

THINK AGAIN

Insights & Perspectives

The seesaw between normal function and protein aggregation: How functional interactions may increase protein solubility

Piero Andrea Temussi¹ | Gian Gaetano Tartaglia^{2,3} | Annalisa Pastore¹ 

¹ UK Dementia Research Institute at King's College London, The Maurice Wohl Institute, London, UK

² Center for Human Technologies, Central RNA laboratory, Istituto Italiano di Tecnologia, Genova, Italy

³ Charles Darwin Department of Biology and Biotechnology, Sapienza University of Rome, Rome, Italy

Correspondence

Annalisa Pastore and Piero Andrea Temussi, UK Dementia Research Institute at King's College London, The Maurice Wohl Institute, 5 Cutcombe Rd, London SE5 9RT, UK.
Email: annalisa.1.pastore@kcl.ac.uk and temussi@unina.it

Abstract

Protein aggregation has been studied for at least 3 decades, and many of the principles that regulate this event are relatively well understood. Here, however, we present a different perspective to explain why proteins aggregate: we argue that aggregation may occur as a side-effect of the lack of one or more natural partners that, under physiologic conditions, would act as chaperones. This would explain why the same surfaces that have evolved for functional purposes are also those that favour aggregation. In the course of reviewing this field, we substantiate our hypothesis with three paradigmatic examples that argue for the generality of our proposal. An obvious corollary of this hypothesis is, of course, that targeting the physiological partners of a protein could be the most direct and specific approach to designing anti-aggregation molecules. Our analysis may thus inform a different strategy for combating diseases of protein aggregation and misfolding.

KEYWORDS

misfolding diseases, normal-to-aberrant function relationship, protein aggregation

INTRODUCTION: WHY DO PROTEINS AGGREGATE?

It is widespread knowledge that protein aggregation and misfolding are general properties of polypeptide chains, and that protein aggregates share common structural features, often resulting in fibrillar species widely known as amyloid aggregates.^[1] Amyloids are characterised by increased resistance to hydrolysis and a β -rich conformation, the amyloid structure, in which the sheets run perpendicularly to the main axis of the fibre (cross- β assembly). Since aggregate deposition and amyloid fibres have been associated with a number of quite different degenerative diseases,^[2] this observation led to the hypothesis that aggregates may be the direct cause of toxicity and pathology through a gain-of-function mechanism. An alternative hypothesis is that amyloid fibres may simply constitute a defensive mechanism to protect the organism from the presence of soluble species that could be the real cause of toxicity and damage. In both views, pathology would result from species

formed along with aggregation^[3-5]. This suggests that the best strategy to prevent aggregation is to design molecules able to trap species formed along the aggregation pathway. Unfortunately, the strategy has so far been unfruitful and it is now recognised that aggregation causes at least loss-of-function.

While many aspects of amyloid formation have been analysed in the finest details and much is known about the final aggregates and the various species along the aggregation pathway, it is remarkable that, after almost 30 years of protein aggregation studies, we are still unable to prevent or cure misfolding diseases. We know that several factors contribute to aggregation: destabilising conditions, such as unbalanced overproduction of a protein, lack of necessary chaperones, and mutations, are all important elements that will eventually lead to the same result.^[6,7]

Yet, there is another very important, if not sometimes determinant element, which has been almost completely understated: molecular

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Authors. *BioEssays* published by Wiley Periodicals LLC

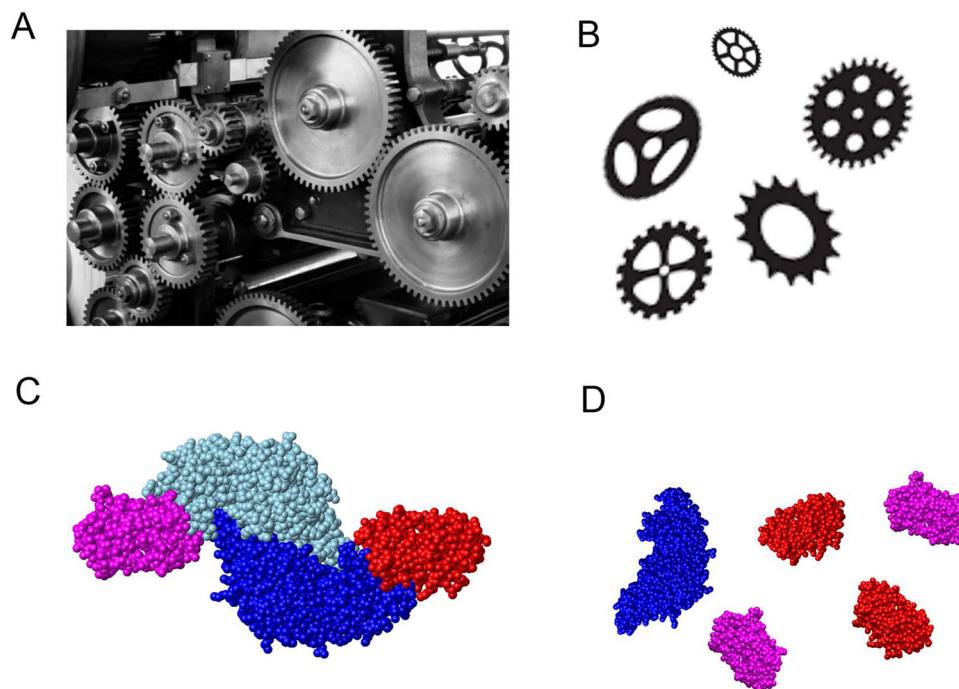


FIGURE 1 Proteins are similar to gear toothed wheels. They must work in groups to carry out their functions. When in isolation their very meaning is lost. (A) A group of wheels that, by turning together, allow the movement of the whole mechanism. (B) Isolated (disconnected), toothed wheels. (C) The ternary complex of two components of iron-sulphur cluster biogenesis from bacteria (pdb id 4eb5). The complex is formed by the dimeric desulphurase IscS (in two blue tones) and the scaffold protein IscU (red and magenta). The complex is dynamic but stable. (D) Isolated IscU and IscS. When isolated, IscU is marginally unstable

interactions. Proteins are, by their very nature, promiscuous molecules that form complex networks of interactions within their (often crowded) environments. Interactions are important, if not crucial, to determine function but also to affect the solubility and stability of each protein. When a protein is not part of its natural functional complex(es), it might unfold, aggregate or precipitate (Figure 1). A direct logical implication of the last statement is that aggregation could be viewed as caused by the lack of natural partners. We could thus say that physiologic interactors are the chaperones that keep proteins in solution.

Different evidence demonstrates a direct link between normal and aberrant functions. A systematic analysis of the protein-protein complexes in the Protein Data Bank (PDB), for instance, showed that protein interface regions are more prone to aggregate than other surfaces. This indicated that many of the interactions that promote formation of functional complexes, including hydrophobic and electrostatic forces, can potentially also cause abnormal self-assembly.^[8] Accordingly, we demonstrated that the same regions of a protein that have evolved to endow the molecule with its function are those that promote aggregation.^[9] This implies that regions involved in protein aggregation may originate from an authentic physiological functional need that, under specific conditions, may backfire and result in pathology.^[10] A direct relation between normal and aberrant function strongly suggests studying the physiologic function to gain a better understanding of aggregation. This is at variance with the common tendency to consider aggregation the key to understanding pathology, often completely neglecting the non-pathologic function(s) of the protein under

study and the environment in which it operates. There seems to be, in other words, a barrier that separates studies aiming at elucidating the non-pathologic function of proteins from investigations of protein aggregation. Another crucial consequence of a partner-orphan model of aggregation is that normal partners could inform drug design.

The perspective of a partner-orphan mechanism as the origin of at least some misfolding diseases may change drastically our perception of protein aggregation: self-assembly would be a late side-effect of the disease and not, as it has often been considered, the cause of many diseases.

In this review, we discuss three paradigmatic examples that argue for the generality of our working hypothesis. We also analyse the case of the A β peptide as a classic example that has been widely studied without questioning why the peptide is present in the brain. Our analysis may inform new strategies to combat protein aggregation and misfolding and demonstrate that the study of normal function holds specific value for understanding the pathological behaviour of proteins.

A STRIKING EXAMPLE OF A PARTNER-ORPHAN MECHANISM: THE ROLE OF UBIQUITIN BINDING IN ATAXIN-3 AGGREGATION

The first example we encountered in which a direct relationship between physiologic function and aggregation was immediately clear is that of ataxin-3 (UniProt ID P54252).^[11] This is a 40–43-kDa cysteine

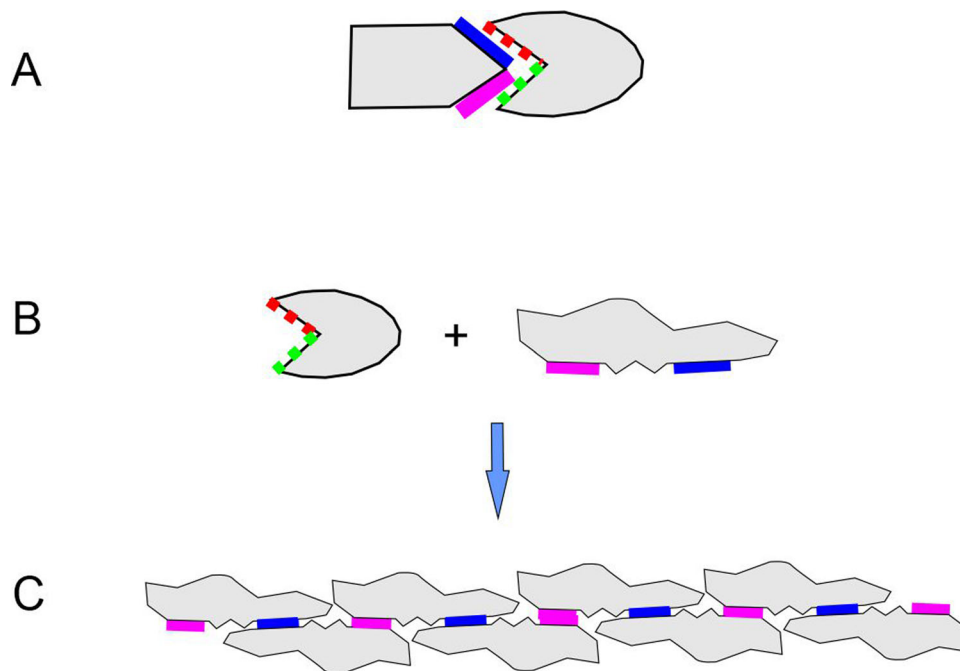


FIGURE 2 Relationship between physiologic function and aggregation. (A) Proteins may evolve to have exposed hydrophobic regions to mediate interactions with their partners, indicated in the figure with magenta and blue patches. (B) When the substrate is absent, the protein may be unstable and/or misfold leaving the exposed surfaces naked. (C) The regions functionally important can then promote aggregation

protease with specific ubiquitin hydrolase enzymatic functions.^[12,13] Ataxin-3 takes part in different cellular pathways and seems to play an important role both in the ubiquitin proteasome machinery^[14] and in gene expression regulation.^[15] Ataxin-3 is ubiquitously expressed in different cell types of peripheral and neuronal tissues.^[16–20] Although not essential, the protein is evolutionarily conserved.^[21] Its structure consists of a globular N terminal Josephin domain, followed by two ubiquitin-interacting motifs (UIMs), a polymorphic tract of polyglutamine (polyQ) repeat, and a more flexible C-terminus that, in some isoforms, contains a third UIM. Two nuclear export signals and one nuclear localisation signal are also present in the C-terminus.^[22] We recently demonstrated that, although generally more flexible, the C-terminus of ataxin-3 contains well-defined secondary structure elements.^[23]

Ataxin-3 is linked to the progressive neurodegenerative Machado–Joseph disease (MJD; MIM catalogue no. 109150), also known as spinocerebellar ataxia type 3 (or SCA3). The disease occurs as the consequence of the pathological expansion of CAG repeats in a coding region of the ATXN3 gene, which results in a tract of polyQ repeats within ataxin-3. The repeat length correlates with disease onset: normal individuals have 12–44 glutamine repeats, whereas MJD patients have expanded polyQ repeats above 53 up to approximately 87 glutamines.^[24,25] A direct link between pathology and polyQ is testified by an inverse correlation between the age at onset and the number of CAG repeats in the ATXN3 gene^[26,27] within a variability of the age at onset^[28] that is probably influenced by a mixture of environmental^[29] and genetic factors.^[30]

A hallmark of MJD is the formation of macromolecular aggregates containing the pathological expanded ataxin-3.^[31] It is still unclear

whether aggregation is initiated by full-length ataxin-3 or its calpain and caspase cleavage fragments, since both are able to aggregate in vitro and in vivo:^[32–34] while polyQ expansion is deemed to be essential for disease onset, the isolated Josephin domain is strongly prone to aggregation and fibrillation in vitro, suggesting that the protein contains multiple aggregation motifs as is also the case for other disease-associated proteins, such as prions (for a comprehensive review see^[35]).

Ataxin-3 binds and cleaves preferentially polyubiquitin chains of at least four subunits. The catalytic triad is in the Josephin domain, which is thus the enzymatically active region. Several ubiquitin-binding sites are observed along the ataxin-3 length, two or possibly three being in the Josephin domain,^[36] two or three more (depending on the isoform) corresponding to the UIMs in the protein C-terminus.^[37] In the course of our analyses on ataxin-3 and Josephin aggregation, we observed that the surfaces of Josephin evolved to bind ubiquitin are the same that promote aggregation.^[9] This was evidenced by mutation of even only one amino acid in the ubiquitin-binding surfaces appreciably reducing self-assembly.^[9] Conversely, addition of mono-ubiquitin resulted in a strong reduction of aggregation, presumably through masking the hydrophobic surfaces that promote aggregation.

From these studies, we deduced that ataxin-3 has exposed hydrophobic surfaces evolved to recognise the substrate and other partners. When the substrate is absent the same surfaces could promote misfolding and aggregation (Figure 2).

These results thus argue for a direct relationship between normal function, physiologic partners and aberrant aggregation.

WHAT ABOUT RNA IN PROTEIN AGGREGATION?

Mislocalisation of the nuclear heterogeneous ribonucleoprotein transactive response DNA binding protein 43 (or TDP-43) to the cytoplasm, and its subsequent aggregation into toxic inclusions is the common hallmark found in over 97% of all cases of amyotrophic lateral sclerosis (ALS).^[38] Increasing evidence supports the hypothesis that these aggregates have a direct toxic effect and are not just a consequence.^[39] TDP-43 is a highly conserved and ubiquitously expressed protein, which was initially identified thanks to its ability to bind HIV-1 TAR DNA and act as a transcriptional repressor.^[40] When non-pathologic, TDP-43 is a nuclear protein,^[41] but, under pathologic conditions, the protein becomes predominant in the cytoplasm where it forms the inclusions.^[42–44] Altered subcellular localisation has thus been suggested as an integral part of the disease pathogenesis.

TDP-43 is a modular protein that contains an N-terminus with a nuclear localisation signal,^[45,46] two tandem highly conserved RNA-recognition motifs (RRM), which preferentially bind single-stranded UG or TG-rich nucleotide sequences,^[46] and a disordered glycine-rich low complexity C-terminus. This region contains most of the pathologically significant mutations^[47–49] and was for this reason mistakenly considered the only region important for TDP-43 aggregation. This view was supported by early studies that had shown that C-terminus was sufficient for aggregation.^[50]

Taking on board the hypothesis of normal interactions as the basis of the design of specific anti-aggregants, we reasoned that we could perhaps use RNA as a way to prevent aggregation. We first demonstrated that the isolated tandem RRM and a longer construct excluding only the C-terminus are able to aggregate and misfold *in vitro* under mildly denaturant conditions; these were generated by subjecting the proteins to a temperature scan or simply incubating them at physiologic temperature.^[51,52] This study ensured that, even if not the main promoter of aggregation, other regions, such as the RNA binding motifs, are part of the aggregation process of the full-length protein. In agreement with our data, clinically important TDP-43 mutations have been found in the RNA binding motifs.^[53]

We then selected RNA aptamers that could mimic the properties of the natural RNA partners of the protein. These are small molecules that are increasingly being used in different applications as an alternative to antibodies. They have the advantage of being small, manageable, and easily obtainable. We used an aptamer (RNA12) already described in the literature as able to bind the RRM domains of TDP-43 with nanomolar affinities^[54] and its negative control, taken as the reverse and complementary sequence of RNA12. We found that stoichiometric ratios of RNA12 have a strong effect on aggregation as seen by the kinetics followed by thioflavin T fluorescence and quantification of the soluble protein.^[52] Conversely the control had no effect or, if any, a negative one, an observation that indicates how relevant the role of a specific partner is in determining the solubility of a protein. We also demonstrated that the effect is not purely electrostatic since heparin, the polyanion often used to mimic nucleic acids did not have any effect. Scrambled sequences of RNA12 were found to have a lower binding constant and an effect ca. 20% lower than the original aptamer,

indicating that the effect is not only compositional but also sequence dependent. The reverse and complementary scrambled sequences did not have any effect.

Our findings suggest a model for the disease mechanism of TDP-43 associated pathologies (Figure 3). In the normal cell, TDP-43 shuttles in and out the nucleus bringing out its cargoes. Under pathological conditions, mutations may prevent RNA binding, leaving the protein “naked”. This will thus allow misfolding and aggregation, causing the protein to accumulate in the cytoplasm. Our findings also highlight the dual role of RNA: when RNA-protein interactions are tight and specific, RNA may act as a chaperone that keeps TDP-43 in solution. When binding is weak, mostly electrostatic, and unspecific, RNA functions as a polyanion and induces phase separation. A strong confirmation of our conclusions was recently published: RNA-binding deficient TDP-43, as induced by mutations and or acetylation within the RRM domains, was shown to lead to separation of the protein into intranuclear liquid spherical shells with liquid cores, which were named anisosomes because of their similarities to liquid crystals.^[55] Formation of these droplets when RNA binding is impaired proves the role of this interaction for the physical state of TDP-43.

THE SCENARIO GETS MORE COMPLEX: CALCIUM REGULATED INTERACTIONS KEEP ANNEXIN A11 IN SOLUTION

A more recent perfect example of partner-orphan aggregation is that of annexin A11. This is a candidate gene that was only recently added to SOD1, TDP-43, FUS and many other genes as a putative cause of ALS.^[56] As in all the other cases, annexin A11-related ALS is associated with, and thought to cause pathology by, protein aggregation and misfolding.^[56] Several different mutations were identified in annexin A11 by screening a large cohort of familial ALS patients.^[56,57] An implication of this protein in human disease is not new, as deregulation and mutations in the *ANXA11* gene are known to be associated to autoimmune diseases, sarcoidosis and cancer.^[58] Annexin A11 is a 56 kDa protein that belongs to the large family of Ca²⁺-binding annexins whose function is lipid binding. Annexins are subdivided into twelve different subfamilies, most of which are characterised by a highly conserved C-terminal annexin core formed by four helical repeats of an annexin motif, with the exception only of annexin A6, which contains eight annexin motifs.^[59] These motifs are approximately 70 amino acids long and arranged in five α -helices, called A to E.^[60] Ca²⁺ is coordinated by residues at the loops connecting the AB and DE helical hairpins. Helix C packs orthogonally against the other components of the bundle. Lipid binding involves the core domain. Regulation is mediated in the whole family by a Ca²⁺-induced conformational rearrangement, which may be further regulated by the N-terminus, which has great variability among the various sub-families. The length of the N-terminus is also variable, typically being between 10 and 30 residues. In annexin A1, for instance, the N-terminus is 33 residues. In the Ca²⁺-free form this region of annexin A1 inserts into the core domain and forms a new helix that pushes aside helix B of the third annexin motif.^[61–63]

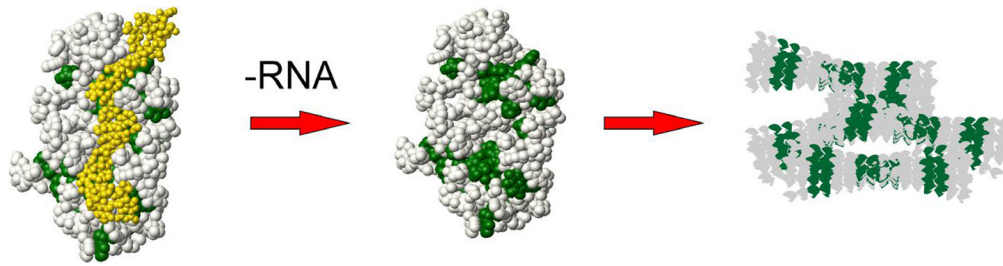


FIGURE 3 Possible mechanism of a TDP-43 associated pathology. (A) Binding of TDP-43 with RNA inside the nucleus, illustrated by the structure of two RRM motives with a small RNA (pdb id 4bs2). In the atom representation apolar residues are coloured green. (B) Dissociation of TDP-43 and RNA in the cytosol. (C) Fibers of TDP-43

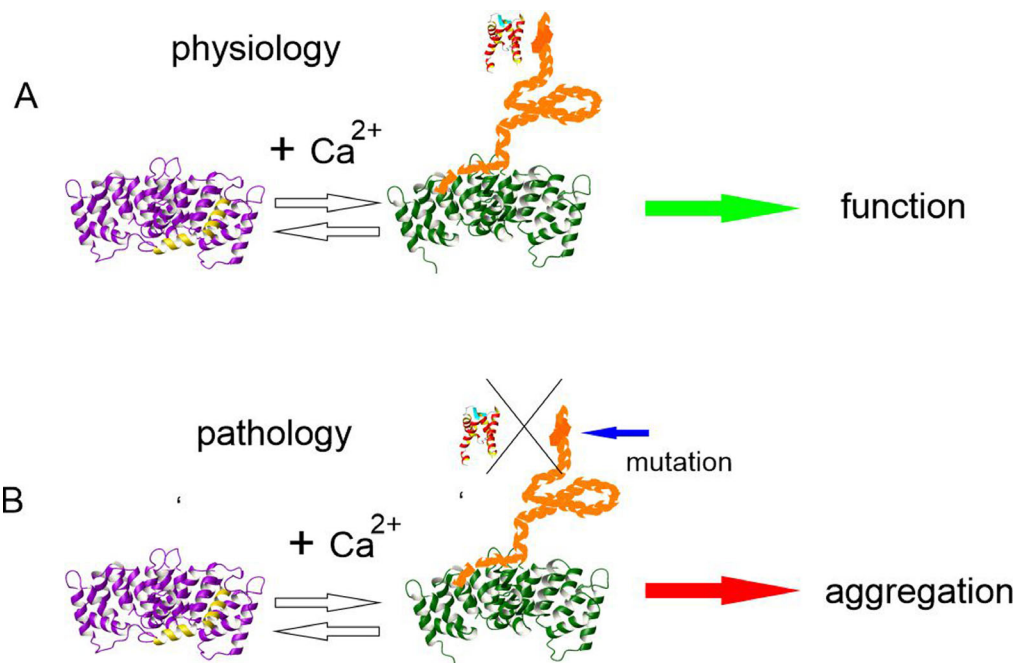


FIGURE 4 Model of annexin A11 aggregation as proposed by Smith et al.^[56] (A) Under physiologic conditions, annexin A11 forms calcium-mediated interactions with calyculin. In the absence of calcium the low-complexity N-terminal domain would fold back and interact with the C-terminal annexin domain, in analogy with what happens with annexin A1. In the presence of calcium, the N-terminus would become available for interaction with calyculin through a helix encompassing G38 and D40. This regulation would prevent protein aggregation and allow normal function. (B) When these residues are mutated, calyculin interaction would be impaired and the protein would become susceptible to aggregation, almost certainly promoted by the low-complexity domain

Ca^{2+} binding forces a remodelling of the domain that causes helix B to displace the N-terminal helix and come out from the core domain where it unfolds. Annexin A11 has a much longer N-terminus than the rest of the family members, comprising ~200 residues. The sequence of this region is enriched in Gly, Tyr, and Pro, suggesting intrinsic disorder,^[64] and seems to be important for ALS in that the most common ALS-associated mutations, p.D40G and p.G38R, map there.^[56] The N-terminus mediates interactions with a number of other proteins among which are the apoptosis-linked gene-2 protein (ALG-2) and S100A6 (calyculin).^[65]

On the basis of a weak sequence homology, it was suggested that the N-terminus of annexin A11 could contain a putative helical motif

around residues 38–59 that could dictate, by analogy with annexin A1, a similar regulation.^[56] The suggested model of regulation would thus be that also in annexin A11 the helix could interact with the core in a calcium dependent fashion and, when released by Ca^{2+} binding would then facilitate binding to calyculin. Both the annexin A11 p.G38R and p.D40G variants have been proven experimentally to abolish binding to calyculin.

This evidence makes it clear that the involvement of annexin A11 in ALS is consistent with the concept of partner orphan disease: mutations that abolish interaction with calyculin would expose surfaces that would normally be protected. This in turn would induce aggregation and misfold as seen in disease (Figure 4).

NATURAL INTERACTIONS AND BEYOND: WHAT ROLE DO QUINARY INTERACTIONS PLAY?

A question that arises spontaneously from these analyses is about the non-specific interactions that occur in the crowding environment of the cell. Do they play a role in protein solubility? An aspect of physiological aggregation that has gained appreciable attention in the last 2 decades is the influence of the cellular environment on protein stability. It is now commonly accepted that the cellular milieu is crowded and confined,^[66] where these terms were predicted to result in the stabilisation of folded proteins owing to the reduction of available volume.^[67] This interpretation was probably optimistic, possibly due to an overestimation of the volume difference between folded and unfolded species.^[68] Recently, a review on the chemical organisation of biological systems has proposed a more complex but possibly more realistic model according to which the same macromolecule may be affected in different ways, depending on the cell cycle, the specific compartment where they are, or the forces acting on the cell.^[69] These authors stressed the concept that the factors shaping protein interaction networks can affect the free energy of marginally stable proteins and sparsely-populated states, directing the evolutionary pathways in an organism.

More in general, it has been pointed out that the presence of many macromolecular components in the cell media may give rise to stabilisation or destabilisation effects under the general concept of quinary structure (the fifth level of protein architecture) in addition to primary, secondary, tertiary and quaternary structures. The concept was coined to refer to macromolecular interactions that are transient *in vivo*.^[70] In the words of the author: "Such interactions will not be evident from the composition of purified proteins, but they may constitute an important source of constraints on changes in primary structure." It is now believed that quinary interactions are shaped, in the course of evolution, by adaptation to the physiology of cells^[65] and can lead both to stabilisation or destabilisation of proteins, depending on relative interactions.^[71] This means that, at least in some cases, a weaker and probably less durable effect on preventing protein aggregation might be exerted by these transient and unspecific interactions. A possible consequence of quinary interactions in the case of misfolding diseases might be a contribution to the long lag times observed in some diseases: the presence of quinary interactions might stabilise, at least temporarily, species with strong intrinsic tendency to aggregate.

All these considerations add a further layer of complexity to the problem of which interactions are "chaperoning" and which ones detrimental, but, at the same time, they contribute to a deeper understanding of the field.

DISEASES IN SEARCH OF A PARTNER: THE EXAMPLE OF A β

The presented examples strongly suggest that a way to approach diseases whose therapy has remained stagnant for years might be to search for natural interactors. An example that cries out for a change

in perspective is that of the A β peptides associated with Alzheimer disease (AD), a pathology that remains out of the reach of treatment. The mechanism of AD^[72] is inextricably linked to the presence of the A β peptides that are the proteolytic products of the amyloid precursor protein through the action of secretases. The non-symptomatic phase of AD lasts often several years, a period during which the peptides form plaques containing amyloid fibrils of A β ^[73]. These species were originally regarded as the main toxic determinant causing AD according to what is now known as the "amyloid hypothesis."^[72] It is interesting to note that, although the aggregates were first observed by Alois Alzheimer in 1906 and considered responsible for the disease, they were soon discarded as possible culprits by the same scientist.^[74] In fact, although the amyloid hypothesis has dominated the AD literature for many years, it became more and more evident that the toxicity of the aggregates is probably just marginal. The main cause of cell damage became the A β oligomers of different stoichiometries formed during the aggregation pathway.^[75] Unfortunately, oligomer isolation in quantities large enough for structural studies hampers a better understanding of this hypothesis, as elegantly described in a recent review.^[76]

It is thus essential to ask whether A β peptides have a non-pathological function that could enable us to find beneficial interactions to stabilise the peptides. Itzhaki and coworkers have for years advocated a role of A β in innate immunity.^[77] In support of their hypothesis, it has been shown that A β peptides are highly conserved across vertebrates^[78] and present in all healthy individuals.^[76] Their complex biosynthesis is inconsistent with the just occasional formation of protein debris. We recently demonstrated that A β peptides are structurally and functionally similar to anti-microbial peptides.^[79] Getting inspiration from the already rich interactome of A β might thus provide a better answer for AD treatment. Among the several interacting proteins, a particularly interesting one could be transthyretin, whose levels have been shown to enhance A β solubility.^[80,81]

CONCLUSIONS AND OUTLOOK

In this review, we have provided a new perspective based on the natural functions of proteins and their interactions to explain the old concept of protein aggregation. The examples discussed suggest what we have called the partner-orphan model in which aggregation is caused by the absence of natural partners. We have supported our working hypothesis with three examples taken from the neurodegeneration field. Many more examples may reiterate the same message: tau, for instance, a protein whose function is to bind microtubules and maintain their stability in axons, aggregates when mutations prevent interaction with microtubules.^[82] Other interesting examples, such as insulin^[83] or macroglobulin^[84], are associated with diseases other than neurodegenerative, demonstrating the general validity of the concept. The new model shifts our perception of protein aggregation considerably.

Many important open questions remain for the future: what are the partners with more appreciable effect on solubility? Which roles do specific physico-chemical properties of the polypeptide chains such as structurally disordered regions play? Do other events including

macromolecular crowding influence the events leading to loss of solubility? The new paradigm has important consequences in that it suggests that providing the partner either fully or mimicked by small molecules may be a better strategy to prevent protein aggregation. This aspect has important consequences for future strategies in drug design. A shift of the paradigm also suggests the importance of a holistic view of proteins, seen not just as the cause of aberrant function but, more generally, as entities that need to be understood in full before starting therapeutical interventions.

ACKNOWLEDGMENTS

This work was supported by Dementia Research UK (RE1 3556) and by Alzheimer Research UK (ARUK-PG2019B-020).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Annalisa Pastore  <https://orcid.org/0000-0002-3047-654X>

REFERENCES

- Luheshi, L. M., Crowther, D. C., & Dobson, C. M. (2008). Protein misfolding and disease: From the test tube to the organism. *Current Opinion in Chemical Biology*, 12, 25–31.
- Stefani, M. (2004). Protein misfolding and aggregation: New examples in medicine and biology of the dark side of the protein world. *Biochimica Et Biophysica Acta*, 1739, 5–25.
- Lansbury, P. T., & Lashuel, H. A. (2006). A century-old debate on protein aggregation and neurodegeneration enters the clinic. *Nature*, 443, 774–779.
- Aguzzi, A., & De Cecco, E. (2020). Shifts and drifts in prion science. *Science*, 370, 32–34.
- Winklhofer, K. F., Tatzelt, J., & Haass, C. (2008). The two faces of protein misfolding: Gain- and loss-of-function in neurodegenerative diseases. *Embo Journal*, 27, 336–349.
- Hipp, M. S., Kasturi, P., & Hartl, F. U. (2019). The proteostasis network and its decline in ageing. *Nature Reviews Molecular Cell Biology*, 20, 421–435.
- Hipp, M. S., Park, S. - H., & Hartl, F. U. (2014). Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends in Cell Biology*, 24, 506–514.
- Pechmann, S., Levy, E. D., Tartaglia, G. G., & Vendruscolo, M. (2009). Physicochemical principles that regulate the competition between functional and dysfunctional association of proteins. *Proc. Natl Acad. Sci. U.S.A.*, 106, 10159–10164.
- Masino, L., Nicastro, G., Calder, L., Vendruscolo, M., & Pastore, A. (2011). Functional interactions as a survival strategy against abnormal aggregation. *Faseb Journal*, 25, 45–54.
- Pastore, A., & Temussi, P. A. (2012). The two faces of Janus: Functional interactions and protein aggregation. *Current Opinion in Structural Biology*, 22, 30–37.
- Costa, M. D. O. C., & Paulson, H. L. (2012). Toward understanding Machado–Joseph Disease. *Progress in Neurobiology*, 97, 239–257.
- Burnett, B. (2003). The polyglutamine neurodegenerative protein ataxin-3 binds polyubiquitylated proteins and has ubiquitin protease activity. *Human Molecular Genetics*, 12, 3195–3205.
- Tzvetkov, N., & Breuer, P. (2007). Josephin domain-containing proteins from a variety of species are active de-ubiquitination enzymes. *Biological Chemistry*, 388, 973–978.
- Schmitt, I., Linden, M., Khazneh, H., Evert, B. O., Breuer, P., Klockgether, T., & Wuellner, U. (2007). Inactivation of the mouse *Atxn3* (ataxin-3) gene increases protein ubiquitination. *Biochemical and Biophysical Research Communications*, 362, 734–739.
- Evert, B. O., Araujo, J., Vieira-Saecker, A. M., De Vos, R. A. I., Harendza, S., Klockgether, T., & Wullner, U. (2006). Ataxin-3 represses transcription via chromatin binding, interaction with histone deacetylase 3, and histone deacetylation. *Journal of Neuroscience*, 26, 11474–11486.
- Paulson, H. L., Das, S. S., Crino, P. B., Perez, M. K., Patel, S. C., Gotsdiner, D., Fischbeck, K. H., & Pittman, R. N. (1997). Machado–Joseph disease gene product is a cytoplasmic protein widely expressed in brain. *Annals of Neurology*, 41, 453–462.
- Trottier, Y., Cancel, G., An-Gourfinkel, I., Lutz, Y., Weber, C., Brice, A., Hirsch, E., & Mandel, J. - L. (1998). Heterogeneous intracellular localization and expression of ataxin-3. *Neurobiology of Disease*, 5, 335–347.
- Schmidt, T., Landwehrmeyer, G. B., Schmitt, I., Trottier, Y., Auburger, G., Laccone, F., Klockgether, T., Völpel, M., Epplen, J. T., Schöls, L., & Riess, O. (1998). An isoform of ataxin-3 accumulates in the nucleus of neuronal cells in affected brain regions of SCA3 patients. *Brain Pathology*, 8, 669–679.
- Ichikawa, Y., Goto, J., Hattori, M., Toyoda, A., Ishii, K., Jeong, S. - Y., Hashida, H., Masuda, N., Ogata, K., Kasai, F., Hirai, M., Maciel, P., Rouleau, G. A., Sakaki, Y., & Kanazawa, I. (2001). The genomic structure and expression of MJD, the Machado–Joseph disease gene. *Journal of Human Genetics*, 46, 413–422.
- Do Carmo Costa, M., Gomes-Da-Silva, J., Miranda, C. J., Sequeiros, J., Santos, M. M., & Maciel, P. (2004). Genomic structure, promoter activity, and developmental expression of the mouse homologue of the Machado–Joseph disease (MJD) gene. *Genomics*, 84, 361–373.
- Rodrigues, A. - J., Coppola, G., Santos, C., Costa, M. Do. C., Ailion, M., Sequeiros, J., Geschwind, D. H., & Mariel, P. (2007). Functional genomics and biochemical characterization of the *C. elegans* orthologue of the Machado–Joseph disease protein ataxin-3. *Faseb Journal*, 21, 1126–1136.
- Antony, P. M. A., Mäntele, S., Mollenkopf, P., Boy, J., Kehlenbach, R. H., Riess, O., & Schmidt, T. (2009). Identification and functional dissection of localization signals within ataxin-3. *Neurobiology of Disease*, 36, 280–292.
- Sicorello, A., Rózycki, B., Konarev, P. V., Svergun, D. I., & Pastore, A. (2020). Capturing the conformational ensemble of the mixed folded polyglutamine protein ataxin-3. *Structure (London, England)*, 29, 70–81.e5.e5.
- Lima, M., Costa, M. C., Montiel, R., Ferro, A., Santos, C., Silva, C., Betencourt, C., Sousa, A., Sequeiros, J., Coutinho, P., & Maciel, P. (2005). Population genetics of wild-type CAG repeats in the Machado–Joseph disease gene in Portugal. *Human Heredity*, 60, 156–163.
- Maciel, P. (2001). Improvement in the molecular diagnosis of Machado–Joseph disease. *Archives of Neurology*, 58, 1821–1827.
- Maciel, P., Gaspar, C., DeStefano, A. L., Silveira, I., Coutinho, P., Radvany, J., Dawson, D. M., Sudarsky, L., Guimarães, J., & Loureiro, J. E. (1995). Correlation between CAG repeat length and clinical features in Machado–Joseph disease. *American Journal of Human Genetics*, 57, 54–61.
- Schöls, L., Vieira-Saecker, A. M. M., Schöls, S., Przuntek, H., Epplen, J. T., & Riess, O. (1995). Trinucleotide expansion within the MJD1 gene presents clinically as spinocerebellar ataxia and occurs most frequently in german SCA patients. *Human Molecular Genetics*, 4, 1001–1005.
- Van De Warrenburg, B. P. C., Hendriks, H., Dürr, A., Van Zuijlen, M. C. A., Stevanin, G., Camuzat, A., Sinke, R. J., Brice, A., & Kremer, B. P. H. (2005). Age at onset variance analysis in spinocerebellar ataxias: A study in a Dutch-French cohort. *Annals of Neurology*, 57, 505–512.
- Junck, L., & Fink, J. K. (1996). Machado–Joseph disease and SCA3: The genotype meets the phenotypes. *Neurology*, 46, 4–8.

30. DeStefano, A. L., Cupples, L. A., Maciel, P., Gaspar, C., Radvany, J., Dawson, D. M., Sudarsky, L., Corwin, L., Coutinho, P., & MacLeod, P. (1996). A familial factor independent of CAG repeat length influences age at onset of Machado-Joseph disease. *American Journal of Human Genetics*, *59*, 119–127.
31. Paulson, H. L., Perez, M. K., Trottier, Y., Trojanowski, J. Q., Subramony, S. H., Das, S. S., Vig, P., Mandel, J. - L., Fischbeck, K. H., & Pittman, R. N. (1997). Intranuclear inclusions of expanded polyglutamine protein in spinocerebellar ataxia type 3. *Neuron*, *19*, 333–344.
32. Wellington, C. L., Ellerby, L. M., Hackam, A. S., Margolis, R. L., Trifiro, M. A., Singaraja, R., Mccutcheon, K., Salvesen, G. S., Propp, S. S., Bromm, M., Rowland, K. J., Zhang, T., Rasper, D., Roy, S., Thornberry, N., Pinsky, L., Kakizuka, A., Ross, C. A., Nicholson, D. W., ... Hayden, M. R. (1998). Caspase cleavage of gene products associated with triplet expansion disorders generates truncated fragments containing the polyglutamine tract. *Journal of Biological Chemistry*, *273*, 9158–9167.
33. Weber, J. J., Golla, M., Guaitoli, G., Wanichawan, P., Hayer, S. N., Hauser, S., Krahl, A. - C., Nagel, M., Samer, S., Aronica, E., Carlson, C. R., Schöls, L., Riess, O., Gloeckner, C. J., Nguyen, H. P., & Hübener-Schmid, J. (2017). A combinatorial approach to identify calpain cleavage sites in the Machado-Joseph disease protein ataxin-3. *Brain*, *140*, 1280–1299.
34. Hubener, J., Weber, J. J., Richter, C., Honold, L., Weiss, A., Murad, F., Breuer, P., Wullner, U., Bellstedt, P., Paquet-Durand, F., Takano, J., Saido, T. C., Riess, O., & Nguyen, H. P. (2013). Calpain-mediated ataxin-3 cleavage in the molecular pathogenesis of spinocerebellar ataxia type 3 (SCA3). *Human Molecular Genetics*, *22*, 508–518.
35. Carvalho, A. L., Silva, A., & Macedo-Ribeiro, S. (2018). Polyglutamine-independent features in ataxin-3 aggregation and pathogenesis of Machado-Joseph disease. *Advances in Experimental Medicine and Biology*, *1049*, 275–288.
36. Faggiano, S., Menon, R. P., Kelly, G. P., Todi, S. V., Scaglione, K. M., Konarev, P. V., Svergun, D. I., Paulson, H. L., & Pastore, A. (2015). Allosteric regulation of deubiquitylase activity through ubiquitination. *Front Mol Biosci*, *2*, 2.
37. Weishäupl, D., Schneider, J., Peixoto Pinheiro, B., Ruess, C., Dold, S. M., Von Zweydford, F., Gloeckner, C. J., Schmidt, J., Riess, O., & Schmidt, T. (2019). Physiological and pathophysiological characteristics of ataxin-3 isoforms. *Journal of Biological Chemistry*, *294*(2), 644–661.
38. Scotter, E. L., Chen, H. - J., & Shaw, C. E. (2015). TDP-43 proteinopathy and ALS: Insights into disease mechanisms and therapeutic targets. *Neurotherapeutics*, *12*(2), 352–363.
39. Hergesheimer, R. C., Chami, A. A., De Assis, D. R., Vourc'h, P., Andres, C. R., Corcia, P., Lanznaster, D., & Blasco, H. (2019). The debated toxic role of aggregated TDP-43 in amyotrophic lateral sclerosis: A resolution in sight? *Brain*, *142*, 1176–1194.
40. Ayala, Y. M., Zago, P., D'ambrogio, A., Xu, Y. - F., Petrucelli, L., Buratti, E., & Baralle, F. E. (2008). Structural determinants of the cellular localization and shuttling of TDP-43. *Journal of Cell Science*, *121*, 3778–3785.
41. Winton, M. J., Igaz, L. M., Wong, M. M., Kwong, L. K., Trojanowski, J. Q., & Lee, V. M.-Y. (2008). Disturbance of nuclear and cytoplasmic TAR DNA-binding protein (TDP-43) induces disease-like redistribution, sequestration, and aggregate formation. *Journal of Biological Chemistry*, *283*, 13302–13309.
42. Bodansky, A., Kim, J. M. 'H.', Tempest, L., Velagapudi, A., Libby, R., & Ravits, J. (2010). TDP-43 and ubiquitinated cytoplasmic aggregates in sporadic ALS are low frequency and widely distributed in the lower motor neuron columns independent of disease spread. *Amyotroph. Lateral Scler.*, *11*, 321–327.
43. Van Eersel, J., Ke, Y. D., Gladbach, A., Bi, M., Götz, J., Kril, J. J., & Iltner, L. M. (2011). Cytoplasmic accumulation and aggregation of TDP-43 upon proteasome inhibition in cultured neurons. *Plos One*, *6*, e22850.
44. Romano, V., Quadri, Z., Baralle, F. E., & Buratti, E. (2015). The structural integrity of TDP-43 N-terminus is required for efficient aggregate entrapment and consequent loss of protein function. *Prion*, *9*, 1–9.
45. Zhang, Y. - J., Caulfield, T., Xu, Y.-F., Gendron, T. F., Hubbard, J., Stetler, C., Sasaguri, H., Whitelaw, E. C., Cai, S., Lee, W. C., & Petrucelli, L. (2013). The dual functions of the extreme N-terminus of TDP-43 in regulating its biological activity and inclusion formation. *Human Molecular Genetics*, *22*, 3112–3122.
46. Kuo, P. - H., Doudeva, L. G., Wang, Y. - T., Shen, C.-K. J., & Yuan, H. S. (2009). Structural insights into TDP-43 in nucleic-acid binding and domain interactions. *Nucleic Acids Res.*, *37*, 1799–1808.
47. Pesiridis, G. S., Lee, V. M.-Y., & Trojanowski, J. Q. (2009). Mutations in TDP-43 link glycine-rich domain functions to amyotrophic lateral sclerosis. *Human Molecular Genetics*, *18*, R156–R162.
48. Yang, C., Tan, W., Whittle, C., Qiu, L., Cao, L., Akbarian, S., & Xu, Z. (2010). The C-terminal TDP-43 fragments have a high aggregation propensity and harm neurons by a dominant-negative mechanism. *Plos One*, *5*, e15878.
49. Gendron, T. F., Rademakers, R., & Petrucelli, L. (2013). TARDBP mutation analysis in TDP-43 proteinopathies and deciphering the toxicity of mutant TDP-43. *J. Alzheimers. Dis.*, *33*(Suppl 1), S35–45.
50. Chen, A. K.-H., Lin, R. Y.-Y., Hsieh, E. Z.-J., Tu, P. - H., Chen, R. P.-Y., Liao, T. - Y., Chen, W., Wang, C. - H., & Huang, J. J.-T. (2010). Induction of amyloid fibrils by the C-terminal fragments of TDP-43 in amyotrophic lateral sclerosis. *Journal of the American Chemical Society*, *132*(4), 1186–1187.
51. Zacco, E., Martin, S. R., Thorogate, R., & Pastore, A. (2018). The RNA-recognition motifs of TAR DNA-binding protein 43 may play a role in the aberrant self-assembly of the protein. *Front. Mol. Neurosci.*, *11*, 372.
52. Zacco, E., Graña-Montes, R., Martin, S. R., De Groot, N. S., Alfano, C., Tartaglia, G. G., & Pastore, A. (2019). RNA as a key factor in driving or preventing self-assembly of the TAR DNA-binding protein 43. *Journal of Molecular Biology*, *43*, 1671–1688.
53. Chen, H. - J., Topp, S. D., Hui, H. O. S., Zacco, E., Katarya, M., Mcloughlin, C., King, A., Smith, B. N., Troakes, C., Pastore, A., & Shaw, C. E. (2019). RRM adjacent TARDBP mutations disrupt RNA binding and enhance TDP-43 proteinopathy. *Brain*, *142*(12), 3753–3770.
54. Lukavsky, P. J., Daujotyte, D., Tollervey, J. R., Ule, J., Stuani, C., Buratti, E., Baralle, F. E., Damberger, F. F., & Allain, F. H.-T. (2013). Molecular basis of Ug-rich RNA recognition by the human splicing factor Tdp-43. *Nat.Struct.Mol.Biol.*, *20*, 1443.
55. Yu, H., Lu, S., Gasior, K., Singh, D., Vazquez-Sanchez, S., Tapia, O., Toprani, D., Beccari, M. S., Yates, J. R., Da Cruz, S., Newby, J. M., Lafarga, M., Gladfelter, A. S., Villa, E., & Cleveland, D. W. (2020). HSP70 chaperones RNA-free TDP-43 into anisotropic intranuclear liquid spherical shells. *Science*, *371*, eabb4309.
56. Smith, B. N., Topp, S. D., Fallini, C., Shibata, H., Chen, H. - J., Troakes, C., King, A., Ticozzi, N., Kenna, K. P., Soragia-Gkazi, A., Miller, J. W., Sato, A., Dias, D. M., Jeon, M., Vance, C., Wong, C. H., De Majo, M., Kattuah, W., Mitchell, J. C., ... & Shaw, C. E. (2017). Mutations in the vesicular trafficking protein annexin A11 are associated with amyotrophic lateral sclerosis. *Science Translational Medicine*, *9*(388), ead9157.
57. Teyssou, E., Muratet, F., Amador, M. D., Ferrien, M., Lautrette, G., Machat, S., Boillée, S., Larmonier, T., Saker, S., Leguern, E., Cazeneuve, C., Marie, Y., Guegan, J., Gyorgy, B., Cintas, P., Meininger, V., Le Forestier, N., Salachas, F., Couratier, P., ... & Millecamps, S. (2020). Genetic screening of ANXA11 revealed novel mutations linked to amyotrophic lateral sclerosis. *Neurobiology of Aging*, *99*, 102.e11–102.e20.
58. Wang, J., Guo, C., Liu, S., Qi, H., Yin, Y., Liang, R., Sun, M. - Z., & Greenaway, F. T. (2014). Annexin A11 in disease. *Clinica Chimica Acta*, *431*, 164–168.
59. Kourie, J. I., & Wood, H. B. (2000). Biophysical and molecular properties of annexin-formed channels. *Progress in Biophysics and Molecular Biology*, *73*, 91–134.
60. Gerke, V., & Moss, S. E. (2002). Annexins: From structure to function. *Physiological Reviews*, *82*, 331–371.

61. Weng, X., Luecke, H., Song, I. n S., Kang, D. S., Kim, S. - H., & Huber, R. (1993). Crystal structure of human annexin I at 2.5 Å resolution. *Protein Science*, 2, 448–458.
62. Rosengarth, A., Gerke, V., & Luecke, H. (2001). X-ray structure of full-length annexin 1 and implications for membrane aggregation. *Journal of Molecular Biology*, 306, 489–498.
63. Rosengarth, A., & Luecke, H. (2003). A calcium-driven conformational switch of the N-terminal and core domains of annexin A1. *Journal of Molecular Biology*, 326, 1317–1325.
64. Weisz, J., & Uversky, V. N. (2020). Zooming into the dark side of human annexin-S100 complexes: Dynamic alliance of flexible partners. *International Journal of Molecular Sciences*, 21, 5879.
65. Rezvanpour, A., & Shaw, G. S. (2009). Unique S100 target protein interactions. *General Physiology and Biophysics*, 28, F39–F46.
66. Zhou, H. - X., Rivas, G., & Minton, A. P. (2008). Macromolecular crowding and confinement: Biochemical, biophysical, and potential physiological consequences. *Annual Review of Biophysics*, 37, 375–397.
67. Minton, A. P. (2000). Effect of a concentrated “inert” macromolecular cosolute on the stability of a globular protein with respect to denaturation by heat and by chaotropes: A statistical-thermodynamic model. *Biophysical Journal*, 78(1), 101–109.
68. Politou, A., & Temussi, P. A. (2015). Revisiting a dogma: The effect of volume exclusion in molecular crowding. *Curr. Opin. Struct. Biol.*, 30, 1–6.
69. Guin, D., & Gruebele, M. (2019). Weak chemical interactions that drive protein evolution: crowding, sticking, and quinary structure in folding and function. *Chem. Rev.*, 119(18), 10691–10717.
70. Mcconkey, E. H. (1982). Molecular evolution, intracellular organization, and the quinary structure of proteins. *Proc. Natl. Acad. Sci. U.S.A.*, 79(10), 3236–3240.
71. Cohen, R. D., & Pielak, G. J. (2017). A cell is more than the sum of its (dilute) parts: A brief history of quinary structure. *Protein Science*, 26(3), 403–413.
72. Hardy, J. (2002). The amyloid hypothesis of Alzheimer's disease: Progress and problems on the road to therapeutics. *Science*, 297, 353–356.
73. Spinney, L. (2014). Alzheimer's disease: The forgetting gene. *Nature*, 510, 26–28.
74. Alzheimer, A., Förstl, H., & Levy, R. (1991). On certain peculiar diseases of old age. *Hist. Psychiatry*, 2(5 Pt 1), 71–73.
75. Doig, A. J., & Derreumaux, P. (2015). Inhibition of protein aggregation and amyloid formation by small molecules. *Current Opinion in Structural Biology*, 30, 50–56.
76. Ayton, S., & Bush, A. I. (2020). β -amyloid: The known unknowns. *Ageing Research Reviews*, 65, 101212.
77. Itzhaki, R. F., Lathe, R., Balin, B. J., Ball, M. J., Bearer, E. L., Braak, H., Bullido, M. J., Carter, C., Clerici, M., Cosby, S. L., Del Tredici, K., Field, H., Fulop, T., Grassi, C., Griffin, W. S. T., Haas, J., Hudson, A. P., Kamer, A. R., Kell, D. B., ..., & Whittum-Hudson, J. A. (2016). Microbes and Alzheimer's disease. *Journal of Alzheimers Disease*, 51(4), 979–984.
78. Moir, R. D., & Tanzi, R. E. (2019). Low evolutionary selection pressure in senescence does not explain the persistence of A β in the vertebrate genome. *Front. Aging. Neurosci.*, 11, 70.
79. Pastore, A., Raimondi, F., Rajendran, L., & Temussi, P. A. (2020). Why does the A β peptide of Alzheimer share structural similarity with antimicrobial peptides? *Commun. Biol.*, 3(1), 135.
80. Schwarzman, A. L., Gregori, L., Vitek, M. P., Lyubski, S., Strittmatter, W. J., Enghilde, J. J., Bhasin, R., Silverman, J., Weisgraber, K. H., & Coyle, P. K. (1994). Transthyretin sequesters amyloid β protein and prevents amyloid formation. *Proc Natl Acad Sci U. S. A.*, 91, 8368–8372.
81. Stein, T. D., & Johnson, J. A. (2002). Lack of neurodegeneration in transgenic mice overexpressing mutant amyloid precursor protein is associated with increased levels of transthyretin and the activation of cell survival pathways. *Journal of Neuroscience*, 22, 7380–7388.
82. Iqbal, K., Liu, F., & Gong, C. - X. (2016). Tau and neurodegenerative disease: The story so far. *Nature Reviews Neurology*, 12, 15–27.
83. Akbarian, M., Yousefi, R., Farjadian, F., & Uversky, V. N. (2021). Insulin fibrillation: Toward strategies for attenuating the process. *Chemical Communications*, 56, 11354–11373.
84. Stoppini, M., & Bellotti, V. (2015). Systemic amyloidosis: Lessons from β 2-microglobulin. *Journal of Biological Chemistry*, 290, 9951–9958.

How to cite this article: Temussi P. A., Tartaglia, G. G., & Pastore, A. (2021). The seesaw between normal function and protein aggregation: How functional interactions may increase protein solubility. *BioEssays*, e2100031.
<https://doi.org/10.1002/bies.202100031>