Evolutionary selection for protein aggregation

Natalia Sanchez de Groot*†, Marc Torrent*, Anna Villar-Piqué‡‡, Benjamin Lang*, Salvador Ventura‡‡, Jörg Gsponer§ and M. Madan Babu*†

* MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, U.K., †Department of Biochemistry and Molecular Biology, Faculty of Science, Autonomous University of Barcelona, E-08193 Bellaterra, Spain, ‡Institute of Biotechnology and Biomedicine, Autonomous University of Barcelona, E-08193 Bellaterra, Spain, and §Centre for High-Throughput Biology, University of British Columbia, Vancouver, Canada, V6T 1Z4

Abstract

Protein aggregation is being found to be associated with an increasing number of human diseases. Aggregation can lead to a loss of function (lack of active protein) or to a toxic gain of function (cytotoxicity associated with protein aggregates). Although potentially harmful, protein sequences predisposed to aggregation seem to be ubiquitous in all kingdoms of life, which suggests an evolutionary advantage to having such segments in polypeptide sequences. In fact, aggregation-prone segments are essential for protein folding and for mediating certain protein–protein interactions. Moreover, cells use protein aggregates for a wide range of functions. Against this background, life has adapted to tolerate the presence of potentially dangerous aggregation-prone sequences by constraining and counteracting the aggregation process. In the present review, we summarize the current knowledge of the advantages associated with aggregation-prone stretches in proteomes and the strategies that cellular systems have developed to control the aggregation process.

Introduction

During evolution, cells have been equipped with a wide array of proteins that perform diverse functions. Generally, to accomplish their function, polypeptide chains need to adopt a defined three-dimensional structure (native conformation). The failure to acquire or maintain this conformation might lead to protein aggregation, in which the polypeptide chains self-assemble to form insoluble deposits [1,2]. This event typically involves a loss of function and/or the gain of different toxic effects, which include jamming of chaperones and proteases, sequestration of proteins with essential functions, depletion of protein synthesis resources and the formation of toxic polypeptide species [3–5]. Accordingly, protein aggregation has been associated with more than 40 different human diseases, such as Alzheimer’s disease or Type 2 diabetes [1,6].

A large number of proteins have been found to form insoluble deposits. Although it is now widely accepted that aggregation is a general property of proteins, in most cases, aggregation can only be triggered under non-physiological conditions. Interestingly, these aggregates can display different organization levels, from mostly amorphous to defined macromolecular structures, with some of them, such as amyloid fibrils, being more stable than the functional native state [1,6,7].

The benefits of protein aggregation

Two sides of the same coin: functional interactions and aggregation

The diverse monomeric and multimeric states for a given protein constitute two linked energy landscapes: the first contains different combinations of intramolecular interactions and aggregation.
functional protein aggregates

Although protein aggregation has been extensively associated with pathological conditions, there are many examples of how Nature employs protein aggregation for beneficial purposes [8]. Indeed, protein aggregates offer exceptional stability, compactness and forms of organization that could not be achieved by monomeric or oligomeric conformations.

Different organisms use protein aggregates for adaptation to diverse environments. For example, Escherichia coli and other Enterobacteriaceae make use of the rigid structure of amyloid aggregates to form fibrillar extracellular deposits that facilitate bacterial adhesion, biofilm development and host invasion [24]. Similarly, Streptomyces coelicolor uses fibrils to build hyphae to colonize aerial environments [25]. There are also examples in higher organisms, such as the spidroin amyloid aggregates [26], which give spider silk a strength similar to steel [27], as well as resistance to proteases and detergents [8].

More recently, it has been shown that the polyglutamine protein pqn-41, whose overexpression causes protein aggregation and a phenotype similar to polyglutamine-induced neurodegeneration, is crucial for programmed cell death during the development of Caenorhabditis elegans [28].

Several examples of functional amyloid aggregates have also been reported in humans [29–32]. For example, amyloids of Pmel17 serve as a template for melanin formation in mammalian melanocytes [29]. Also, aggregation is required for the assembly of some cellular bodies, such as stress granules comprised of specific peptide hormones [31].

It has been suggested that nearly all organisms may have taken advantage of the toxic properties associated with protein aggregates to develop weapons against pathogenic micro-organisms [33–35]. One example involves antimicrobial peptides that kill micro-organisms by destabilizing their plasma membrane either by aggregation [36] or by opening channels that penetrate the lipid bilayer, the latter resembling the toxic mechanism of amyloid oligomers associated with diseases such as Alzheimer’s or Parkinson’s disease [35].

In this context, it has been suggested that evolution may have developed cation-rich and amphipathic regions from aggregation-prone sequences in order to generate membrane-directed antimicrobial activity [33].

One of the most diverse and remarkable groups of functional aggregates is the prion proteins. They are characterized by the ability to self-replicate and transmit structural conformations. In this sense, they are able to develop functional macromolecular arrangements that act as transmissible genetic elements, to promote the transmission of encoded in its sequence and modulated by the surrounding conformational environment.

Figure 1 | Scenarios favouring or disfavouring protein aggregation

The conservation of an aggregation-prone region is subject to a competition between the beneficial (up arrow) and detrimental (down arrow) effects that an aggregation-prone sequence can have on the cell.
Strategies that prevent protein aggregation

Aggregation-prone sequences are conserved only if they can provide an evolutionary advantage. In these cases, natural strategies have emerged to constrain the protein aggregation process. We group these strategies into two classes: those that constrain protein sequences and conformations (upper box), and those that mediate the maintenance of cellular proteostasis (lower box).

Safeguarding strategies

- Protein evolution
  - Protein stability
  - Interface contacts
  - Gatekeepers
  - Disulphide bridges
  - No aggregation clusters
  - Beta-breakers (e.g. Pro)

Cellular regulation
- Compartmentalization
- Chaperones
- Proteases
- Phagocytosis
- Protein abundance

- of diseases and to encode heritable phenotypic traits and memory [8,37,38]. Prions exist in both soluble and aggregated states, generating a wide diversity of phenotypes, which, under adverse conditions, can provide a fitness advantage [39,40]. Prions are also associated with memory storage. Si and co-workers have extensively studied the prion-like properties of two neuronal isoforms of the CPEB (cytoplasmic polyadenylation-element-binding protein) and observed that both CPEB isoforms form amyloid-like aggregates that are necessary for long-term synaptic persistence. In this sense, when aggregation is prevented by sequence mutation or inhibited using specific antibodies, persistence of long-term memory is lost [41]. Taken together, this illustrates that, during the course of evolution, the regulated assembly of protein aggregates has become a powerful mechanism that has allowed different life forms to achieve amazing feats.

Sequence and regulatory constraints to minimize aggregation

Pressure exerted on protein sequence evolution

The cell needs to fine-tune aggregation events, and thus this process must be tightly regulated. To achieve this, protein aggregation can be modulated at the sequence level to reduce aggregation propensity or certain folds may effectively hide the aggregation-prone regions in the core (Figure 2).

Analysis at the proteome level has revealed that essential proteins tend to be less predisposed to aggregation and contain less aggregation-prone stretches than non-essential ones [42,43]. This suggests that a strong pressure exists to avoid aggregation in essential proteins [15,44], since their depletion due to aggregation would result in loss of function, leading to cell death. Similar trends are also seen in IDPs (intrinsically disordered proteins), where the likelihood of exposing aggregation-prone residues to solvent is increased. Because disordered regions are intrinsically more susceptible to aggregation, stronger evolutionary pressure is required to avoid the presence of aggregation-prone stretches in such regions [16,44]. As a result, IDPs generally avoid hydrophobic stretches, and are rich in hydrophilic and charged residues to avoid the formation of a hydrophobic core, and destabilize any compact state [45]. In addition to IDPs, globular proteins also use charged residues to remain soluble by increasing local charge to reduce aggregation propensity, even at the cost of folding efficiency [46]. Only a small number of proteins have been reported to aggregate under native conditions [47,48], suggesting that the native state is generally protected against aggregation by various means. As mentioned above, aggregation-prone regions in globular proteins are usually found in the hydrophobic core, where they are unable to spontaneously form interactions that are conducive to aggregation [49].

There is now convincing evidence from a number of studies that, in protein complexes, aggregation-prone regions are located on binding interfaces, and complex formation may serve to prevent interactions leading to aggregation [21,23,50]. For example, Ataxin-3 exhibits two highly aggregation-prone patches involved in binding to ubiquitin and other partners [50]. The analysis of protein aggregation in vitro shows decelerated protein deposition after the addition of ubiquitin into the solution. Therefore these strategies to stabilize functional contacts also reduce the propensity for initiating aggregation [21]. In this light, Pechmann et al. [21] studied the factors that determine the formation of macromolecular assemblies by analysing the structure of 797 different protein complexes. They show that more than 80% of the complexes studied have disulfide bonds and salt bridges at or close to aggregation-prone interfaces, thereby favouring the correct orientation and contact formation associated with their native conformation. They also find that all complexes studied have interfaces with charged residues close to hydrophobic regions. Similarly, previous studies have shown that charged residues (aspartate, glutamate, arginine and lysine) and proline can act as gatekeepers when flanking aggregation-prone regions by promoting repulsive electrostatic interactions or by destabilizing regular structures respectively [15,16,51].

In humans, it has been shown that specific mutations disturbing the interface region can lead to diseases, either by impeding the intermolecular interactions or by favouring aggregation [21,22]. For instance, the analysis of four mutant variants associated with familial transthyretin amyloidosis disease (V122I, L55P, A25T and V30M) reveals that, for...
the first three, the quaternary homotetrameric structure of transthyretin is destabilized, whereas the fourth mutation leads to an unusual dissociation mechanism that predisposes the protein to aggregation [52]. Similarly, it was shown that several mutations related to ALS (amyotrophic lateral sclerosis) provoke aggregation of SOD1 (superoxide dismutase 1) by increasing its propensity to expose hydrophobic surfaces [53].

**Cellular regulation of protein aggregation**

As discussed above, cells need to preserve aggregation-prone regions, but, at the same time, they need to prevent the toxic consequences associated with protein aggregation. With this purpose, cells have developed mechanisms to prevent aggregation by various means in the first place, and minimize their damage by either actively removing or dissolving the aggregates (Figure 2).

Protein aggregation is a second- or higher-order reaction and is therefore highly dependent on protein concentration [54–56]. As a result, one strategy to avoid aggregation includes strong regulation of the amount of protein in the cell by preserving a fine balance between protein expression and degradation [57]. The cellular network of mechanisms designed to achieve this equilibrium is known as the proteostasis network [58]. In this context, Tartaglia et al. [59] compared the expression levels of human genes with proteostasis network [58]. They found that the mRNA abundance correlates negatively with protein aggregation propensity, suggesting that the most abundant proteins have been selected to be less aggregation-prone [59]. Along these lines, it has been hypothesized that these proteins can have a global effect in decreasing aggregation propensities in the cell due to their high concentrations and low aggregation propensity [44]. Also, De Baets et al. [11] observed that short-lived proteins are, on average, more prone to aggregate and display fewer interactions with chaperones than long-lived proteins. Consistently, a higher number of short-lived proteins are linked to conformational diseases [11]. Also, secretory mechanisms such as dedicated transporters can prevent the accumulation of potentially toxic deposits within the cell by transporting them outside of the cell [60]. For example, it is known that a set of transporters are involved in the clearance of Aβ (amyloid β-peptide) through the blood–brain barrier. Consistent with this idea, a decrease in the activity of the ABC (ATP-binding cassette) transporter ABCC1 has been associated with an increase of cerebral Aβ levels [60].

In addition to these mechanisms, cells are equipped with a versatile quality control machinery, chiefly composed of chaperones, proteasome and proteases [10,61–64]. Chaperones assist and accelerate the protein folding process, hence reducing the time that proteins remain completely or partially unfolded [10]. Additionally, chaperones can protect nascent polypeptide chains from making non-native interactions and hold proteins, ensuring their solubility, during their lifetime [10]. From an evolutionary point of view, chaperone activity permits evolvability by tolerating protein mutations and minimizing their effect on protein misfolding [65,66]. However, under certain circumstances (e.g. stress conditions), such mechanisms may be insufficient or excessive [67], and the resultant deposits must be removed to ensure cell viability. In some cases, these deposits can be disaggregated [68], resolubilized and later degraded by the proteasome [69]. However, when aggregates cannot be disassembled through all of these mechanisms, cells may employ autophagy to remove the internal protein deposits [70,71].

Another robust strategy employed by eukaryotic cells to prevent misfolding and aggregation is compartmentalization, which provides specific environments and restricts spurious interactions [9,50]. For example, the oxidizing environment and special chaperones of the ER (endoplasmic reticulum) lumen are necessary for the correct folding of certain proteins [9].

Especially in unicellular organisms such as bacteria, cell division results in an asymmetric segregation of protein aggregates which generates one cell inheriting the majority of deposits, with limited reproductive ability, and a rejuvenated cell mostly free from aggregates [72]. However, in higher organisms, despite all the aforementioned strategies, unwanted protein products may accumulate with age, principally due to a gradual increase in damaged material and inability of the proteostasis machinery to deal with increased damage over the cell’s lifespan, especially in post-mitotic cells such as neurons [70]. In this context, it is worth mentioning that maintenance of the proteostasis network incurs an energetic cost to the organism and may result in delayed maturation to the reproductive phase, which suggests that the requirements of reproduction and somatic maintenance must be balanced to maximize fitness, particularly in multicellular organisms [70]. In humans, modulating the proteostasis network using drugs that target its components has been suggested as a promising therapeutic strategy for the treatment of conformational diseases and to favour a longer lifespan [58,70,73].

**Conclusions**

In the present paper, we have discussed how undesired protein aggregation can have negative consequences for the cell. However, functional aggregation-prone regions and aggregates are essential for life. The importance of both in concert becomes evident when examining the Arf (ADP-ribosylation factor) and Hdm2 (human double minute 2) pair of proteins, which interact through aggregation-prone unstructured regions. This results in the formation of amyloid-like fibrils with the Arf–Hdm2 complex as a unit. This depletes the cell of soluble Hdm2, which is then not able to perform its function as a p53 inhibitor, resulting in a regulatory change [74]. Hence, life appears to have taken the risk to conserve the aggregation-prone sequences by investing in safeguarding strategies to constrain the protein aggregation process.

It is now well established that the prevention of protein aggregation is an important evolutionary force that shapes
Funding

This work was supported by the Medical Research Council [grant number U105185859]. We also thank the Federation of European Biochemical Societies (FEBS) (N.S.d.G.), Human Frontier Science Program (HFSP) (M.M.B.) [grant number RGY0073/2010], EraSysBio + (M.M.B.), Ministerio de Ciencia e Innovacion-Spain [grant number BFU2010-14901 (to S.V.)], from Agencia de Gestio d’Ajuts Universitaris i de Recerca (AGAUR) [grant number 2009-SGR 760 (to S.V.)] and Institucio Catalana de Recerca i Estudis Avancats (ICREA)-Academia (S.V.), Herchel Smith Research Studentship Fund (B.L.). PriorNet Canada and Canadian Institutes of Health Research (CIHR) (J.G.) for funding our research.

References