Biophysical Chemistry, Biomolecules, and Biomaterials; Surfactants and Membranes

Proteins Are Solitary! Pathways of Protein Folding and Aggregation in Protein Mixtures

Valentino Bianco, Miren Alonso-Navarro, Desirè Di Silvio, Sergio E. Moya, Aitziber L. Cortajarena, and Ivan Coluzza

J. Phys. Chem. Lett., **Just Accepted Manuscript** • DOI: 10.1021/acs.jpclett.9b01753 • Publication Date (Web): 02 Aug 2019

Proteins are Solitary! Pathways of Protein Folding and Aggregation in Protein Mixtures

Valentino Bianco,[*,†] Miren Alonso-Navarro,[‡] Desire Di Silvio,[‡] Sergio Moya,[‡]

Aitziber L. Cortajarena,[‡,¶] and Ivan Coluzza[*,‡,¶]

†*Faculty of Chemistry, Chemical Physics Deprtment, Universidad Complutense de Madrid, Plaza de las Ciencias, Ciudad Universitaria, Madrid 28040, Spain*

‡*CIC biomaGUNE, Paseo Miramon 182, 20014 San Sebastian, Spain.*

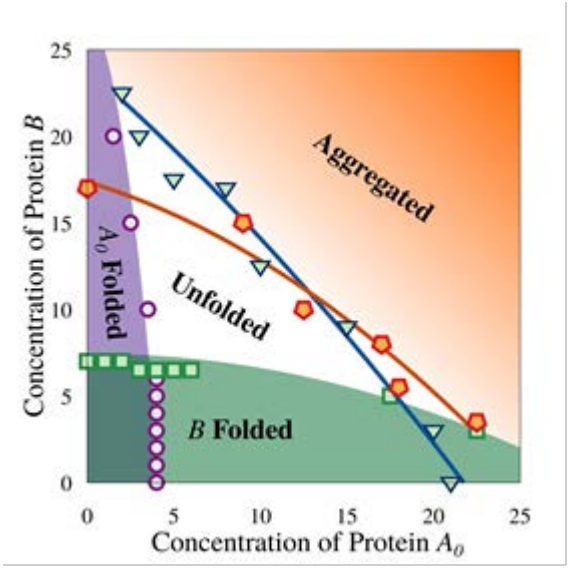
¶*IKERBASQUE, Basque Foundation for Science, 48013 Bilbao, Spain.*

E-mail: vabianco@ucm.es; icoluzza@cicbiomagune.es

Abstract

We present a computational and experimental study on the folding and aggregation in solutions of multiple protein mixtures at different concentrations. We show how in protein mixtures, each component is capable of maintaining its folded state at desensitises higher then the one at which they would precipitate in single species solutions. We demonstrate the generality of our observation over many different proteins using computer simulations capable of fully characterising the cross-aggregation phase diagram of all the mixtures. Dynamic light Scattering experiments were performed to evaluate the aggregation of two proteins, the bovine serum albumin (BSA) and the consensus tetratricopeptide repeat (CTPR), in solutions of one or both proteins. The experiment confirm our hypothesis and the simulations. These findings elucidate critical aspects on the cross-regulation of expression and aggregation of proteins exerted by the cell and on the evolutionary selection of folding and not-aggregating protein sequences, paving the way for new experimental tests.

Graphical TOC Entry



Introduction

Proteins are involved in a wide range of physiological processes crucial for all living organism .[1] After the synthesis at the ribosome, polypeptide chain are exposed to a highly crowded cellular environment,[2] where, despite many non-specific interactions,[3–5] the chain is capable to select for a subset of amino acid contacts which funnel the free energy landscape toward a unique native/folded state.[6–10] In fact, cells evolved complex post-translational processes that include chaperones to facilitate proteins to fold.

However, protein aggregation is mostly unavoidable when proteins are over-expressed at concentrations higher than the physiological one.[11–13] The linear anti-correlation between expression levels and aggregation propensity is a remarkable experimental fact Fig. 1[121], especially considering the large number of protein species present in cells.[14,15]

Provided that the expression levels are kept below the precipitation concentration of the isolated species, the result suggests the intriguing possibility that protein folding is unperturbed by the presence of the other proteins. The alternative would require a complex regulatory process that would suppress all the potential interacting partners upon the increase of the expression level of a specific protein. However, such a condition is incompatible with the linear dependence mentioned above. Hence, it is natural to assume that, in order to guarantee the correct biological functions, proteins have evolved to have a low enough propensity to aggregate within a range of protein expression required for their biological activity. However, also this hypothesis requires a remarkably complex process of selection where each protein mutation was acceptable only if it did not introduce unwanted interactions with all the other potential partners in the organism. Such a selection process is often referred to as negative design[17–22] and it is used to optimise the folding and the interactions of specific proteins. From such studies, it becomes apparent that the negative design against

[1]Protein propensity to aggregate is also correlated with the sub-cellular volume where the protein resides: proteins tend to be abundant and highly soluble when localised in small sub-cellular environment like the ribosome, while they show a higher propensity to aggregate at lower concentration when localised in larger volumes like the cytosol.[13]

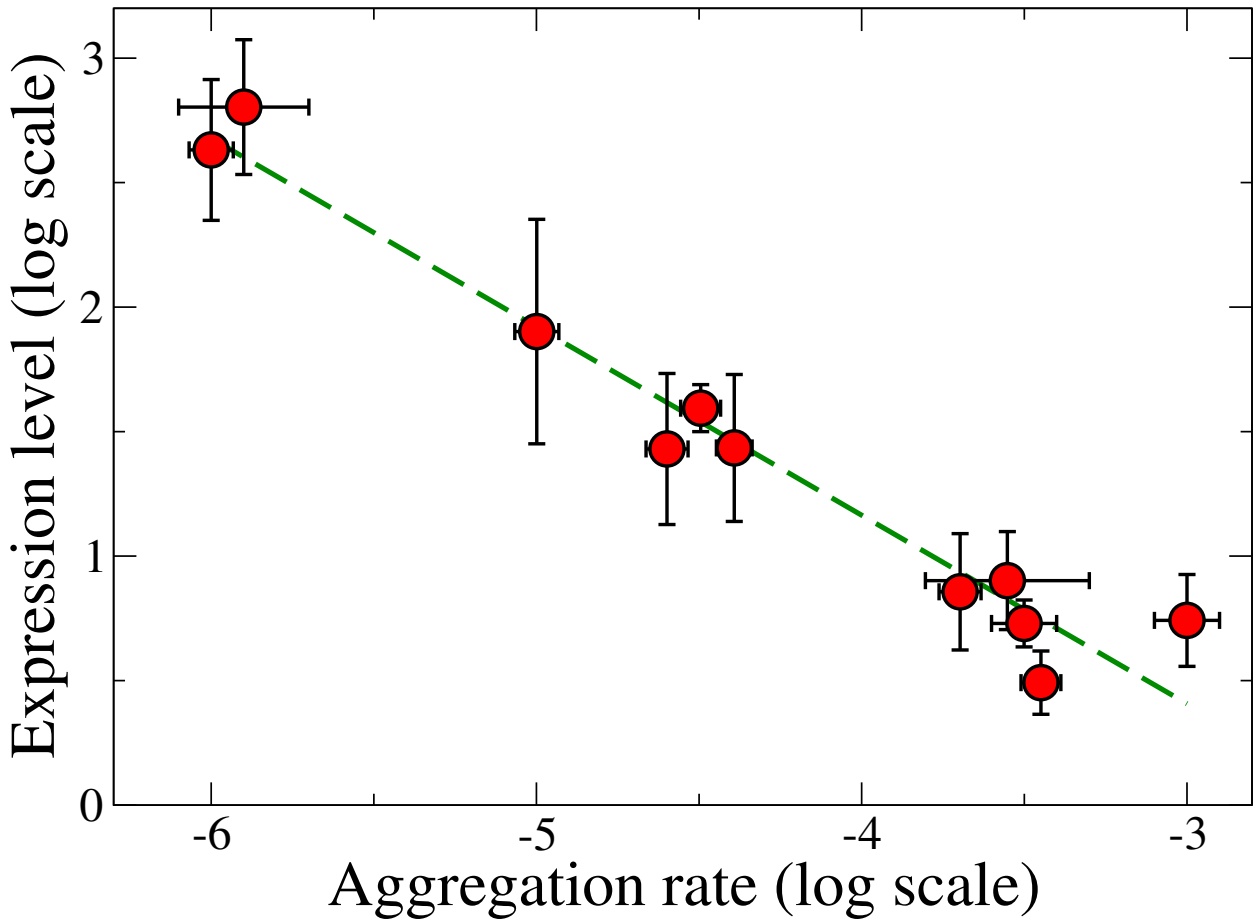


Figure 1: Expression levels of proteins measured by the cellular mRNA concentrations,[16] plotted versus their aggregation rates in seconds and plotted on a base 10 log scale (adapted from[12]). Each point corresponds to an independent experiment performed on a specific protein. For details on the specific experiments we refer to Table 1 of the review of Tartaglia et al.[12] The linear anti-correlation trend suggests each protein is regulated up to its aggregation concentration independently of the other proteins.

thousand of potential partners is a problem of incredible complexity.

Here we present a computational and experimental study to support a different hypothesis to explain the uncorrelated behaviour of protein folding. With our model, we imitate a simple cellular environment, where each protein coexists in its native state with a large number of different proteins. We show how phylogenic unrelated proteins (artificial and natural) are unperturbed by each other precisely up to their corresponding aggregation concentration. We observe that our proteins can keep their native conformation at higher concentration when solvated in heterogeneous solutions (multiple copies of different proteins) compared

to their aggregation concentration (determined in our previous study [23]) in homogeneous solutions (multiple copies of the same protein). Finally, we have verified these conclusions experimentally by the evaluation of the aggregation behaviour of two model proteins, the bovine serum albumin (BSA) and the consensus tetratricopeptide repeat (CTPR). The experimental results fully support the theoretical studies, since heterogeneous proteins mixtures at total protein concentration above the aggregation threshold of one of the components did not show any aggregation.

Results

To tackle computationally the folding/aggregation problem in solutions involving tens of proteins we adopt a coarse-grain model for protein solvated by water,[23–29] based on a wa-ter model that reproduces–at least qualitatively–the thermodynamic properties of the sol-vent.[30–32] We perform Monte Carlo simulations to design artificial protein structures, with a well tested scheme,[33,34] recently extended to optimize the protein sequence to the water properties at a given temperature and pressure.[28] Once designed, the Monte Carlo fold-ing/aggregation simulations are performed with a statics involving from ~ 20 to ~ 200 independent runs–each up to $\sim 4 \times 10^7$ steps–according to the number of proteins (see the Supplementary Informations for further details).

In a recent study[23] we have studied 8 designed proteins (named A_0, A_1, A_2, B, C, D, E and F) capable of folding into aqueous environment[27–29] (see Fig. S1 in the SI). For each protein, we confirmed the proper refolding at low concentration and we studied the folding behaviour as a function of the concentration. In particular, we identified three main regions: FOL where the proteins are completely folded, in UNF the proteins are neither fully folded neither aggregated, and in AGG the proteins are unfolded and aggregated. The UNF is extremely interesting and is the subject of our previous study[23] so we will not focus our attention on it here. In Tab. 1 we list the concentrations $c_{UNF \to AGG}$ of the $UNF \to AGG$

transition.

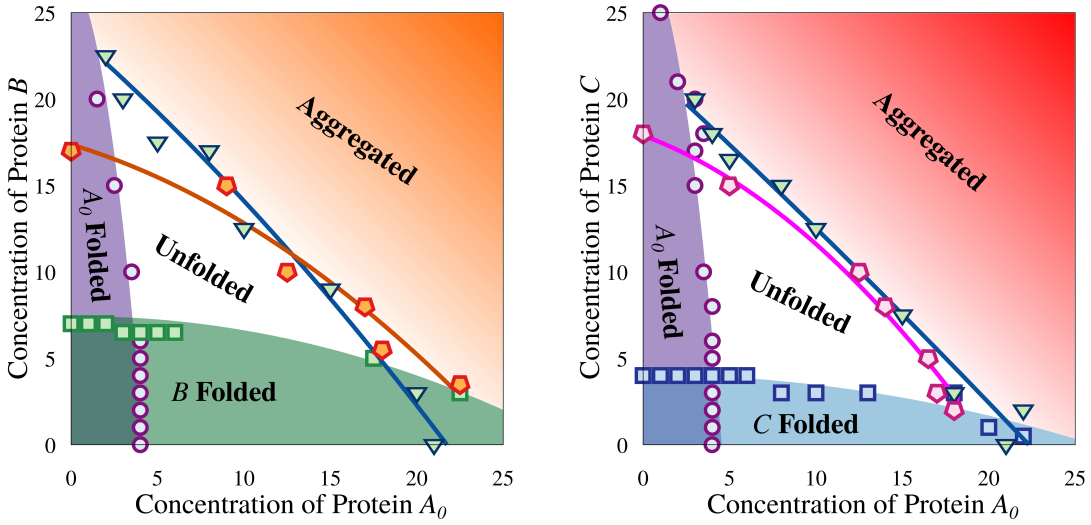


Figure 2: Phase diagrams of heterogeneous solution of proteins a) A_0 vs B and b) A_0 vs C respectively. The folded regions of protein of A_0 is orthogonal to B and C with a small dependence from each others concentrations. The circles represent the $c^{(A_0)}_{FOL\to UNF}$ transition line estimated by our computer simulations, while the squares represent the $c^{(B)}_{FOL\to UNF}$ and the $c^{(C)}_{FOL\to UNF}$ respectively. The triangles represent the $c^{(A_0)}_{UNF\to AGG}$ transition line, while the pentagons the $c^{(B)}_{UNF\to AGG}$ and the $c^{(C)}_{UNF\to AGG}$ respectively. The simulation grid points are in Fig. S7 in the SI.

In this work we describe how the same proteins fold in heterogeneous mixtures with the following scenarios i) A_0 and B; ii) A_0 and C; iii) A_0 and A_1; iv) A_0 and A_2; vi) A_0, B, C, D and E ; where each species is expressed up to the 25% (hence the total protein concentration considered is up to 50%). In what follow we discuss in detail the behaviour of proteins A_0 and B, and the "vi" scenario, while data for the other cases are reported in the SI as show a similar trend (see Fig. S4).

Our findings are summarised in the phase diagram (for the specific case of proteins A_0 vs B and A_0 vs C), computed by our simulations and shown in Fig. 2.

Our results show that any tested protein i in a solution of proteins i and j, is capable to fold into the native conformation as long as the protein i is expressed below the

$FOL \rightarrow UNF$ threshold, i.e. whenever $c^{(i)} < c^{(i)}_{FOL \rightarrow UNF}$. In fact, correct folding occurs even if the total protein concentration $c_{\rm tot} \equiv c^{(i)} + c^{(j)}$ overcomes the threshold concentration $c^{(i)}_{FOL \rightarrow UNF}$. Moreover, the free energy of the protein i remains unaffected (within the numerical error) by the presence of the other protein species j (Fig. S5a,b in the SI). These results demonstrate that the folding of binary protein mixtures depends essentially on their individual concentrations.

At high concentrations $c^{(i)} > c^{(i)}_{FOL \rightarrow UNF}$ proteins undergo the $FOL \rightarrow UNF$ transition, but the free energy profile is slightly affected by the presence of the other proteins and keeps its minimum (Fig. S5c,d in the SI). This holds until $c_{\rm tot}$ roughly overcomes $c^{(i)}_{UNF \rightarrow AGG}$, when the $UNF \rightarrow AGG$ transition is induced in the protein i although $c^{(i)} < c^{(i)}_{UNF \rightarrow AGG}$ (Fig. S5f,g in the SI) and the minimum of the free energy shifts at lower value of N_c. By further increasing $c_{\rm tot}$ the aggregation is induced. In this case, we found that the aggregates are mostly formed between proteins of the same species. We quantified such observation measuring the average number of inter-proteins contacts between the same protein species (HOMO) and different species (HETERO). In Fig. S9 of the SI we compare the number of HETERO and HOMO contacts for various protein mixtures. The result demonstrate that HOMO contacts dominates over the the HETERO ones.

Moreover, our results indicate that the propensity to aggregate is mostly dictated by similarities in protein sequence more than in protein structure. Indeed, as shown if Fig. S4, the free energy profiles of proteins A_0, A_1 and A_2 (all sharing the same native structure but with different sequences) mixed in protein mixtures, are essentially unaffected by the presence of the other proteins, resulting in a picture analogous to what observed in mixtures of proteins differing in structure and sequence, shown in Fig. S5a,b,c,d.

Our results remain valid even if we push our model closer to a typical cell environment where the number of species is much larger than two. We then studied the behaviour of proteins solvated in a heterogeneous solution composed of more than two species. In particular, we consider a solution composed by proteins A_0, B, C, D and E, each expressed

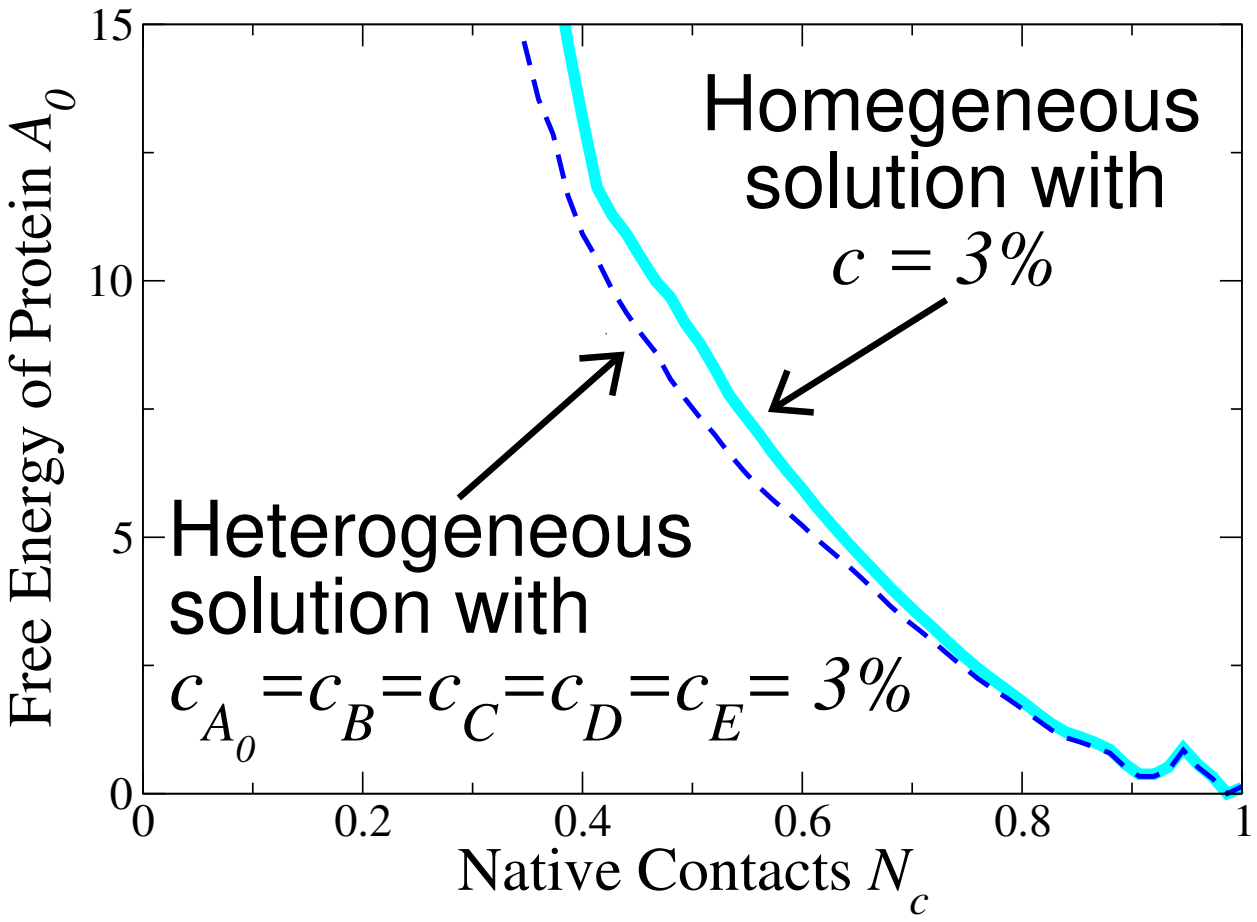


Figure 3: Free energy profiles of proteins in a heterogeneous solution of proteins A_0, B, C, D and E, where each protein is expressed at $c = 3\%$. Clear blue lines show the profile of the free energy for each protein i, when solvated in a homogeneous solution at $c^{(i)} = 3\%$.

at concentration $c = 3\%$ (therefore $c_{\text{tot}} = 15\%$). In Fig. 3 we report the free energy profiles of protein A_0 (the others are in Fig. S6 in the SI), in comparison with the profile of a homogeneous solution at $c^{(i)} = 3\%$. We observe that, also, in this case, the free energy remains substantially unaffected close to the minimum. Hence, we can conclude that even in the general scenario of heterogeneous proteins mixtures correct folding occurs in the condition that each protein is expressed below their isolated precipitation concentration as observed in cells.[12]

In order to experimentally verify the qualitative predictions of the phase diagrams in Fig. 2, we selected two proteins with distinct physicochemical properties. The first selected protein is the bovine serum albumin (BSA), while the second selected is a designed consensus

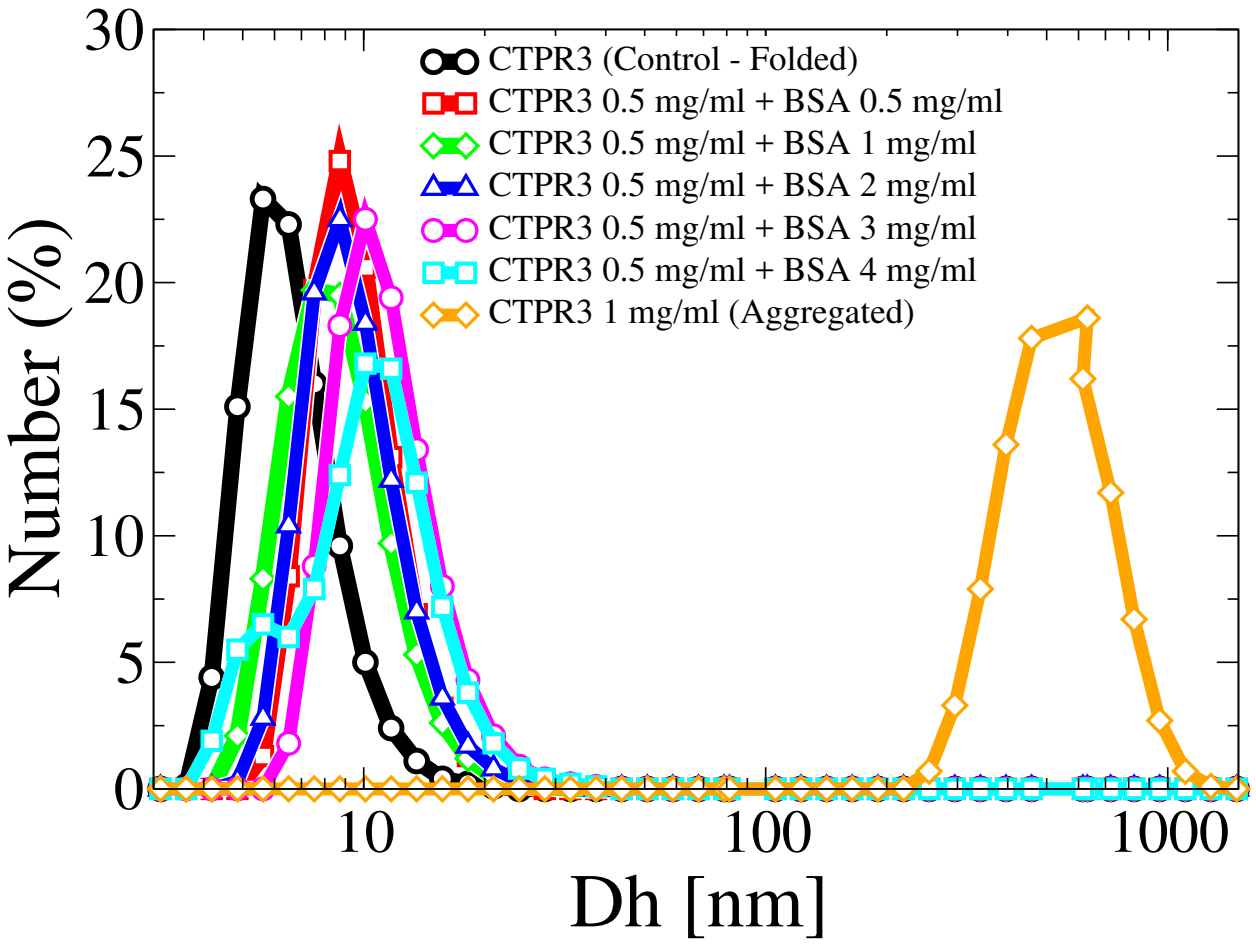


Figure 4: Folding vs. aggregation in heterogeneous protein solutions. Hydrodynamic diameter (Dh) of BSA and CTPR3 solutions determined after heat denaturation of the heterogeneous protein mixtures at a constant concentration of CTPR3 (C3) (0.5 mg/ml) and increasing concentrations of BSA (0.5-4 mg/ml).

tetratricopeptide repeat (CTPR) with three identical repeats, CTPR3 as an example of an entirely different class of protein. Once the aggregation concentration thresholds have been established (see Fig. S8 in the SI) for BSA (5 mg/ml) and CTPR (1 mg/ml), heterogeneous proteins solutions were evaluated. In this experiment, the concentration for CTPR3 is fixed below its aggregation threshold to 0.5 mg/ml. Then increasing concentrations of BSA are tested to pass the CTPR aggregation threshold when the total protein concentration is considered. In all cases, the total concentration is below the concentration aggregation threshold for BSA. The experimental results of the folding vs aggregation in heterogeneous proteins solutions show that CTPR protein can fold into its native conformation at total

protein concentrations above its aggregation threshold (Fig. 4). Neither the BSA nor the CTPR are amyloid forming proteins so we do not expect the aggregates to be fibrils. However, further analysis of the aggregates is the subject of current investigation.

Conclusions

We have presented a computational and experimental study on the competition between folding and aggregation of proteins in homogeneous and heterogeneous solutions. Inappropriate protein aggregation represents a crucial issue in biology and medicine, being associated with a growing number of diseases such as Alzheimer's and Parkinson's disease.[35–38] The main conclusion of this work is that proteins tend to fold uninfluenced by the presence of other proteins provided that their single concentration is below their specific unfolding concentration $c_{FOL\rightarrow UNF}$. In other words proteins do not aggregate in heterogeneous mixtures even if the total protein concentration is much higher than the aggregation threshold for a single protein. It is important to stress, that such robustness of the folding in heterogeneous mixtures occurs without any particular design procedure against aggregated or misfolded states, or any external help during the refolding. In order to verify our hypothesis, we performed experiments on the aggregation of a binary mixture of two proteins without any phylogeny correlation, namely the BSA and CTPR3. The proteins have different unfolding/aggregation concentrations, and in particular, the CTPR3 clear aggregation threshold allows for unambiguous identification of any transition events. Hence, we could identify very precisely the $c_{UNF\rightarrow AGG}$ for CTPR3. When we combined the BSA and the CTPR3 we did not observe and aggregated states of CTPR3 even at total protein concentrations much higher than $c_{UNF\rightarrow AGG}$. This result confirms our hypothesis that proteins fold mainly ignoring the other proteins in solutions.

Experimental observation shows that protein regulation is linearly correlated with the aggregation tendency confirming our hypothesis.[12] Considering that in cells there are many

different proteins, the combinatorial problem of cross regulating proteins to prevent aggregation would be extremely complicated. Our results instead show that the problem can be simplified to the matter of making sure that each protein is never over-expressed above its precipitation concentration. Of course, any post-translational modifications or chaperones action will have a positive contribution that further help against aggregation, but it does not change the baseline set by our results.

Acknowledgements

V.B. acknowledges the support from the Austrian Science Fund (FWF) project M 2150-N36 and from the European Commission through the Marie Skodowska-Curie Fellowship No. 748170 ProFrost. A.L.C. gratefully acknowledges support from the Ministerio de Economià y Competitividad (MINECO) (BIO2016-77367-C2-1-R) and the Basque Government (Elkartek KK-2017/00008). I.C. gratefully acknowledges support from the Ministerio de Economià y Competitividad (MINECO) (FIS2017-89471-R). This work was performed under the Maria de Maeztu Units of Excellence Program from the Spanish State Research Agency Grant No. MDM-2017-0720. gratefully acknowledges support from the Ministerio de Economià y Competitividad (MINECO) (FIS2017-89471-R). V.B. and I.C. acknowledge the support from the FWF project P 26253-N27. Simulations have been performed using the Vienna Scientific Cluster (VSC-3).

References

(1) Finkelstein, A. V.; Ptitsyn, O. B. *Protein physics*; Elsevier, 2016.

(2) Feig, M. Crowding in the Cellular Context: Tales of Clusters and Dynamics. *Biophys. J.* **2019**, *116*, 151a.

(3) Huang, K. Y.; Kingsley, C. N.; Sheil, R.; Cheng, C. Y.; Bierma, J. C.; Roskamp, K. W.;

Khago, D.; Martin, R. W.; Han, S. Stability of Protein-Specific Hydration Shell on Crowding. *J. Am. Chem. Soc.* **2016**, *138*, 5392–5402.

(4) Harada, R.; Tochio, N.; Kigawa, T.; Sugita, Y.; Feig, M. Reduced native state stability in crowded cellular environment due to protein-protein interactions. *J. Am. Chem. Soc.* **2013**, *135*, 3696–3701.

(5) Harada, R.; Sugita, Y.; Feig, M. Protein crowding affects hydration structure and dynamics. *J. Am. Chem. Soc.* **2012**, *134*, 4842–4849.

(6) Onuchic, J. N.; Wolynes, P. G. Theory of protein folding. *Current Opinion in Structural Biology* **2004**, *14*, 70–75.

(7) Horwich, A. L. Chaperonin-mediated protein folding. *Journal of Biological Chemistry* **2013**, *288*.

(8) Kim, Y. C.; Best, R. B.; Mittal, J. Macromolecular crowding effects on protein-protein binding affinity and specificity. *Journal of Chemical Physics* **2010**, *133*.

(9) Sosnick, T. R.; Hinshaw, J. R. How Proteins Fold. *Science* **2011**, *334*, 464–465.

(10) Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Shaw, D. E. How Fast-Folding Proteins Fold. *Science* **2011**, *334*, 517–520.

(11) Schröder, M.; Schäfer, R.; Friedl, P. Induction of protein aggregation in an early secretory compartment by elevation of expression level. *Biotechnology and bioengineering* **2002**, *78*, 131–140.

(12) Tartaglia, G. G.; Pechmann, S.; Dobson, C. M.; Vendruscolo, M. Life on the edge: a link between gene expression levels and aggregation rates of human proteins. *Trends in Biochemical Sciences* **2007**, *32*, 204–206.

(13) Tartaglia, G. G.; Vendruscolo, M. Correlation between mRNA expression levels and protein aggregation propensities in subcellular localisations. *Molecular BioSystems* **2009**, *5*, 1873.

(14) van den Berg, B.; Ellis, R. J.; Dobson, C. M. Effects of macromolecular crowding on protein folding and aggregation. *The EMBO Journal* **1999**, *18*, 6927–6933.

(15) Feig, M.; Yu, I.; Wang, P. H.; Nawrocki, G.; Sugita, Y. Crowding in Cellular Environments at an Atomistic Level from Computer Simulations. *Journal of Physical Chemistry B* **2017**, *121*, 8009–8025.

(16) Su, A. I.; Wiltshire, T.; Batalov, S.; Lapp, H.; Ching, K. A.; Block, D.; Zhang, J.; Soden, R.; Hayakawa, M.; Kreiman, G. et al. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl. Acad. Sci.* **2004**, *101*, 6062–6067.

(17) Micheletti, C.; Seno, F.; Maritan, A.; Banavar, J. R. Design of proteins with hydrophobic and polar amino acids. *Proteins: Structure, Function, and Genetics* **1998**, *32*, 80–87.

(18) Havranek, J. J.; Harbury, P. B. Automated design of specificity in molecular recognition. *Nature Structural Biology* **2003**, *10*, 45–52.

(19) Sear, R. P. Highly specific proteinprotein interactions, evolution and negative design. *Physical Biology* **2004**, *1*, 166–172.

(20) Berezovsky, I. N.; Zeldovich, K. B.; Shakhnovich, E. I. Positive and Negative Design in Stability and Thermal Adaptation of Natural Proteins. *PLoS Computational Biology* **2007**, *3*, e52.

(21) Noivirt-Brik, O.; Horovitz, A.; Unger, R. Trade-off between positive and negative design of protein stability: From lattice models to real proteins. *PLoS Computational Biology* **2009**, *5*, e1000592.

(22) Kurnik, M.; Hedberg, L.; Danielsson, J.; Oliveberg, M. Folding without charges. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, *109*, 5705–5710.

(23) Bianco, V.; Franzese, G.; Coluzza, I. In silico evidence that protein unfolding is as a precursor of the protein aggregation. *arXiv* **2019**, *1906.05010*.

(24) Franzese, G.; Bianco, V.; Iskrov, S. Water at Interface with Proteins. **2011**, *6*, 186–198.

(25) Bianco, V.; Iskrov, S.; Franzese, G. Understanding the role of hydrogen bonds in water dynamics and protein stability. *J. Biol. Phys.* **2012**, *38*, 27–48.

(26) Franzese, G.; Bianco, V. Water at Biological and Inorganic Interfaces. *Food Biophysics* **2013**, *8*, 153–169.

(27) Bianco, V.; Franzese, G. Contribution of Water to Pressure and Cold Denaturation of Proteins. *Physical Review Letters* **2015**, *115*, 1–12.

(28) Bianco, V.; Franzese, G.; Dellago, C.; Coluzza, I. Role of Water in the Selection of Stable Proteins at Ambient and Extreme Thermodynamic Conditions. *Physical Review X* **2017**, *7*, 021047.

(29) Bianco, V.; Pagès-Gelabert, N.; Coluzza, I.; Franzese, G. How the stability of a folded protein depends on interfacial water properties and residue-residue interactions. *Journal of Molecular Liquids* **2017**, *245*, 129–139.

(30) Stokely, K.; Mazza, M. G.; Stanley, H. E.; Franzese, G. Effect of hydrogen bond cooperativity on the behavior of water. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, *107*, 1301–1306.

(31) Bianco, V.; Franzese, G. Critical behavior of a water monolayer under hydrophobic confinement. *Scientific Reports* **2014**, *4*.

(32) Bianco, V.; Franzese, G. Hydrogen bond correlated percolation in a supercooled water monolayer as a hallmark of the critical region. *Journal of Molecular Liquids* **2019**, *285*, 727 – 739.

(33) Shakhnovich, E.; Gutin, A. A new approach to the design of stable proteins. *"Protein Engineering, Design and Selection"* **1993**, *6*, 793–800.

(34) Cardelli, C.; Bianco, V.; Rovigatti, L.; Nerattini, F.; Tubiana, L.; Dellago, C.; Coluzza, I. The role of directional interactions in the designability of generalized heteropolymers. *Scientific Reports* **2017**, *7*, 4986.

(35) Ross, C. A.; Poirier, M. A. Opinion: What is the role of protein aggregation in neurodegeneration? *Nature Reviews Molecular Cell Biology* **2005**, *6*, 891–898.

(36) Chiti, F.; Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annual Review of Biochemistry* **2006**, *75*, 333–366.

(37) Aguzzi, A.; O'Connor, T. Protein aggregation diseases: pathogenicity and therapeutic perspectives. *Nature Reviews Drug Discovery* **2010**, *9*, 237–248.

(38) Knowles, T. P. J.; Vendruscolo, M.; Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nature Reviews Molecular Cell Biology* **2014**, *15*, 384–396.