PART I

PRINCIPLES OF PROTEIN MISFOLDING
WHY PROTEINS MISFOLD

SILVIA CAMPIONI,* ELODIE MONSELLIER,* AND FABRIZIO CHITI
Dipartimento di Scienze Biochimiche, Università di Firenze, Firenze, Italy

INTRODUCTION

Protein misfolding, which literally means “incorrect folding,” is associated with a number of pathological states in humans, collectively termed protein misfolding diseases. In some cases, the disease arises because a specific protein is no longer functional when adopting a misfolded state or undergoes a severe trafficking impairment (loss-of-function diseases). In most diseases, however, the pathological state originates because misfolding occurs concomitantly with aggregation, and the underlying aggregates are detrimental (gain-of-function diseases). The question of why proteins misfold and self-assemble into well-organized fibrillar aggregates characterized by an extensive β-sheet structure, generally referred to as amyloid or amyloidlike fibrils, is very complex and cannot be answered exhaustively even in several pages. Despite these difficulties, it is fascinating to try to address this issue, particularly by describing a few concepts that are well established, have general validity, and benefit from a general consensus in the scientific community. In this challenge we start by describing our current knowledge of why and how proteins misfold in vitro. Considerable attention will be paid to the conformational changes that represent the starting point of misfolding phenomena and to the sequence determinants that enable the resulting misfolded, yet conformational states to

* These authors contributed equally and are listed in alphabetical order.
self-assemble into ordered fibrillar aggregates. Finally, we mention all the factors that are thought to influence protein misfolding in an in vivo context.

WHY PROTEINS MISFOLD IN VITRO

A newly synthesized polypeptide chain usually folds into a functional, globular conformational state characterized by a well-defined, persistent secondary and tertiary structure (Fig. 1). In this conformation, generally termed the **folded or native state**, a protein is generally stable and soluble and has a minimal propensity to undergo aberrant aggregation, particularly because most of its hydrophobic moieties and a large portion of the backbone are sequestered inside the protein (Fig. 1). Even the regions located on the protein surface that may potentially retain a residual ability to trigger an undesired intermolecular association between folded protein molecules, such as the peripheral strands of β-sheets, are protected by structural adaptations developed during evolution [1]. Therefore, the first and necessary event that generally triggers the aggregation process of a normally globular protein into structured amyloidlike fibrils is the adoption of a nonnative partially or globally unfolded state [2,3].

Aggregation from Equilibrium Partially Folded States

It is well known to experimentalists that the easiest way to induce amyloid fibril formation of a globular protein in vitro is to place the protein into environmental conditions that promote partial or substantial unfolding of the native state but still allow the formation of noncovalent bonds [4–10]. These conditions, which include extreme pH values, high temperatures, high pressures, and aqueous solutions with organic cosolvents, give rise to partially unfolded states that are significantly populated at equilibrium, in which most of the hydrophobic groups and backbone amide and carbonyl groups, normally buried in the interior of the protein, become solvent exposed. Such newly exposed moieties possess a sufficiently high conformational flexibility and are available for intermolecular interactions with other protein molecules to form the cross-β structure typical of amyloid fibrils.

Amyloid fibril formation can also be promoted, under conditions that are optimal for the conformational stability of the native state, by mutations or truncations that destabilize the folded state and thus cause the formation of analogous partially unfolded states [11–16]. Along the same lines, it has been shown that the propensity of a protein to form amyloidlike fibrils can be reduced significantly by stabilization of the native structure via specific binding to ligands or antibodies [17–21]. These observations have increased the interest in a detailed in vitro structural characterization of equilibrium misfolded states.

Among the partially folded states characterized in more detail are those populated at acidic pH at equilibrium, before aggregation occurs, by β₂-microglobulin, the protein involved in dialysis-related amyloidosis.
The combination of multiple techniques has allowed a deep analysis of the structural properties of partially folded states that form fibrils with different morphologies [22]. These studies have highlighted the complex, dynamic, and heterogeneous nature of these states, with different conformations interconverting into each other [23,24]. They have also shown that the amyloidogenic precursor states of this protein are hydrophobically collapsed species that contain different degrees of secondary and tertiary structure and lack the cooperativity typical of the native state [22].

**Folding Intermediates as Precursors of Protein Deposition Diseases**

Advances in experimental methods able to detect rare species and monitor folding events on a microsecond time scale have shown that partially folded...
states normally accumulate during folding before the major energy barrier, even in small proteins [25,26]. It has recently been shown that these species play a key role in determining the diversion of molecules between the folding and aggregation regions of the combined energy landscape (Fig. 1) [27,29]. Indeed, if the minima occupied by intermediates in the aggregation landscape become too deep, the likelihood of off-pathway events such as protein aggregation increases [25,30].

Recent data on the human prion protein (hPrP), whose aggregation induces fatal and infectious neurodegenerative disorders, suggest that the conformational conversion into the infectious form could be mediated by a folding intermediate [31,32]. This transient species was found to have a higher stability than the fully unfolded state and to be populated to a greater extent, relative to the wild-type protein, in many variants carrying mutations associated with inherited forms of prion diseases, including those that do not alter the global stability of the prion protein [31].

Native-State Fluctuations Can Trigger Amyloid Aggregation

The native state of a protein is not a single, rigid conformation, but rather, undergoes multiple localized unfolding reactions that give rise to an ensemble of rapidly interconverting conformations fluctuating around a free-energy minimum [30,33–35]. Some of these conformations, or ensembles of them, can also be regarded as “late intermediates” formed after the rate-limiting step of folding and remaining accessible by conformational fluctuations after the native state is fully reached [36–38]. Fluctuations that expose significant regions of the polypeptide main chain and its hydrophobic core could represent an initial step in the development of strong intermolecular interactions [39].

Two amyloidogenic variants of human lysozyme that cause autosomal-dominant hereditary amyloidosis, I56T and D67H, alter the cooperativity of the folded protein and enhance the rates of fluctuations toward partially unstructured conformations while preserving a global nativelike structure [11,39,40]. The most frequent sampling of such species, in which only the β-domain and the adjacent C-helix appear predominantly unfolded, makes the variants susceptible to aggregation. Fluctuations of the native state of monomeric transthyretin (TTR) and superoxide dismutase type I (SOD-1), associated with familial amyloid polyneuropathy and amyotrophic lateral sclerosis, respectively, cause the edge β-strands of these all-β proteins to unfold locally [41,42]. The resulting conformations have been shown to occupy the right side of the major free-energy barrier for folding and were therefore considered both as late-folding intermediates and conformational states freely accessible from the native structures through thermal fluctuations [42].

A late-folding intermediate of β2-microglobulin, populated under native conditions and in equilibrium with the native structure, has been shown to be competent for amyloid fibril elongation [27,28,43]. The transition from the native state to this late intermediate is caused by the isomerization of a single
cis-proline residue, but causes local unfolding of the A and D edge \(\beta\)-strands [27]. The acylphosphatase from *Sulfolobus solfataricus* (Sso AcP) has also been shown to aggregate at a higher rate than that of unfolding under conditions in which the native state is populated [44]. Such conditions have been shown to increase the fluctuations of the native state, as revealed by hydrogen/deuterium exchange detected with mass spectrometry, indicating that aggregation requires fluctuations within the native-state ensemble, but not necessarily unfolding [45].

**Intrinsically Disordered Proteins Are at Risk for Aggregation**

It has been recognized that numerous proteins lack a specific globular structure, at least when they are not associated with their specific molecular targets, and that many others contain disordered regions in addition to folded domains [46–48]. Such intrinsically disordered proteins (IDPs) have developed sequence adaptations to escape from aggregation, including a low hydrophobic residue content and a high net charge, a high content of proline residues, and a low number of aggregation-promoting regions [47,49,50]. In addition, induced folding upon binding to target molecules and interaction with membranes or macromolecules can provide further protection from aberrant aggregation [51,52]. Despite these protective mechanisms, IDPs are usually more prone than globular proteins to aggregate, as they cannot take advantage of folding. Therefore, it is not surprising to see that many proteins and peptides associated with protein-deposition diseases are indeed disordered under physiological conditions. Examples include the amyloid-beta peptide (A\(\beta\)), \(\alpha\)-synuclein, tau, islet amyloid polypeptide (IAPP), calcitonin, and gelsolin fragments [53]. Even the monomeric precursor protein of prion diseases (PrP\(^c\)) is composed by a long N-terminal disordered region in addition to a compact C-terminal domain [54].

**THE DETERMINANTS OF PROTEIN AGGREGATION FROM LARGELY OR PARTIALLY UNFOLDED STATES**

We have shown that some degree of unfolding is the first prerequisite for initiating aggregation. But how do largely or partially unfolded states self-assemble into fibrillar aggregates? While it was being realised that most polypeptide chains have an inherent ability to form amyloidlike fibrils from their partially unfolded states, it was also becoming clear that this conversion occurs at different rates and with different efficiencies. Today it is well known that the tendency to form aggregates of similarly unstructured chains is indeed sequence dependent and that specific stretches of a sequence can facilitate or counteract amyloid formation [53]. Great effort has been expended to find simple rules that govern the aggregation process in order to rationalize and even predict aggregation rates and/or propensities and identify aggregation-prone segments within unfolded polypeptide chains [55–70].
Aggregation Is Promoted by High Hydrophobicity and β-Sheet Propensity and Low Net Charge

In a first attempt to analyze quantitatively the effects of mutations on aggregation, it was observed that the aggregation rates of variants of a model protein (AcP), determined under conditions in which an equilibrium denatured state is populated, depend on the changes caused by these mutations on simple physicochemical properties of the polypeptide chain: hydrophobicity, net charge, and propensity to convert from an α-helical to a β-sheet conformation [53]. On this basis, an empirical equation to predict the effect of a mutation on the aggregation rate of an unstructured peptide or protein was proposed and validated on a set of 27 single-point mutants of various unfolded proteins [53]. A conceptually similar algorithm was developed by Tartaglia and co-workers [59]. This algorithm uses charge, β-propensity, and parameters related to hydrophobicity (the polar and nonpolar water-accessible surface areas, the dipole moment, and the π-stacking interactions of aromatic residues) as determinants of amyloid aggregation. Their equation was able to reproduce the relative aggregation propensity of a set of variants of both disease-related and model proteins or peptides [59]. Fernandez-Escamilla and co-workers independently developed a statistical mechanics algorithm based on secondary structure propensities and estimation of a desolvatation penalty (TANGO) to predict β-sheet aggregating regions of a protein sequence as well as mutational effects [57]. TANGO evaluates the percentage of occupancy of the aggregate conformation for every residue of the chain by considering that the secondary structure of the aggregate is a β-structure and that the regions involved in aggregation are fully buried [57].

The established concept, that simple physicochemical parameters of the amino acid residues of an unstructured protein influence its aggregation propensity has encouraged the search for novel and increasingly accurate algorithms. The algorithms that are available so far can predict not just the effect of a mutation on the aggregation rate of a protein, but also the absolute aggregation rate or propensity of an unstructured system [56,57,63] and, more important, the regions of the sequence that promote aggregation and form the β-core of the resulting fibrils [57,62,66,69]. All these algorithms are based on the same physicochemical parameters, or related ones, as those used in the original formula and seem to be successful in yielding predictions with reasonable accuracy.

From these studies it emerges that the sequence segments promoting aggregation of unstructured polypeptide chains have a hydrophobicity and β-sheet propensity higher than those of other regions. Mutations located in such regions can increase the aggregation propensity if they increase such factors. Interestingly, the locomotor efficiency and longevity of transgenic flies expressing mutant human Aβ(1–42) have been shown to correlate with the aggregation propensity of the expressed Aβ(1–42) variants, as deduced from one of these algorithms, indicating that these principles are largely applicable to in vivo systems [71].
Other Promising Predictive Methods

Predictive methods based on physicochemical parameters are not the only computational tools available to predict aggregation propensities and aggregation-promoting regions in a sequence. Other methods exploit empirical scales of aggregation propensities of individual amino acids based on purely experimental data [61,72], consensus sequences extracted from systematic mutagenesis of peptides [58], structural properties of protein databases [60,64,66], ensembles of structural templates derived from high-resolution amyloid fibril structures [65], or molecular dynamics simulations of protein aggregation [73,74]. Interestingly, in cases where the aggregation propensity scales of individual amino acids were determined, clear analogies exist between the results of very different approaches, ranging from empirically derived to statistical approaches [62,64,72]. This general agreement further confirms that common generic principles determine the molecular basis of amyloid aggregation.

WHY PROTEINS MISFOLD IN VIVO

Studies of protein aggregation in vitro take place in defined and extremely simple environments where important parameters such as pH, temperature, and ionic strength are controlled and modulated at order, the time scale is limited to a few weeks at most, and the only molecular species present in solution is the protein under study. However, the natural environment of a polypeptide chain—the cellular cytosol, the organelles, or the extracellular matrix—is far more complex and variable. We mention briefly the factors that are thought to contribute to protein misfolding in vivo. Several chapters of the book are dedicated to these factors. We show that although some factors causing misfolding in vitro are also able to promote aggregation in vivo, a further level of complexity lies in the biological environment in which proteins are produced and operate (Fig. 2).

Mutations Induce Aggregation

Many of the protein-deposition diseases consist of familial forms in which a mutation leads to aggregation of the mutated protein and the subsequent onset of disease. Examples include lysozyme amyloidosis and familial amyloidotic polyneuropathy [53]. Other diseases, such as amyotrophic lateral sclerosis and Creutzfeldt–Jakob disease, are mainly sporadic but include a few cases of familial forms [53]. Some of these mutations destabilize the native state of a globular protein, resulting in the formation of partially unstructured, aggregation-prone conformations (Fig. 2) [11,12,15,16,40]. Others increase the propensity to aggregate of the fully or partially unfolded state of a globular protein or of the natively unfolded state of an intrinsically disordered protein [55,75].
The pathogenicity of other mutations lies in the biological context of the protein that undergoes aggregation and of its in vivo processing. For example, it has been suggested that some mutations of tau cause its dissociation from microtubules, resulting in an increase in free, intrinsically disordered tau that is susceptible to aggregation [76]. Mutations in the amyloid precursor protein APP increase the propensity for cleavage by \( \gamma \)-secretase, yielding a higher ratio of the more amyloidogenic \( \alpha\beta(1–42) \) with respect to the less amyloidogenic \( \alpha\beta(1–40) \) [77]. The D187N and D187Y mutations of gelsolin abolish the ability of domain 2 to bind calcium; the destabilized domain is more susceptible to proteolysis by furin, contributing to yield the amyloidogenic fragments of gelsolin [78].

**Increase in Protein Concentration Contributes to Aggregation**

As aggregation is a second- or higher-order reaction, its rate is very sensitive to the concentration of interacting chains. In vitro, increasing protein concentration decreases the lag time of aggregation and increases the elongation rate of the fibrils [79,80]. Different lines of evidence demonstrate that an increase in the in vivo concentration of a protein can also promote its aggregation. Neurodegeneration and expression levels of certain disease-related proteins, such as \( \alpha\beta \) and \( \alpha\)-synuclein, are clearly correlated [81]. In dialysis-related amyloidosis, the concentration of \( \beta_2 \)-microglobulin increases up to 50-fold due to kidney failure, and this contributes to amyloid deposition [82]. Increase in the local concentration of \( \beta_2 \)-microglobulin by collagen is thought to induce amyloid fibril formation on the surface of collagen fibrils and to explain the tissue specificity of dialysis-related amyloidosis [83]. Secondary systemic amyloidosis, also known as AA or reactive amyloidosis, is caused by increased levels of the serum amyloid A protein as a consequence of chronic inflammatory
reactions [84]. In addition, increase in the local concentration of a membrane-bound protein is one mechanism by which biological membranes promote protein aggregation [85].

**Unfolded States Form Obligatorily for Globular Proteins In Vivo**

The occurrence of fully or partially unfolded states during the lifetime of a normally folded protein is certainly more frequent than that for the same protein in a physiological buffer and is thus a risk factor for protein aggregation in vivo. As long as translation is not terminated, a nascent polypeptide chain is susceptible to remaining unfolded in the milieu. Ongoing translation therefore offers a continuous supply of aggregation-prone species [86]. The presence of highly organized chaperone machinery coupled directly to translation witnesses to the potential deleterious effects of translation [87,88]. Secretory, mitochondrial, and membrane proteins are translocated through narrow channels and therefore need to be unfolded beforehand. Again, dedicated chaperone machinery exists to assist in these processes [89,90]. Many other circumstances can induce the transient unfolding of a protein, including the rise of stress conditions, and interaction with, or dissociation from, molecular targets (Fig. 2).

**Oxidative Stress Promotes Aggregation Through a Variety of Mechanisms**

Another biological factor that is potentially triggering aggregation in vivo is oxidative stress, defined as an imbalance between oxidant generation and antioxidant systems. Reactive oxygen species (ROS) are among the most commonly formed oxidants in biology and are generated as a by-product of normal metabolism and/or by exogenous stimuli such as ultraviolet light [91]. ROS can damage all biological polymers, including proteins. Oxidation of a protein in vitro can promote its aggregation via a variety of mechanisms, including destabilisation of the native state, whose ultimate effects were described earlier in this chapter (Fig. 2) [92]. Oxidation of a protein in vivo can also promote aggregation by additional mechanisms, such as by causing cross-linking [97] or by modifying the susceptibility to proteolysis [94]. However, the exact relationship among oxidative stress, amyloid deposition, and disease onset remains unclear.

**Macromolecular Crowding Favors Self-Assembly**

In a typical eukaryotic cell protein, RNAs and others biopolymers occupy 30% of the volume, a situation known as macromolecular crowding. The direct consequence of crowding, which also exists in the extracellular space, albeit to a lesser extent, is that little space is left for additional macromolecules, which reduces their configurational entropy and therefore increases their free energy [95]. As a consequence, aggregation is theoretically favored, both kinetically and thermodynamically (Fig. 2) [95]. Macromolecular crowding has been
shown to accelerate the aggregation of a wide range of proteins in vitro, leading to the suggestion that natural aggregation processes could be more favored than actually anticipated from studies performed in the test tube [95].

**Interaction with Natural Compounds Promotes Fibril Formation**

The excluded-volume effect is not the only mechanism by which macromolecules and other natural compounds influence protein aggregation in living organisms. Macromolecules such as glycosaminoglycans (GAGs), the serum amyloid protein (SAP), apolipoprotein E, and collagen fibers are frequently associated with amyloid deposits under pathological conditions [83,96–100]. Such species alter the kinetics of fibril formation via a direct interaction with the aggregating protein, thus sometimes being qualified as “pathological chaperones” [83,101–104]. They can also alter the susceptibility of the resulting fibrils to dissociation or proteolytic degradation [105–108]. Biological membranes also play a catalytic role in the aggregation of numerous amyloidogenic proteins (Fig. 2) [85,109].

These effects are not limited to macromolecules. Metal ions, particularly Cu$^{2+}$ and Zn$^{2+}$ ions, are known to accelerate fibril formation by the Aβ peptide [110], α-synuclein [111], and β2-microglobulin [28]. In some cases, this aggregation is reversible upon chelation [110,112]. Fe$^{3+}$ is found at abnormally high levels in Lewy bodies, suggesting that this metal ion also plays a significant role in Parkinson disease [113].

**Aging Enhances the Probability of Protein Aggregation**

Most protein-deposition diseases are of late onset. It is therefore obvious to draw the conclusion that aging increases the occurrence of physiological protein aggregation and is a key determinant of the pathological conditions associated with this phenomenon. Experimental evidence has been collected in animal models on the link between aging and protein aggregation [114–116]. A range of biological alterations occur, with the progression of a human being’s life, that favor protein aggregation. These include a decline in the proteasome, lysosome, and chaperone activities and in the ability of these cellular machineries to respond to stress, dysregulation of metal-ion homeostasis, diminution of the antioxidant defenses, and so on. (Fig. 2) [116–118].

Although plausible, a decrease in the biological defenses against aggregation is not the only explanation that links pathology with aging. An aging person faces the progressive overwhelming of the cellular folding capacity and quality control systems by the inexorable accumulation of misfolded proteins (Fig. 2). In a Caenorhabditis elegans model of polyglutamine (polyQ) pathologies, polyQ aggregation increases the misfolding of otherwise soluble proteins, which in turn enhances further polyQ aggregation [119]. Therefore, it is likely that the protective mechanisms become insufficient with aging and increase the probability of further misfolding events.
CONCLUSIONS

Understanding why proteins misfold into amyloid or structurally related aggregates is of extreme importance. Although much remains to be understood, considerable progress has been achieved over the past 20 years. Some of the factors causing protein misfolding and aggregation in vitro are actually operating in vivo as well and are likely to play important roles in the pathogenesis of disease. These include destabilization of the folded states of globular proteins, increase in the intrinsic propensity of the unfolded states to aggregate, and increase in protein concentration. However, the existence of additional factors in vivo, such as oxidative stress, macromolecular crowding, interaction with biological macromolecules, and aging, adds complexity to the description of the causes of protein aggregation in vivo and of each of the pathological conditions associated with it.

REFERENCES


REFERENCES


