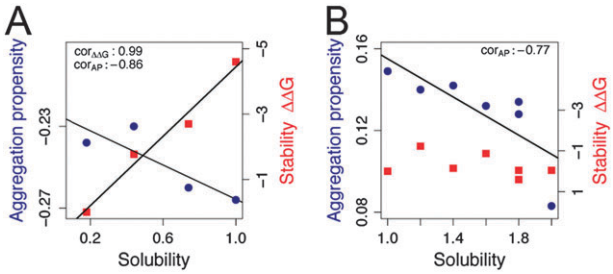

Fig. 6 Cellular protein solubility as a function of protein aggregation propensity (circles) and stability (squares); we compare the relative solubility of the wild type over changes of stability upon mutation $\Delta\Delta G$ in kcal mol^{-1}, and predicted intrinsic aggregation propensity. (A) Mutations that decrease the stability and increase the aggregation propensity decrease *in vivo* solubility of the abundant maltose-binding protein. (B) Surface residue mutations that only marginally change the stability, but clearly decrease the aggregation propensity of the core domain of the enzyme HIV integrase, heterologously expressed in *E. coli*, increase the *in vivo* solubility.

Correlation between protein expression levels and rates of protein sequence evolution

This correlation, which was observed recently,[17] can be rationalized in the following way. While the amino acid sequences of proteins are maintained at a mutation-selection equilibrium,[61] the majority of mutations, and thus the random genetic drift, on average decrease protein stability[52] and increase the aggregation propensity.[62] Thus, an increase in protein stability, or a reduction of the aggregation propensity, should be directly proportional to the strength of the evolutionary pressure. To maintain homeostasis, the ratio of protein stability and aggregation propensity should scale proportionally with the cellular protein concentration (eqn (3)). As a result, a higher protein abundance implies higher evolutionary pressure, and as a result a lower rate of protein sequence evolution in order to avoid aggregation.[17]

In addition, the mistranslation-induced misfolding hypothesis[63] suggests that mRNA sequences have a translational robustness proportional to their level of expression in order to avoid misfolding and aggregation of the products. Further, since after synthesis in the ribosome, nascent chains are particularly vulnerable to misfolding and aggregation,[64] additional safety mechanisms are in place. For example, the presence of rare codons forces the translation process to slow down, giving long polypeptide chains time to fold,[65] and if a mistake is detected after peptide bond formation abortive termination is initiated.[66] It is also known that structural and functional constraints have a comparable effect on the rate of protein sequence evolution than the rate of expression,[67] or kinetic evolutionary constraints.[68] Consistent with these observations, the solubility condition indicates that the avoidance of aggregation is indeed a major constraint on the evolutionary design of protein sequences.[50,69]

Heterologous protein expression

The failure to obtain soluble products from heterologous expression is a major bottleneck in structural studies and biotechnology. An often successful strategy to increase yield of recombinant protein production is the co-expression with chaperones.[70,71] This option is consistent with the solubility condition (eqn (3)), as increasing the capacity Q of the protein homeostasis network enhances the concentration of soluble proteins.

Role of molecular chaperones

As these molecules assist protein folding and help the degradation of misfolded proteins by the ubiquitin–proteasome system,[27] they contribute significantly to the overall protein homeostasis capacity. Chaperone substrates tend to have low expression,[72] low propensity to fold[73] and low solubility.[74] Furthermore, chaperone substrates display low noise at the transcript level,[72] but unusually high noise levels at the protein level.[2] By enhancing the capacity Q, molecular chaperones can buffer this noise by lowering the solubility condition (eqn (3)), for example to facilitate protein evolvability.[75]

Conclusions

We have addressed the problem of understanding the consequence of the competition between folding and aggregation, and have shown that it is possible to derive an expression, which we call the 'solubility condition', defining a close relationship between the abundance of proteins in the cell and their folding stability, aggregation propensity and degradation rates. These results provide directly testable hypotheses and offer a rationalization of the recent 'life on the edge' observation, according to which protein abundance and aggregation are strongly correlated.

Methods

Derivation of the solubility condition

To avoid aggregation and maintain solubility, the system parameters (folding, unfolding and degradation rates, aggregation propensity and degradation capacity) must be selected so that the aggregated state is avoided and that only the folded state is present. From the steady state conditions, $\mathrm{d}F/\mathrm{d}t = \mathrm{d}A/\mathrm{d}t = 0$, we obtain

$$F = \frac{k_\mathrm{f}/k_\mathrm{u}}{1 + k_\mathrm{f}/k_\mathrm{u}}(P_\mathrm{tot} - A)$$
$$F = (P_\mathrm{tot} - A) - \frac{\phi_\mathrm{d}}{\Lambda A} \quad (5)$$

Stationary solutions (steady states or fixed points) are calculated by finding the intersections of the nullclines (eqn (5)). Steady states are stable if the real parts of all eigenvalues of the Jacobian, evaluated at steady state, are negative.

In the case of unlimited first order degradation $\phi_\mathrm{d}^\mathrm{I}$, under the assumption that no aggregation takes place ($A = 0$), the only stable steady state is given by

$$P = \frac{Sd}{\Lambda}$$

Under the assumption of a limited degradation capacity the system can have only 1 or 3 steady states. Protein homeostasis is guaranteed, if (i) the folded steady state exists, and (ii) the aggregated steady state as well as the unstable steady state that destabilizes the system and separates the stable steady states do not exist.

The solution of eqn (5) for $A < Q$ is $P - A = Sd/\Lambda$, and hence

$$P - Q < \frac{Sd}{\Lambda}$$

which is the condition that the folded state exists as the stable one. A similar solution of eqn (5) is obtained for the non-existence of the aggregated steady state for $A \geq Q$ and $\phi_\mathrm{d}^\mathrm{II}$.

For $\phi_\mathrm{d}^\mathrm{III}$ one has instead

$$P - Q < \frac{Sd_0}{\Lambda}$$

The solutions for $\phi_\mathrm{d}^\mathrm{II}$ and $\phi_\mathrm{d}^\mathrm{III}$ differ only by a constant. To model a more realistic transition from first order to zeroth order degradation kinetics, the degradation flux is expressed with decreasing and increasing Hill functions as pre-factors, where the Hill coefficient n describe the sharpness of the transition

$$\phi_\mathrm{d} = \frac{1}{1 + \left(\frac{A}{Q}\right)^n}\mathrm{d}A + \frac{1}{1 + \left(\frac{Q}{A}\right)^n}\mathrm{d}Q$$

Bifurcation analysis

The software packages AUTO and Mathematica were used to perform bifurcation analyses. Qualitative or topological changes in the systems behavior were analyzed in dependence on the aggregation propensity as bifurcation parameter by mapping the steady states as a function of the aggregation propensity. Numerical integration of the set of ordinary differential equations was performed in MATLAB with standard solvers.

Stochastic simulations

Because the underlying reaction mechanism of the competition between protein folding and aggregation is formulated in terms of elementary fluxes described by the law of mass action, stochastic simulations can be implemented without further modifications by the Gillespie algorithm.[45] Aggregation and degradation rates are updated after every time step as a function of the number of proteins in the aggregated state. Noise is modeled explicitly by implementing a two-component noise generator that simulates transcription and translation, as described in ref. 42. Independent simulations were run to verify that the generated levels of noise fell within a normal distribution with a standard deviation of 1 around the desired mean noise level. The simulations were repeated 1000 times for each setting and the percentage recorded that ended in high numbers of aggregated proteins, far exceeding the initial steady state concentration as well as the quality-control capacity.

Acknowledgements

We would like to thank Tuomas Knowles, Thomas Jahn and Gian Gaetano Tartaglia for helpful discussions. Support from EMBO (SP and MV), EPSRC (SP) and the Royal Society (MV) is gratefully acknowledged.

References

1 W. E. Balch, R. I. Morimoto, A. Dillin and J. W. Kelly, *Science*, 2008, **319**, 916–919.
2 E. T. Powers, R. I. Morimoto, A. Dillin, J. W. Kelly and W. E. Balch, *Annu. Rev. Biochem.*, 2009, **78**, 959–991.
3 M. Vendruscolo and C. M. Dobson, *Faraday Discuss.*, 2009, **143**, 277–291.
4 D. J. Selkoe, *Nature*, 2003, **426**, 900–904.
5 F. Chiti and C. M. Dobson, *Annu. Rev. Biochem.*, 2006, **75**, 333–366.
6 J. S. Richardson and D. C. Richardson, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **99**, 2754–2759.
7 G. G. Tartaglia, A. P. Pawar, S. Campioni, C. M. Dobson, F. Chiti and M. Vendruscolo, *J. Mol. Biol.*, 2008, **380**, 425–436.
8 S. Pechmann, E. D. Levy, G. G. Tartaglia and M. Vendruscolo, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 10159–10164.
9 F. Chiti and C. M. Dobson, *Nat. Chem. Biol.*, 2009, **5**, 15–22.
10 T. R. Jahn and S. E. Radford, *Arch. Biochem. Biophys.*, 2008, **496**, 100–117.
11 Q. Ding, V. Cacarini and J. N. Keller, *Trends Neurosci.*, 2007, **30**, 31–36.
12 S. Lindquist, *Annu. Rev. Biochem.*, 1986, **55**, 1151–1191.
13 D. Ron and P. Walter, *Nat. Rev. Mol. Cell Biol.*, 2007, **8**, 519–529.
14 E. Dekel and U. Alon, *Nature*, 2005, **436**, 588–592.
15 R. Bedford and D. L. Hartl, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 1133–1138.
16 C. Pal, B. Papp and L. D. Hurst, *Genetics*, 2001, **158**, 927–931.
17 D. A. Drummond, J. D. Bloom, C. Adami, C. O. Wilke and F. H. Arnold, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 14338–14343.
18 D. A. Drummond, A. Raval and C. O. Wilke, *Mol. Biol. Evol.*, 2005, **23**, 327–337.
19 M. A. DePristo, D. M. Weinreich and D. L. Hartl, *Nat. Rev. Genet.*, 2005, **6**, 678–687.
20 T. R. Rieger, R. I. Morimoto and V. Hatzimanikatis, *Biophys. J.*, 2006, **90**, 886–895.
21 N. Barkai and S. Leibler, *Nature*, 1997, **387**, 913–917.
22 D. Segre, D. Vitkup and G. M. Church, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 15112–15117.
23 T. P. J. Knowles, C. A. Waudby, G. L. Devlin, a. Aguzzi, M. Vendruscolo, E. M. Terentjev, M. E. Welland and C. M. Dobson, *Science*, 2009, **326**, 1533–1537.
24 A. Wagner, *J. Exp. Zool., Part B*, 2007, **308**, 322–324.
25 A. Ciechanover, *Nat. Rev. Mol. Cell Biol.*, 2005, **6**, 79–86.
26 A. L. Goldberg, *Nature*, 2003, **426**, 895–899.
27 A. J. McClellan, S. Tam, D. Kaganovich and J. Frydman, *Nat. Cell Biol.*, 2005, **7**, 736–741.
28 A. Belle, A. Tanay, L. Bitincka, R. Shamir and E. K. O'Shea, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 13004–13009.
29 W. W. Wong, T. Y. Tsai and J. C. Liao, *Mol. Syst. Biol.*, 2007, **3**, 130–137.
30 N. F. Bence, R. M. Sampat and R. R. Kopito, *Science*, 2001, **292**, 1552–1555.
31 E. Cohen, J. Bieschke, R. M. Perciavalle, J. W. Kelly and A. Dillin, *Science*, 2006, **313**, 1604–1610.
32 J. W. Veening, W. K. Smits and O. P. Kuipers, *Annu. Rev. Microbiol.*, 2008, **62**, 193–210.
33 K. Sneppen, M. A. Micheelsen and I. B. Dodd, *Mol. Syst. Biol.*, 2008, **4**, 182.
34 S. Bershtein, M. Segal, R. Bekerman, N. Tokuriki and D. S. Tawfik, *Nature*, 2006, **444**, 929–932.
35 J. D. Bloom, F. H. Arnold and C. O. Wilke, *Mol. Syst. Biol.*, 2007, **3**, 76.
36 D. Kaganovich, R. Kopito and J. Frydman, *Nature*, 2008, **454**, 1088–1095.
37 R. R. Kopito, *Trends Cell Biol.*, 2000, **10**, 524–530.
38 M. B. Elowitz, A. J. Levine, E. D. Siggia and P. S. Swain, *Science*, 2002, **297**, 1183–1186.
39 M. Samoilov, S. Plyasunov and A. P. Arkin, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **102**, 2310–2315.
40 J. M. Pedraza and J. Paulsson, *Mol. Syst. Biol.*, 2007, **3**, 81.
41 G. M. Süel, R. P. Kulkarni, J. Dworkin, J. Garcia-Ojalvo and M. B. Elowitz, *Science*, 2007, **315**, 1716–1719.
42 T. Lu, M. Ferry, R. Weiss and J. Hasty, *Phys. Biol.*, 2008, **5**, 036006.
43 M. Kaern, T. C. Elston, W. J. Blake and J. J. Collins, *Nat. Rev. Genet.*, 2005, **6**, 451–464.
44 T. Lu, T. Shen, M. Bennett, P. G. Wolynes and J. Hasty, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 18982–18987.
45 D. T. Gillespie, *Annu. Rev. Phys. Chem.*, 2007, **58**, 35–55.
46 K. Sneppen, L. Lizana, M. H. Jensen, S. Pigolotti and D. Otzen, *Phys. Biol*, 2009, **6**, 036005.
47 B. Ravikumar, R. Duden and D. D. Rubinsztein, *Hum. Mol. Genet.*, 2002, **11**, 1107–1117.
48 M. Thattai and A. V. Oudenaarden, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 8614–8619.
49 J. R. S. Newman, S. Ghaemmaghami, J. Ihmels, D. K. Breslow, M. Noble, J. L. DeRisi and J. S. Weissman, *Nature*, 2006, **441**, 840–846.
50 G. G. Tartaglia, S. Pechmann, C. M. Dobson and M. Vendruscolo, *Trends Biochem. Sci.*, 2007, **32**, 204–206.
51 G. G. Tartaglia and M. Vendruscolo, *Mol. Biosyst.*, 2009, **5**, 1873–1876.
52 P. Yue, Z. Li and J. Moult, *J. Mol. Biol.*, 2005, **353**, 459–473.
53 H. C. Yen, Q. Xu, D. M. Chou, Z. Zao and S. J. Elledge, *Science*, 2008, **322**, 918–923.
54 W. C. Wigley, R. D. Stidham, N. M. Smith, J. F. Hunt and P. J. Thomas, *Nat. Biotechnol.*, 2001, **19**, 131–136.
55 J. M. Betton and M. Hofnung, *J. Biol. Chem.*, 1996, **271**, 8046–8052.
56 E. Capriotti, P. Fariselli and R. Casadio, *Nucleic Acids Res.*, 2005, **33**, W306–W310.
57 K. L. Maxwell, A. K. Mittermaier, J. D. Forman-Kay and A. R. Davidson, *Protein Sci.*, 1999, **8**, 1908–1911.
58 N. Eswar, D. Eramian, B. Webb, M. Y. Shen and A. Sali, *Methods Mol. Biol.*, 2008, **426**, 145–159.
59 R. Guerois, J. E. Nielsen and L. Serrano, *J. Mol. Biol.*, 2002, **320**, 369–387.
60 N. Tokuriki, F. Stricher, J. Schymkowitz, L. Serrano and D. S. Tawfik, *J. Mol. Biol.*, 2007, **369**, 1318–1332.
61 J. W. Drake, B. Charlesworth, D. Charlesworth and J. F. Crow, *Genetics*, 1998, **148**, 1667–1686.
62 W. Kim and M. H. Hecht, *J. Mol. Biol.*, 2008, **377**, 565–574.
63 D. A. Drummond and C. O. Wilke, *Cell (Cambridge, Mass.)*, 2008, **134**, 341–352.
64 D. A. Drummond and C. O. Wilke, *Nat. Rev. Genet.*, 2009, **10**, 715–724.
65 A. A. Komar, *Trends Biochem. Sci.*, 2009, **34**, 16–24.
66 H. S. Zaher and R. Green, *Nature*, 2009, **457**, 161–166.
67 M. Y. Wolf, Y. I. Wolf and E. V. Koonin, *Biol. Direct*, 2008, **3**, 40.
68 C. T. Friel, D. A. Smith, M. Vendruscolo, J. Gsponer and S. E. Radford, *Nat. Struct. Mol. Biol.*, 2009, **16**, 318–324.
69 C. M. Dobson, *Trends Biochem. Sci.*, 1999, **24**, 329–332.
70 A. de Marco, E. Deuerling, a. Mogk, T. Tomoyasu and B. Bukau, *BMC Biotechnol.*, 2007, **7**, 32.
71 A. Haacke, G. Fendrich, P. Ramage and M. Geiser, *Protein Expression Purif.*, 2009, **64**, 185–193.
72 O. Noivirt-Brik, R. Unger and A. Horovitz, *Bioinformatics*, 2007, **23**, 3276–3279.
73 A. Y. Yam, Y. Xia, H. T. Lin, A. Burlingame, M. Gerstein and J. Frydman, *Nat. Struct. Mol. Biol.*, 2008, **15**, 1255–1262.
74 T. Niwa, B.-W. Ying, K. Saito, W. Z. Jin, S. Takada, T. Ueda and H. Taguchi, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 4201–4206.
75 N. Tokuriki and D. S. Tawfik, *Nature*, 2009, **459**, 668–673.