Protein structure and aggregation: a marriage of necessity ruled by aggregation gatekeepers

Bert Houben, Frederic Rousseau* and Joost Schymkowitz*

VIB Center for Brain and Disease Research, Leuven, Belgium & Switch Laboratory, Department of Cellular and Molecular Medicine KULeuven, Leuven, Belgium.

* Corresponding authors: Frederic.rousseau@kuleuven.vib.be, Joost.schymkowitz@kuleuven.vib.be

ORCID

JS: 0000-0003-2020-0168
FR: 0000-0002-9189-7399
BH: 0000-0002-6750-011X

Abbreviations

APR: Aggregation-prone region **–** **GK:** aggregation gatekeeper **–** **IDP:** Intrinsically Disordered Protein – **PN:** Proteostasis Network

Keywords

protein structure and stability, protein aggregation, amyloid, proteostasis, kinetic partitioning

Abstract

Protein aggregation propensity is a pervasive and seemingly inescapable property of proteomes. Strikingly, a significant fraction of the proteome is supersaturated, meaning that for these proteins, the native conformation is less stable than the aggregated state. Maintaining the integrity of a proteome under such conditions is precarious and requires energy-consuming proteostatic regulation. Why then is aggregation propensity maintained at such high levels during long evolutionary timescales? We argue that the conformational stability of the native and aggregated states are correlated thermodynamically and that codon usage strengthens this correlation. As a result, the folding of stable proteins requires kinetic control to avoid aggregation, provided by aggregation gatekeepers. These unique residues are evolutionarily selected to kinetically favour native folding, either on their own or by co-opting chaperones.

Protein aggregation propensity is a constant threat to cellular health

The most widely studied aggregation mechanism is the formation of intermolecular β-sheets by short **Aggregation-Prone Regions** (APRs, see Glossary). Prediction software can detect APRs based on their physicochemical properties directly in primary protein sequences (more below). The assembly mechanism of APRs can give rise to highly structured amyloid fibrils or to more amorphous aggregates. The specific outcome of this process depends to a large extent on experimental and/or physiological conditions. As we have pointed out before, many amorphous aggregates still show an enrichment in β-sheet structure and are thus based on the same basic assembly mechanism, although such amorphous structures may also form in other ways [1]. In this Opinion, we focus on the cross-β aggregation mechanism, irrespective of whether it leads to higher order structure such as amyloid and we assume that β-sheet-enriched amorphous aggregates consist of shorter stretches of β-sheets clustered into less defined entities. The combined output of the myriad of computational methods that are currently available for the prediction of APRs in entire proteomes [2] suggests that most likely less than 1% of proteins in any proteome have no APR and are hence unaffected by protein **aggregation propensity** (see Glossary). In fact, on average about 20% of residues in a protein sequence have tendency to misfold into β-structured aggregates [3]. The ensemble of APRs in a polypeptide have been called its **intrinsic aggregation propensity** (see Glossary). The intrinsic aggregation propensity of a protein sequence can further be modulated by other factors, such as conformation, concentration, and environmental conditions to result in its actual aggregation propensity (Box 1) [4-7]. In recent years, we have come to realize that 10% to 30% of proteins are supersaturated under physiological conditions meaning they are expressed at abundances exceeding their intrinsic solubility [8-10]. Hence, a significant amount of metabolic energy has to be invested in proteostatic control ensuring proteins get and remain properly folded [11]. The erosion of this proteostatic control is also why ageing organisms are increasingly at risk of aggregation-associated diseases [12].

In this Opinion article, we discuss how this precarious situation came to be. Firstly, we argue that intrinsic aggregation propensity is directly correlated to globular fold stability. The intricate stereochemical packing of the hydrophobic core required for the thermodynamic stability of the native state severely limits the extent to which protein sequences can be optimized to avoid intrinsic aggregation propensity without critically destabilizing the native fold (Figure 1A). This thermodynamically imposes a solubility limit on most proteins, and causes a large fraction of the proteome to be **metastable** (see Glossary) at physiological concentrations. Secondly, we review mechanisms that help avoid this thermodynamic trap by **kinetically partitioning** (see Glossary) the native fold from the aggregated state [13], allowing proteins to be expressed for biologically relevant timescales at supersaturated concentrations.

Widespread aggregation propensity causes proteome metastability

The classical image of a folded protein is that of a chain of amino acids folded in on itself, forming local secondary structures such as α-helices, β-sheets, and loops, which arrange further into a predefined three-dimensional structure, i.e. the functional form or native fold. It has long been clear that some proteins can also adopt a drastically different structure, known as the aggregated state. In this configuration, hydrophobic segments of the protein with a high propensity for β-sheet formation and low net charge engage in extended intermolecular β-sheets with identical counterparts in a sequence-specific manner [14] (Figure 1B). Multiple sheets can align length-wise and as such extensively interact through the interdigitation of their sidechains, perpendicularly to the β-sheet axis. The resulting "cross-β" conformation is highly stable, both mechanically and physico-chemically. Indeed, the combination of the regular stacking of hydrophobic side chains in subsequent layers in the fibril core, combined with the extensive network of backbone hydrogen bonds connecting the layers renders the mature aggregate highly stable, certainly when compared to the well-documented marginal stability of biologically functional native states, which often require flexibility for function. A more complete discussion on amyloid stability can be found elsewhere [15, 16], but one prominent difference is the fact that the polypeptide fulfils its full backbone hydrogen bonding potential in the amyloid state, where this is not true for globular proteins that typically contain a mix of secondary structure element and loops. However, this should be mitigated by the recent realisation that regions of suboptimal H-bond geometry and hence structural frustration also occur in the β-aggregated state [17].

The amyloid conformation is most well-known as the pathological hallmark of more than 30 degenerative diseases, in which specific proteins adopt this intermolecular β-conformation. As a result, amyloid formation is sometimes considered a rare off-pathway event affecting only a select group of proteins. However, intense research has made clear that most, if not all, proteins carry within them an inherent tendency to form amyloid – in the form of short segments with the right conjunction of physicochemical properties. The most important factors that keep these regions from actually initiating aggregation is native protein folding and the cellular **proteostasis** machinery (see Glossary) [18]. Many proteins are obligate **chaperone** (See Glossary) substrates and aggregate when translated *in vitro* in the absence of these factors [19-22]. Importantly, with some exceptions (discussed below) most chaperones are not classic catalysts that stabilize the transition state of the folding reaction. Rather by binding to APRs they prevent or reverse the interactions of exposed hydrophobic regions thereby inhibiting aggregation [23]. In doing so they not only increase folding yields but can also increase folding rates, which we believe they achieve by destabilizing local conformational minima resulting from erroneous hydrophobic collapse. Indeed, modern proteins are riddled with intrinsic aggregation propensity: over 90 % of globular proteins contain at least one region with a tendency to form β-structured intermolecular aggregates [3] making APRs universal handles to partition aggregation from folding and to regulate hydrophobic collapse. Moreover, aggregated states are generally more stable than their native counterparts, even at common cellular concentrations, effectively making the native fold a metastable conformation that is only kinetically protected from converting to the aggregated state [13, 24]. This situation imposes a thermodynamic limit on the concentration at which most proteins can be functionally expressed, as is seen from the relation between protein aggregation and both mRNA levels [25-27] and cellular protein abundance [28]. The consequences for proteome stability are profound: in recent years, it has become clear that a large portion of the proteome exists in a supersaturated state under physiological conditions, meaning their abundance exceeds their intrinsic solubility, giving rise to a metastable sub-proteome that likely plays an important role in age-related disease [8-10].

Why do high levels of intrinsic aggregation propensity persist during evolution?

Following our recent work [18], we propose that the most straightforward explanation for the evolutionary persistence of protein aggregation propensity is that it is a co-evolutionary side-effect of globular protein structure. Stable globular protein folds require both secondary structure propensity and extensive hydrophobic cores. Furthermore, proteins are synthesised as linear polymers, and globular folds require hydrophobic segments of sufficient length to traverse the core (Figure 1A). Sequence segments where hydrophobicity coincides with high β-sheet propensity have the emergent property of aggregation propensity (they are called Aggregation-Prone Regions or APRs [4, 6, 7, 29, 30]).

Importantly, as long as the protein maintains its native state such APRs cannot engage in alternative interactions, which typically requires at least some degree of unfolding [18, 31]. However, there is a deeper link between native fold stability and aggregation propensity that is becoming clear: a survey of point mutations showed that mutations that decrease aggregation propensity tend to decrease native state stability and *vice versa* [32, 33]. The same concept was explored in a systematic computational mutational analysis of amyloid structures of proteins for which the structure of the native fold was also known: it was again found that mutations that disrupt the amyloid state also tend to decrease the stability of the native fold [34]. Moreover, the aggregation propensity of aggregation-prone segments correlates to their contribution to native state stability: the segments that make up the most stable parts of the native structure tend to have the highest aggregation propensity. Furthermore, aggregation propensity is higher in proteomes of extremophiles, whose proteins by definition require more thermodynamic stability [34]. And finally, at the other end of the spectrum, intrinsically disordered protein (IDP) domains, which are by definition devoid of stable three-dimensional structures, are the sole class of naturally occurring polypeptides that harbour significantly fewer predicted aggregation-prone regions [3, 33]. This suggests that the only evolutionary pathway to lower aggregation propensity is through the loss of globular protein structure. Of importance, there are two main sources of APRs: the majority stem from hydrophobic core formation as discussed above, but the second type finds its origins in functional sites, such as protein-protein interaction sites [35]. Whereas many disordered regions have successfully shed the APRs that arise as a result of globular structure, they do still contain the second class of APRs associated with functional interactions [36]. In IDPs however, these take their own more polar flavour and the aggregation propensity is driven more by β-sheet propensity and less by hydrophobicity with Tyr, Gln and Asn as typical enriched amino acids [37, 38]. The aggregation propensity of these regions is suppressed by being embedded in highly charged sequences that act as so-called **entropic bristles** (see

Glossary)[39], but can still lead to aggregation, often in an age-dependent manner. In fact, several of the most intensely studied amyloid-forming proteins, tau, Aβ, TDP43 and α-synuclein [40, 41], are disordered or have substantial intrinsically disordered regions [42]. Intriguingly, aggregation propensity is conserved even down to the genetic level, as mutations that potentially abrogate amylogenic stretches are often inaccessible through single point mutations as a result of the genetic code [34]. Preservation of amyloid propensity thus appears to be deeply embedded within in the genetic code, most likely as a side effect of favouring the conservation of native protein structure. As a result, it is almost impossible to evolve globular structure without also increasing aggregation propensity: it appears as if globular protein structure is addicted to aggregation propensity and the strongest aggregate-forming sequences are among the most deeply conserved in the core of globular protein structures. These considerations explain why so many proteins end up being supersaturated [8-10].

Kinetic partitioning of globular structure to the rescue of the Anfinsen hypothesis

Anfinsen's famous thermodynamic hypothesis stated that proteins fold spontaneously because the biologically active native state is the point of the lowest energy in the conformational landscape [43]. This concept was already put into perspective by the realisation that many proteins require chaperone intervention in order to fold, but the idea of a supersaturated sub-proteome puts even larger question marks by the Anfinsen postulate [44]. If, as we argue, most proteins are indeed aggregation-prone and thermodynamically fated to form aggregates, how is globular protein folding secured? To a large extent, this appears to be achieved through kinetic partitioning, in which the rate of protein folding is made to exceed that of aggregation at relevant concentrations, and the native state has a sufficiently long lifetime by virtue of a slow unfolding rate, so that even if proteins are destined to aggregate eventually, they are able to adopt and maintain their native fold for a physiologically relevant timespan. In fact, many proteins that are involved in aggregation pathologies have a shorter than average lifetime, suggesting they are protected from aggregation by a fast turnover (i.e. they are degraded before they can aggregate) [45], but this hinges on efficient protein degradation which notoriously declines during ageing [46]. The so-called kinetic partitioning, where the native state is metastably trapped for as long as it is required for function, is achieved through both protein-intrinsic features as well as protein-extrinsic factors (Figure 2).

Protein-intrinsic kinetic partitioning by aggregation gatekeepers

"**Aggregation gatekeepers**" (GKs; see Glossary and Figure 2B) are charged residues and β-structure breakers that directly flank APRs, thereby reducing aggregation propensity [3]. Once again, this underpins the tight link between fold stability and aggregation propensity, and shows that evolution had to stop short of completely abrogating APRs since this would require introduction of charged residues or the disruption of secondary structure in the hydrophobic core of the protein. Instead, GKs are found at the first position where the polypeptide emerges from hydrophobic core, often still at some depth from the protein surface. As a result, 'aggregation gatekeeping' comes at a significant cost to native state stability as each GK on average reduces the thermodynamic stability of the native fold by about 0.5 kcal/mol [47]. Moreover, the conservation of GKs scales to the aggregation propensity of the region they are flanking [48]. Such evolutionary conservation despite a negative effect on protein stability is typically seen in functionally important residues, such as in active sites, leading us to propose that GKs are functional class of residues unto themselves. Apart from their thermodynamic effects, GKs likely also slow down the kinetics of native protein folding, as removing charges altogether is known to increase protein folding rates [49]. However, GKs slow down the aggregation reaction *more* than they slow down native folding, making them a quintessential example of kinetic partitioning to circumvent the constraints that arise from the entanglement between aggregation propensity and fold stability.

Interestingly, we recently demonstrated that even within the class of the charged GK residues, there is an important distinction between positively and negatively charged GKs [50]. The positively charged moieties on Lys and Arg are more readily dehydrated than their negatively charged counterparts Asp and Glu, and they have longer and more hydrophobic sidechains. As a result, positively charged GKs are more readily incorporated into a globular protein but unfortunately are also more compatible with amyloid structure and thus poorer aggregation breakers. In fact, positively charged GKs barely destabilize the aggregated state and only marginally slow down the aggregation process. As a compensatory mechanism, positively charged GKs are specifically recognized and assisted by molecular chaperones, which augment the kinetic partitioning capacity of the gatekeepers [51-60]. The positively charged GKs have therefore been referred to as "non-autonomous". Negatively charged GKs, on the other hand, both strongly disrupt amyloid structure and severely slow down its formation and were therefore termed "autonomous" GKs [50]. However, because of the entanglement between the stability of the native and aggregated states, the negatively charged GKs are less compatible with native protein structure and can hence not always be accommodated.

Molecular chaperones: protein-extrinsic partitioners

Molecular chaperones are a diverse group of major effectors of the Proteostasis Network (PN). Some, like the Proline-Prolyl isomerases catalyse protein folding, while others (such as the Hsp70 and chaperonin family members) prevent aggregation (e.g. the small heat shock proteins by virtue of their holdase activity), disaggregate aggregated species, or direct terminally misfolded proteins towards appropriate degradation pathways [61]. Other chaperones, like the Hsp90 family members, help maintain the integrity of the native state using extensive interaction surfaces. Although molecular chaperones come in many varieties with distinct modes of action, a recurring theme is that they recognize and bind their substrates through exposed hydrophobic regions [56, 62-73]. Not only is exposed hydrophobicity, and by extension APRs, a sign of incomplete folding or misfolding, but such regions are also at risk of engaging in aberrant intermolecular interactions. By engaging their clients, chaperones shield these hydrophobic regions, thereby preventing aggregation. In effect, this mode of action is a form of kinetic partitioning, in which a large energetic barrier is maintained between the native fold and the aggregated state (**Error! Reference source not found.**C). Moreover, the interaction of chaperones with their clients results in an excluded volume, decreasing local protein concentration and favouring the formation of intramolecular interactions over intermolecular ones [74]. Finally, chaperones partially unfold their client proteins, potentially resolving kinetically trapped misfolded states and accelerating the folding process.

In effect, molecular chaperones constitute the ultimate evolutionary measure that maintains modern proteomes in a metastable state in the face of widespread aggregation propensity. It could be argued that chaperones are folding catalysts, in that they are not a part of the final folded protein, and therefore do not affect the thermodynamics of protein folding. However, classic enzyme catalysts generally increase reaction rates by binding to and therefore stabilizing the rate-limiting transition state of a reaction. Except for the Proline-Prolyl isomerases, it is unlikely that most chaperones increase the kinetics of protein folding in the same manner by directly binding to and stabilizing the transition state of folding, given the diversity in protein topology and sequence. By binding to APRs however, we think that chaperones not only prohibit aggregate assembly (thereby increasing the folding yield) but can probably in some cases also increase folding rates by destabilizing the ground state of unfolded and partially (mis)folded conformations. Modifying the folding landscape in this manner equally results in a lower kinetic barrier and hence faster folding. By binding APRs and controlling hydrophobic surfaces chaperones are therefore not only the ultimate kinetic partitioners between native folding and amyloid-like aggregation but also by smoothing out the native folding landscape thereby both improving protein folding yields and rates. As a result, large portions of modern proteomes depend on them for their solubility [75, 76].

As mentioned above, several classes of chaperones specifically prioritise hydrophobic segments when they are flanked by positively charged residues, such as APRs flanked by positive GKs [56, 63-73, 77]. We recently showed that because of this binding preference, chaperones are able to recognize APRs most at risk of aggregating because of poor gatekeeping [50]. This points towards a coevolution between GKs and molecular chaperones, allowing even proteins with insufficiently protected APRs to reach appropriate cellular concentrations through the concerted effects of positively charged GKs and molecular chaperones.

Co-translational folding: temporal partitioning?

For small proteins, folding *in vitro* takes place on the µsecond timescale. Translation of mRNA into protein at the ribosomes however, is a slower process, as the prokaryotic translation machinery produces 15-20 amino acids/s, and eukaryotic ribosomes work even slower, at 1-5 amino acids/s on average [78]. This discrepancy in timing makes it highly likely that substantial protein folding occurs before a protein is fully translated and is therefore still physically attached to the ribosome, which has profound effects on the folding landscape

(Figure 2A and **Error! Reference source not found.**). Indeed, it has become abundantly clear that many proteins do fold co-translationally, and that translational kinetics are optimized for this very purpose: translational pause sites were found to be enriched in interdomain regions over two decades ago, and, more recently, it was observed that pause sites are enriched 20-60 aa downstream of sequence segments predicted to form subdomain co-translational folding intermediates [79]. Simulations confirm that such pause sites allow domains to fold more efficiently by preventing potential non-native interactions with not yet formed C-terminal regions [80]. Aggregation-prone segments on the other hand, tend to be enriched in optimal codons. Although not yet fully understood, this could suggest the necessity for the regions containing APRs to be rapidly translated, allowing for at least partial co-translational folding and descending into the native folding basin before aggregation has a chance to occur. In line with this, protein interaction sites, which are often hydrophobic and tend to contain APRs, are depleted near the N-terminus of proteins [81]. This allows protein domains to progress down the native folding funnel directly upon the emergence of an APR, temporally partitioning native folding from aggregation (**Error! Reference source not found.**). It is tempting to speculate that recent findings regarding the proximal translation of interacting proteins [82] fit into the same framework: since interaction sites often require exposed APRs in the monomeric subunits, coordinating the translation of the interacting proteins so that the time of APR exposure in the subunits is minimized could be another form of kinetic partitioning.

Ribosome association has an added benefit: physical linkage to the large ribosomes creates an excluded volume around the nascent chain, effectively instituting a low local concentration of exposed aggregation-prone regions (which is why the nascent chains being translated by the ribosome are depicted on the “intramolecular” side of the folding landscape in Figure 2). Indeed, it was recently shown that interaction with any soluble protein can indirectly increase folding efficiency by preventing aggregation [74].

These factors make the translation process an effective form of kinetic partitioning. By allowing folding to happen co-translationally, the chances of aberrant interactions and misfolding are reduced, increasing the rate of the native folding reaction. Concurrently, association with the ribosome creates an excluded volume which decreases the rate of intermolecular interactions and hence aggregate formation.

When partitioning fails

The efficacy of the proteostasis network declines with age. This has long been viewed as one of the major reasons why age is the predominant risk factor in many of the neurodegenerative amyloidoses plaguing modern society [12]. Given the proteome metastability discussed above, decreasing proteostasis logically results in aggregation of supersaturated proteins: indeed, proteins known to precipitate in protein misfolding disorders are significantly more supersaturated than the remainder of the proteome, and therefore more dependent on kinetic partitioning to remain soluble [44, 83]. Moreover, it has become abundantly clear that chaperones have an important role in keeping misfolding-disease-associated proteins soluble, and ridding the cell of inadvertently aggregated material [84].

As mentioned above, many of the proteins associated with aggregation-related diseases carry at least some degree of intrinsic disorder. Although intrinsically disordered protein domains have less aggregation propensity overall, the energetic basins associated with their native folds are relatively shallow at best [85, 86]. Lacking thermodynamic stabilization, these proteins are more reliant on kinetic partitioning by external factors (i.e. chaperones and proteases) for their solubility. Indeed, recent work shows how the so-called supersaturation barrier needs to be broken in order to induce aggregation of folded proteins [87], and that this is easier in specific protein types, particularly short peptides and intrinsically disordered ones, both groups of proteins with a shallow native state energetic basin [88]. This might explain why many intrinsically disordered proteins are often stabilized by clusters of (often negative) non-neutralized charges [89]. Indeed, Aβ, α-Synuclein and Tau are all stabilized by highly charged clusters, and the removal or neutralization thereof results in their aggregation [90-93]. Such charge clusters likely constitute a radical form of intrinsic kinetic partitioning, whereby strong charge repulsion prevents amyloid nucleation.

Clearly, some proteins are intrinsically at risk of forming amyloid deposits because of their inherent characteristics and the specific tissues they are expressed in. This situation can be exacerbated by genetic alterations both in these proteins themselves, or in the PN that ensures their kinetic partitioning [94]. Some familial mutations associated with misfolding disorders even cause proteins to escape recognition by the PN, effectively removing kinetic partitioning and leading to aggregation, as is the case for the SOD1 A4V mutant [95].

Concluding remarks

The propensity to misfold and aggregate into amyloid-like assemblies is a universal property of proteins in all kingdoms of life. Protein aggregation is unfavourable, resulting in protein functional dysregulation and disease. Maintaining proteostasis under these conditions requires an extensive protein quality control machinery representing a high metabolic cost. It is therefore remarkable that the protein aggregation propensity of proteomes is maintained at such high levels.

In this opinion piece we discussed how protein aggregation is under continuous selective pressure yet cannot be reduced below the levels observed in proteomes. While protein aggregation decreases the efficiency of protein folding it also favours protein stability. Even more, we argue it is almost impossible to increase the conformational stability of a protein without increasing its aggregation propensity and conversely reducing the aggregation propensity of proteins generally results in protein destabilisation. Remarkably, we found that the entanglement between protein stability and aggregation is further increased by the universal genetic code: protein sequence segments that both strongly contribute to protein stability and have a high aggregation propensity are also strongly conserved appearing as if proteins are addicted to those amyloidogenic sequences.

The global result of coupling between protein stability and aggregation is that the Anfinsen postulate of thermodynamic determination is only a local property of the native folding basin and that globally protein folding requires mechanisms of kinetic control to ensure native protein folding is favoured over aggregation. The presence of such mechanisms also explains why a substantial fraction of proteins are in fact supersaturated under physiological conditions.

Kinetic control of protein folding is enforced in two interdependent ways by gatekeeper residues and chaperones. Hydrophobic aggregation-prone protein segments are flanked by charged residues that function as aggregation gatekeepers: these residues disfavour protein aggregation by electrostatic repulsion, favouring kinetic partitioning towards the native state. Short, negatively charged residues such as Asp and Glu are particularly good at inhibiting aggregation, allowing protein folding without the help of chaperones. Due to their short side chains they are however difficult to incorporate into native protein structures. Positively charged residues Arg and Lys can be used instead but they are less efficient gatekeepers and are incapable of fully inhibiting aggregation. This is compensated by the fact that chaperones evolved to favour binding to aggregation-prone regions that are flanked by positive residues.

The insights outlined in this Opinion of course raise additional question, which have been highlighted in the Open Questions section.

Acknowledgements

The Switch Laboratory was supported by grants from the European Research Council under the European Union's Horizon 2020 Framework Programme ERC Grant agreement 647458 (MANGO) to JS, the Flanders institute for biotechnology (VIB), the University of Leuven (“Industrieel Onderzoeksfonds”), and the Funds for Scientific Research Flanders (FWO (G045920N). BH was supported by a Post-doctoral Mandate (PDM) from KU Leuven.

References

1. Rousseau, F. et al. (2006) Protein aggregation and amyloidosis: confusion of the kinds? Curr Opin Struct Biol 16 (1), 118-26.

2. Santos, J. et al. (2020) Computational prediction of protein aggregation: Advances in proteomics, conformation-specific algorithms and biotechnological applications. Comput Struct Biotechnol J 18, 1403-1413.

3. Rousseau, F. et al. (2006) How evolutionary pressure against protein aggregation shaped chaperone specificity. J Mol Biol 355, 1037-1047.

4. Pawar, A.P. et al. (2005) Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. J Mol Biol 350 (2), 379-92.

5. Chiti, F. et al. (2003) Rationalization of the effects of mutations on peptide and protein aggregation rates. Nature 424 (6950), 805-8.

6. Fernandez-Escamilla, A.M. et al. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins. Nature Biotechnology 22 (10), 1302-1306.

7. Sanchez de Groot, N. et al. (2005) Prediction of "hot spots" of aggregation in disease-linked polypeptides. BMC Struct Biol 5, 18.

8. Vecchi, G. et al. (2020) Proteome-wide observation of the phenomenon of life on the edge of solubility. Proc Natl Acad Sci U S A 117 (2), 1015-1020.

9. Ciryam, P. et al. (2015) Supersaturation is a major driving force for protein aggregation in neurodegenerative diseases. Trends Pharmacol Sci 36 (2), 72-77.

10. Ciryam, P. et al. (2019) A metastable subproteome underlies inclusion formation in muscle proteinopathies. Acta Neuropathologica Communications 7 (1).

11. Swovick, K. et al. (2021) Interspecies Differences in Proteome Turnover Kinetics Are Correlated With Life Spans and Energetic Demands. Molecular & Cellular Proteomics 20, 100041.

12. Hipp, M.S. et al. (2019) The proteostasis network and its decline in ageing. Nature Reviews Molecular Cell Biology 20 (7), 421-435.

13. Chiti, F. et al. (2002) Kinetic partitioning of protein folding and aggregation. Nat Struct Biol 9 (2), 137-43.

14. Gallardo, R. et al. (2020) Amyloid structures: much more than just a cross-β fold. Curr Opin Struct Biol 60, 7-16.

15. Makin, O.S. et al. (2005) Molecular basis for amyloid fibril formation and stability. Proc Natl Acad Sci U S A 102 (2), 315-20.

16. Fitzpatrick, A.W. et al. (2011) Inversion of the Balance between Hydrophobic and Hydrogen Bonding Interactions in Protein Folding and Aggregation. Plos Computational Biology 7 (10).

17. van der Kant, R.a.L., Nikolaos and Schymkowitz, Joost and Rousseau, Frederic (2021) A structural analysis of amyloid polymorphism in disease: clues for selective vulnerability? bioRxiv.

18. Langenberg, T. et al. (2020) Thermodynamic and Evolutionary Coupling between the Native and Amyloid State of Globular Proteins. Cell Rep 31 (2), 107512.

19. Ramakrishnan, R. et al. (2020) Protein Homeostasis Database: protein quality control in E.coli. Bioinformatics 36 (3), 948-949.

20. Ramakrishnan, R. et al. (2019) Differential proteostatic regulation of insoluble and abundant proteins. Bioinformatics.

21. Niwa, T. et al. (2012) Global analysis of chaperone effects using a reconstituted cell-free translation system. Proc Natl Acad Sci U S A 109 (23), 8937-42.

22. Fujiwara, K. et al. (2010) A systematic survey of in vivo obligate chaperonin-dependent substrates. EMBO J 29 (9), 1552-64.

23. Priya, S. et al. (2013) Molecular chaperones as enzymes that catalytically unfold misfolded polypeptides. FEBS Lett 587 (13), 1981-7.

24. Baldwin, A.J. et al. (2011) Metastability of Native Proteins and the Phenomenon of Amyloid Formation. J Am Chem Soc 133 (36), 14160-14163.

25. Ganesan, A. et al. (2016) Structural hot spots for the solubility of globular proteins. Nat Commun 7, 10816.

26. Tartaglia, G.G. and Vendruscolo, M. (2009) Correlation between mRNA expression levels and protein aggregation propensities in subcellular localisations. Mol Biosyst 5 (12), 1873-6.

27. Tartaglia, G.G. et al. (2009) A relationship between mRNA expression levels and protein solubility in E. coli. J Mol Biol 388 (2), 381-9.

28. Castillo, V. et al. (2011) The aggregation properties of Escherichia coli proteins associated with their cellular abundance. Biotechnol J 6 (6), 752-60.

29. Linding, R. et al. (2004) A comparative study of the relationship between protein structure and beta-aggregation in globular and intrinsically disordered proteins. J Mol Biol 342 (1), 345-53.

30. De Baets, G. et al. (2014) Predicting aggregation-prone sequences in proteins. Essays Biochem 56, 41-52.

31. Dobson, C.M. (2001) The structural basis of protein folding and its links with human disease. Philos Trans R Soc Lond B Biol Sci 356 (1406), 133-45.

32. Sanchez, I.E. et al. (2006) Point mutations in protein globular domains: contributions from function, stability and misfolding. J Mol Biol 363 (2), 422-32.

33. Linding, R. et al. (2004) A Comparative Study of the Relationship Between Protein Structure and β-Aggregation in Globular and Intrinsically Disordered Proteins. J Mol Biol 342, 345-353.

34. Langenberg, T. et al. (2020) Thermodynamic and Evolutionary Coupling between the Native and Amyloid State of Globular Proteins. Cell Reports 31 (2), 107512.

35. Castillo, V. and Ventura, S. (2009) Amyloidogenic Regions and Interaction Surfaces Overlap in Globular Proteins Related to Conformational Diseases. Plos Computational Biology 5 (8).

36. Ali, M. et al. (2020) Screening Intrinsically Disordered Regions for Short Linear Binding Motifs. In Intrinsically Disordered Proteins: Methods and Protocols (Kragelund, B.B. and Skriver, K. eds), pp. 529-552, Springer US.

37. Maurer-Stroh, S. et al. (2010) Exploring the sequence determinants of amyloid structure using position-specific scoring matrices. Nature Methods 7 (3), 237-U109.

38. Louros, N. et al. (2020) Structure-based machine-guided mapping of amyloid sequence space reveals uncharted sequence clusters with higher solubilities. Nat Commun 11 (1), 3314.

39. Santner, A.A. et al. (2012) Sweeping away protein aggregation with entropic bristles: intrinsically disordered protein fusions enhance soluble expression. Biochemistry 51 (37), 7250-62.

40. Nguyen, P.H. and Derreumaux, P. (2020) Structures of the intrinsically disordered Abeta, tau and alpha-synuclein proteins in aqueous solution from computer simulations. Biophys Chem 264, 106421.

41. Lim, L. et al. (2016) ALS-Causing Mutations Significantly Perturb the Self-Assembly and Interaction with Nucleic Acid of the Intrinsically Disordered Prion-Like Domain of TDP-43. PLoS Biol 14 (1), e1002338.

42. Das, S. and Mukhopadhyay, D. (2011) Intrinsically unstructured proteins and neurodegenerative diseases: Conformational promiscuity at its best. IUBMB Life 63 (7), 478-488.

43. Anfinsen, C.B. (1973) Principles that Govern the Folding of Protein Chains. Science 181, 223-230.

44. Ciryam, P. et al. (2013) Widespread aggregation and neurodegenerative diseases are associated with supersaturated proteins. Cell Rep 5 (3), 781-90.

45. De Baets, G. et al. (2011) An Evolutionary Trade-Off between Protein Turnover Rate and Protein Aggregation Favors a Higher Aggregation Propensity in Fast Degrading Proteins. Plos Computational Biology 7 (6).

46. Sun-Wang, J.L. et al. (2020) The dialogue between the ubiquitin-proteasome system and autophagy: Implications in ageing. Ageing Res Rev 64, 101203.

47. De Baets, G. et al. (2014) A genome-wide sequence-structure analysis suggests aggregation gatekeepers constitute an evolutionary constrained functional class. J Mol Biol 426, 2405-12.

48. De Baets, G. et al. (2014) A Genome-Wide Sequence-Structure Analysis Suggests Aggregation Gatekeepers Constitute an Evolutionary Constrained Functional Class. J Mol Biol 426 (12), 2405-2412.

49. Kurnik, M. et al. (2012) Folding without charges. 109 (15), 5705-5710.

50. Houben, B. et al. (2020) Autonomous aggregation suppression by acidic residues explains why chaperones favour basic residues. EMBO J, e102864.

51. Rudiger, S. et al. (1997) Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. Embo Journal 16 (7), 1501-1507.

52. Van Durme, J. et al. (2009) Accurate prediction of DnaK-peptide binding via homology modelling and experimental data. PLoS Comput Biol 5 (8), e1000475.

53. Deuerling, E. et al. (2003) Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. Mol Microbiol 47 (5), 1317-28.

54. Rudiger, S. et al. (2001) Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone. EMBO J 20 (5), 1042-50.

55. Patzelt, H. et al. (2001) Binding specificity of Escherichia coli trigger factor. Proc Natl Acad Sci U S A 98 (25), 14244-9.

56. Schlieker, C. et al. (2004) Substrate recognition by the AAA+ chaperone ClpB. Nat Struct Mol Biol 11 (7), 607-15.

57. Doring, K. et al. (2017) Profiling Ssb-Nascent Chain Interactions Reveals Principles of Hsp70-Assisted Folding. Cell 170 (2), 298-+.

58. Karagoz, G.E. et al. (2017) An unfolded protein-induced conformational switch activates mammalian IRE1. Elife 6.

59. Flynn, G.C. et al. (1991) Peptide-binding specificity of the molecular chaperone BiP. Nature 353 (6346), 726-30.

60. Fourie, A.M. et al. (1994) Common and divergent peptide binding specificities of hsp70 molecular chaperones. J Biol Chem 269 (48), 30470-8.

61. Jayaraj, G.G. et al. (2020) Functional Modules of the Proteostasis Network. Cold Spring Harbor Perspectives in Biology 12 (1), a033951.

62. Bose, D. and Chakrabarti, A. (2017) Substrate specificity in the context of molecular chaperones. IUBMB Life 69, 647-659.

63. Rudiger, S. (1997) Substrate specificity of the DnaK chaperone determined by screening cellulose-bound peptide libraries. 16 (7), 1501-1507.

64. Van Durme, J. et al. (2009) Accurate prediction of DnaK-peptide binding via homology modelling and experimental data. PLoS Comp Biol 5, e1000475.

65. Deuerling, E. et al. (2003) Trigger Factor and DnaK possess overlapping substrate pools and binding specificities. Mol Microbiol 47 (5), 1317-1328.

66. Knoblauch, N.T.M. et al. (1999) Substrate Specificity of the SecB Chaperone. J Biol Chem 274 (48), 34219-34225.

67. Patzelt, H. et al. (2001) Binding specificity of Escherichia coli trigger factor. Proc Natl Acad Sci U S A 98, 14244-9.

68. de Crouy-Chanel, A. et al. (1996) Specificity of DnaK for Arginine/Lysine and Effect of DnaJ on the Amino Acid Specificity of DnaK. J Biol Chem 271 (26), 15486-15490.

69. Döring, K. et al. (2017) Profiling Ssb-Nascent Chain Interactions Reveals Principles of Hsp70-Assisted Folding. Cell 170 (2), 298-311.e20.

70. Karagöz, G.E. et al. (2017) An unfolded protein-induced conformational switch activates mammalian IRE1. eLife 6, e30700.

71. Karagöz, G.E. and Rüdiger, S.G.D. (2015) Hsp90 interaction with clients. Trends Biochem Sci 40, 117-125.

72. Flynn, G.C. et al. (1991) Peptide-binding specificity of the molecular chaperone BiP. Nature 353, 726-730.

73. Fourie, A.M. et al. (1994) Common and divergent peptide binding specificities of hsp70 molecular chaperones. J Biol Chem 269, 30470-30478.

74. Kwon, S.B. et al. (2019) Conversion of a soluble protein into a potent chaperone in vivo. Scientific Reports 9 (1).

75. Ramakrishnan, R. et al. (2019) Differential proteostatic regulation of insoluble and abundant proteins. Bioinformatics 35 (20), 4098-4107.

76. Ryu, S.W. et al. (2020) Proteome-wide identification of HSP70/HSC70 chaperone clients in human cells. PLoS Biol 18 (7), e3000606.

77. Rüdiger, S. et al. (2001) Its substrate specificity characterizes the DnaJ co-chaperone as a scanning factor for the DnaK chaperone. The EMBO journal 20, 1042-50.

78. Zhang, G. and Ignatova, Z. (2011) Folding at the birth of the nascent chain: coordinating translation with co-translational folding. Curr Opin Struct Biol 21 (1), 25-31.

79. Jacobs, W.M. and Shakhnovich, E.I. (2017) Evidence of evolutionary selection for cotranslational folding. Proc Natl Acad Sci U S A 114 (43), 11434-11439.

80. Bitran, A. et al. (2020) Cotranslational folding allows misfolding-prone proteins to circumvent deep kinetic traps. Proceedings of the National Academy of Sciences 117 (3), 1485-1495.

81. Natan, E. et al. (2018) Cotranslational protein assembly imposes evolutionary constraints on homomeric proteins. Nat Struct Mol Biol 25 (3), 279-288.

82. Bertolini, M. et al. (2021) Interactions between nascent proteins translated by adjacent ribosomes drive homomer assembly. Science 371 (6524), 57-64.

83. Ciryam, P. et al. (2019) A metastable subproteome underlies inclusion formation in muscle proteinopathies. Acta Neuropathol Commun 7 (1), 197.

84. Ciechanover, A. and Kwon, Y.T. (2017) Protein Quality Control by Molecular Chaperones in Neurodegeneration. Front Neurosci 11, 185.

85. Chong, S.H. and Ham, S. (2019) Folding Free Energy Landscape of Ordered and Intrinsically Disordered Proteins. Sci Rep 9 (1), 14927.

86. Yang, F. et al. (2018) The Cost of Long Catalytic Loops in Folding and Stability of the ALS-Associated Protein SOD1. J Am Chem Soc 140 (48), 16570-16579.

87. Noji, M. et al. (2019) Heating during agitation of β2-microglobulin reveals that supersaturation breakdown is required for amyloid fibril formation at neutral pH. J Biol Chem 294 (43), 15826-15835.

88. Noji, M. et al. (2021) Breakdown of supersaturation barrier links protein folding to amyloid formation. Communications Biology 4 (1).

89. Uversky, V.N. (2019) Intrinsically Disordered Proteins and Their “Mysterious” (Meta)Physics. Frontiers in Physics 7.

90. Levitan, K. et al. (2011) Conserved C-Terminal Charge Exerts a Profound Influence on the Aggregation Rate of α-Synuclein. J Mol Biol 411 (2), 329-333.

91. Lin, T.-W. et al. (2017) Alzheimer’s amyloid-β A2T variant and its N-terminal peptides inhibit amyloid-β fibrillization and rescue the induced cytotoxicity. PLOS ONE 12 (3), e0174561.

92. Seuma, M. et al. (2021) The genetic landscape for amyloid beta fibril nucleation accurately discriminates familial Alzheimer's disease mutations. Elife 10.

93. Gu, J. et al. (2020) Truncation of Tau selectively facilitates its pathological activities. J Biol Chem 295 (40), 13812-13828.

94. Hohn, A. et al. (2020) Proteostasis Failure in Neurodegenerative Diseases: Focus on Oxidative Stress. Oxid Med Cell Longev 2020, 5497046.

95. Claes, F. et al. (2020) Exposure of a cryptic Hsp70 binding site determines the cytotoxicity of the ALS-associated SOD1-mutant A4V. Protein Engineering, Design and Selection.

96. Sawaya, M.R. et al. (2007) Atomic structures of amyloid cross-beta spines reveal varied steric zippers. Nature 447 (7143), 453-7.

97. Marshall, K.E. et al. (2016) A critical role for the self-assembly of Amyloid-beta1-42 in neurodegeneration. Sci Rep 6, 30182.

98. von Bergen, M. et al. (2005) Tau aggregation is driven by a transition from random coil to beta sheet structure. Biochim Biophys Acta 1739 (2-3), 158-66.

99. Teng, P.K. and Eisenberg, D. (2009) Short protein segments can drive a non-fibrillizing protein into the amyloid state. Protein Engineering Design & Selection 22 (8), 531-536.

100. Ventura, S. et al. (2004) Short amino acid stretches can mediate amyloid formation in globular proteins: the Src homology 3 (SH3) case. Proc Natl Acad Sci U S A 101 (19), 7258-63.

101. Fitzpatrick, A.W.P. et al. (2013) Atomic structure and hierarchical assembly of a cross-β amyloid fibril. Proceedings of the National Academy of Sciences 110 (14), 5468-5473.

Glossary

- **Amyloid**: a type of cross-β protein aggregation typified by a highly structured, elongated, fibrous nature.
- **Aggregation**: the coagulation of proteins into mostly dysfunctional conglomerates. This process is mainly driven by Aggregation-Prone Regions that engage in intermolecular interactions in a sequence-specific manner.
- **Aggregation-prone region (APR)**: short (5-15 amino acids) stretches of “sticky”, usually hydrophobic amino acids with a strong tendency to form homomeric intermolecular β-sheets, thereby driving protein aggregation.
- **Intrinsic Aggregation Propensity:** The inherent propensity of a polypeptide to form aggregates irrespective of folding or external factors (e.g. at elevated temperatures). It has been shown that this is directly determined by the presence of APRs in the primary sequence.
- **Aggregation gatekeeper (GK):** Aggregation-inhibiting residues that directly flank aggregation-prone regions, thereby reducing aggregation tendency. The most common GK types are the charged residues and Pro, and they function through charge repulsion and/or an incompatibility with β-structure.
- **Chaperone**: A class of proteins dedicated to catalyzing folding, translocation and assembly of their substrate proteins. Chaperones are vital parts of the proteostasis network.
- **Proteostasis**: short for protein homeostasis. This term encompasses all cellular factors and processes that maintain proteins in the proper functional states necessary for cellular health. The proteostasis network encompasses the translation machinery, molecular chaperones and degradation pathways.
- **Aggregation Propensity:** The actual aggregation propensity of a protein is determined by the balance between its intrinsic aggregation propensity, its conformational stability and external factors, such as solution conditions, concentration and the state of the proteostasis network.
- **Supersaturation:** Many proteins solubly accumulate at levels above their intrinsic solubility, and these proteins are hence said to be supersaturated. Supersaturation is a metastable state maintained through kinetic partitioning. so that in contradiction to the Anfinsen postulate, the thermodynamically most stable state of these protein is not their biologically active native state, but their aggregated state.
- **Proteome metastability:** Denotes the fact that many proteins in any given cell are supersatured.
- **Kinetic Partitioning:** Denotes the fact that for supersaturated proteins, for which the native state is metastable, the lifetime of the native state is determined by the kinetic barriers separating that state from the unfolded and aggregated states. The higher the energy barrier for unfolding and aggregation, the better the kinetic entrapment of the native state. Molecular chaperones are extrinsic factors acting directly on this, whereas aggregation gatekeepers are a protein-intrinsic factor that shapes kinetic partitioning.
- **Entropic bristle:** Intrinsically disordered regions are enriched in charged residues and the disordered chain is highly flexible, creating large excluded volume effect for intermolecular interactions, as well as a high degree of solvent interactions. This has a strong solubilizing effect on sequences fused to these ‘entropic bristles’ [39].

Elements

Text Box 1: The difference between Intrinsic and Actual Aggregation Propensity

In order to be able to form amyloid-like aggregates, proteins require short polypeptide segments capable of nucleating the formation of intermolecular β-sheet structures, called APRs. Despite the fact that the amyloid fibrils of full-length proteins contain much longer stretches of the sequence in the amyloid conformation, the importance of APRs for the formation of amyloids is beyond doubt:

- Isolated as peptides, these regions are capable of independently forming amyloid-like aggregates with similar properties as those formed by the full-length proteins [96].
- Mutational suppression or deletion of these regions strongly reduces the aggregation propensity of a protein [97, 98].
- Grafting of an APR from one protein to another is sufficient to render the chimera aggregation-prone [99, 100].
- Computational analysis of the architecture of the high resolution structures of amyloid fibrils of full length proteins shows the APRs to be the most stable regions in the amyloid, acting as a framework that compensates for the poor fit of the rest of the sequence [17].

APRs can be distinguished from non-aggregation-prone sequences through the physicochemical properties of their constituting residues. These properties are mainly high β-sheet propensity and hydrophobicity and low net charge. As these properties are readily quantifiable, it is possible to computationally identify APRs based on primary sequence alone [30]. The ensemble of APRs in a polypeptide has been called its 'intrinsic aggregation propensity'. The intrinsic aggregation propensity of a protein sequence can further be modulated by environmental conditions, concentration and, importantly, the conformational landscape of the protein, resulting in an 'actual aggregation propensity': Protein conformations that bury the APRs away from the solvent (folding, binding) are aggregation-resistant, whereas those that expose APRs are aggregation-prone. It is for this reason that destabilisation of the native state of a protein by e.g. mutation or heat exposure increases its aggregation propensity: the APRs are neatly buried inside the hydrophobic core of the native state of the protein, rendering it aggregation-resistant, but in the (partially) denatured state, the APRs come to the surface and start the aggregation process. The equivalent for intrinsically disordered protein is the entropic bristle effect of the rest of the sequence. This can potentially lead to confusion, as a protein with a high intrinsic aggregation propensity that buries its APRs because of its high conformational stability or tight binding to an interaction partner, may be aggregation-resistant in conditions where a protein with much lower intrinsic aggregation propensity may aggregate due to the absence of such protective interactions.

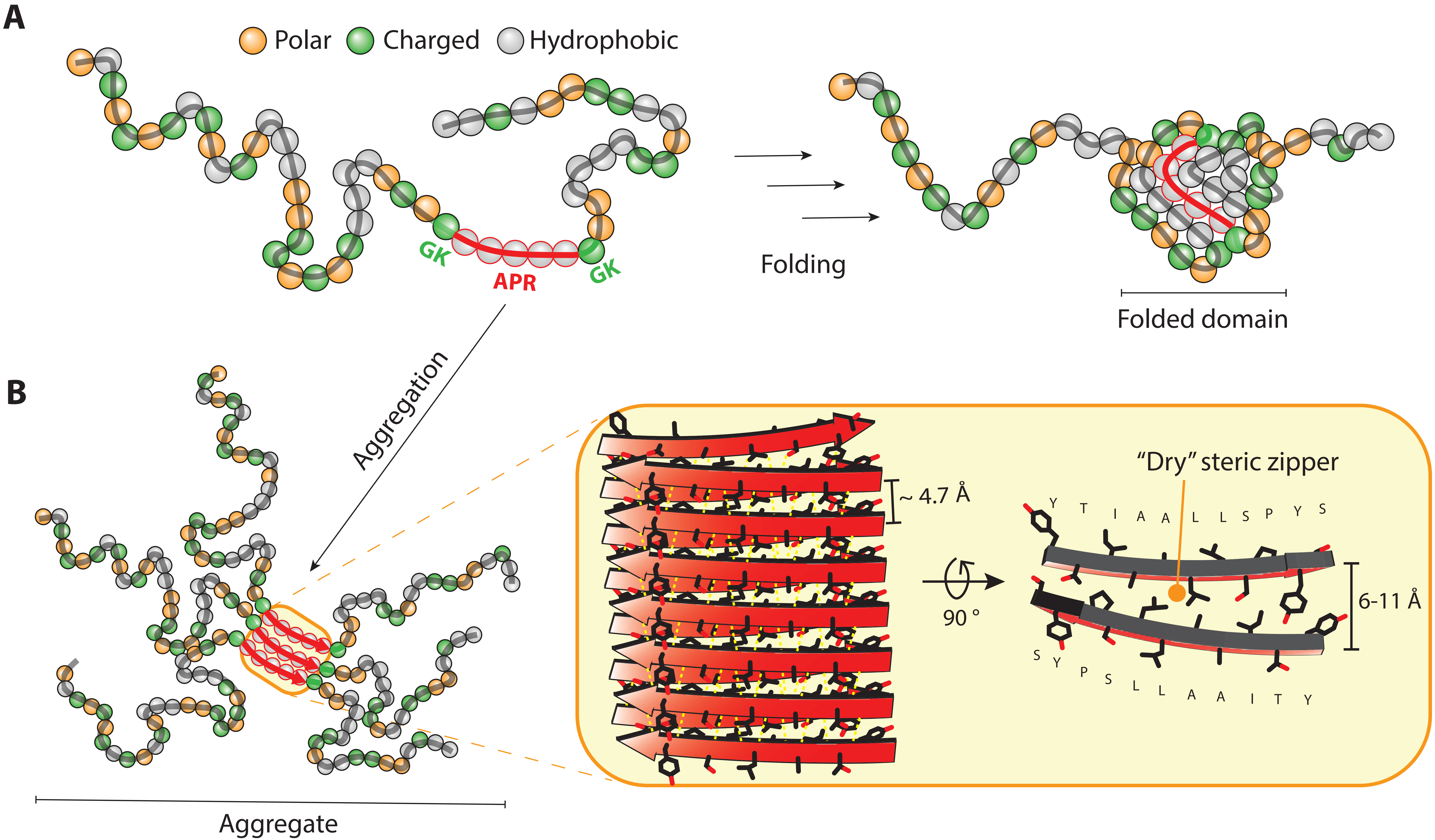
Figure captions

Figure 1: Aggregation propensity is a consequence of the dependence of protein core stability on core-spanning hydrophobic stretches. (A) Proteins are linear concatenations of amino acids that must adopt a predefined shape in order to be functional. In the aqueous cellular environment, this folding process is driven by the tight packing of hydrophobic amino acids into protein cores. To achieve this, proteins contain extended stretches of hydrophobic amino acids capable of spanning the hydrophobic core. Some of these stretches have a tendency to adopt non-native, intermolecular β-sheet conformations, causing their parent proteins to aggregate. These stretches are commonly known as Aggregation-Prone Regions (APRs). APRs are systematically flanked by Gatekeeper residues (GKs), charged residues and β-breakers which slow down the aggregation process while leaving core-spanning hydrophobic stretches intact. **(B)** Protein aggregates share a common core structure, comprised of elongated intermolecular β-sheets with interdigitating sidechains that form an expansive hydrophobic core known as a “dry steric zipper”. The structure is further stabilized by the precise stacking of amino acids in consecutive β-strands, making this process highly sequence-specific. Its repetitiveness gives this structure its typical “cross-β” X-ray diffraction pattern. Amyloid structure depicted here is based on PDB structure 2M5N [101].

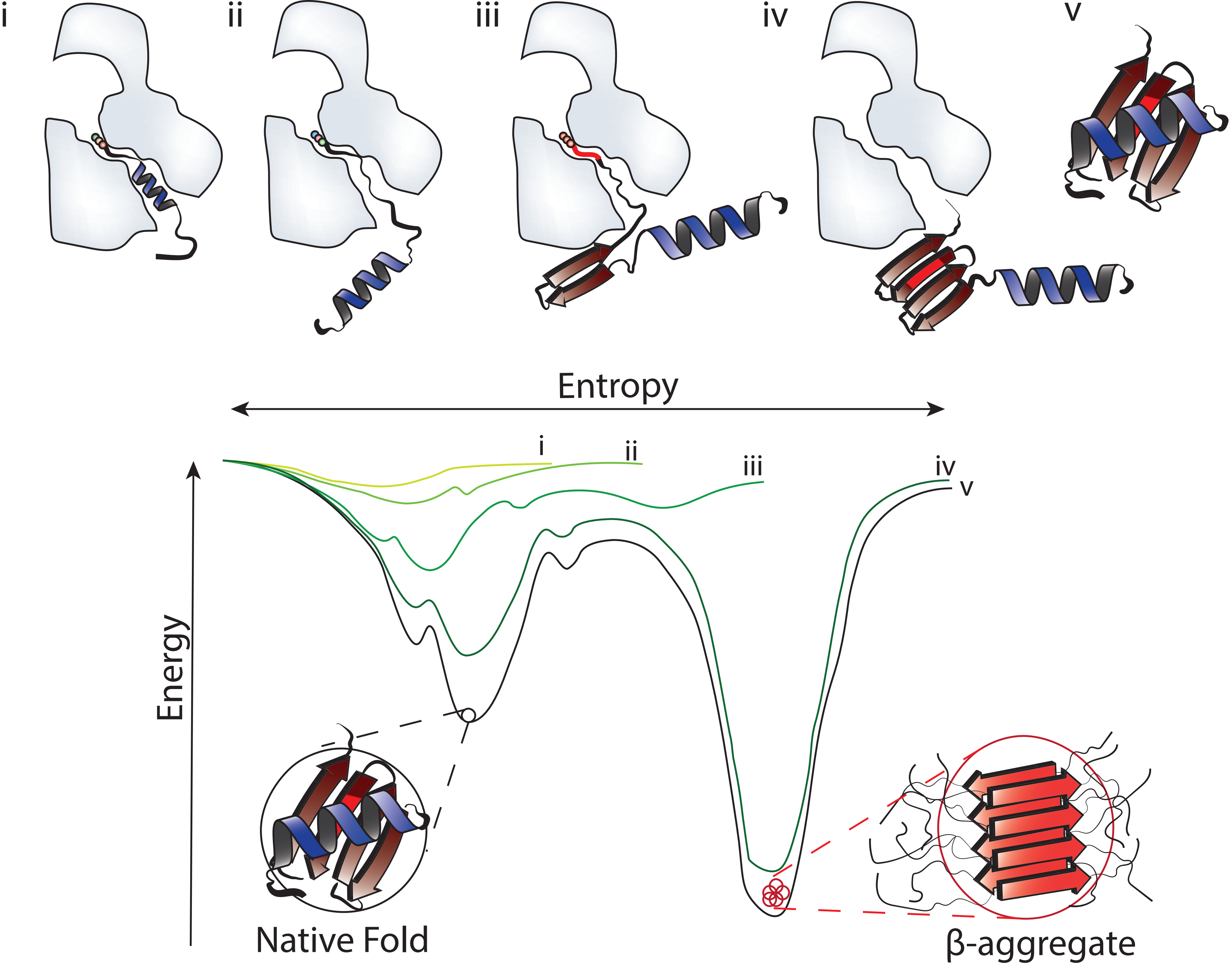
Figure 2: Kinetic partitioning by GKs and molecular chaperones allows supersaturated proteins to fold. (A) 2D representation of a generic folding landscape. Each point on the funnel-shaped surface represents a specific conformation, the energy of which is represented by the landscape depth, while the width of the funnel represents the entropy – i.e. the number of possible conformations – at each energetic level. The folding landscape for globular proteins is typically dominated by two separate basins: a native fold basin that can be navigated down by individual molecules through intramolecular interactions (indicated by the green shaded area), and an aggregation basin, in which multiple molecules engage in intermolecular interactions (indicated by the red shaded area) through their Aggregation-Prone regions (APRs; red stretches in both the native fold and β-aggregated state). Since their association with the ribosome places nascent chains in an excluded volume with a low local protein concentration, the ribosome is depicted towards the intramolecular end of the landscape. This is in fact a method of temporal kinetic partitioning discussed in more detail in the main text and in Figure 3. For many globular proteins at their physiological expression levels, the thermodynamic stability of the aggregated state exceeds that of the native fold, creating a deeper and virtually inescapable basin. Potential pathways proteins can take down the folding landscape are indicated by arrows. Green arrows indicate folding reactions in diluted conditions, while red arrows indicate folding in (super)saturated conditions. In the

latter case, proteins are more likely to engage in APR-driven intermolecular interactions, causing them to descend the aggregate basin. The chain-linked weight in the folding landscape indicates the link between native fold stability and aggregation propensity, both of which are stabilized by APRs and therefore interdependent. **(B)** Given its stability, the aggregated state is thermodynamically favoured, especially at concentrations close to or exceeding the critical concentration. The only way for proteins to be stably expressed at such concentrations, is therefore to kinetically separate the native and aggregation basins by way of an energetic barrier. This is partially achieved through Gatekeepers (GKs; indicated in green), charged residues and β-breakers that directly flank APRs. GKs decrease the APR burden, thereby destabilizing the aggregated state, but also the native fold (as indicated by their increased energies). However, GKs also increase the energetic barrier between folding and aggregation, slowing down the latter process and favouring the native folding reaction. In doing so, GKs allow for higher concentrations of proteins to be stably expressed, at least temporarily. **(C)** Another powerful method of kinetic partitioning is interaction with molecular chaperones. These engage APRs or even entire proteins, creating a huge energetic barrier as these contacts would need to be broken for aggregation to ensue. Most chaperones consume ATP and hence cellular energy for their functional cycle. For most chaperones, ATP consumption results in substrate binding-release cycles, each cycle giving proteins another chance at obtaining the native fold. As is the case for ribosome attachment, chaperone binding and release results in an excluded volume, in which chances for a protein to descend the native funnel are increased.

Figure 3: Co-translational folding temporally partitions the native from the aggregated state. The rates of protein translation are slow enough to allow for co-translational folding to occur. This means the folding landscape, i.e. the conformations available to a nascent protein chain, expands as translation progresses, depicted here for 5 distinct timepoints i – v. In the initial stages of translation, the folding landscape is rather shallow because of the limited number of native stabilizing interactions available. As translation progresses, the landscape deepens, and co-translational folding allows proteins to start descending the native funnel, before APR-driven intermolecular interactions become available (stages i – iii). Placement of an Aggregation-Prone Region (APR; indicated in red) towards the C-terminus of an emerging domain means it can be instantly buried upon its emergence from the ribosome (stages iv and v). In this way the kinetics of translation combined with proper placement of APRs can effectively partition protein folding from aggregation and increase the probability of the former, even though the latter is thermodynamically favored.


A
Polar
Charged
Hydrophobic
GK
APR
GK
Folding
Folded domain
Aggregation
B
Aggregate
~ 4.7 Å
90 °
"Dry" steric zipper
Y T I A A L L S P Y S
S Y P S L L A A I T Y
6-11 Å

A
Intramolecular
Intermolecular
Ribosome
Energy
Native Fold
APRs
β-aggregate
B
Intramolecular
Intermolecular
Energy
GKs
GKs
Native Fold
APRs
β-aggregate
C
Intramolecular
Intermolecular
Chaperones
ATP
Energy
Native Fold
APRs
β-aggregate


i
ii
iii
iv
v
Entropy
Energy
i
ii
iii
iv
v
Native Fold
β-aggregate