Chemistry & Biology
Review

# Protein Misfolded Oligomers: Experimental Approaches, Mechanism of Formation, and Structure-Toxicity Relationships

Francesco Bemporad[1,*] and Fabrizio Chiti[2,*]
[1]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK
[2]Dipartimento di Scienze Biochimiche, Università degli Studi di Firenze, Viale Morgagni 50, 50134 Firenze, Italy
*Correspondence: fb330@cam.ac.uk (F.B.), fabrizio.chiti@unifi.it (F.C.)
DOI 10.1016/j.chembiol.2012.02.003

The conversion of proteins from their native state to misfolded oligomers is associated with, and thought to be the cause of, a number of human diseases, including Alzheimer's disease, Parkinson's disease, and systemic amyloidoses. The study of the structure, mechanism of formation, and biological activity of protein misfolded oligomers has been challenged by the metastability, transient formation, and structural heterogeneity of such species. In spite of these difficulties, in the past few years, many experimental approaches have emerged that enable the detection and the detailed molecular study of misfolded oligomers. In this review, we describe the basic and generic knowledge achieved on protein oligomers, describing the mechanisms of oligomer formation, the methodologies used thus far for their structural determination, and the structural elements responsible for their toxicity.

## Introduction

Polypeptide chains have a high intrinsic propensity to self-assemble into a variety of misfolded aggregates, ranging from dimers to highly organized fibrils consisting of thousands of protein molecules (Chiti and Dobson, 2006; Dobson, 2003; Eichner and Radford, 2011; Jahn and Radford, 2008). Protein aggregation can occur starting from any of the conformational states adopted by the initially monomeric protein, including the fully unfolded state, the folded state, and any partially folded states, although the latter have the highest propensity to self-assemble (Bemporad et al., 2006; Chiti and Dobson, 2006). Protein aggregation is deleterious for any living organism as it prevents a protein from adopting its functional state and because the resulting protein oligomers are inherently toxic (Chiti and Dobson, 2006; Walsh and Selkoe, 2007). In fact, the failure of proteins to remain soluble—occurring as a consequence of mutation, aging, local increases of protein concentration, medical treatment, or other circumstances—results in pathological states that are very diverse and depend on the protein undergoing aggregation and on the tissues involved (Chiti and Dobson, 2006).

The end product of protein aggregation processes occurring in pathology is generally represented by extracellular amyloid fibrils or structurally equivalent intracellular inclusions, often associating further to form larger assemblies that are visible with optical microscopy. However, protein oligomers forming early during the process of amyloid fibril formation or, alternatively, released by mature fibrils, have acquired increasing importance over the past 10–15 years. This is first due to the fact that formation of protein oligomers is a key event of the overall process of amyloid fibril formation and has been long regarded as the rate-limiting step, responsible for the lag phase in aggregation kinetics (Morris et al., 2009; Orte et al., 2008). Second, oligomers are thought to be the pathogenic species associated with the formation of amyloid in diseases (Billings et al., 2005; Bucciantini et al., 2002; Cleary et al., 2005; Koffie et al., 2009; Lesné et al., 2006; Winner et al., 2011). As the importance of protein oligomers was increasingly realized, many reports have appeared with information on their mechanism of formation and structure, with the ultimate goal of identifying the structural determinants of their pathogenicity, the molecular events of disease onset, and the molecular targets for therapeutic intervention. In this review, we describe how oligomers form, the structure of such species, including the technological progress recently achieved to gain insight into their molecular structure and the structural elements responsible for their toxicity.

## Mechanism of Formation of Amyloid Oligomers

One of the most widely accepted mechanisms proposed for the assembly of monomers into oligomers is the so-called "nucleation growth" mechanism (Jarrett and Lansbury, 1993; Lomakin et al., 1996). According to this mechanism, monomers convert into a nucleus through a thermodynamically unfavorable process taking place in the lag phase of amyloid aggregation kinetics (pathway A → C/D → H → G; red arrows in Figure 1). The nucleus can be defined as the least thermodynamically stable species in solution (i.e., the multimer of minimal size able to initiate assembly; Morris et al., 2009). Alternatively, the nucleus can also be defined as the aggregate size after which the association rate exceeds the dissociation rate for the first time (Ferrone, 1999). The nucleus could even be a monomer that acts as a template for the rapid growth of the amyloid aggregate through the association of further monomers (H → G → I in Figure 1) (Lomakin et al., 1996).

After the nucleation growth model was proposed, the concept of nucleus was investigated in further detail and other more accurate models were proposed. In the nucleated conformational conversion, native monomers initially convert into misfolded conformations, which initiate self-assembly through a

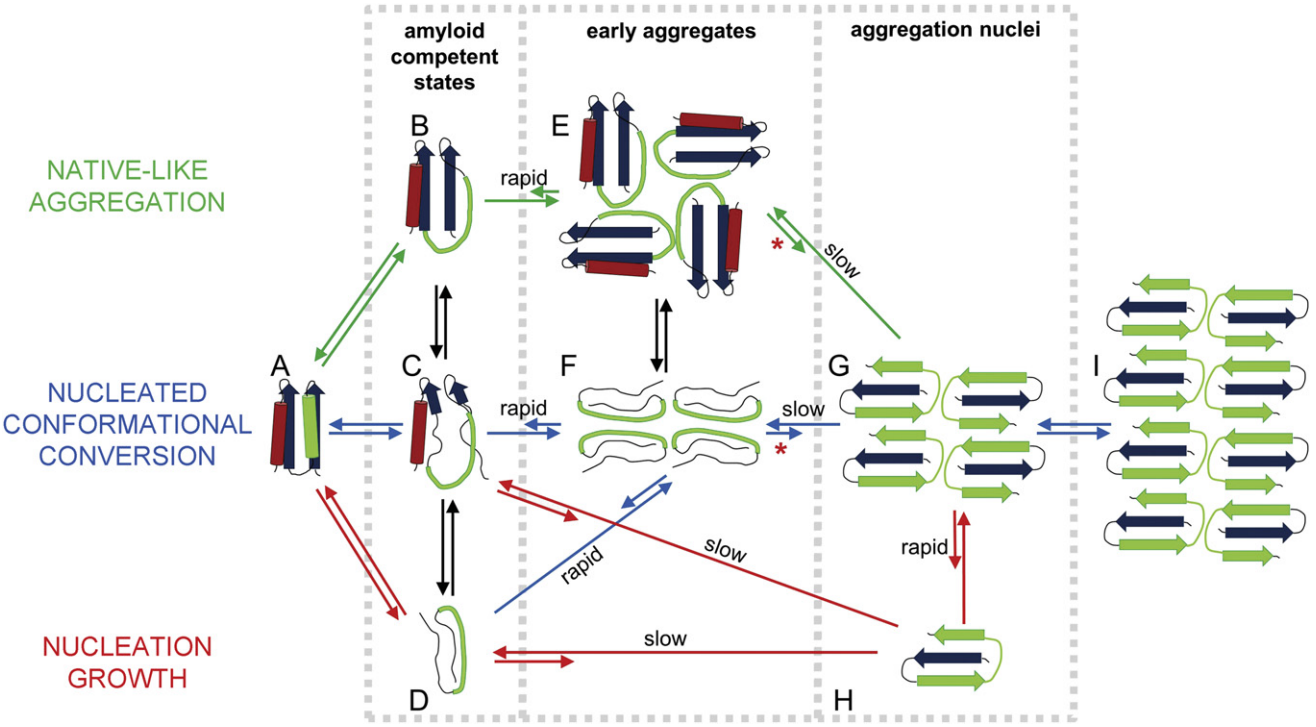


**Figure 1. Mechanisms Leading to the Formation of Protein Oligomers**

Proteins initially populate a native conformation (A) in which amyloid-prone segments (in green) are structured/buried and unable to initiate polymerization. However, native states can convert, under certain conditions, into aggregation-prone states, a native-like (B), a partially folded (C), and an unfolded (D) monomer. In these conformational ensembles, aggregation-prone segments become exposed to the solvent and can establish intermolecular interactions, with the resulting formation of early aggregates, which can be native-like aggregates (E) or molten aggregates of misfolded monomers (F). The early aggregates convert later into amyloid oligomers competent for fibril nucleation (G). The pathway that transits through the native-like state (A → B → E → G, indicated by green arrows) is referred to as native-like aggregation. The pathway that transits through fully or partially unfolded monomers (A → C/D → F → G, indicated by light blue arrows) is the nucleated conformational conversion. There is finally another possible pathway, in which a misfolded monomer converts into an amyloid-competent monomer (A → C/D → H, indicated by red arrows). This acts as a template for formation of the oligomer (G). This pathway is usually referred to as the nucleation growth mechanism. The red asterisks denote two steps (E → G and F → G) whose removal makes oligomer formation an off-pathway process that leads to species unable to further convert into amyloid fibrils.

template-independent mechanism, with formation of a molten oligomer lacking persistent structure (A → C/D → F; light blue arrows in Figure 1) (Serio et al., 2000). This aggregated species then undergoes a structural reorganization into an amyloid-like oligomer, which acts as a nucleus (F → G). The nucleus rapidly triggers aggregation as other molten oligomers acquire the amyloid conformation through a templating or induced-fit mechanism at the aggregate ends (Serio et al., 2000). This leads to the formation of higher order oligomers and eventually fibrils (G → I).

These two models differ for a number of aspects. While in the nucleation growth mechanism, the monomer directly interacts with the nucleus and the nucleus could be as small as a monomer, in the nucleated conformational conversion both the building block and the nucleus are described as oligomeric species. Moreover, in the first model the rate-limiting step is represented by the formation of the nucleus (C/D → H in Figure 1), whereas in the second model oligomers form rapidly, with the rate-limiting step being the conversion of the misfolded oligomer into the amyloid oligomer (F → G in Figure 1). Different reports support either model. In many cases, the nucleation growth mechanism seems to account for the observed aggregation kinetics (Bhattacharyya et al., 2005; Morris et al., 2008). Furthermore, proteins populating amyloid conformations directly at the monomer level have been described (Sandal et al., 2008), while some authors have reported (e.g., in the case of insulin) template-dependent conversion of monomers at the growing fibril end (Pease et al., 2010). However, it was recently shown that the monomers can rapidly form small globular oligomers that are kinetically competent to slowly convert to amyloid oligomers and later to amyloid fibrils following the nucleated conformational conversion. This was observed for example in ADA2h (Cerdà-Costa et al., 2007), Aβ (Lee et al., 2011; Petty and Decatur, 2005), huntingtin (Thakur et al., 2009), and IAPP (Wei et al., 2011).

In fact, it has been observed that formation of oligomers by a protein can occur through different pathways. In some cases, a certain pathway can be enhanced by changes in solution parameters such as pH and protein concentration (Bader et al., 2006; Gosal et al., 2005) or by introducing mutations (Bitan et al., 2003; Kumar and Udgaonkar, 2009). In other cases, aggregation can proceed via competing pathways occurring concomitantly in the same sample (Jain and Udgaonkar, 2011; Kayed et al., 2007; Kaylor et al., 2005). Differences in pathways imply differences in morphology (Bader et al., 2006; Gosal et al., 2005), dimensions (Bitan et al., 2003), and compactness (Kaylor et al., 2005) of the resulting oligomers. It is important to note that

alternative pathways differ for the stage at which conformational conversion from a nonamyloid to an amyloid conformation occurs (Kumar and Udgaonkar, 2009). Thus, the two models described earlier could be the limit cases of a more complicated scenario where multiple pathways are accessible and selected depending on the conditions.

A possible unifying view able to combine the two apparently competing models described earlier is that the pathway prevalently followed depends on the structural nature of both the amyloid-competent monomer and the oligomer (Figure 1). If the amyloid-competent conformational ensemble has a sufficient degree of dynamical fluctuations and hydrophobic clusters exposed to the solvent, then formation of molten oligomers lacking persistent structure is fast, while conversion of the molten oligomer into an amyloid oligomer takes place only later and is accelerated by the template effect (nucleated conformational conversion; light blue arrows in Figure 1). If the misfolded conformational ensemble has not sufficient hydrophobic clusters exposed to the solvent and/or is prone to hydrogen bonding, then the other mechanism is likely to take place (nucleation growth; red arrows in Figure 1). Consistently, it was recently shown that short hydrophobic peptides give rise prevalently to misfolded oligomers, while peptides prone to establish intermolecular hydrogen bonds with their side chains rapidly form β-rich nuclei (Bleiholder et al., 2011).

So far, we have considered that a folded protein needs to unfold to generate a partially or fully unfolded state to become competent for aggregation. However, it was recently shown that normally folded proteins retain a small but significant tendency to form amyloid without the need of reactions that cross the major energy barrier for unfolding. In these cases, the amyloid-competent state is best described as a monomeric conformation that possesses an extent of secondary and tertiary structure still comparable with that of the fully folded state yet being endowed with aberrant features that enable self-assembly (A → B in Figure 1, indicated by green arrows). Formation of this conformational ensemble, usually referred to as a native-like state, can be induced either as a consequence of a local unfolding event, as in the case of Cu,Zn-uperoxide dismutase (Banci et al., 2005), $\beta_2$-microglobulin (Eakin et al., 2006), and transthyretin (Olofsson et al., 2004; Quintas et al., 2001) or due to cooperative increase in thermal fluctuations, as it was shown in the case of lysozyme (Canet et al., 2002) and Sso AcP (Pagano et al., 2010). In either case, aggregation-prone segments that are normally buried or structured in the fully folded state become exposed to the solvent or gain flexibility, triggering the formation of native-like aggregates (B → E); these then convert directly into amyloid-like oligomers and fibrils (E → G → I).

## Are Oligomers On- or Off-Pathway Species?

One long debated issue is whether protein oligomers represent on-pathway structures that must be necessarily populated along the pathway leading to fibril formation or rather off-pathway particles that are placed at a dead end of the reaction scheme and act only as a reservoir of monomers, while fibril formation proceeds directly through association of monomers to the amyloid nucleus. In the case of the immunoglobulin light chain LEN, it was shown that oligomers represent off-pathway species: Fibril formation occurs only after the amyloid-incompetent oligomer dissociates and monomers convert into a more amyloidogenic conformation (Souillac et al., 2003; Souillac et al., 2002). Other off-pathway oligomers have been reported for $\beta_2$-microglobulin (Gosal et al., 2005) and albebetin (Gosal et al., 2005; Morozova-Roche et al., 2004; Souillac et al., 2002). A recent computational study has shown that the amyloid-β peptide (Aβ) can form off-pathway oligomers; these assemblies are too curvy, are too compact, and display an amount of β-structure not enough for the conversion into larger amyloid oligomers (Yu and Zheng, 2011). Despite these reports, many proteins, such as the PI3-SH3 domain, barstar, sup35p, IAPP, and Aβ itself have been shown to form on-pathway oligomers that represent either the building block or the nucleus for fibril formation (Bader et al., 2006; Ehrnhoefer et al., 2008; Harper et al., 1997; Kumar and Udgaonkar, 2009; Serio et al., 2000; Wei et al., 2011).

Of note, by varying the solution conditions, one can modulate the nature of the oligomers, switching from β-rich toxic on-pathway oligomers to unstructured nontoxic off-pathway assemblies (Ehrnhoefer et al., 2008; Ladiwala et al., 2010). Changes from off- to on-pathway oligomers can be also obtained by mutating the peptide sequence (Jain and Udgaonkar, 2011). Furthermore, it was shown, using short polyglutamine-containing huntingtin fragments, that the same sample can contain contemporaneously on-pathway and off-pathway oligomers, with the latter acting as a reservoir of monomers that support fibril elongation (Jayaraman et al., 2011). Thus, these observations may suggest a scenario where structural features of the oligomers determine their off-pathway or on-pathway nature. For example, highly stable oligomers may populate a deep minimum in the energy landscape; consequently, their conversion into an amyloid-like conformation may be slower than their dissociation. Consistently with this idea, the structural rearrangement from off-pathway to on-pathway oligomers was shown to consist in a significant increase of disordered secondary structure, an increase in solvent accessibility, and a decrease in intrinsic stability of the soluble oligomeric species (Souillac et al., 2003).

Protein misfolded oligomers do not form only as on- or off-pathway species. They can also form as species released directly by mature fibrils. In particular, it has been recently shown that mature fibrils are not necessarily highly stable end products populating the deepest minima in the energy landscape (Eichner and Radford, 2011) and that amyloid fibrils can directly release monomers and oligomers (I → G in Figure 1). It was shown, for example, that fibrils of the Aβ peptide are destabilized by the presence of lipid vesicles and brain lipid extracts, releasing toxic oligomers (Martins et al., 2008). It is interesting that these "backward" oligomers are biophysically and biochemically similar to the "forward" oligomers (i.e., to the oligomers formed along the pathway going from monomers to fibrils) (Martins et al., 2008). Leakage of Aβ oligomers from fibrils has also been observed in vivo (Koffie et al., 2009; Lesné et al., 2006). The toxic effect observed in a mouse model expressing the Aβ peptide was shown to be proportional to the amount of oligomers released from amyloid plaques surrounding cells (Lesné et al., 2006). Solubilization of transthyretin, β2-microglobulin, and insulin fibrils by doxycycline, tetracyclins, and weakly basic pH, respectively, also led to toxic oligomers (Cardoso and Saraiva, 2006; Giorgetti et al., 2011; Heldt et al., 2011b). Hence, these observations show that amyloid fibrils can be secondarily

**Table 1. Techniques Used to Study Protein Oligomers at the Molecular Level**

| Technique | Obtained information | Advantages | References |
|---|---|---|---|
| H/D exchange coupled to fragmentation methods and MS | Involvement of different protein segments in secondary structure formation | Noninvasive detection of multiple oligomers, good resolution (if fragmentation is obtained with top-down techniques) | (Kheterpal et al., 2006; Kheterpal and Wetzel, 2006; Pan et al., 2011) |
| H/D exchange coupled to both MS and NMR | Involvement of individual residues in secondary structure formation | Detection of multiple oligomers, information at the residue level, high resolution | (Carulla et al., 2009) |
| Solution NMR | Secondary and tertiary structures in low-molecular-weight oligomers | High resolution, information on both secondary and tertiary structures at the residue level | (Pagano et al., 2010; Yu et al., 2009) |
| Solid-state NMR | Secondary and tertiary structure in high-molecular-weight oligomers | Information at the residue level | (Chimon et al., 2007) |
| DEST solution NMR | Secondary and tertiary structures in oligomers | Information at the residue level, applicable to low- and high-molecular-weight oligomers | (Fawzi et al., 2011) |
| Labeling with fluorescent probes | Intermolecular interactions between labeled residues, burial of labeled residues | Information at the residue level, standard lab instrumentation required | (Campioni et al., 2010; Krishnan and Lindquist, 2005) |
| Scanning proline mutagenesis | Regions of the sequence involved in oligomer formation | Standard lab instrumentation required, good resolution | (Williams et al., 2005) |
| Small angle X-ray scattering | Simultaneous detection of size and shape | Detection of multiple oligomers | (Langkilde and Vestergaard, 2009) |
| Single-molecule fluorescence (TCCD or FRET) | Compactness, stability and size of oligomers | Direct monitoring of single oligomer particles, detection of multiple oligomers | (Orte et al., 2008, 2011) |
| Ion mobility MS | Mass and shape of oligomers | Direct monitoring of single oligomer particles, detection of multiple oligomers | (Bernstein et al., 2009; Bleiholder et al., 2011; Pease et al., 2010) |
| FRET sensor | Direct observation of oligomer formation in vivo | Noninvasive, applicable in vivo | (Kaminski Schierle et al., 2011) |

toxic via the release of cytotoxic oligomers and that fibrils could speed up oligomer formation directly acting as a reservoir of amyloid nuclei.

## Experimental Approaches to Study the Oligomer Structure

A number of biophysical and biochemical methods have been used to obtain structural information on protein oligomers. These include far-UV circular dichroism (CD), Fourier transform infrared (FTIR), and hydrogen/deuterium (H/D) exchange coupled to mass spectrometry (MS) to determine the content of β sheet and α-helical structure; ultracentrifugation, size exclusion chromatography, photo-induced crosslinking of unmodified proteins (PICUP), and dynamic light scattering to determine the oligomeric state (i.e., the amount of protein molecule in the oligomers); ANS and acrylodan binding to determine the solvent exposure of hydrophobic clusters; Thioflavine T (ThT) and Congo Red (CR) binding to determine the level of structural order; imaging techniques such as transmission electron microscopy (TEM) and atomic force microscopy (AFM) to determine the morphology and size; and so forth. Although such techniques provide valuable information on the average structure and oligomeric state of the aggregates, they do not reveal their molecular details at the residue level. This is a challenging problem as protein misfolded oligomers, similarly to amyloid fibrils, cannot be crystallized and generally have a size prohibitive for solution nuclear magnetic resonance (NMR), thus preventing the use of techniques commonly used in structural biology. In addition to the complexities encountered for fibrils, oligomers are structurally more heterogeneous and metastable than fibrils, thus raising further problems.

In the past 5 years, however, a number of approaches have been introduced for a deep molecular characterization of the oligomers (Table 1). H/D exchange followed by peptide fragmentation and MS has proven a valuable technique, exploiting the principle that backbone amide hydrogen atoms are protected to H/D exchange if they are engaged in stable β sheet or α-helical contacts. Oligomers are subjected to H/D exchange upon incubation in deuterated water ($D_2O$), fragmented with limited proteolysis or top-down techniques (e.g., electron capture dissociation) and then analyzed with MS (Kheterpal et al., 2006; Kheterpal and Wetzel, 2006; Pan et al., 2011). The determination of the extent of protection to deuterium incorporation in the fragments detected in the mass spectra has revealed the extent to which the fragments are engaged in secondary structure formation in the intact oligomers, often with a resolution down to

 

individual residues (Kheterpal et al., 2006; Pan et al., 2011). MS methods coupled to top-down techniques have the advantage of being noninvasive and photographing the solution of interest (no purification, chemical modification, or treatment of the oligomers is required), of detecting and structurally determining a multiplicity of coexisting species rather than their average structural properties, and of obtaining a spatial resolution down to individual residues (Pan et al., 2011).

In another study, H/D exchange and MS methods were coupled to NMR (Carulla et al., 2009). The PI3-SH3 protein domain was incubated at low pH in the presence of $D_2O$, under conditions favoring its rapid unfolding and subsequent aggregation. After a variable time interval ($\Delta t_{agg}$), the sample was transferred into $H_2O$ for a fixed time and the H/D exchange was then quenched by freeze drying and dissolution in DMSO, which is a denaturing and disaggregating solvent preventing H/D exchange. Samples left to aggregate for variable $\Delta t_{agg}$ time values were then analyzed with both electrospray ionization-MS (ESI-MS) and NMR. This approach allowed the concomitant detection of various oligomeric species during aggregation to follow their kinetics of appearance and disappearance and to determine their structure in terms of amide protection to H/D exchange at the residue level.

Other H/D independent studies aimed at determining the degree of structural flexibility and solvent exposure of side chains in the aggregates. Twenty mutants of the HypF-N protein containing a single cysteine residue at various positions were labeled with pyrene and allowed to aggregate under two different conditions; the fluorescence spectra of the resulting samples were acquired, revealing in each case whether the pyrene moiety was buried inside the oligomers and in contact with another pyrene moiety or rather flexible and solvent exposed (Campioni et al., 2010). This study allowed the identification of the regions of the sequence that were most structured and buried in the interior of the protein oligomers, revealing significant differences between the two oligomer populations that could explain differences in their cytotoxicity (discussed later).

Low-molecular-weight oligomers have also been studied with conventional solution NMR (Pagano et al., 2010; Yu et al., 2009), whereas large aggregates have been successfully studied with solid state NMR using approaches normally applied to amyloid fibrils (Chimon et al., 2007). A novel, very interesting solution NMR technique, named dark-state exchange saturation transfer (DEST), has been very recently presented to probe the structure of $A\beta_{40}$ and $A\beta_{42}$ protofibrils and circumvent the problems of limiting solution NMR to small oligomers (Fawzi et al., 2011). In all such cases, secondary and/or tertiary information of the molecules forming the oligomers have been obtained for individual residues.

Detailed molecular structure of protein oligomers was also obtained for $A\beta_{40}$ protofibrils using scanning proline mutagenesis (Williams et al., 2005). This approach consisted in the systematic substitution of all peptide residues with proline and in the subsequent evaluation of the equilibrium between soluble and aggregated peptide for all mutants having a single substitution. Mutants that produced the largest changes in such equilibrium relative to the wild-type peptide were meant to indicate the involvement of the mutated residues in the protofibrillar structure. Another interesting method is small angle X-ray scattering. This technique gives information about shape and dimensions of oligomers ranging from 1 to about 100 nm (Langkilde and Vestergaard, 2009) and has been used in different systems such as insulin (Vestergaard et al., 2007), $\alpha$-synuclein (Giehm et al., 2011), and the immunoglobulin light chain LEN (Souillac et al., 2002).

Many of the approaches described here to study protein oligomers were inspired by previous studies successfully applied to stable amyloid-like fibrils. The complexities arising from the structural heterogeneity and metastability of the oligomers has been circumvented with various strategies; for example, by using compounds stabilizing the oligomers (Williams et al., 2005), protocols based on the disaggregation of fibrils at high pH (Heldt et al., 2011a), proteins forming stable oligomers (Campioni et al., 2010) or by using MS methods that allow the concomitant detection of various species in a sample (Carulla et al., 2009; Pan et al., 2011). The study performed on the PI3-SH3 domain provides a nice example of an approach to effectively circumvent such problems, as the oligomeric species could be studied in spite of their transient appearance and coexistence with other species (Carulla et al., 2009).

The problem of structural heterogeneity is also being overcome by a number of emerging cutting-edge methodologies that allow the oligomers to be studied at the single molecule level. In two-color coincident detection (TCCD), two subpopulations of a protein sample are labeled with two different probes. By inducing the coaggregation of the resulting subpopulations, it is possible to observe the formation of individual oligomers through the coincident detection of the two fluorescent probes (Orte et al., 2008). Intensity of the observed two fluorescence signals provided information about oligomer size distribution. The same approach was extended to perform single-molecule fluorescence resonance energy transfer (FRET) studies (i.e., monitoring the energy transferred from a donor probe to an acceptor probe through measurements of the fluorescence emitted by the latter); the observed fluorescence arises from FRET events within individual oligomers and provided information about the distance between the two labeled positions (Orte et al., 2011). Finally, in ion mobility MS, the oligomers were separated in two steps: (1) according to their mobility in a chamber containing a carrier buffer gas that opposes their ion motion and (2) according to their mass-to-charge ratio. This technique gave direct information about mass, size, and shape for individual oligomer species, detecting concomitantly the presence of different oligomer subpopulations, and was applied to A$\beta$ (Bernstein et al., 2009; Kłoniecki et al., 2011), $\beta_2$-microglobulin (Smith et al., 2010), and insulin (Pease et al., 2010), as well as to a set of synthetic peptides (Bleiholder et al., 2011).

It is interesting that a noninvasive FRET sensor has been recently set up to monitor in vivo the formation of oligomers and their interconversion (Kaminski Schierle et al., 2011). In this approach, changes in fluorescence lifetime reflect changes in the oligomer status and can be related to differences in toxicity, providing a new tool to study noninvasively oligomer-related toxicity in vivo.

## Structure of the Oligomers

The existence of parallel pathways in protein aggregation, and the metastability of the various species accumulating during

the process, creates a multiplicity of oligomers, often with very diverse characteristics. This complexity has led to the utilization of different criteria for their classification; for example, on the grounds of their size, β sheet content (or other purely structural characteristics), stability to SDS solubilization or other treatments, productive role in amyloid fibril formation, reactivity to conformation-specific antibodies, and so forth. Even by focusing on a single system such as Aβ, a full description of the various oligomers described in the literature would require a review or book chapter per se. Therefore, in this section, we try to describe the key structural features shared by protein oligomers from different peptides/proteins and how such structures change as aggregation proceeds.

When aggregation is initiated by fully or largely unfolded monomers in the nucleated conformational conversion, the initial oligomers exhibit a large variety of conformations, with monomers still adopting a disordered structure. For example, using PICUP and CD spectroscopy, it was found that the early aggregates formed by $A\beta_{40}$ and $A\beta_{42}$ are dimers-tetramers and pentamers-hexamers, respectively, with a poor level of structure; these acquire β sheet structure later on in the process at the level of protofibrils (Bitan et al., 2003). The appearance of early unstructured aggregates of $A\beta_{40}$ with unstable β sheet structure was also detected in other two more recent studies, performed independently with H/D exchange (coupled to MS) and FlAsH labeling, respectively, with larger oligomers containing stable β sheet structure forming later (Lee et al., 2011; Qi et al., 2008). Furthermore, with tryptophan fluorescence, it was shown that the initial aggregates formed by α-synuclein display unfolded monomers that expose hydrophobic residues to the solvent (Dusa et al., 2006). Characterization of the aggregation pathway of the pH-unfolded PI3-SH3 domain by pulse-labeling H/D exchange coupled to MS and NMR showed that species detected early in the aggregation process are disordered on the basis of the low degree of amide hydrogen exchange protection (Carulla et al., 2009). Very recently, it has been shown, by means of ion mobility MS, that a set of peptides form initially unfolded assemblies (Bleiholder et al., 2011). Molten globule-like self-associating oligomers have been found to be populated prior to fibril formation also for IAPP (Wei et al., 2011).

If aggregation starts from native-like states, the early aggregates formed at the beginning of the process display monomers populating native-like states (Banci et al., 2005; Olofsson et al., 2004; Pagano et al., 2010). In the case of the model protein Sso AcP, aggregation is not just initiated by a native-like state; the native-like structure persists when the protein forms the initial aggregates (Pagano et al., 2010). In a mutant form of superoxide dismutase type-1, native dimers interact to form larger oligomers stabilized by transient interactions between electrostatic loops from different dimers; these oligomers retain monomers in a native-like conformation (Banci et al., 2005). Transthyretin assembles into aggregates in which monomers exhibit six β strands in a native-like conformation while two edge strands are misfolded (Olofsson et al., 2004). Early oligomers from insulin have been shown to be rich in α-helical structure (Bouchard et al., 2000; Vestergaard et al., 2007).

It therefore emerges that initial aggregates are far from the amyloid structure. In fact, regardless of the aggregation pathway followed by a protein, the initial aggregates display the same conformational features observed in the aggregation-competent monomers. The aggregates do not bind amyloid specific dyes, nor do they exhibit a significant content of stable β sheet structure (Lee et al., 2011; Plakoutsi et al., 2005).

As aggregation proceeds, oligomers undergo a structural rearrangement into species stabilized by β sheet structure, able to bind ThT and CR. In addition, oligomers undergo an increase in dimensions, compactness, stability, and order, still retaining a nonfibrillar morphology. It was proposed that fibril elongation of the NM region of the prion Sup35p is initiated following formation of ordered nuclei by conformational rearrangements of less structured, molten, oligomeric intermediates (Serio et al., 2000). During aggregation of human muscle acylphosphatase (mAcP), oligomers increase their dimensions and disaggregation induced by dilution into nonamyloidogenic conditions becomes slower as aggregation proceeds, consistently with an increase in oligomer stability (Calamai et al., 2005). The study on the pH-unfolded PI3-SH3 domain mentioned earlier also indicated the late appearance of oligomeric intermediates with highly ordered β sheet structure (Carulla et al., 2009). Single-molecule studies on the same protein provided direct evidence that the stable cross-β structure of the late aggregates emerges via internal reorganization of disordered oligomers formed during the lag phase of the self-assembly reaction (Orte et al., 2008). FRET studies carried out on α-synuclein oligomers showed a decrease in the distance between Tyr39 and Trp125 from the early oligomers to the late oligomers, suggesting an increase in compactness (Kaylor et al., 2005). During aggregation, the NNQQNY peptide undergoes a transition, beginning near the octamer, from a natively unstructured assembly to a highly ordered β sheet assembly (Bleiholder et al., 2011). Finally, as mentioned earlier, an increase of order and β sheet structure content was also observed for $A\beta_{40}$ and $A\beta_{42}$ oligomers by different authors and using different methodologies (Bitan et al., 2003; Lee et al., 2011; Qi et al., 2008; Sandberg et al., 2010).

Such a structural conversion into β sheet containing amyloid-like aggregates was also observed for native-like aggregates containing significant levels of α-helical structure. For example, the native-like aggregates formed by Sso AcP convert, with no need of disaggregation, into amyloid-like protofibrils with a higher level of order, as deduced with ThT fluorescence, FTIR, and far-UV CD (Plakoutsi et al., 2005). Similarly, α-helical native-like oligomers of insulin were also shown to convert into β sheet containing protofibrils, suggesting again a similar process (Bouchard et al., 2000). The direct conversion from α-helical to β sheet oligomers without the need of disaggregation and reassociation was also observed in many molecular dynamics simulations; for example, for a 17-residue peptide designed to form a coiled coil trimer (Strodel et al., 2008). In cases where native-like aggregation involves all-β proteins, such a structural conversion from early to amyloid-like aggregates has not been yet reported, maybe because of inherent difficulties for the detection of such a transition where β sheet structure remains the dominant secondary structure type. Alternatively, the early oligomers may possess already the structural characteristics to act as nuclei.

In addition to the observation of a transition from unstructured (or native-like) oligomers into amyloid oligomers with β sheet content, the β-structure that forms in early aggregates, when

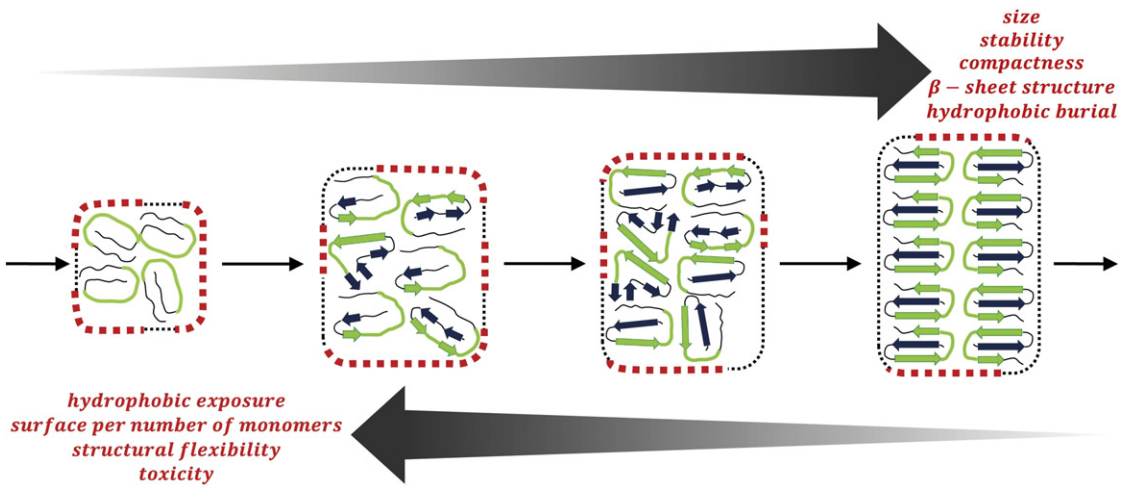


**Figure 2. A Schematic Representation of the Structural Rearrangements Occurring during Oligomer Formation**
For simplicity, only aggregation starting from fully or largely unfolded monomers is considered (reaction F → G in Figure 1). Amyloidogenic/hydrophobic segments are in green. The oligomer surface is drawn as a thin black and a thick red dotted line when amyloidogenic/hydrophobic segments are buried and exposed to the solvent, respectively. While aggregation proceeds (left to right), a set of structural rearrangements takes place: The top and bottom arrows show the parameters that increase and decrease, respectively. Binding of monomers to early oligomers is isotropic, whereas late oligomers can bind to monomers only at the edges. This leads to growth of thin filaments, which eventually originate amyloid fibrils.

present, tends to be variable and unstable. Aβ(16–22) peptide molecules rearrange, for example, through a realignment of β strands, from a less regular β sheet structure into a β sheet structure more stable in register (Petty and Decatur, 2005). While fibrils and large spherical amyloid intermediates (diameter, 15–35 nm) contain stable parallel β sheets, as determined with solid-state NMR (Chimon et al., 2007), solution NMR data indicate that small globulomers by Aβ (38–48 kDa) have a mixed parallel and antiparallel β sheet structure (Yu et al., 2009). In another study, it was found that initial low-molecular-weight oligomers of $A\beta_{40}$ have a less extended and more unstable β sheet structure, judging from H/D exchange data, than larger intermediate oligomers forming later (Qi et al., 2008). Consistently with these data, the oligomeric interfaces of oligomers formed by a set of model peptides display a variety of sheet-to-sheet pairing angles (Liu et al., 2011). Such internal conversions from variable to regular sheets produce oligomers more productive in terms of fibril formation, as fibrils have a stable and a regular β sheet structure (Carulla et al., 2009; Cerdà-Costa et al., 2007; Orte et al., 2008). This was found directly in a paper where oligomers exhibiting aligned strands were found to be capable of forming thermostable, long, rigid, and twisted fibrils, whereas oligomers without this strand alignment aggregate to form thin, flexible, and smooth protofibrils (Petty et al., 2005).

Thus, the general picture emerging from all such studies is that, as aggregation proceeds through a nucleated conformational conversion process, oligomers undergo a continuous rearrangement of structure (Figure 2). This reorganization involves an increase in size, stability compactness, regularity of the β sheet structure, and hydrophobic burial. It also implies a decrease in dynamical fluctuations, exposure of hydrophobic clusters, and oligomer surface per number of monomers. We show in a following section that this trend is also related with a decrease in oligomer toxicity. The conversion may occur through a multiplicity of oligomeric states or just as a two-state process, not necessarily through four oligomeric states, as depicted in Figure 2. In addition, each oligomeric state may be considered as an ensemble of oligomers with distributions of size, structure, and so forth.

## The Oligomers Are Polymorphic in Size and Structure

The aforementioned existence of parallel competing pathways and the presence of multiple species even for the same pathway, imply that oligomers exist as a number of species with different morphological and structural properties. This phenomenon is usually referred to as oligomer polymorphism (Kodali and Wetzel, 2007; Stefani, 2010). Oligomers exhibit polymorphism in terms of their size (Mastrangelo et al., 2006), shape (Pountney et al., 2005), compactness (Kaylor et al., 2005), stability (Calamai et al., 2005; Souillac et al., 2003), and secondary and tertiary structure content (Bleiholder et al., 2011; Ehrnhoefer et al., 2008; Ladiwala et al., 2010; Lee et al., 2011; Serio et al., 2000). The spectrum of oligomer polymorphism extends even further if we consider that small oligomers containing both Aβ and α-synuclein molecules form in patients with mixed Alzheimer's disease and Parkinson's disease and in transgenic mice coexpressing both proteins (Tsigelny et al., 2008).

Different types of oligomers can coexist in solution at the same time (Goldsbury et al., 2005; Gosal et al., 2005; Jain and Udgaonkar, 2011; Mastrangelo et al., 2006; Relini et al., 2010) and even in vivo (Winner et al., 2011). The predominance of some particular species can be determined by a number of factors. First, mutations can alter the pathway by which oligomers form and, consequently, the morphology/structure of the oligomers. For example, in the case of Sso AcP, a set of point mutations leads to protein variants that aggregate following a pathway different from that of the wild-type protein, leading to ThT-binding β-structured oligomers and native-like aggregates, respectively (Soldi et al., 2008). Addition of two residues at the C terminus of the Aβ peptide is sufficient to change the

population of oligomers in equilibrium with the monomer, with a distribution of monomers to tetramers typical for $A\beta_{40}$ shifting to a mixture of pentamers and hexamers for $A\beta_{42}$ (Bitan et al., 2003). In the presence of membrane extracts, different mutants of $A\beta_{40}$ having single amino acid substitutions also formed different distributions of polymorphic aggregates (Pifer et al., 2011). A similar mutation dependence of oligomer formation was also observed for α-synuclein (Ono et al., 2011).

A second determinant of oligomer polymorphism is the solution conditions under which aggregation is initiated. In the case of HypF-N, it was shown that oligomers formed under two different solution conditions, differing in terms of pH and cosolvent composition, exhibit different toxicities (Campioni et al., 2010), and this can be attributed to differences in compactness and solvent exposure of hydrophobic clusters (discussed later). Two different products of lipid peroxidation have recently been shown to be able to induce α-synuclein oligomers that differ in terms of morphology, dimensions, compactness, and stability (Näsström et al., 2011). Finally, the type of seeding can affect the final fibril morphology/structure, and, consequently, it is expected to alter the properties of the oligomers that accumulate prior to fibril formation (Paravastu et al., 2009; Yamaguchi et al., 2005).

Oligomer polymorphism is not just a relevant phenomenon for a full description of the structure and mechanism of formation of oligomers and fibrils but also has many implications in biology. It was indeed found that expression of different mutants of Aβ in *Drosophila* qualitatively led to different pathologies (Iijima et al., 2008) and that different oligomers of α-synuclein and Aβ caused toxicity in cell cultures through different mechanisms (Danzer et al., 2007; Deshpande et al., 2006). On top of that, it is well established that polymorphism is a central theme to explain the propagation of prion strain infectivity (Jones and Surewicz, 2005).

## Oligomers In Vitro and In Vivo

An important question is whether the large number of protein oligomers formed in vitro and described in the literature are relevant for amyloid fibril formation processes in vivo and their associated diseases. Using approaches aimed at detecting oligomers directly in vivo, a number of studies have revealed the presence of such species in human patients suffering from protein deposition diseases or related animal models. This has been observed, for example, for Aβ (Lesné et al., 2006; Shankar et al., 2008), tau (Lasagna-Reeves et al., 2012; Patterson et al., 2011), huntingtin (Nekooki-Machida et al., 2009), and α-synuclein (Tsigelny et al., 2008). Such oligomers, when isolated from the living specimen and administered to normal rats, have been shown to cause cognitive impairment in the animals (Lesné et al., 2006; Shankar et al., 2008). Models for oligomerization in vivo are now also being proposed (Larson and Lesné, 2012).

In addition, conformation-specific antibodies raised against particular types of exogenous Aβ or tau oligomers and showing no specificity for monomeric or fibrillar protein have been used to reveal the presence of the same types of oligomers in Alzheimer's disease patients, indicating the existence of such in vitro-formed oligomers in human patients as well (Hillen et al., 2010; Kayed et al., 2007, 2003; Lacor et al., 2004; Lasagna-Reeves et al., 2012, 2011; Noguchi et al., 2009; Patterson et al., 2011). Thus, although our understanding of oligomer formation mechanisms in vivo remains limited, the existence of protein oligomers with structural characteristics similar to those found and characterized in vitro suggests that we can capitalize on our current knowledge obtained in vitro.

## The Structural Determinants of Oligomer-Induced Cytotoxicity

The progressive elucidation of oligomer structure is starting to reveal the structural determinants of oligomer toxicity; i.e., the structural elements that are responsible for the ability of protein oligomers to interact with the cells and cause their dysfunction. An important source of information is represented by the large body of studies on the Aβ peptides. Various oligomers have been described for both the $A\beta_{40}$ and $A\beta_{42}$, which have been attributed different names, such as Aβ-derived diffusible ligands (ADDLs), protofibrils, prefibrillar oligomers, fibrillar oligomers, annular protofibrils, amylospheroids, globulomers, spherical amyloid intermediates, and so forth. The toxicity of many of such oligomeric species has been measured on cultured cells using the MTT reduction assay, providing an opportunity to compare their toxicities.

Figure 3A reports the MTT reduction values measured by various authors for their studied oligomers versus the size of the oligomers, expressed as mean molecular weight. The MTT reduction values reported in the figures were all measured at a peptide monomer concentration of 2.0–2.7 μM and are thus comparable between different oligomeric species. The data refer to both $A\beta_{40}$ and $A\beta_{42}$ as it was shown that the same type of oligomers formed by the two species cause similar decreases of MTT reduction following their addition to the cells (Kayed et al., 2003; Kayed et al., 2009). A clear trend is present in the plot, with toxicity decreasing with the size of the oligomers until a well-defined plateau. It is interesting that the plateau obtained with the best fitting procedure was not found as 100% but as 83% ± 5%, which is similar to the values reported by different authors for amyloid fibrils (molecular weight > 2,000 kDa) formed by both $A\beta_{40}$ and $A\beta_{42}$ (Chafekar et al., 2008; Kayed et al., 2003; Walsh et al., 1999). The analysis therefore indicates that oligomer size is an important determinant of oligomer toxicity. In agreement with this analysis, it was found that three classes of small aromatic molecules can inhibit $A\beta_{42}$ oligomer toxicity by converting the small oligomers into large aggregates, fibrils, and monomers, respectively (Ladiwala et al., 2011).

Hydrophobic exposure on the aggregate surface seems to be another important determinant of oligomer-mediated toxicity. A pioneristic study that used reconstructed model membranes rather than cells reported a correlation between hydrophobic exposure in $A\beta_{40}$ aggregates and membrane fluidity, measured with bis-ANS fluorescence and membrane fluorescence anisotropy, respectively (Kremer et al., 2000). Later on, a correlation between hydrophobicity of homopolymeric amino acid stretches and cytotoxicity of their aggregates has been observed (Oma et al., 2005). In a more recent study, two types of spherical oligomers formed by the HypF-N protein using different protocols and shown to have indistinguishable size and morphology, as detected with AFM, were found to have very different toxicities, with one species found to be nontoxic altogether (Campioni et al., 2010). With site-directed pyrene labeling, it was found

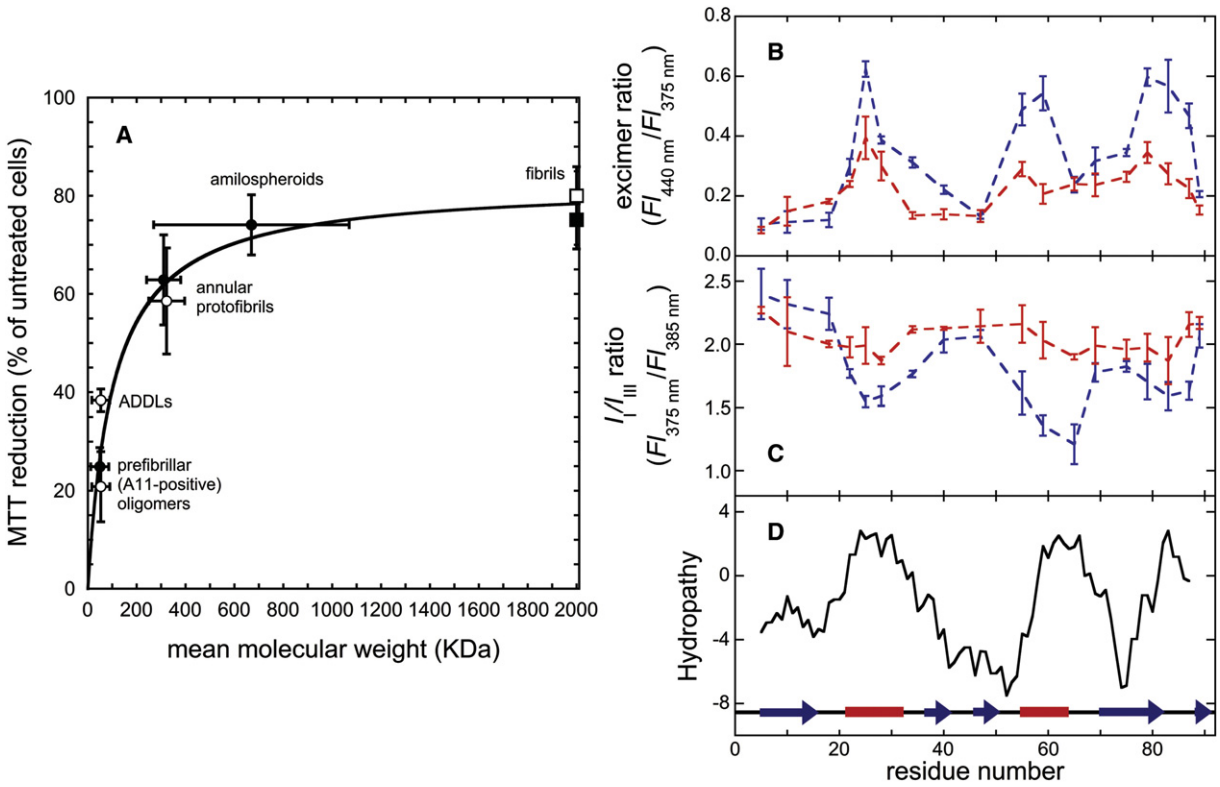


**Figure 3. Structural Determinants of Oligomer-Induced Toxicity**

(A) Toxicity versus size of $A\beta_{40}$ and $A\beta_{42}$ aggregates. Toxicity is measured by determining MTT reduction by cultured cells following their exposure to oligomers added to the extracellular medium. Aggregate toxicity was expressed as percentage of MTT reduction relative to untreated cells, where 0% and 100% values are two extremes of full cell death and full viability, respectively. Values and error bars are from the original papers: prefibrillar oligomers (Kayed et al., 2003), ADDLs (Lambert et al., 2001), annular protofibrils (Kayed et al., 2009), and amylospheroids (Hoshi et al., 2003). All data were obtained at a peptide concentration in the range of 2.0-2.7 μM. Aggregate size was expressed as mean molecular weight of the reported distributions in the original papers, and error bars refer to the width of the distributions, not SD or SEM: prefibrillar oligomers (Kayed et al., 2007), ADDLs (Gong et al., 2002), annular protofibrils (Kayed et al., 2009), and amylospheroids (Hoshi et al., 2003). Only data for which both molecular weight and MTT reduction values at ca. 2.0–2.7 μM Aβ are reported. Data for both $A\beta_{40}$ (filled circles) and $A\beta_{42}$ (empty circles) are presented, as the same type of oligomers or fibrils formed by the two species cause similar decreases of MTT reduction (Kayed et al., 2003; Kayed et al., 2009). All data points were fitted to a hyperbolic function of the form $y = a * x/(b + x)$. MTT reduction induced by Aβ fibrils (filled and empty squares for $A\beta_{40}$ and $A\beta_{42}$, respectively) are not taken into account in the fitting procedure and are shown for comparison (Kayed et al., 2003). It is implicit that their molecular weight is often >2000 kDa.

(B–D) Structural differences between toxic and nontoxic aggregates of HypF-N. (B) Excimer ratio of pyrene (related to the degree of structure formation) versus number of labeled residue for toxic (red lines) and nontoxic (blue lines) oligomers of HypF-N. (C) $I_I/I_{III}$ ratio of pyrene (a correlate of the degree of solvent exposure) versus number of labeled residue for toxic (red lines) and nontoxic (blue lines) oligomers. (D) Hydropathy profile of HypF-N sequence. The three panels show that in nontoxic aggregates, unlike the toxic aggregates, the three hydrophobic regions of the sequence are structured and buried inside the oligomers.

that the three most hydrophobic regions of the protein sequence are structured and buried in the nontoxic oligomers, whereas in the toxic oligomers the same regions are more solvent exposed and flexible (Figures 3B–3D).

Other studies have appeared very recently on the importance of hydrophobic exposure for protein aggregate toxicity. Striking correlations have been found for a number of peptides/proteins between the toxicity of various forms of aggregates formed in vitro and added extracellularly to cell cultures, measured with propidium iodide incorporation, and their solvent exposure determined with ANS binding (Bolognesi et al., 2010). Another study has demonstrated that highly amyloidogenic proteins expressed intracellularly in human embryonic kidney 293T cells have levels of toxicity, measured with the MTT test, that increase with the exposure of hydrophobic clusters on the aggregate surface measured with ANS binding (Olzscha et al., 2011). In the latter study, the toxicity of the intracellularly expressed proteins was attributed to the ability of their aggregates to interact with a number of multifunctional cellular proteins and alter their function (Olzscha et al., 2011), whereas in the HypF-N study, toxicity was attributed to the ability of the extracellularly added oligomers to interact with the cell membrane and cause an uptake of calcium (Campioni et al., 2010; Zampagni et al., 2011). In addition, many other mechanisms of toxicity have been proposed. Hydrophobic exposure on oligomeric surface thus seems an important determinant of toxicity, independent of the mechanism by which oligomers cause cell dysfunction.

Another proposed determinant of oligomer toxicity is the shape of the aggregates. In particular, it has been proposed that monomers may associate to form pore-like oligomers that bind membranes or, alternatively, that monomers directly self-assemble into pores at the membrane interface (Giehm et al., 2011; Lashuel et al., 2002a, 2002b; Last et al., 2011). The proposed consequence of pore formation is that normally

intracellular or extracellular components can freely circulate between the two environments, leading to cellular dysfunction. Ion-channel oligomers have been observed by TEM for a number of systems, including Aβ, IAPP, α-synuclein, ABri, and ADan (Lashuel et al., 2002a; Last et al., 2011; Quintas et al., 2001). It should be emphasized, however, that alternative mechanisms of oligomer-mediated toxicity have been proposed for all systems found so far to form annular oligomers. For α-synuclein, for example, it has recently been proposed that spherical oligomers growing at the membrane edge captures lipids from the membrane, eventually resulting in membrane disruption (Reynolds et al., 2011). Furthermore, the correlation between size and toxicity (Figure 3A) includes annular protofibrils (Kayed et al., 2009), and these oligomers do not significantly deviate from the observed trend. It is therefore evident that a role for this specific morphology—and, more generally, for oligomer shape—in oligomer-mediated toxicity remains elusive.

Although the results reviewed here indicate that protein oligomer toxicity is determined by a well-defined set of parameters such as size and hydrophobic exposure of the oligomers, it is important to emphasize that toxicity does not reside in one or limited number of oligomeric forms of a given protein. Different oligomers from the same protein have been shown to affect cell viability to some degree. Moreover, the observation that protein oligomers formed by different proteins share similar levels of toxicity is itself a strong suggestion that toxicity is a shared property of protein misfolded oligomers rather than a characteristic of a specific structural pattern.

## Conclusions

Although the increasing awareness of the existence of a myriad of different metastable oligomers seemed to make their characterization an impossible task, in the past few years considerable efforts expended by different researchers has led to the emergence of innovative methodologies that are starting to unveil the multiple structures of protein misfolded oligomers and the structural features underlying their toxicity. Oligomer particles can now be detected at the single molecule level and can be observed not only in the test tube but even in vivo. The general emerging picture is that oligomers cannot be described as a finite number of protein structures, each of them identified by a defined set of parameters. Oligomers are best described as a number of conformational ensembles, each containing an indefinite number of different assemblies that vary, almost without solution of continuity, in secondary and tertiary structure, compactness, shape, size, number of monomers, and so forth. All these parameters contribute to a complex energy landscape and determine, directly or indirectly, the toxicity of the oligomers through relationships that are slowly being revealed.

Of course, much remains to be determined. Key points that need to be addressed in future research include (1) developing new and more accurate methodologies for oligomer structure determination; (2) identifying additional molecular determinants of oligomer toxicity and the mechanisms of action of the resulting aggregates so that new molecular targets for therapeutic intervention can be identified; (3) understanding the reasons why, where, and when functional proteins convert into such aggregates, again with the aim of interfering therapeutically with their formation; and (4) developing new methodologies to detect reliably and routinely toxic oligomers in vivo in a noninvasive way and before disease onset.

It is hoped that such studies will make it possible to fully describe the "protein aggregation side" of the protein (mis-)folding landscape, to link our in vitro knowledge with systems in vivo and to reach a complete understanding of the relationships between structure and toxicity of protein misfolded oligomers.

## ACKNOWLEDGMENTS

F.B. is supported by an EU-FP7 Marie Curie fellowship.

## REFERENCES

Bader, R., Bamford, R., Zurdo, J., Luisi, B.F., and Dobson, C.M. (2006). Probing the mechanism of amyloidogenesis through a tandem repeat of the PI3-SH3 domain suggests a generic model for protein aggregation and fibril formation. J. Mol. Biol. *356*, 189–208.

Banci, L., Bertini, I., D'Amelio, N., Gaggelli, E., Libralesso, E., Matecko, I., Turano, P., and Valentine, J.S. (2005). Fully metallated S134N Cu,Zn-superoxide dismutase displays abnormal mobility and intermolecular contacts in solution. J. Biol. Chem. *280*, 35815–35821.

Bemporad, F., Calloni, G., Campioni, S., Plakoutsi, G., Taddei, N., and Chiti, F. (2006). Sequence and structural determinants of amyloid fibril formation. Acc. Chem. Res. *39*, 620–627.

Bernstein, S.L., Dupuis, N.F., Lazo, N.D., Wyttenbach, T., Condron, M.M., Bitan, G., Teplow, D.B., Shea, J.E., Ruotolo, B.T., Robinson, C.V., and Bowers, M.T. (2009). Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. Nat. Chem. *1*, 326–331.

Bhattacharyya, A.M., Thakur, A.K., and Wetzel, R. (2005). polyglutamine aggregation nucleation: thermodynamics of a highly unfavorable protein folding reaction. Proc. Natl. Acad. Sci. USA *102*, 15400–15405.

Billings, L.M., Oddo, S., Green, K.N., McGaugh, J.L., and LaFerla, F.M. (2005). Intraneuronal Abeta causes the onset of early Alzheimer's disease-related cognitive deficits in transgenic mice. Neuron *45*, 675–688.

Bitan, G., Kirkitadze, M.D., Lomakin, A., Vollers, S.S., Benedek, G.B., and Teplow, D.B. (2003). Amyloid beta -protein (Abeta) assembly: Abeta 40 and Abeta 42 oligomerize through distinct pathways. Proc. Natl. Acad. Sci. USA *100*, 330–335.

Bleiholder, C., Dupuis, N.F., Wyttenbach, T., and Bowers, M.T. (2011). Ion mobility-mass spectrometry reveals a conformational conversion from random assembly to β-sheet in amyloid fibril formation. Nat. Chem. *3*, 172–177.

Bolognesi, B., Kumita, J.R., Barros, T.P., Esbjorner, E.K., Luheshi, L.M., Crowther, D.C., Wilson, M.R., Dobson, C.M., Favrin, G., and Yerbury, J.J. (2010). ANS binding reveals common features of cytotoxic amyloid species. ACS Chem. Biol. *5*, 735–740.

Bouchard, M., Zurdo, J., Nettleton, E.J., Dobson, C.M., and Robinson, C.V. (2000). Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. Protein Sci. *9*, 1960–1967.

Bucciantini, M., Giannoni, E., Chiti, F., Baroni, F., Formigli, L., Zurdo, J., Taddei, N., Ramponi, G., Dobson, C.M., and Stefani, M. (2002). Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. Nature *416*, 507–511.

Calamai, M., Canale, C., Relini, A., Stefani, M., Chiti, F., and Dobson, C.M. (2005). Reversal of protein aggregation provides evidence for multiple aggregated States. J. Mol. Biol. *346*, 603–616.

Campioni, S., Mannini, B., Zampagni, M., Pensalfini, A., Parrini, C., Evangelisti, E., Relini, A., Stefani, M., Dobson, C.M., Cecchi, C., and Chiti, F. (2010). A causative link between the structure of aberrant protein oligomers and their toxicity. Nat. Chem. Biol. *6*, 140–147.

 

Canet, D., Last, A.M., Tito, P., Sunde, M., Spencer, A., Archer, D.B., Redfield, C., Robinson, C.V., and Dobson, C.M. (2002). Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. Nat. Struct. Biol. *9*, 308–315.

Cardoso, I., and Saraiva, M.J. (2006). Doxycycline disrupts transthyretin amyloid: evidence from studies in a FAP transgenic mice model. FASEB J. *20*, 234–239.

Carulla, N., Zhou, M., Arimon, M., Gairí, M., Giralt, E., Robinson, C.V., and Dobson, C.M. (2009). Experimental characterization of disordered and ordered aggregates populated during the process of amyloid fibril formation. Proc. Natl. Acad. Sci. USA *106*, 7828–7833.

Cerdà-Costa, N., Esteras-Chopo, A., Avilés, F.X., Serrano, L., and Villegas, V. (2007). Early kinetics of amyloid fibril formation reveals conformational reorganisation of initial aggregates. J. Mol. Biol. *366*, 1351–1363.

Chafekar, S.M., Baas, F., and Scheper, W. (2008). Oligomer-specific Abeta toxicity in cell models is mediated by selective uptake. Biochim. Biophys. Acta *1782*, 523–531.

Chimon, S., Shaibat, M.A., Jones, C.R., Calero, D.C., Aizezi, B., and Ishii, Y. (2007). Evidence of fibril-like β-sheet structures in a neurotoxic amyloid intermediate of Alzheimer's β-amyloid. Nat. Struct. Mol. Biol. *14*, 1157–1164.

Chiti, F., and Dobson, C.M. (2006). Protein misfolding, functional amyloid, and human disease. Annu. Rev. Biochem. *75*, 333–366.

Cleary, J.P., Walsh, D.M., Hofmeister, J.J., Shankar, G.M., Kuskowski, M.A., Selkoe, D.J., and Ashe, K.H. (2005). Natural oligomers of the amyloid-beta protein specifically disrupt cognitive function. Nat. Neurosci. *8*, 79–84.

Danzer, K.M., Haasen, D., Karow, A.R., Moussaud, S., Habeck, M., Giese, A., Kretzschmar, H., Hengerer, B., and Kostka, M. (2007). Different species of alpha-synuclein oligomers induce calcium influx and seeding. J. Neurosci. *27*, 9220–9232.

Deshpande, A., Mina, E., Glabe, C., and Busciglio, J. (2006). Different conformations of amyloid beta induce neurotoxicity by distinct mechanisms in human cortical neurons. J. Neurosci. *26*, 6011–6018.

Dobson, C.M. (2003). Protein folding and misfolding. Nature *426*, 884–890.

Dusa, A., Kaylor, J., Edridge, S., Bodner, N., Hong, D.P., and Fink, A.L. (2006). Characterization of oligomers during alpha-synuclein aggregation using intrinsic tryptophan fluorescence. Biochemistry *45*, 2752–2760.

Eakin, C.M., Berman, A.J., and Miranker, A.D. (2006). A native to amyloidogenic transition regulated by a backbone trigger. Nat. Struct. Mol. Biol. *13*, 202–208.

Ehrnhoefer, D.E., Bieschke, J., Boeddrich, A., Herbst, M., Masino, L., Lurz, R., Engemann, S., Pastore, A., and Wanker, E.E. (2008). EGCG redirects amyloidogenic polypeptides into unstructured, off-pathway oligomers. Nat. Struct. Mol. Biol. *15*, 558–566.

Eichner, T., and Radford, S.E. (2011). A diversity of assembly mechanisms of a generic amyloid fold. Mol. Cell *43*, 8–18.

Fawzi, N.L., Ying, J., Ghirlando, R., Torchia, D.A., and Clore, G.M. (2011). Atomic-resolution dynamics on the surface of amyloid-β protofibrils probed by solution NMR. Nature *480*, 268–272.

Ferrone, F. (1999). Analysis of protein aggregation kinetics. Methods Enzymol. *309*, 256–274.

Giehm, L., Svergun, D.I., Otzen, D.E., and Vestergaard, B. (2011). Low-resolution structure of a vesicle disrupting α-synuclein oligomer that accumulates during fibrillation. Proc. Natl. Acad. Sci. USA *108*, 3246–3251.

Giorgetti, S., Raimondi, S., Pagano, K., Relini, A., Bucciantini, M., Corazza, A., Fogolari, F., Codutti, L., Salmona, M., Mangione, P., et al. (2011). Effect of tetracyclines on the dynamics of formation and destructuration of beta2-microglobulin amyloid fibrils. J. Biol. Chem. *286*, 2121–2131.

Goldsbury, C., Frey, P., Olivieri, V., Aebi, U., and Müller, S.A. (2005). Multiple assembly pathways underlie amyloid-beta fibril polymorphisms. J. Mol. Biol. *352*, 282–298.

Gong, J.S., Sawamura, N., Zou, K., Sakai, J., Yanagisawa, K., and Michikawa, M. (2002). Amyloid beta-protein affects cholesterol metabolism in cultured neurons: implications for pivotal role of cholesterol in the amyloid cascade. J. Neurosci. Res. *70*, 438–446.

Gosal, W.S., Morten, I.J., Hewitt, E.W., Smith, D.A., Thomson, N.H., and Radford, S.E. (2005). Competing pathways determine fibril morphology in the self-assembly of beta2-microglobulin into amyloid. J. Mol. Biol. *351*, 850–864.

Harper, J.D., Lieber, C.M., and Lansbury, P.T., Jr. (1997). Atomic force microscopic imaging of seeded fibril formation and fibril branching by the Alzheimer's disease amyloid-beta protein. Chem. Biol. *4*, 951–959.

Heldt, C.L., Sorci, M., Posada, D., Hirsa, A., and Belfort, G. (2011a). Detection and reduction of microaggregates in insulin preparations. Biotechnol. Bioeng. *108*, 237–241.

Heldt, C.L., Kurouski, D., Sorci, M., Grafeld, E., Lednev, I.K., and Belfort, G. (2011b). Isolating toxic insulin amyloid reactive species that lack β-sheets and have wide pH stability. Biophys. J. *100*, 2792–2800.

Hillen, H., Barghorn, S., Striebinger, A., Labkovsky, B., Müller, R., Nimmrich, V., Nolte, M.W., Perez-Cruz, C., van der Auwera, I., van Leuven, F., et al. (2010). Generation and therapeutic efficacy of highly oligomer-specific beta-amyloid antibodies. J. Neurosci. *30*, 10369–10379.

Hoshi, M., Sato, M., Matsumoto, S., Noguchi, A., Yasutake, K., Yoshida, N., and Sato, K. (2003). Spherical aggregates of beta-amyloid (amylospheroid) show high neurotoxicity and activate tau protein kinase I/glycogen synthase kinase-3beta. Proc. Natl. Acad. Sci. USA *100*, 6370–6375.

Iijima, K., Chiang, H.C., Hearn, S.A., Hakker, I., Gatt, A., Shenton, C., Granger, L., Leung, A., Iijima-Ando, K., and Zhong, Y. (2008). Abeta42 mutants with different aggregation profiles induce distinct pathologies in Drosophila. PLoS One *3*, e1703.

Jahn, T.R., and Radford, S.E. (2008). Folding versus aggregation: polypeptide conformations on competing pathways. Arch. Biochem. Biophys. *469*, 100–117.

Jain, S., and Udgaonkar, J.B. (2011). Defining the pathway of worm-like amyloid fibril formation by the mouse prion protein by delineation of the productive and unproductive oligomerization reactions. Biochemistry *50*, 1153–1161.

Jarrett, J.T., and Lansbury, P.T., Jr. (1993). Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie? Cell *73*, 1055–1058.

Jayaraman, M., Kodali, R., Sahoo, B., Thakur, A.K., Mayasundari, A., Mishra, R., Peterson, C.B., and Wetzel, R. (2011). Slow amyloid nucleation via α-helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments. J. Mol. Biol. *415*, 881–899.

Jones, E.M., and Surewicz, W.K. (2005). Fibril conformation as the basis of species- and strain-dependent seeding specificity of mammalian prion amyloids. Cell *121*, 63–72.

Kaminski Schierle, G.S., Bertoncini, C.W., Chan, F.T.S., van der Goot, A.T., Schwedler, S., Skepper, J., Schlachter, S., van Ham, T., Esposito, A., Kumita, J.R., et al. (2011). A FRET sensor for non-invasive imaging of amyloid formation in vivo. ChemPhysChem *12*, 673–680.

Kayed, R., Head, E., Thompson, J.L., McIntire, T.M., Milton, S.C., Cotman, C.W., and Glabe, C.G. (2003). Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. Science *300*, 486–489.

Kayed, R., Head, E., Sarsoza, F., Saing, T., Cotman, C.W., Necula, M., Margol, L., Wu, J., Breydo, L., Thompson, J.L., et al. (2007). Fibril specific, conformation dependent antibodies recognize a generic epitope common to amyloid fibrils and fibrillar oligomers that is absent in prefibrillar oligomers. Mol. Neurodegener. *2*, 18.

Kayed, R., Pensalfini, A., Margol, L., Sokolov, Y., Sarsoza, F., Head, E., Hall, J., and Glabe, C. (2009). Annular protofibrils are a structurally and functionally distinct type of amyloid oligomer. J. Biol. Chem. *284*, 4230–4237.

Kaylor, J., Bodner, N., Edridge, S., Yamin, G., Hong, D.P., and Fink, A.L. (2005). Characterization of oligomeric intermediates in alpha-synuclein fibrillation: FRET studies of Y125W/Y133F/Y136F alpha-synuclein. J. Mol. Biol. *353*, 357–372.

Kheterpal, I., and Wetzel, R. (2006). Hydrogen/deuterium exchange mass spectrometry—a window into amyloid structure. Acc. Chem. Res. *39*, 584–593.

Kheterpal, I., Chen, M., Cook, K.D., and Wetzel, R. (2006). Structural differences in Abeta amyloid protofibrils and fibrils mapped by hydrogen exchange—mass spectrometry with on-line proteolytic fragmentation. J. Mol. Biol. *361*, 785–795.

Kłoniecki, M., Jabłonowska, A., Poznański, J., Langridge, J., Hughes, C., Campuzano, I., Giles, K., and Dadlez, M. (2011). Ion mobility separation coupled with MS detects two structural states of Alzheimer's disease Aβ1-40 peptide oligomers. J. Mol. Biol. *407*, 110–124.

Kodali, R., and Wetzel, R. (2007). Polymorphism in the intermediates and products of amyloid assembly. Curr. Opin. Struct. Biol. *17*, 48–57.

Koffie, R.M., Meyer-Luehmann, M., Hashimoto, T., Adams, K.W., Mielke, M.L., Garcia-Alloza, M., Micheva, K.D., Smith, S.J., Kim, M.L., Lee, V.M., et al. (2009). Oligomeric amyloid beta associates with postsynaptic densities and correlates with excitatory synapse loss near senile plaques. Proc. Natl. Acad. Sci. USA *106*, 4012–4017.

Kremer, J.J., Pallitto, M.M., Sklansky, D.J., and Murphy, R.M. (2000). Correlation of beta-amyloid aggregate size and hydrophobicity with decreased bilayer fluidity of model membranes. Biochemistry *39*, 10309–10318.

Krishnan, R., and Lindquist, S.L. (2005). Structural insights into a yeast prion illuminate nucleation and strain diversity. Nature *435*, 765–772.

Kumar, S., and Udgaonkar, J.B. (2009). Conformational conversion may precede or follow aggregate elongation on alternative pathways of amyloid protofibril formation. J. Mol. Biol. *385*, 1266–1276.

Lacor, P.N., Buniel, M.C., Chang, L., Fernandez, S.J., Gong, Y., Viola, K.L., Lambert, M.P., Velasco, P.T., Bigio, E.H., Finch, C.E., et al. (2004). Synaptic targeting by Alzheimer's-related amyloid beta oligomers. J. Neurosci. *24*, 10191–10200.

Ladiwala, A.R.A., Lin, J.C., Bale, S.S., Marcelino-Cruz, A.M., Bhattacharya, M., Dordick, J.S., and Tessier, P.M. (2010). Resveratrol selectively remodels soluble oligomers and fibrils of amyloid Abeta into off-pathway conformers. J. Biol. Chem. *285*, 24228–24237.

Ladiwala, A.R.A., Dordick, J.S., and Tessier, P.M. (2011). Aromatic small molecules remodel toxic soluble oligomers of amyloid beta through three independent pathways. J. Biol. Chem. *286*, 3209–3218.

Lambert, M.P., Viola, K.L., Chromy, B.A., Chang, L., Morgan, T.E., Yu, J., Venton, D.L., Krafft, G.A., Finch, C.E., and Klein, W.L. (2001). Vaccination with soluble Abeta oligomers generates toxicity-neutralizing antibodies. J. Neurochem. *79*, 595–605.

Langkilde, A.E., and Vestergaard, B. (2009). Methods for structural characterization of prefibrillar intermediates and amyloid fibrils. FEBS Lett. *583*, 2600–2609.

Larson, M.E., and Lesné, S.E. (2012). Soluble Aβ oligomer production and toxicity. J. Neurochem. *120* (*Suppl. 1*), 125–139.

Lasagna-Reeves, C.A., Glabe, C.G., and Kayed, R. (2011). Amyloid-β annular protofibrils evade fibrillar fate in Alzheimer disease brain. J. Biol. Chem. *286*, 22122–22130.

Lasagna-Reeves, C.A., Castillo-Carranza, D.L., Sengupta, U., Sarmiento, J., Troncoso, J., Jackson, G.R., and Kayed, R. (2012). Identification of oligomers at early stages of tau aggregation in Alzheimer's disease. FASEB J. Published online January 17, 2012. 10.1096/fj.11-199851.

Lashuel, H.A., Hartley, D., Petre, B.M., Walz, T., and Lansbury, P.T., Jr. (2002a). Neurodegenerative disease: amyloid pores from pathogenic mutations. Nature *418*, 291.

Lashuel, H.A., Petre, B.M., Wall, J., Simon, M., Nowak, R.J., Walz, T., and Lansbury, P.T., Jr. (2002b). Alpha-synuclein, especially the Parkinson's disease-associated mutants, forms pore-like annular and tubular protofibrils. J. Mol. Biol. *322*, 1089–1102.

Last, N.B., Rhoades, E., and Miranker, A.D. (2011). Islet amyloid polypeptide demonstrates a persistent capacity to disrupt membrane integrity. Proc. Natl. Acad. Sci. USA *108*, 9460–9465.

Lee, J., Culyba, E.K., Powers, E.T., and Kelly, J.W. (2011). Amyloid-β forms fibrils by nucleated conformational conversion of oligomers. Nat. Chem. Biol. *7*, 602–609.

Lesné, S., Koh, M.T., Kotilinek, L., Kayed, R., Glabe, C.G., Yang, A., Gallagher, M., and Ashe, K.H. (2006). A specific amyloid-beta protein assembly in the brain impairs memory. Nature *440*, 352–357.

Liu, C., Sawaya, M.R., Cheng, P.N., Zheng, J., Nowick, J.S., and Eisenberg, D. (2011). Characteristics of amyloid-related oligomers revealed by crystal structures of macrocyclic β-sheet mimics. J. Am. Chem. Soc. *133*, 6736–6744.

Lomakin, A., Chung, D.S., Benedek, G.B., Kirschner, D.A., and Teplow, D.B. (1996). On the nucleation and growth of amyloid beta-protein fibrils: detection of nuclei and quantitation of rate constants. Proc. Natl. Acad. Sci. USA *93*, 1125–1129.

Martins, I.C., Kuperstein, I., Wilkinson, H., Maes, E., Vanbrabant, M., Jonckheere, W., Van Gelder, P., Hartmann, D., D'Hooge, R., De Strooper, B., et al. (2008). Lipids revert inert Abeta amyloid fibrils to neurotoxic protofibrils that affect learning in mice. EMBO J. *27*, 224–233.

Mastrangelo, I.A., Ahmed, M., Sato, T., Liu, W., Wang, C., Hough, P., and Smith, S.O. (2006). High-resolution atomic force microscopy of soluble Abeta42 oligomers. J. Mol. Biol. *358*, 106–119.

Morozova-Roche, L.A., Zamotin, V., Malisauskas, M., Ohman, A., Chertkova, R., Lavrikova, M.A., Kostanyan, I.A., Dolgikh, D.A., and Kirpichnikov, M.P. (2004). Fibrillation of carrier protein albebetin and its biologically active constructs. Multiple oligomeric intermediates and pathways. Biochemistry *43*, 9610–9619.

Morris, A.M., Watzky, M.A., Agar, J.N., and Finke, R.G. (2008). Fitting neurological protein aggregation kinetic data via a 2-step, minimal/"Ockham's razor" model: the Finke-Watzky mechanism of nucleation followed by autocatalytic surface growth. Biochemistry *47*, 2413–2427.

Morris, A.M., Watzky, M.A., and Finke, R.G. (2009). Protein aggregation kinetics, mechanism, and curve-fitting: a review of the literature. Biochim. Biophys. Acta *1794*, 375–397.

Näsström, T., Fagerqvist, T., Barbu, M., Karlsson, M., Nikolajeff, F., Kasrayan, A., Ekberg, M., Lannfelt, L., Ingelsson, M., and Bergström, J. (2011). The lipid peroxidation products 4-oxo-2-nonenal and 4-hydroxy-2-nonenal promote the formation of α-synuclein oligomers with distinct biochemical, morphological, and functional properties. Free Radic. Biol. Med. *50*, 428–437.

Nekooki-Machida, Y., Kurosawa, M., Nukina, N., Ito, K., Oda, T., and Tanaka, M. (2009). Distinct conformations of in vitro and in vivo amyloids of huntingtin-exon1 show different cytotoxicity. Proc. Natl. Acad. Sci. USA *106*, 9679–9684.

Noguchi, A., Matsumura, S., Dezawa, M., Tada, M., Yanazawa, M., Ito, A., Akioka, M., Kikuchi, S., Sato, M., Ideno, S., et al. (2009). Isolation and characterization of patient-derived, toxic, high mass amyloid beta-protein (Abeta) assembly from Alzheimer disease brains. J. Biol. Chem. *284*, 32895–32905.

Olofsson, A., Ippel, J.H., Wijmenga, S.S., Lundgren, E., and Ohman, A. (2004). Probing solvent accessibility of transthyretin amyloid by solution NMR spectroscopy. J. Biol. Chem. *279*, 5699–5707.

Olzscha, H., Schermann, S.M., Woerner, A.C., Pinkert, S., Hecht, M.H., Tartaglia, G.G., Vendruscolo, M., Hayer-Hartl, M., Hartl, F.U., and Vabulas, R.M. (2011). Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. Cell *144*, 67–78.

Oma, Y., Kino, Y., Sasagawa, N., and Ishiura, S. (2005). Comparative analysis of the cytotoxicity of homopolymeric amino acids. Biochim. Biophys. Acta *1748*, 174–179.

Ono, K., Ikeda, T., Takasaki, J.I., and Yamada, M. (2011). Familial Parkinson disease mutations influence α-synuclein assembly. Neurobiol. Dis. *43*, 715–724.

Orte, A., Birkett, N.R., Clarke, R.W., Devlin, G.L., Dobson, C.M., and Klenerman, D. (2008). Direct characterization of amyloidogenic oligomers by single-molecule fluorescence. Proc. Natl. Acad. Sci. USA *105*, 14424–14429.

Orte, A., Clarke, R.W., and Klenerman, D. (2011). Single-molecule fluorescence coincidence spectroscopy and its application to resonance energy transfer. ChemPhysChem *12*, 491–499.

Pagano, K., Bemporad, F., Fogolari, F., Esposito, G., Viglino, P., Chiti, F., and Corazza, A. (2010). Structural and dynamics characteristics of acylphosphatase

from Sulfolobus solfataricus in the monomeric state and in the initial native-like aggregates. J. Biol. Chem. *285*, 14689–14700.

Pan, J., Han, J., Borchers, C.H., and Konermann, L. (2011). Conformer-specific hydrogen exchange analysis of Aβ(1-42) oligomers by top-down electron capture dissociation mass spectrometry. Anal. Chem. *83*, 5386–5393.

Paravastu, A.K., Qahwash, I., Leapman, R.D., Meredith, S.C., and Tycko, R. (2009). Seeded growth of beta-amyloid fibrils from Alzheimer's brain-derived fibrils produces a distinct fibril structure. Proc. Natl. Acad. Sci. USA *106*, 7443–7448.

Patterson, K.R., Remmers, C., Fu, Y., Brooker, S., Kanaan, N.M., Vana, L., Ward, S., Reyes, J.F., Philibert, K., Glucksman, M.J., and Binder, L.I. (2011). Characterization of prefibrillar Tau oligomers in vitro and in Alzheimer disease. J. Biol. Chem. *286*, 23063–23076.

Pease, L.F., 3rd, Sorci, M., Guha, S., Tsai, D.H., Zachariah, M.R., Tarlov, M.J., and Belfort, G. (2010). Probing the nucleus model for oligomer formation during insulin amyloid fibrillogenesis. Biophys. J. *99*, 3979–3985.

Petty, S.A., and Decatur, S.M. (2005). Experimental evidence for the reorganization of beta-strands within aggregates of the Abeta(16-22) peptide. J. Am. Chem. Soc. *127*, 13488–13489.

Petty, S.A., Adalsteinsson, T., and Decatur, S.M. (2005). Correlations among morphology, beta-sheet stability, and molecular structure in prion peptide aggregates. Biochemistry *44*, 4720–4726.

Pifer, P.M., Yates, E.A., and Legleiter, J. (2011). Point mutations in Aβ result in the formation of distinct polymorphic aggregates in the presence of lipid bilayers. PLoS ONE *6*, e16248.

Plakoutsi, G., Bemporad, F., Calamai, M., Taddei, N., Dobson, C.M., and Chiti, F. (2005). Evidence for a mechanism of amyloid formation involving molecular reorganisation within native-like precursor aggregates. J. Mol. Biol. *351*, 910–922.

Pountney, D.L., Voelcker, N.H., and Gai, W.P. (2005). Annular alpha-synuclein oligomers are potentially toxic agents in alpha-synucleinopathy. Hypothesis. Neurotox. Res. *7*, 59–67.

Qi, W., Zhang, A., Patel, D., Lee, S., Harrington, J.L., Zhao, L., Schaefer, D., Good, T.A., and Fernandez, E.J. (2008). Simultaneous monitoring of peptide aggregate distributions, structure, and kinetics using amide hydrogen exchange: application to Abeta(1-40) fibrillogenesis. Biotechnol. Bioeng. *100*, 1214–1227.

Quintas, A., Vaz, D.C., Cardoso, I., Saraiva, M.J., and Brito, R.M. (2001). Tetramer dissociation and monomer partial unfolding precedes protofibril formation in amyloidogenic transthyretin variants. J. Biol. Chem. *276*, 27207–27213.

Relini, A., Torrassa, S., Ferrando, R., Rolandi, R., Campioni, S., Chiti, F., and Gliozzi, A. (2010). Detection of populations of amyloid-like protofibrils with different physical properties. Biophys. J. *98*, 1277–1284.

Reynolds, N.P., Soragni, A., Rabe, M., Verdes, D., Liverani, E., Handschin, S., Riek, R., and Seeger, S. (2011). Mechanism of membrane interaction and disruption by α-synuclein. J. Am. Chem. Soc. *133*, 19366–19375.

Sandal, M., Valle, F., Tessari, I., Mammi, S., Bergantino, E., Musiani, F., Brucale, M., Bubacco, L., and Samorì, B. (2008). Conformational equilibria in monomeric alpha-synuclein at the single-molecule level. PLoS Biol. *6*, e6.

Sandberg, A., Luheshi, L.M., Söllvander, S., Pereira de Barros, T., Macao, B., Knowles, T.P., Biverstål, H., Lendel, C., Ekholm-Petterson, F., Dubnovitsky, A., et al. (2010). Stabilization of neurotoxic Alzheimer amyloid-beta oligomers by protein engineering. Proc. Natl. Acad. Sci. USA *107*, 15595–15600.

Serio, T.R., Cashikar, A.G., Kowal, A.S., Sawicki, G.J., Moslehi, J.J., Serpell, L., Arnsdorf, M.F., and Lindquist, S.L. (2000). Nucleated conformational conversion and the replication of conformational information by a prion determinant. Science *289*, 1317–1321.

Shankar, G.M., Li, S., Mehta, T.H., Garcia-Munoz, A., Shepardson, N.E., Smith, I., Brett, F.M., Farrell, M.A., Rowan, M.J., Lemere, C.A., et al. (2008). Amyloid-beta protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. Nat. Med. *14*, 837–842.

Smith, D.P., Radford, S.E., and Ashcroft, A.E. (2010). Elongated oligomers in beta2-microglobulin amyloid assembly revealed by ion mobility spectrometry-mass spectrometry. Proc. Natl. Acad. Sci. USA *107*, 6794–6798.

Soldi, G., Bemporad, F., and Chiti, F. (2008). The degree of structural protection at the edge beta-strands determines the pathway of amyloid formation in globular proteins. J. Am. Chem. Soc. *130*, 4295–4302.

Souillac, P.O., Uversky, V.N., Millett, I.S., Khurana, R., Doniach, S., and Fink, A.L. (2002). Elucidation of the molecular mechanism during the early events in immunoglobulin light chain amyloid fibrillation. Evidence for an off-pathway oligomer at acidic pH. J. Biol. Chem. *277*, 12666–12679.

Souillac, P.O., Uversky, V.N., and Fink, A.L. (2003). Structural transformations of oligomeric intermediates in the fibrillation of the immunoglobulin light chain LEN. Biochemistry *42*, 8094–8104.

Stefani, M. (2010). Structural polymorphism of amyloid oligomers and fibrils underlies different fibrillization pathways: immunogenicity and cytotoxicity. Curr. Protein Pept. Sci. *11*, 343–354.

Strodel, B., Fitzpatrick, A.W., Vendruscolo, M., Dobson, C.M., and Wales, D.J. (2008). Characterizing the first steps of amyloid formation for the ccbeta peptide. J. Phys. Chem. B *112*, 9998–10004.

Thakur, A.K., Jayaraman, M., Mishra, R., Thakur, M., Chellgren, V.M., Byeon, I.J., Anjum, D.H., Kodali, R., Creamer, T.P., Conway, J.F., et al. (2009). Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. Nat. Struct. Mol. Biol. *16*, 380–389.

Tsigelny, I.F., Crews, L., Desplats, P., Shaked, G.M., Sharikov, Y., Mizuno, H., Spencer, B., Rockenstein, E., Trejo, M., Platoshyn, O., et al. (2008). Mechanisms of hybrid oligomer formation in the pathogenesis of combined Alzheimer's and Parkinson's diseases. PLoS ONE *3*, e3135.

Vestergaard, B., Groenning, M., Roessle, M., Kastrup, J.S., van de Weert, M., Flink, J.M., Frokjaer, S., Gajhede, M., and Svergun, D.I. (2007). A helical structural nucleus is the primary elongating unit of insulin amyloid fibrils. PLoS Biol. *5*, e134.

Walsh, D.M., Hartley, D.M., Kusumoto, Y., Fezoui, Y., Condron, M.M., Lomakin, A., Benedek, G.B., Selkoe, D.J., and Teplow, D.B. (1999). Amyloid beta-protein fibrillogenesis. Structure and biological activity of protofibrillar intermediates. J. Biol. Chem. *274*, 25945–25952.

Walsh, D.M., and Selkoe, D.J. (2007). A beta oligomers—a decade of discovery. J. Neurochem. *101*, 1172–1184.

Wei, L., Jiang, P., Xu, W., Li, H., Zhang, H., Yan, L., Chan-Park, M.B., Liu, X.W., Tang, K., Mu, Y., and Pervushin, K. (2011). The molecular basis of distinct aggregation pathways of islet amyloid polypeptide. J. Biol. Chem. *286*, 6291–6300.

Williams, A.D., Sega, M., Chen, M., Kheterpal, I., Geva, M., Berthelier, V., Kaleta, D.T., Cook, K.D., and Wetzel, R. (2005). Structural properties of Abeta protofibrils stabilized by a small molecule. Proc. Natl. Acad. Sci. USA *102*, 7115–7120.

Winner, B., Jappelli, R., Maji, S.K., Desplats, P.A., Boyer, L., Aigner, S., Hetzer, C., Loher, T., Vilar, M., Campioni, S., et al. (2011). In vivo demonstration that alpha-synuclein oligomers are toxic. Proc. Natl. Acad. Sci. USA *108*, 4194–4199.

Yamaguchi, K., Takahashi, S., Kawai, T., Naiki, H., and Goto, Y. (2005). Seeding-dependent propagation and maturation of amyloid fibril conformation. J. Mol. Biol. *352*, 952–960.

Yu, L., Edalji, R., Harlan, J.E., Holzman, T.F., Lopez, A.P., Labkovsky, B., Hillen, H., Barghorn, S., Ebert, U., Richardson, P.L., et al. (2009). Structural characterization of a soluble amyloid beta-peptide oligomer. Biochemistry *48*, 1870–1877.

Yu, X., and Zheng, J. (2011). Polymorphic structures of Alzheimer's β-amyloid globulomers. PLoS ONE *6*, e20575.

Zampagni, M., Cascella, R., Casamenti, F., Grossi, C., Evangelisti, E., Wright, D., Becatti, M., Liguri, G., Mannini, B., Campioni, S., et al. (2011). A comparison of the biochemical modifications caused by toxic and nontoxic protein oligomers in cells. J. Cell Mol. Med. *15*, 2106–2116.

Point of VIEW

ACS chemical biology

# When Conjugated Polymers Meet Amyloid Fibrils

**Cliff I. Stains and Indraneel Ghosh***

Department of Chemistry, University of Arizona, 1306 E. University Boulevard, Tucson, Arizona 85721



**ABSTRACT** In the early 1900s, Alois Alzheimer diagnosed one of his patients with a devastating neurological impairment, and this form of dementia became known as Alzheimer's disease (AD). Much research over the past century has clearly established that numerous human diseases, ranging from AD and Parkinson's disease to dialysis-related amyloidosis, are best characterized by the abnormal aggregation of specific proteins. However, in the case of AD, the true toxic molecular species is still debated. Thus, the recent development of new diagnostic agents capable of distinguishing between different morphologies of aggregated proteins is of much interest.

Alzheimer's disease (AD) is characterized by the extracellular deposits of the 39–42 amino acid amyloid-β (Aβ) peptides along with neurofibrillary tangles that stain with Congo red in the brains of patients. The Aβ peptides arise from cleavage of the extracellular portion of the transmembrane amyloid-precursor protein (APP) by β- and γ-secretases (*1–3*). The toxicity of the processed Aβ peptide and its aggregates (Figure 1, panel a) may result from a combination of apoptosis, disrupted $Ca^{2+}$ homeostasis, toxic radicals, and complement formation. Over the past decade, Selkoe, Lansbury, Teplow, Kelly, Dobson, Prusiner, and many others have helped establish a nucleation-dependent paradigm for fibril formation that appears to be general for most proteins tested (Figure 1, panel a). Numerous experiments have also demonstrated that the rate of fibrillization and the morphology of the final fibrillar state of "amyloidogenic" proteins are strongly influenced by environmental factors (pH, salt, temperature, agitation, *etc.*), by chemicals (proteins, lipids, cholesterol, metals, *etc.*), and by the nature of the seeding agent (*4*). Recent evidence strongly suggests that soluble oligomeric intermediates rather than the final fibrils are responsible for neurological toxicity (*5–7*), though equilibrium between these different states is certainly possible (Figure 1, panel a) (*6, 8*). Perhaps most important, a stronger correlation exists between soluble Aβ (monomer and oligomers of Aβ) in the brain and early cognitive dysfunction than between the Aβ deposits stained by Congo red and the clinical severity of AD (*9–12*). Even with this growing knowledge, the origin and progression of AD remain a complex and formidable challenge because no genetic markers, diagnostic agents, or drugs directly address the progress of AD. A recent review by Kodali and Wetzel discusses the numerous structural polymorphisms possible for Aβ (*13*). Clearly, it would be an important breakthrough to provide new chemically tuned reagents that are selective for different classes of oligomers and fibers and to correlate them to disease outcome. Thus, Nilsson *et al.*'s (*14*) recent identification of a class of thiophene-based polymers that distinguish between different classes of fibrils is very significant.

**Current Clinical Status of AD.** The diagnosis of AD-associated dementia is currently based on clinical diagnosis rather than chemical or biological tests. Pathological findings have until recently been postmortem through the selective staining of senile plaques, primarily β-sheet fibrillar aggregates. Postmortem diagnosis unfortunately provides little relief for those afflicted with AD and perhaps surprisingly does not provide a measure of the severity of the disease; similarly stained plaques have also been observed in normal aging. Current medications for AD, primarily cholinesterase inhibitors (*15*) or *N*-methyl D-aspartate antagonists (*16*), treat some of the symptoms of the disease. However, new avenues are being actively explored that include secretase inhibitors as well as direct Aβ

*Corresponding author, ghosh@email.arizona.edu.

Published online August 17, 2007
10.1021/cb700165v CCC: $37.00

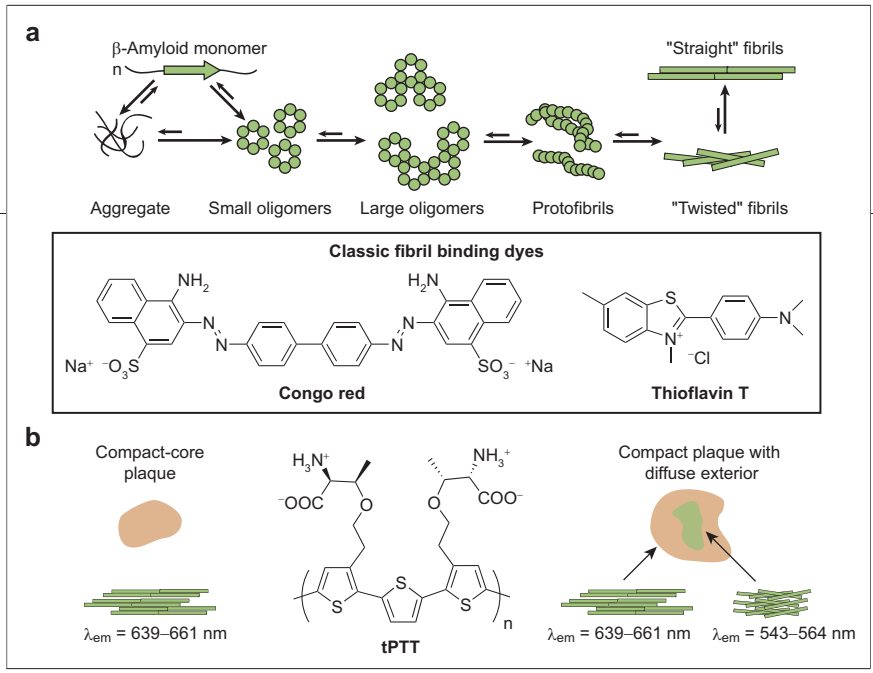


**Figure 1. Model of fibrillogenesis and illustration of tPTT dye activity. a) A possible model for the Aβ fibrillogenesis, where two final fibrillar forms are shown. The straight fibrillar isoform is stained orange by tPTT, and the twisted isoform is stained green. The classic dyes Congo red and ThT do not distinguish between these isoforms unless they are under polarized light. b) The thiophene-based conjugated polymer, tPTT, selectively stains morphologically different plaques in brain sections of transgenic mice (tg-APP$_{swe}$). In one class, the plaques are stained completely orange, whereas in another, they have diffuse green centers and orange exteriors.**

targeting therapeutics (*17*). Therapeutics go hand-in-hand with diagnostics, and recent advances in SPECT probes (*18, 19*) for *in vivo* brain imaging are based on analogues of the classic Congo red and thioflavin T (ThT) dyes. However, this method will likely suffer from intrinsic problems because of its inability to distinguish between plaques from healthy and diseased patients. Thus, new chemical reagents that may be truly useful in the early diagnosis of AD can perhaps distinguish between normal and disease-associated fibrils or by directly imaging the soluble "toxic" oligomers. Any advance in these areas has the potential to significantly impact how AD is diagnosed and treated. Moreover, such reagents would be of immense benefit in testing new therapeutic modalities both *in vivo* and *in vitro* if they can be shown to correlate specific Aβ species with the clinical symptoms of AD.

Congo red was the first small molecule shown to bind to amyloid in tissue sections and exhibited a yellow-green birefringence under cross polarizers (*20, 21*). Several decades later, ThT and ThS were also shown to characteristically stain amyloid deposits (*22, 23*). These dyes have remained the classic reagents for determining β-sheet-mediated fibrillization, although they bind different sites (*24*). In the case of Congo red, beautiful polarized light microscopy studies have shown that Congo red is oriented along the long axes of fibrils in Aβ plaques and can be utilized to recognize plaques from different protein aggregates (*25*). Despite the knowledge gained from many experiments with these dyes, a facile method for distinguishing between Aβ aggregates has yet to emerge.

**Polymers and Fibrils.** This brings us to the current work from Nilsson *et al.*, which describes a new class of conjugated polythiophene (PT)-based dyes that resolve differences between fibril conformations (Figure 1, panel b). This study is exciting because it connects the conducting polymer field (*26*) popularized by Heeger, MacDiarmid, and Shirakawa with protein-misfolding diseases. PTs are polymers of conjugated sulfur heterocycles and possess novel electrical and optical properties resulting from electron delocalization along the polymer backbone. The synthesis of PTs was first reported in the late 1980s (*27*); these polymers have since been exploited in various sensor applications (*28, 29*). Of interest is that Swager and coworkers synthesized 2,5-diphenylthiophene derivatives that bind Aβ with high affinity and have demonstrated that they can be used to image amyloid deposits in the brain of a transgenic mouse model of AD by multiphoton spectroscopy (*30*). More recently, Kung and coworkers have also synthesized numerous thiophene derivatives to target Aβ with high affinities, which they hope to develop as novel positron emission tomography tracers in patients with AD (*31*). Although these studies are promising, neither reported on the use of thiophene derivatives for the selective recognition of different types of amyloid fibers.

Nilsson and coworkers now provide a new reagent for the direct fluorescence imaging of amyloid fibrils, which can parse them into either "straight" or "twisted" species (*14*). Several classes of thiophene derivatives were investigated, and the most useful analogue, poly[(5,5′′)tertiophene -(2*S*,3*R*)-2-amino-3-(2-{3′′-[2-((1*R*,2*S*)-2-amino-2-carboxy-1-methyl-ethoxy)-ethyl]-[2,2′;5′,2′′]terthiophen-3-yl}-ethoxy)-butyric acid] (tPTT), contains chiral amino acid "side chains" (Figure 1, panel b) that likely aid in both solubility and its amyloid recognition properties. They clearly observe in *in vitro* experiments that tPTT stains fibrils obtained from unstirred Aβ solutions orange (639–661 nm), whereas it stains fibrils obtained from agitated Aβ solutions green (543–564 nm). It is likely that the polymer chains of tPTT are aligned along the fibrillar axis as documented for a related polymer, PT acetic acid (*32*). Thus, the orange luminescence likely arises from a linear arrangement of the dye in straight fibers, whereas the green luminescence arises from binding twisted fibers that disrupt conjugation and dye alignment. Even more interesting were the results when these dyes were used to investigate brain sections of transgenic mouse (tg-APP$_{swe}$) models of AD. The fluorescence microscopy images showed two distinct classes of amyloid plaque, one the authors dub a compact-core plaque that primarily

stained orange, whereas a more frequently found plaque form stained orange at the edges and green at the center (Figure 1, panel b). These data agree with those from Jin and others (*25, 33*), who have also shown that Aβ plaques contain a diffuse center and fibrillar exterior. However, this is the first example where two different classes of amyloid plaque in an AD mouse model have been directly identified, and it opens many new doors into investigating AD pathophysiology. Numerous questions arise from this work, some of which will certainly be addressed in the near future. For example, are the two plaque isoforms observed *in vitro* interconvertible, or are they kinetically trapped (Figure 1, panel b)? Can seeding experiments with one fibrillar isoform result in the selective enrichment of that particular species? Are different classes of fibrillar species also observed for other amyloidogenic proteins if those proteins are stained with tPTT? Does tPTT differentially stain the alternative nontoxic amyloid isoforms reported by Kiessling, Murphy, and coworkers(*34, 35*)? Can the two types of *in vivo* and *ex vivo* plaques be isolated and tested for their toxicity against neuronal cells? And finally and perhaps most important, can this approach be utilized to demonstrate a correlation between the type of plaque and the severity of AD?

Future work in this area will certainly involve combinatorial optimization studies on the thiophene core, which may provide more effective binders for different classes of fibrils. It may even be possible to develop reagents that selectively stain soluble oligomeric species that have been shown to be sequestered by antibodies (*36*) and small proteins (*37*). It would also be of interest to establish the maximum effective conjugation length necessary for selective luminescence for tPTT-like molecules. Smaller, appropriately functionalized, but still selectively luminescent thiophenes may cross the blood–brain barrier and be utilized in powerful multiphoton imaging modalities for *in vivo* imaging that can guide AD treatment. Clearly, much can be gained in the chemistry and biology of amyloid imaging by going afield and borrowing from the rich chemistry and physics available in the conducting polymer arena.

## REFERENCES


1. Fraering, P. C., Ye, W. J., Strub, J. M., Dolios, G., LaVoie, M. J., Ostaszewski, B. L., van Dorsselaer, A., Wang, R., Selkoe, D. J., and Wolfe, M. S. (2004) Purification and characterization of the human gamma-secretase complex, *Biochemistry 43*, 9774–9789.
2. Kimberly, W. T., Esler, W. P., Ye, W. J., Ostaszewski, B. L., Gao, J., Diehl, T., Selkoe, D. J., and Wolfe, M. S. (2003) Notch and the amyloid precursor protein are cleaved by similar gamma-secretase(s), *Biochemistry 42*, 137–144.
3. Selkoe, D. J., and Schenk, D. (2003) Alzheimer's disease: molecular understanding predicts amyloid-based therapeutics, *Annu. Rev. Pharmacol. Toxicol. 43*, 545–584.
4. Rochet, J. C., and Lansbury, P. T. (2000) Amyloid fibrillogenesis: themes and variations, *Curr. Opin. Struct. Biol. 10*, 60–68.
5. Dahlgren, K. N., Manelli, A. M., Stine, W. B., Baker, L. K., Krafft, G. A., and LaDu, M. J. (2002) Oligomeric and fibrillar species of amyloid-beta peptides differentially affect neuronal viability, *J. Biol. Chem. 277*, 32046–32053.
6. Hartley, D. M., Walsh, D. M., Ye, C. P. P., Diehl, T., Vasquez, S., Vassilev, P. M., Teplow, D. B., and Selkoe, D. J. (1999) Protofibrillar intermediates of amyloid beta-protein induce acute electrophysiological changes and progressive neurotoxicity in cortical neurons, *J. Neurosci. 19*, 8876–8884.
7. Walsh, D. M., Klyubin, I., Fadeeva, J. V., Cullen, W. K., Anwyl, R., Wolfe, M. S., Rowan, M. J., and Selkoe, D. J. (2002) Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation *in vivo*, *Nature 416*, 535–539.
8. Lambert, M. P., Barlow, A. K., Chromy, B. A., Edwards, C., Freed, R., Liosatos, M., Morgan, T. E., Rozovsky, I., Trommer, B., Viola, K. L., Wals, P., Zhang, C., Finch, C. E., Krafft, G. A., and Klein, W. L. (1998) Diffusible, nonfibrillar ligands derived from A beta(1–42) are potent central nervous system neurotoxins, *Proc. Natl. Acad. Sci. U.S.A. 95*, 6448–6453.
9. Lemere, C. A., Blusztajn, J. K., Yamaguchi, H., Wisniewski, T., Saido, T. C., and Selkoe, D. J. (1996) Sequence of deposition of heterogeneous amyloid beta-peptides and APO E in Down syndrome: implications for initial events in amyloid plaque formation, *Neurobiol. Dis. 3*, 16–32.
10. McLean, C. A., Cherny, R. A., Fraser, F. W., Fuller, S. J., Smith, M. J., Beyreuther, K., Bush, A. I., and Masters, C. L. (1999) Soluble pool of A beta amyloid as a determinant of severity of neurodegeneration in Alzheimer's disease, *Ann. Neurol. 46*, 860–866.
11. Naslund, J., Haroutunian, V., Mohs, R., Davis, K. L., Davies, P., Greengard, P., and Buxbaum, J. D. (2000) Correlation between elevated levels of amyloid beta-peptide in the brain and cognitive decline, *JAMA, J. Am. Med. Assoc. 283*, 1571–1577.
12. Terry, R. D., Masliah, E., Salmon, D. P., Butters, N., Deteresa, R., Hill, R., Hansen, L. A., and Katzman, R. (1991) Physical basis of cognitive alterations in Alzheimers-disease—synapse loss is the major correlate of cognitive impairment, *Ann. Neurol. 30*, 572–580.
13. Kodali, R., and Wetzel, R. (2007) Polymorphism in the intermediates and products of amyloid assembly, *Curr. Opin. Struct. Biol. 17*, 48–57.
14. Nilsson, K. P. R., Aslund, A., Berg, I., Nystrom, S., Konradsson, P., Herland, A., Inganas, O., Stabo-Eeg, F., Lindgren, M., Westermark, G. T., Lannfelt, L., Nilsson, L. N. G., and Hammarstrom, P. (2007) Imaging distinct conformational states of amyloid-beta fibrils in Alzheimer's disease using novel luminescent probes, *ACS Chem. Biol. 2*, 553–560.
15. Trinh, N. H., Hoblyn, J., Mohanty, S. U., and Yaffe, K. (2003) Efficacy of cholinesterase inhibitors in the treatment of neuropsychiatric symptoms and functional impairment in Alzheimer disease—a meta-analysis, *JAMA, J. Am. Med. Assoc. 289*, 210–216.
16. Rogawski, M. A., and Wenk, G. L. (2003) The neuropharmacological basis for the use of memantine in the treatment of Alzheimer's disease, *CNS Drug Rev. 9*, 275–308.
17. Aisen, P. S. (2005) The development of anti-amyloid therapy for Alzheimer's disease-from secretase modulators to polymerisation inhibitors, *CNS Drugs 19*, 989–996.
18. Cai, L. S., Chin, F. T., Pike, V. W., Toyama, H., Liow, J. S., Zoghbi, S. S., Modell, K., Briard, E., Shetty, H. U., Sinclair, K., Donohue, S., Tipre, D., Kung, M. P., Dagostin, C., Widdowson, D. A., Green, M., Gao, W., Herman, M. M., Ichise, M., and Innis, R. B. (2004) Synthesis and evaluation of two F-18-labeled 6-iodo-2-(4′-*N*,*N*-dimethylamino)phenylimidazo[1,2-*a*]pyridine derivatives as prospective radioligands for β-amyloid in Alzheimer's disease, *J. Med. Chem. 47*, 2208–2218.
19. Mathis, C. A., Wang, Y., and Klunk, W. E. (2004) Imaging beta-amyloid plaques and neurofibrillary tangles in the aging human brain, *Curr. Pharm. Des. 10*, 1469–1492.
20. Bennhold, H. (1923) Über die Ausscheidung intravenös einverleibten kongorotes bei den verschiedensten erkrankungen insbesondere bei amyloidosis, *Dtsch. Arch. Klin. Med. 142*, 32–46.
21. Divry, P. (1927) Etude histochimique des plaques seniles, *J. Neurol. Psychiatry 27*, 643–657.
22. Kelenyi, G. (1967) On the histochemistry of azo group-free thiazole dyes, *J. Histochem. Cytochem. 15*, 172–180.
23. Vassar, P. S., and Culling, C.F.A. (1959) Fluorescent stains with special reference to amyloid and connective tissues, *Arch. Pathol. 68*, 487–498.
24. Levine, H. (2005) Multiple ligand binding sites on A beta(1–40) fibrils, *Amyloid 12*, 5–14.
25. Jin, L. W., Claborn, K. A., Kurimoto, M., Geday, M. A., Maezawa, I., Sohraby, F., Estrada, M., Kaminsky, W., and Kahr, B. (2003) Imaging linear birefringence and dichroism in cerebral amyloid pathologies, *Proc. Natl. Acad. Sci. U.S.A. 100*, 15294–15298.

26. Argun, A. A., Aubert, P. H., Thompson, B. C., Schwendeman, I., Gaupp, C. L., Hwang, J., Pinto, N. J., Tanner, D. B., MacDiarmid, A. G., and Reynolds, J. R. (2004) Multicolored electrochromism polymers: structures and devices, *Chem. Mater. 16*, 4401–4412.
27. Patil, A. O., Ikenoue, Y., Wudl, F., and Heeger, A. J. (1987) Water-soluble conducting polymers, *J. Am. Chem. Soc. 109*, 1858–1859.
28. McQuade, D. T., Pullen, A. E., and Swager, T. M. (2000) Conjugated polymer-based chemical sensors, *Chem. Rev. 100*, 2537–2574.
29. Thomas, S. W., Joly, G. D., and Swager, T. M. (2007) Chemical sensors based on amplifying fluorescent conjugated polymers, *Chem. Rev. 107*, 1339–1386.
30. Nesterov, E. E., Skoch, J., Hyman, B. T., Klunk, W. E., Bacskai, B. J., and Swager, T. M. (2005) *In vivo* optical imaging of amyloid aggregates in brain: design of fluorescent markers, *Angew. Chem., Int. Ed. 44*, 5452–5456.
31. Chandra, R., Kung, M. P., and Kung, H. F. (2006) Design, synthesis, and structure-activity relationship of novel thiophene derivatives for beta-amyloid plaque imaging, *Bioorg. Med. Chem. Lett. 16*, 1350–1352.
32. Herland, A., Bjork, P., Hania, P. R., Scheblykin, I. G., and Inganas, O. (2007) Alignment of a conjugated polymer onto amyloid-like protein fibrils, *Small 3*, 318–325.
33. Schmidt, M. L., Robinson, K. A., Lee, V. M. Y., and Trojanowski, J. Q. (1995) Chemical and immunological heterogeneity of fibrillar amyloid in plaques of Alzheimers-disease and Downs-syndrome brains revealed by confocal microscopy, *Am. J. Pathol. 147*, 503–515.
34. Lowe, T. L., Strzelec, A., Kiessling, L. L., and Murphy, R. M. (2001) Structure-function relationships for inhibitors of beta-amyloid toxicity containing the recognition sequence KLVFF, *Biochemistry 40*, 7882–7889.
35. Pallitto, M. M., Ghanta, J., Heinzelman, P., Kiessling, L. L., and Murphy, R. M. (1999) Recognition sequence design for peptidyl modulators of β-amyloid aggregation and toxicity, *Biochemistry 38*, 3570–3578.
36. Kayed, R., Head, E., Thompson, J. L., McIntire, T. M., Milton, S. C., Cotman, C. W., and Glabe, C. G. (2003) Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis, *Science 300*, 486–489.
37. Smith, T. J., Stains, C. I., Meyer, S. C., and Ghosh, I. (2006) Inhibition of β-amyloid fibrillization by directed evolution of a β-sheet presenting miniature protein, *J. Am. Chem. Soc. 128*, 14456–14457.

# Protein folding and misfolding

**Christopher M. Dobson**

*University of Cambridge, Department of Chemistry, Lensfield Road, Cambridge CB2 1EW, UK (e-mail: cmd44@cam.ac.uk)*

**The manner in which a newly synthesized chain of amino acids transforms itself into a perfectly folded protein depends both on the intrinsic properties of the amino-acid sequence and on multiple contributing influences from the crowded cellular milieu. Folding and unfolding are crucial ways of regulating biological activity and targeting proteins to different cellular locations. Aggregation of misfolded proteins that escape the cellular quality-control mechanisms is a common feature of a wide range of highly debilitating and increasingly prevalent diseases.**

One of the defining characteristics of a living system is the ability of even the most intricate of its component molecular structures to self-assemble with precision and fidelity. Uncovering the mechanisms through which such processes take place is one of the grand challenges of modern science[1]. The folding of proteins into their compact three-dimensional structures is the most fundamental and universal example of biological self-assembly; understanding this complex process will therefore provide a unique insight into the way in which evolutionary selection has influenced the properties of a molecular system for functional advantage. The wide variety of highly specific structures that result from protein folding and that bring key functional groups into close proximity has enabled living systems to develop astonishing diversity and selectivity in their underlying chemical processes. In addition to generating biological activity, however, we now know that folding is coupled to many other biological processes, including the trafficking of molecules to specific cellular locations and the regulation of cellular growth and differentiation[2]. In addition, only correctly folded proteins have long-term stability in crowded biological environments and are able to interact selectively with their natural partners. It is therefore not surprising that the failure of proteins to fold correctly, or to remain correctly folded, is the origin of a wide variety of pathological conditions. In this article we explore the underlying mechanism of protein folding and of the nature and consequences of misfolding and its links with disease.

## The fundamental mechanism of protein folding

### The concept of an energy landscape

The mechanism by which a polypeptide chain folds to a specific three-dimensional protein structure has until recently been shrouded in mystery. Native states of proteins almost always correspond to the structures that are most thermodynamically stable under physiological conditions[3]. Nevertheless, the total number of possible conformations of any polypeptide chain is so large that a systematic search for this particular structure would take an astronomical length of time. However, it is now clear that the folding process does not involve a series of mandatory steps between specific partly folded states, but rather a stochastic search of the many conformations accessible to a polypeptide chain[3–5]. The inherent fluctuations in the conformation of an unfolded or incompletely folded polypeptide chain enable even residues that are highly separated in the amino-acid sequence to come into contact with one other. Because, on average, native-like interactions between residues are more stable than non-native ones, they are more persistent and the polypeptide chain is able to find its lowest-energy structure by a process of trial and error. Moreover, if the energy surface or 'landscape' has the right shape (see Fig. 1) only a small number of all possible conformations needs to be sampled by any given protein molecule during its transition from a random coil to a native structure[3–6]. Because the landscape is encoded by the amino-acid sequence, natural selection has enabled proteins to evolve so that they are able to fold rapidly and efficiently.

Such a description, based more on the ideas of statistical mechanics and polymer physics than on those of classic chemical dynamics, is often referred to as the 'new view' of protein folding[7]. As well as providing a firm conceptual basis for folding, it has shown that many of the earlier phenomenological descriptions of the folding process are important limiting cases of a general mechanism. These ideas are stimulating the investigation of the most elementary steps in the folding process by both experimental and theoretical procedures. For example, biophysical measurements and computer simulations have revealed that many of the local elements of protein structures can be generated very rapidly; for example, individual α-helices are able to form in less than 100 ns, and β-turns in as little as 1 μs (refs 8, 9). Indeed, the folding *in vitro* of some of the simplest proteins, such as small helical bundles, is completed in less than 50 μs (refs 10, 11). Intriguingly, some other small proteins, particularly those based on β-sheet structures, can take many orders of magnitude longer to fold, as we see below, but such rate changes can be understood to a significant extent in terms of the characteristics of the native structures[12].

A key question is how does the correct fold emerge from such fundamental steps; that is, how is the energy landscape unique to a specific protein defined by its amino-acid sequence. The structural transitions taking place during folding *in vitro* can be investigated in detail by a variety of techniques, ranging from optical methods to NMR spectroscopy[3], some of which can now even be used to follow the behaviour of single molecules[13]. The latter capability is of particular significance in the context of probing the stochastic nature of the folding process (see Fig. 1). Studies of a series of small proteins, typically with 60–100 residues, have been crucial for investigating the most basic steps in folding because these proteins convert from their unfolded states to their native states without the complication of highly populated intermediates. For these systems, monitoring the effects of specific mutations on the kinetics of folding and unfolding has proved to be a seminal technique, because of its ability to probe the role of individual residues in the folding process[14]. Particular insight has come from the use of this approach to analyse the transition states for folding, namely the critical regions of energy surfaces through which all molecules must pass to reach the native fold (see Fig. 1). The results of many studies of these species suggest that the

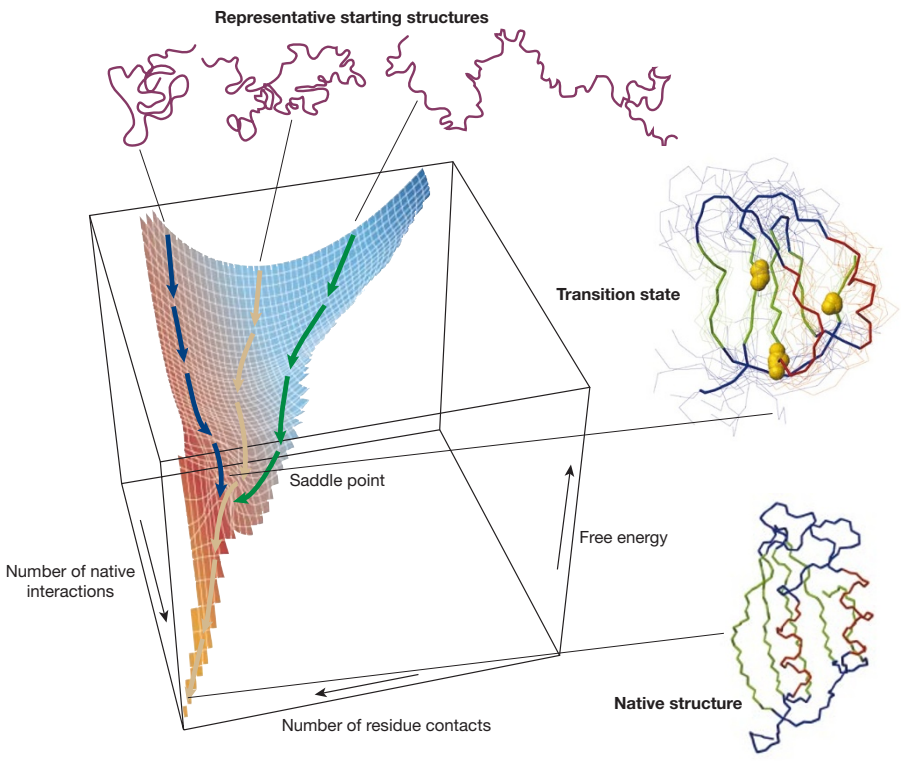


**Figure 1** A schematic energy landscape for protein folding. The surface is derived from a computer simulation of the folding of a highly simplified model of a small protein. The surface 'funnels' the multitude of denatured conformations to the unique native structure. The critical region on a simple surface such as this one is the saddle point corresponding to the transition state, the barrier that all molecules must cross if they are to fold to the native state. Superimposed on this schematic surface are ensembles of structures corresponding to different stages of the folding process. The transition state ensemble was calculated by using computer simulations constrained by experimental data from mutational studies of acylphosphatase[18]. The yellow spheres in this ensemble represent the three 'key residues' in the structure; when these residues have formed their native-like contacts the overall topology of the native fold is established. The structure of the native state is shown at the bottom of the surface; at the top are indicated schematically some contributors to the distribution of unfolded species that represent the starting point for folding. Also indicated on the surface are highly simplified trajectories for the folding of individual molecules. Adapted from ref. 6.

fundamental mechanism of protein folding involves the interaction of a relatively small number of residues to form a folding nucleus, about which the remainder of the structure rapidly condenses[15].

More details of how such a mechanism is able to generate a unique fold have emerged from a range of theoretical studies, particularly involving computer simulation techniques[16]. Of particular significance are investigations that compare the simulation results with experimental observations[6,17]. One approach incorporates experimental measurements directly into the simulations as restraints limiting the regions of conformational space that are explored in each simulation; this strategy has enabled rather detailed structures to be generated for transition states[18] (see Fig. 1). The results suggest that, despite a high degree of disorder, these structures have the same overall topology as the native fold. In essence, interactions involving the key residues force the chain to adopt a rudimentary native-like architecture. Although it is not yet clear exactly how the sequence encodes such characteristics, the essential elements of the fold are likely to be determined primarily by the pattern of hydrophobic and polar residues that favours preferential interactions of specific residues as the structure becomes increasingly compact. Once the correct topology has been achieved, the native structure will then almost invariably be generated during the final stages of folding[18]. Conversely, if these key interactions are not formed, the protein cannot fold to a stable globular structure; this mechanism therefore acts also as a 'quality-control' process by which misfolding can generally be avoided.

### The determinants of protein folds

Secondary structure, the helices and sheets that are found in nearly every native protein structure, is stabilized primarily by hydrogen bonding between the amide and carbonyl groups of the main chain. The formation of such structure is an important element in the overall folding process, although it might not have as fundamental a role as the establishment of the overall chain topology[19]. Perhaps the most dramatic evidence for such a conclusion is the observation of a remarkable correlation between the experimental folding rates of a wide range of small proteins and the complexity of their folds, measured by the contact order[12]. The latter is the average separation in the sequence between residues that are in contact with each other in the native structure. The existence of such a correlation can be rationalized by the argument that a stochastic search process will be more time consuming if the residues that form the nucleus are further away from each other in the sequence. This evidence strongly supports the conclusion that there are relatively simple underlying principles by which the sequence of a protein encodes its structure[20]. Not only will the establishment of such principles reveal in more depth how proteins are able to fold, but it should advance significantly our ability to predict protein folds directly from their sequences and to design sequences that encode novel folds.

For proteins with more than about 100 residues, experiments generally reveal that one (or more) intermediate is significantly populated during the folding process. There has, however, been considerable discussion about the significance of such species: whether they assist the protein to find its correct structure or whether they are traps that inhibit the folding process[21–23]. Regardless of the outcome of this debate, the structural properties of intermediates provide important evidence about the folding of these larger proteins. In particular, they suggest that these proteins generally fold in modules, in other words, folding can take place largely independently in different segments or domains of the protein[6,14]. In such cases, interactions involving key

residues are likely to establish the native-like fold within local regions or domains and also to ensure that the latter then interact appropriately to form the correct overall structure[23,24]. The fully native structure is only acquired when all the native-like interactions have been formed both within and between the domains; this happens in a final cooperative folding step when all the side chains become locked in their unique close-packed arrangement and water is excluded from the protein core[25]. This modular mechanism is appealing because it suggests that highly complex structures might be assembled in manageable pieces. Moreover, such a principle can readily be extended to describe the assembly of other macromolecules, particularly nucleic acids, and even large 'molecular machines' such as the ribosome.

## Protein folding and misfolding in the cell

In a cell, proteins are synthesized on ribosomes from the genetic information encoded in the cellular DNA. Folding *in vivo* is in some cases co-translational; that is, it is initiated before the completion of protein synthesis, whereas the nascent chain is still attached to the ribosome[26]. Other proteins, however, undergo the major part of their folding in the cytoplasm after release from the ribosome, whereas yet others fold in specific compartments, such as mitochondria or the endoplasmic reticulum (ER), after trafficking and translocation through membranes[27,28]. Many details of the folding process depend on the particular environment in which folding takes place, although the fundamental principles of folding, discussed above, are undoubtedly universal. But because incompletely folded proteins must inevitably expose to the solvent at least some regions of structure that are buried in the native state, they are prone to inappropriate interaction with other molecules within the crowded environment of a cell[29]. Living systems have therefore evolved a range of strategies to prevent such behaviour[27–29].

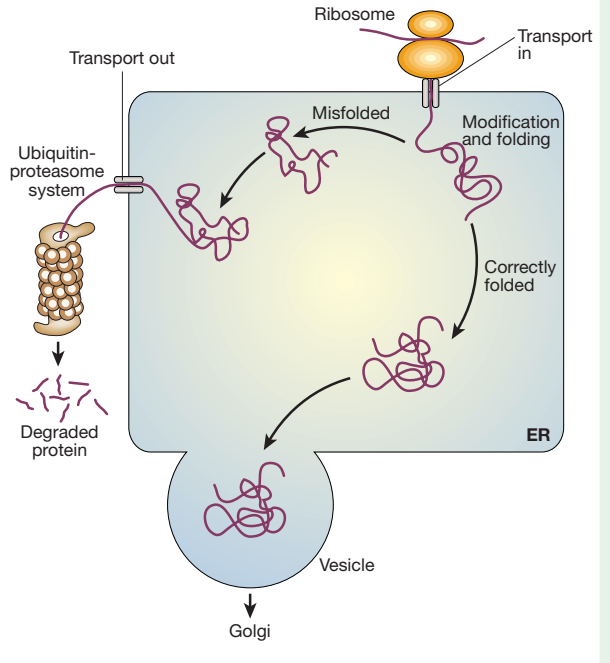


**Figure 2** Regulation of protein folding in the ER. Many newly synthesized proteins are translocated into the ER, where they fold into their three-dimensional structures with the help of a series of molecular chaperones and folding catalysts (not shown). Correctly folded proteins are then transported to the Golgi complex and then delivered to the extracellular environment. However, incorrectly folded proteins are detected by a quality-control mechanism and sent along another pathway (the unfolded protein response) in which they are ubiquitinated and then degraded in the cytoplasm by proteasomes. Adapted from ref. 32.

Of particular importance in this context are the many molecular chaperones that are present in all types of cells and cellular compartments. Some chaperones interact with nascent chains as they emerge from the ribosome, whereas others are involved in guiding later stages of the folding process[27,28]. Molecular chaperones often work in tandem to ensure that the various stages in the folding of such systems are all completed efficiently. Many of the details of the functions of molecular chaperones have been determined from studies of their effects on folding *in vitro*. The best characterized of the chaperones studied in this manner is the bacterial complex involving GroEL, a member of the family of 'chaperonins', and its 'co-chaperone' GroES. Many aspects of the sophisticated mechanism through which this coupled system functions are now well understood[27,28]. Of particular interest is that GroEL, and other members of this class of molecular chaperone, contains a cavity in which incompletely folded polypeptide chains can enter and undergo the final steps in the formation of their native structures while sequestered and protected from the outside world.

Molecular chaperones do not themselves increase the rate of individual steps in protein folding; rather, they increase the efficiency of the overall process by reducing the probability of competing reactions, particularly aggregation. However, there are several classes of folding catalyst that accelerate potentially slow steps in the folding process. The most important are peptidylprolyl isomerases, which increase the rate of *cis–trans* isomerization of peptide bonds involving proline residues, and protein disulphide isomerases, which enhance the rate of formation and reorganization of disulphide bonds[30]. Despite these factors, given the enormous complexity and the stochastic nature of the folding process, it would be remarkable if misfolding never occurred. Clear evidence that molecular chaperones are needed to prevent misfolding and its consequences comes from the fact that the concentrations of many of these species are substantially increased during cellular stress; indeed, the designation of many as heat shock proteins (Hsps) reflects this fact. It is also clear that some molecular chaperones are able not only to protect proteins as they fold but also to rescue misfolded and even aggregated proteins and enable them to have a second chance to fold correctly[27,28]. Active intervention in the folding process requires energy, and ATP is required for most of the molecular chaperones to function with full efficiency.

In eukaryotic systems, many of the proteins that are synthesized in a cell are destined for secretion to the extracellular environment. These proteins are translocated into the ER, where folding takes place before secretion through the Golgi apparatus. The ER contains a wide range of molecular chaperones and folding catalysts, and in addition the proteins that fold here must satisfy a 'quality-control' check before being exported (Fig. 2)[31,32]. Such a process is particularly important because there seem to be few molecular chaperones outside the cell, although one (clusterin), at least, has recently been discovered[33]. This quality-control mechanism involves a remarkable series of glycosylation and deglycosylation reactions that enables correctly folded proteins to be distinguished from misfolded ones[31]. The importance of these regulatory systems is underlined by recent experiments that suggest that a large fraction of all polypeptide chains synthesized in a cell fail to pass this test and are targeted for degradation[34]. Like the 'heat shock response' in the cytoplasm, the 'unfolded protein response' in the ER is also stimulated (upregulated) during stress and, as we shall see below, is strongly linked to the avoidance of misfolding diseases[35].

Folding and unfolding are the ultimate ways of generating and abolishing specific types of cellular activity. In addition, processes as apparently diverse as translocation across membranes, trafficking, secretion, the immune response and regulation of the cell cycle are directly dependent on folding and unfolding events[2]. Failure to fold correctly, or to remain correctly folded, will therefore give rise to the malfunctioning of living systems and hence to disease[36–38]. Some of these diseases (such as cystic fibrosis[36] and some types of cancer[39]) result from proteins folding incorrectly and not being able to exercise

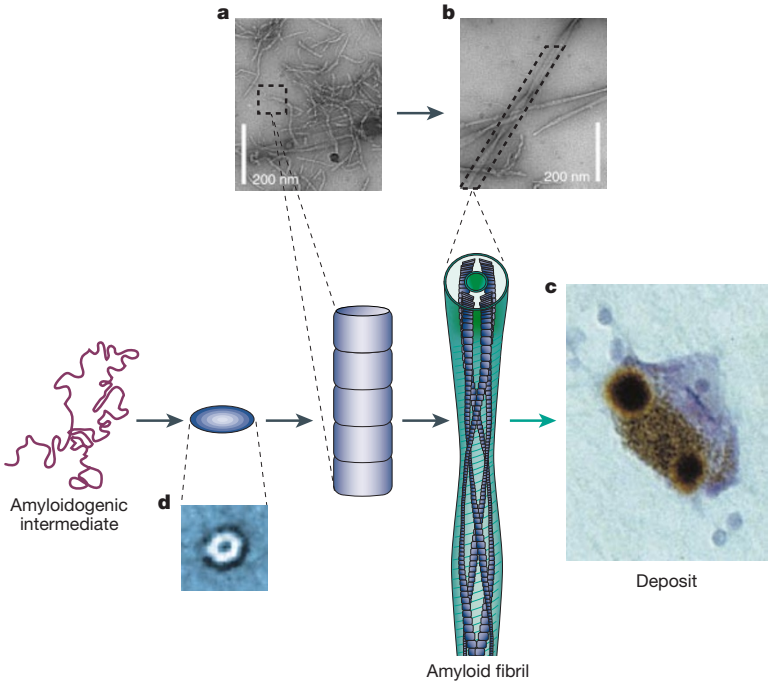


**Figure 3** A schematic representation of the general mechanism of aggregation to form amyloid fibrils. Unfolded or partially unfolded proteins associate with each other to form small, soluble aggregates that undergo further assembly into protofibrils or protofilaments (**a**) and then mature fibrils (**b**, top electron microscope images from ref. 44). The fibrils often accumulate in plaques or other structures such as the Lewy bodies associated with Parkinson's disease (**c**, electron microscope image on right). Some of the early aggregates seem to be amorphous or micellar in nature, although others form ring-shaped species with diameters of approximately 10 nm (**d**, electron microscope image from ref. 53). Adapted from a figure provided by H. A. Lashuel and P. T. Lansbury Jr.

their proper function; many such disorders are familial because the probability of misfolding is often greater in mutational variants. In other cases, proteins with a high propensity to misfold escape all the protective mechanisms and form intractable aggregates within cells or (more commonly) in extracellular space. An increasing number of disorders, including Alzheimer's and Parkinson's diseases, the spongiform encephalopathies and type II diabetes, are directly associated with the deposition of such aggregates in tissues, including the brain, heart and spleen[37,38,40,41]. In the next section we look at the formation of these species.

## Protein aggregation and amyloid formation

Each amyloid disease involves predominantly the aggregation of a specific protein, although a range of other components including additional proteins and carbohydrates are incorporated into the deposits when they form *in vivo*. In neurodegenerative diseases, the quantities of aggregates involved can sometimes be so small as to be almost undetectable, whereas in some systemic diseases literally kilograms of protein can be found in one or more organs[40]. The characteristics of the soluble forms of the 20 or so proteins involved in the well-defined amyloidoses are very varied — they range from intact globular proteins to largely unstructured peptide molecules — but the aggregated forms have many characteristics in common[42]. Amyloid deposits all show specific optical behaviour (such as birefringence) on binding certain dye molecules such as Congo red. The fibrillar structures typical of many of the aggregates have very similar morphologies (long, unbranched and often twisted structures a few nanometres in diameter) and a characteristic 'cross-β' X-ray fibre diffraction pattern. The latter reveals that the organized core structure is composed of β-sheets whose strands run perpendicular to the fibril axis[42]. The ability of polypeptide chains to form amyloid structures is not restricted to the relatively small number of proteins associated with recognized clinical disorders, and it now seems to be a generic feature of polypeptide chains[37,43]. The most compelling evidence for the latter statement is that fibrils can be formed *in vitro* by many other peptides and proteins, including such well-known molecules as myoglobin, and also by homopolymers such as polythreonine or polylysine[37,44].

Although no structure of an amyloid fibril has yet been determined in atomic detail, increasingly convincing models based on data from techniques such as X-ray fibre diffraction[42], cryoelectron microscopy[45] and solid-state NMR[46] are emerging. The core structure of the fibrils seems to be stabilized primarily by interactions, particularly hydrogen bonds, involving the polypeptide main chain. Because the main chain is common to all polypeptides, this observation explains why fibrils formed from polypeptides of very different sequence seem to be so similar[42,43]. In some cases only a handful of the residues of a given protein might be involved in this structure, with the remainder of the chain being associated in some other manner with the fibrillar assembly; in other cases almost the whole polypeptide chain seems to be involved. The generic amyloid structure contrasts strongly with the highly individualistic globular structures of most natural proteins. In these latter structures the interactions associated with the very specific packing of the side chains seem to override the main-chain preferences[43,44].

Even though the ability to form amyloid fibrils seems to be generic, the propensity to do so under given circumstances can vary markedly between different sequences. The relative aggregation rates for a wide range of peptides and proteins correlates with the physicochemical features of the molecules such as charge, secondary-structure propensities and hydrophobicity[47]. In a globular protein the polypeptide main chain and the hydrophobic side chains are largely buried within the folded structure. Only when they are exposed, for example when the protein is partly unfolded (for example, at low pH) or fragmented (for example, by proteolysis), will conversion into amyloid fibrils be possible. Experiments *in vitro* indicate that their formation is then generally characterized by a lag phase, followed by a period of rapid growth[48,49]. Such behaviour is typical of nucleated processes such as crystallization; the lag phase can be eliminated by the addition of preformed aggregates to fresh solutions, a process known as seeding. An interesting recent suggestion is that seeding by chemically modified forms of proteins, resulting for example from deamidation or oxidative stress, might in some cases be an important factor in triggering the aggregation process and the onset of disease[50].

There are striking similarities in the aggregation behaviour of different peptides and proteins (Fig. 3)[48,49]. The first phase in amyloid

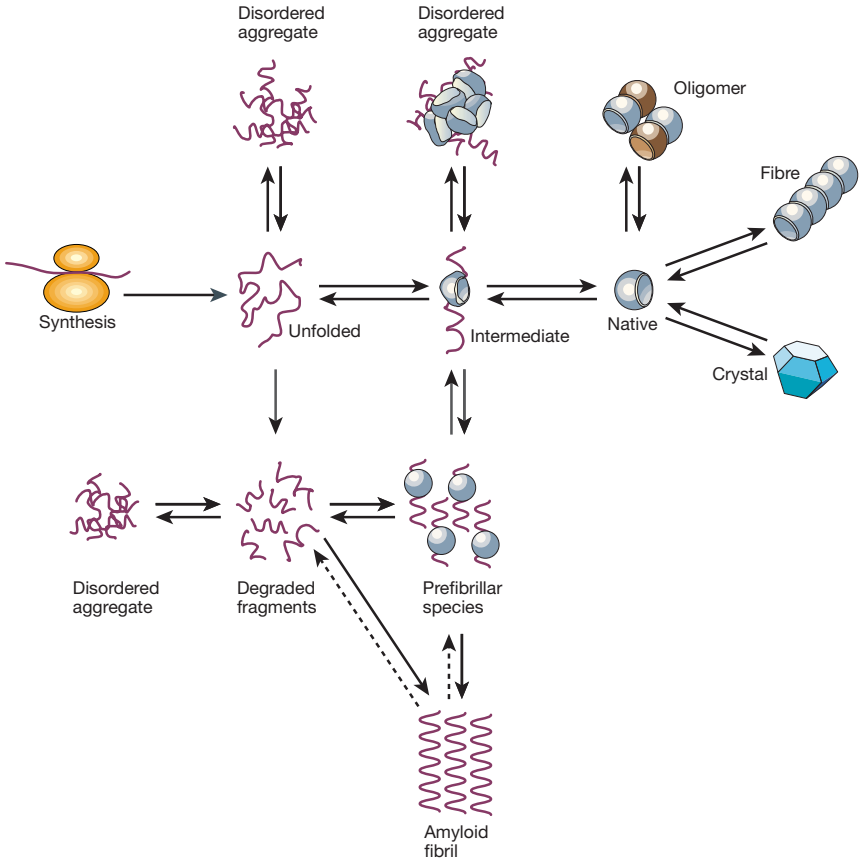


**Figure 4** A unified view of some of the types of structure that can be formed by polypeptide chains. An unstructured chain, for example newly synthesized on a ribosome, can fold to a monomeric native structure, often through one or more partly folded intermediates. It can, however, experience other fates such as degradation or aggregation. An amyloid fibril is just one form of aggregate, but it is unique in having a highly organized 'misfolded' structure, as shown in Fig. 3. Other assemblies, including functional oligomers, macromolecular complexes and natural protein fibres, contain natively folded molecules, as do the protein crystals produced *in vitro* for X-ray diffraction studies of their structures. The populations and interconversions of the various states are determined by their relative thermodynamic and kinetic stabilities under any given conditions. In living systems, however, transitions between the different states are highly regulated by the environment and by the presence of molecular chaperones, proteolytic enzymes and other factors. Failure of such regulatory mechanisms is likely to be a major factor in the onset and development of misfolding diseases. Adapted from ref. 54.

formation seems to involve the formation of soluble oligomers as a result of relatively nonspecific interactions, although, in some cases, specific structural transitions, such as domain swapping, might be important[51]. The earliest species visible by electron or atomic-force microscopy generally resemble small bead-like structures, sometimes linked together, and often described as amorphous aggregates or as micelles. These early 'prefibrillar aggregates' then transform into species with more distinctive morphologies, often called 'protofilaments' or 'protofibrils'. These structures are commonly short, thin, sometimes curly, fibrillar species that are thought to assemble into mature fibrils, perhaps by lateral association accompanied by some degree of structural reorganization. The aggregates that form first are likely to be relatively disorganized structures that expose to the outside world a variety of segments of the protein that are normally buried in the globular state[52]. In some cases, however, these early aggregates appear to adopt quite distinctive structures, including well-defined annular species[53] (see Fig. 3).

## Molecular evolution and the control of protein misfolding

The state of a protein that is adopted under specific conditions depends on the relative thermodynamic stabilities of the various accessible conformations and on the kinetics of their interconversion (Fig. 4)[37,54]. Amyloid fibrils are just one of the types of aggregate that can be formed by proteins, although a significant feature of this particular species is that its highly organized hydrogen-bonded structure is likely to give it unique kinetic stability. Thus, once formed, such aggregates can persist for long periods, allowing a progressive build-up of deposits in tissue, and indeed enabling seeding of the subsequent conversion of additional quantities of the same protein into amyloid fibrils. It is therefore not surprising that biological systems have almost universally avoided the deliberate formation of such material. Nevertheless, there is increasing evidence that the unique properties of amyloid structures have been exploited by some species, including bacteria, fungi and even mammals, for specific (and carefully regulated) purposes[55–57].

There is evidence that evolutionary selection has tended to avoid amino-acid sequences, such as alternating polar and hydrophobic residues, that favour a β-sheet structure of the type seen in amyloid fibrils[58]. Moreover, recent studies suggest that the aggregation process that results in amyloid fibrils is nucleated in a similar manner to that of folding, but that the residues involved might well be located in different regions of the sequence from those that nucleate folding[59]. Such 'kinetic partitioning' means that mutations that occur during evolution could be selected for their ability to enhance folding at the expense of aggregation. However, it is apparent that biological systems have become robust not just by careful manipulation of the sequences of proteins but also by controlling, by means of molecular chaperones and degradation mechanisms, the particular state adopted by a given polypeptide chain at a given time and under given conditions. This process can be thought of as being analogous to the way in which biology regulates and controls the various chemical transformations that take place in the cell by means of enzymes. And just as the aberrant behaviour of enzymes can cause metabolic disease, the aberrant behaviour of the chaperone and other machinery regulating polypeptide conformations can contribute to misfolding and aggregation diseases[35,60].

The ideas encapsulated in Fig. 4 therefore serve as a framework for understanding the fundamental events that underlie misfolding diseases. For example, many of the mutations associated with the familial forms of deposition diseases increase the population of partially

unfolded states, and hence the propensity to aggregate, by decreasing the stability or cooperativity of the native state[37,41,61,62]. Other familial diseases are associated with the accumulation of amyloid deposits whose primary components are fragments of native proteins; such fragments can be produced by aberrant processing or incomplete proteolysis, and are unable to fold into aggregation-resistant states. Other pathogenic mutations enhance the propensities of such species to aggregate, for example by increasing their hydrophobicity or decreasing their charge[47]. And, in the prion disorders such as Kuru or Creutzfeldt–Jakob disease, it seems that ingestion of pre-aggregated states of an identical protein, for example by voluntary or involuntary cannibalism or through the use of contaminated pharmaceuticals or surgical instruments, can markedly increase the inherent rate of aggregation through seeding and hence can generate a mechanism for transmission[48,63].

In some aggregation diseases, the large quantities of insoluble protein involved can physically disrupt specific organs and thereby cause pathological behaviour[40]. But for neurodegenerative disorders, such as Alzheimer's disease, the primary symptoms almost certainly result from a 'toxic gain of function' associated with aggregation[64]. The early prefibrillar aggregates of proteins associated with such diseases are highly damaging to cells; by contrast, the mature fibrils are usually relatively benign[48,65]. Moreover, experiments have recently suggested that similar aggregates of proteins that are not connected with any known diseases could be equally cytotoxic[52]. The generic nature of such aggregates and their effects on cells has recently been supported by the remarkable finding that antibodies can cross-react with early aggregates of different peptides and proteins, and moreover inhibit their toxicity[66]. It is possible that there are specific mechanisms for this toxicity, for example as a result of annular species (Fig. 3) that resemble the toxins produced by bacteria that form pores in membranes and disrupt the ion balance in cells[53]. However, it is likely that the relatively disorganized prefibrillar aggregates are also harmful to cells, probably through a less specific mechanism, for example as a result of the exposure of non-native hydrophobic surfaces stimulating aberrant interactions with cell membranes or other cellular components[67].

## Future directions

In normal circumstances the molecular chaperones and other 'housekeeping' mechanisms are remarkably efficient in ensuring that such potentially toxic species as prefibrillar aggregates are neutralized before they can do any damage[28,68]. This neutralization could result simply from the efficient targeting of misfolded proteins for degradation, but it seems that molecular chaperones are also able to alter the partitioning between harmful and harmless forms of aggregates (Fig. 4)[69]. If the efficiency of these protective mechanisms is impaired, however, the probability of pathogenic behaviour increases[35,68]. Such a process would explain why most of the amyloid diseases are associated with old age, when there is likely to be an increased tendency for proteins to become misfolded or damaged, coupled with a decreased efficiency of the molecular chaperone and unfolded proteins responses[70]. It is ironic that through our success in increasing the life expectancy of the populations of the developed world, we are now seeing the limitations of our proteins and of the regulatory mechanisms that control their behaviour[71]. It is therefore essential that we use our developing understanding of misfolding and aggregation to find effective strategies for combating these increasingly common and highly debilitating diseases[54]. Fortunately, there is now real evidence to suggest that modern science will rise successfully to this tremendous challenge. □

doi:10.1038/nature02261


1. Vendruscolo, M., Zurdo, J., MacPhee, C. E. & Dobson, C. M. Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems. *Phil. Trans. R. Soc. Lond.* **361,** 1205–1222 (2003).
2. Radford, S. E. & Dobson, C. M. From computer simulations to human disease: emerging themes in protein folding. *Cell* **97,** 291–298 (1999).
3. Dobson, C. M., Sali, A. & Karplus, M. Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Ed. Eng.* **37,** 868–893 (1998).
4. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. Navigating the folding routes. *Science* **267,** 1619–1620 (1995).
5. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4,** 10–19 (1997).
6. Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M. & Karplus, M. Understanding protein folding via free energy surfaces from theory and experiment. *Trends Biochem. Sci.* **25,** 331–339 (2000).
7. Baldwin, R. L. Protein folding: matching speed and stability. *Nature* **369,** 183–184 (1994).
8. Eaton, W. A., Munoz, V., Thompson, P. A., Henry, E. R. & Hofrichter, J. Kinetics and dynamics of loops, α-helices, β-hairpins, and fast-folding proteins. *Acc. Chem. Res.* **31,** 745–753 (1998).
9. Snow, C. D., Nguyen, H., Pande, V. S. & Gruebele, M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **420,** 102–106 (2002).
10. Yang, W. Y. & Gruebele, M. Folding at the speed limit. *Nature* **423,** 193–197 (2003).
11. Mayor, U. *et al.* The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **421,** 863–867 (2003).
12. Plaxco, K. W., Simons, K. T. & Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277,** 985–994 (1998).
13. Schuler, B., Lipman, E. A. & Eaton, W. A. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419,** 743–747 (2002).
14. Fersht, A. R. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W.H. Freeman, New York, 1999).
15. Fersht, A. R. Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. *Proc. Natl Acad. Sci. USA* **97,** 1525–1529 (2000).
16. Shea, J. E. & Brooks, C. L. From folding surfaces to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* **52,** 499–535 (2001).
17. Fersht, A. R. & Daggett, V. Protein folding and unfolding at atomic resolution. *Cell* **108,** 573–582 (2002).
18. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. Three key residues form a critical contact network in a transition state for protein folding. *Nature* **409,** 641–646 (2001).
19. Makarov, D. E. & Plaxco, K. W. The topomer search model: a simple, quantitative theory of two-state protein folding kinetics. *Protein Sci.* **12,** 17–26 (2003).
20. Baker, D. A surprising simplicity to protein folding. *Nature* **405,** 39–42 (2000).
21. Roder, H. & Colon, W. Kinetic role of early intermediates in protein folding. *Curr. Opin. Struct. Biol.* **7,** 15–28 (1997).
22. Sanchez, I. E. & Kiefhaber, T. Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325,** 367–376 (2003).
23. Khan, F., Chuang, J. I., Gianni, S. & Fersht, A. R. The kinetic pathway of folding of barnase. *J. Mol. Biol.* **333,** 169–186 (2003).
24. Vendruscolo, M., Paci, E., Karplus, M. & Dobson, C. M. Structures and relative free energies of partially folded states of proteins. *Proc. Natl Acad. Sci. USA* **100,** 14817–14821 (2003).
25. Cheung, M. S., Garcia, A. E. & Onuchic, J. N. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl Acad. Sci. USA* **99,** 685–690 (2002).
26. Hardesty, B. & Kramer, G. Folding of a nascent peptide on the ribosome. *Prog. Nucleic Acid Res. Mol. Biol.* **66,** 41–66 (2001).
27. Bukau, B. & Horwich, A. L. The Hsp70 and Hsp60 chaperone machines. *Cell* **92,** 351 –366 (1998).
28. Hartl, F. U. & Hayer-Hartl, M. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295,** 1852–1858 (2002).
29. Ellis, R. J. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr. Opin. Struct. Biol.* **11,** 114–119 (2001).
30. Schiene, C. & Fischer, G. Enzymes that catalyse the restructuring of proteins. *Curr. Opin. Struct. Biol.* **10,** 40–45 (2000).
31. Hammon, C. & Helenius, A. Quality control in the secretory pathway. *Curr. Opin. Cell. Biol.* **7,** 523–529 (1995).
32. Kaufman, R. J. *et al.* The unfolded protein response in nutrient sensing and differentiation. *Nature Rev. Mol. Cell Biol.* **3,** 411–421 (2002).
33. Wilson, M. R. & Easterbrook Smith, S. B. Clusterin is a secreted mammalian chaperone. *Trends Biochem. Sci.* **25,** 95–98 (2000).
34. Schubert, U. *et al.* Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* **404,** 770–774 (2000).
35. Bence, N. F., Sampat, R. M. & Kopito, R. R. Impairment of the ubiquitin–proteasome system by protein aggregation. *Science* **292,** 1552–1555 (2001).
36. Thomas, P. J., Qu, B. H. & Pedersen, P. L. Defective protein folding as a basis of human disease. *Trends Biochem. Sci.* **20,** 456–459 (1995).
37. Dobson, C. M. The structural basis of protein folding and its links with human disease. *Phil. Trans. R. Soc. Lond. B* **356,** 133–145 (2001).
38. Horwich, A. Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions. *J. Clin. Invest.* **110,** 1221–1232 (2002).
39. Bullock, A. N. & Fersht, A. R. Rescuing the functions of mutant p53. *Nature Rev. Cancer* **1,** 68–76 (2001).
40. Tan, S. Y. & Pepys, M. B. Amyloidosis. *Histopathology* **25,** 403–414 (1994).
41. Kelly, J. W. Alternative conformation of amyloidogenic proteins and their multi-step assembly pathways. *Curr. Opin. Struct. Biol.* **8,** 101–106 (1998).
42. Sunde, M. & Blake, C. C. F. The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv. Protein Chem.* **50,** 123–159 (1997).
43. Dobson, C. M. Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24,** 329–332 (1999).
44. Fändrich, M. & Dobson, C. M. The behaviour of polyamino acids reveals an inverse side-chain effect in amyloid structure formation. *EMBO J.* **21,** 5682–5690 (2002).
45. Jiménez, J. L. *et al.* Cryo-electron microscopy of an SH3 amyloid fibril and model of the molecular packing. *EMBO J.* **18,** 815–821 (1999).
46. Petkova, A. T. *et al.* A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl Acad. Sci. USA* **99,** 16742–16747 (2002).
47. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. Rationalization of mutational effects on protein aggregation rates. *Nature* **424,** 805–808 (2003).
48. Caughey, B. & Lansbury, P. T. Jr. Protofibrils, pores, fibrils, and neurodegeneration: separating the

responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci.* **26,** 267–298 (2003).
49. Bitan, G. *et al.* Amyloid β-protein (Aβ) assembly: Aβ40 and Aβ42 oligomerize through distinct pathways. *Proc. Natl Acad. Sci. USA* **100,** 330–335 (2003).
50. Nilsson, M. R., Driscoll, M. & Raleigh, D. P. Low levels of asparagine deamidation can have a dramatic effect on aggregation of amyloidogenic peptides: implications for the study of amyloid formation. *Protein Sci.* **11,** 342–349 (2002).
51. Schlunegger, M. P., Bennett, M. J. & Eisenberg, D. Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Adv. Protein Chem.* **50,** 61–122 (1997).
52. Bucciantini, M. *et al.* Inherent cytotoxicity of aggregates implies a common origin for protein misfolding diseases. *Nature* **416,** 507–511 (2002).
53. Lashuel, H. A., Hartley, D., Petre, B. M., Walz, T. & Lansbury, P. T. Jr. Neurodegenerative disease: amyloid pores from pathogenic mutations. *Nature* **418,** 291 (2002).
54. Dobson, C. M. Protein folding and disease: a view from the First Horizon Symposium. *Nature Rev. Drug Discov.* **2,** 154–160 (2003).
55. True, H. L. & Lindquist, S. L. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407,** 477–483 (2000).
56. Chapman, M. R. *et al.* Role of *Escherichia coli* curli operons in directing amyloid fiber formation. *Science* **295,** 851–855 (2002).
57. Kelly, J. W. & Balch, W. E. Amyloid as a natural product. *J. Cell Biol.* **161,** 461–462 (2003).
58. Broome, B. M. & Hecht, M. H. Nature disfavours sequences of alternating polar and non-polar amino acids: implications for amyloidogenesis. *J. Mol. Biol.* **296,** 961–968 (2000).
59. Chiti, F. *et al.* Kinetic partitioning of protein folding and aggregation. *Nature Struct. Biol.* **9,** 137–143 (2002).
60. Macario, A. J. L. & Macario, E. C. Sick chaperones and ageing: a perspective. *Ageing Res. Rev.* **1,** 295–311 (2002).
61. Ramirez-Alvarado, M., Merkel, J. S. & Regan, L. A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro. *Proc. Natl Acad. Sci. USA* **97,** 8979–8984 (2000).
62. Dumoulin, M. *et al.* A camelid antibody fragment inhibits amyloid fibril formation by human lysozyme. *Nature* **424,** 783–788 (2003).
63. Prusiner, S. B. Prion diseases and the BSE crisis. *Science* **278,** 245–251 (1997).
64. Taylor, J. P., Hardy, J. & Fischbeck, K. H. Toxic proteins in neurodegenerative disease. *Science* **296,** 1991–1995 (2002).
65. Walsh, D. M. *et al.* Naturally secreted oligomers of amyloid β protein potently inhibit hippocampal long-term potentiation *in vivo*. *Nature* **416,** 535–539 (2002).
66. Kayed, R. *et al.* Common structure of soluble amyloid oligomers implies common mechanisms of pathogenesis. *Science* **300,** 486–489 (2003).
67. Stefani, M. & Dobson, C. M. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J. Mol. Med.* **81,** 678–699 (2003).
68. Sherman, M. Y. & Goldberg, A. L. Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases. *Neuron* **29,** 15–32 (2001).
69. Muchowski, P. J. *et al.* Hsp70 and Hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. *Proc. Natl Acad. Sci. USA* **97,** 7841–7846 (2000).
70. Csermely, P. Chaperone overload is a possible contributor to 'civilization diseases'. *Trends Genet.* **17,** 701–704 (2001).
71. Dobson, C. M. Getting out of shape—protein misfolding diseases. *Nature* **418,** 729–730 (2002).

**Acknowledgements** I should like to thank in particular the Wellcome Trust, the Leverhulme Trust and the UK Research Councils for generous support over many years, without whom my own research activities in this area of science could not have been carried out.

# The Amyloid Phenomenon and Its Significance in Biology and Medicine

**Christopher M. Dobson, Tuomas P.J. Knowles, and Michele Vendruscolo**

Centre for Misfolding Diseases, Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, United Kingdom

*Correspondence:* cmd44@cam.ac.uk; tpjk2@cam.ac.uk; mv245@cam.ac.uk

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

The misfolding of proteins is now recognized to be the origin of a large number of medical disorders. One particularly important group of such disorders is associated with the aggregation of misfolded proteins into amyloid structures, and includes conditions ranging from Alzheimer's and Parkinson's diseases to type II diabetes. Such conditions already affect over 500 million people in the world, a number that is rising rapidly, and at present these disorders cannot be effectively treated or prevented. This review provides an overview of this field of science and discusses recent progress in understanding the nature and properties of the amyloid state, the kinetics and mechanism governing its formation, the origins of its links with disease, and the manner in which its formation may be inhibited or suppressed. This latter topic is of particular importance, both to enhance our knowledge of the maintenance of protein homeostasis in living organisms and also to address the development of therapeutic strategies through which to combat the loss of homeostasis and the associated onset and progression of disease.

Natural proteins are a highly select group of molecules, and their properties have a number of very special characteristics when compared with random sequences of amino acids, including an ability to fold into unique and often highly intricate structures that can remain functional within the complex milieu of living systems (Dobson 2003; Robinson et al. 2007). Such characteristics have enabled biological systems to develop a vast range of functions and with an astonishing degree of specificity. Because proteins are involved in virtually every chemical process taking place within the cellular environment; however, the failure of these molecules to fold correctly or to remain within their functional states can give rise to serious cellular malfunctions (Dobson 2003; Eisenberg and Jucker 2012; Knowles et al. 2014; Chiti and Dobson 2017). Protein misfolding was until relatively recently considered to be a rather esoteric topic except in the context of problems associated with protein production for biotechnological purposes, and of its links with a small number of rather uncommon diseases. This situation has, however, now changed dramatically as misfolding diseases are rapidly becoming the most prevalent and disruptive in terms of health and medical care in the modern world (Chiti and Dobson 2017; World Alzheimer Report 2018).

Editors: Richard I. Morimoto, F. Ulrich Hartl, and Jeffery W. Kelly
Additional Perspectives on Protein Homeostasis available at www.cshperspectives.org

 doi: 10.1101/cshperspect.a033878
Cite this article as *Cold Spring Harb Perspect Biol* 2020;12:a033878

Of particular importance are those disorders in which misfolding results in the conversion through aggregation of normally soluble proteins into intractable and highly stable amyloid structures (Table 1; Eisenberg and Jucker 2012; Dobson 2013; Knowles et al. 2014; Chiti and Dobson 2017). The increase in the significance of this group of diseases results primarily because many of these conditions are strongly associated with aging—notably Alzheimer's disease—and with lifestyle and dietary changes—notably type II diabetes—and because they are currently incurable and largely untreatable. In addition, however, it is now recognized that the study of the process of amyloid formation has profound significance for understanding normal as well as aberrant biological phenomena, and indeed for the generation of important new biomaterials with novel and interesting properties (Knowles and Buehler 2011; Shimanovich et al. 2015; Wei et al. 2017). In this review, we will discuss our increasing knowledge of the amyloid state of proteins, with particular emphasis on the kinetics of its formation, in terms of the mechanisms by which aggregation occurs and can be controlled, and of the significance of such knowledge in the fields of biological and medical sciences.

## THE AMYLOID STATE OF PROTEINS

The amyloid state of a protein typically has the form of threadlike fibrils a few nanometers in diameter and frequently microns in length (Fig. 1; Sunde et al. 1997; Jiménez et al. 2002; Nelson et al. 2005; Eisenberg and Jucker 2012). Regardless of the amino acid sequence, or indeed of the structure of the protein in its native state, such fibrils are rich in β-sheet patterns and are closely packed and highly ordered. Although amyloid structures were originally observed in the context of disease, experiments with a wide range of peptides and proteins in laboratory environments have led to the realization that the ability to form such aggregated states is not a rare phenomenon associated with a small number of disease-related proteins; instead, the amyloid state appears as an alternative well-defined structural form that can be adopted by many, if not all, polypeptide sequences under appropriate conditions, including those of proteins whose native states are globular or intrinsically disordered, and that are cytosolic or membrane bound (Dobson 2003).

Unlike the native states of proteins, however, amyloid fibrils possess a core region having a common or "generic" cross-β architecture (Sunde et al. 1997; Dobson 2003; Eisenberg and Jucker 2012) that is dictated by the properties of the polypeptide backbone, rather than encoded by its sequence, although their propensity to form, and their detailed structures and properties, varies significantly with the amino acid composition and sequence (Fändrich and Dobson 2002; Pawar et al. 2005). Moreover, it is increasingly clear that the amyloid state, particularly for small proteins or proteolytic fragments of larger proteins, with <150 residues, can be thermodynamically more stable than the functional native states even under normal physiological conditions (Gazit 2002; Baldwin et al. 2011). The familiar functional forms of proteins, therefore, do not always represent the global minima on the free energy surfaces defined by their specific polypeptide sequences, but rather can be metastable states that are separated from the amyloid form by high kinetic barriers.

These observations show that biological systems have evolved to enable their functional proteins to remain in solution, or sequestered in membranes, often for prolonged lengths of time under physiological conditions, rather than aggregating and ultimately converting into the amyloid state, except in the relatively small number of cases in which this latter form of protein structure is used functionally, for purposes ranging from structural templates to mechanisms of storage (Fowler et al. 2007; Maji et al. 2009). Some of the kinetic barriers that maintain proteins in solution in their native states are encoded in their sequences, notably through folding into highly cooperative structures with aggregation-prone regions buried in their internal regions (Broome and Hecht 2000; Dobson 2003) and the avoidance of patterns of residues that would favor the amyloid state, for example by stabilizing β-sheet structures (Broome and Hecht 2000; Tartaglia et al. 2008). There is evidence as well that the large size of many protein

**Table 1.** Selection of human diseases associated with protein misfolding and the formation of extracellular amyloid deposits or intracellular inclusions with amyloid-like characteristics

| Disease | Aggregating protein or peptide | Length of protein or peptide[a] | Structure of protein or peptide[b] |
|---|---|---|---|
| **Neurodegenerative diseases** | | | |
| Alzheimer's disease[c] | Amyloid-β peptide | 37–48[f] | Intrinsically disordered |
| | Tau protein | 352–441 | Intrinsically disordered |
| Spongiform encephalopathies[c,e] | Prion protein or fragments | 230 | Intrinsically disordered and α-helical |
| Parkinson's disease[c] | α-Synuclein | 140 | Intrinsically disordered |
| Amyotrophic lateral sclerosis[c] | Superoxide dismutase 1 | 154 | β-sheet and Ig-like |
| Huntington's disease[d] | Fragments of huntingtin 1 | 103–187[g] | Largely intrinsically disordered |
| Familial amyloid polyneuropathy[d] | Mutants of transthyretin | 127 | β-Sheet |
| **Nonneuropathic systemic amyloidoses** | | | |
| AL amyloidosis[c] | Immunoglobulin light chains or fragments thereof | 100[f] | β-Sheet Ig-like |
| AA amyloidosis[c] | Fragments of serum amyloid A protein | 45–104[f] | α-Helical and unknown fold |
| Senile systemic amyloidosis[c] | Wild-type transthyretin | 127 | β-Sheet |
| Hemodialysis-related amyloidosis[c] | β2-Microglobulin | 99 | β-Sheet and Ig-like |
| Lysozyme amyloidosis[d] | Mutants of lysozyme | 130 | α-Helical and β-sheet |
| **Nonneuropathic localized amyloidoses** | | | |
| Apo A-I amyloidosis[d] | Fragments of apolipoprotein A-I | 69–100[f] | Intrinsically disordered |
| Type II diabetes[c] | Amylin (IAPP) | 37 | Intrinsically disordered |
| Medullary carcinoma of the thyroid[c] | Calcitonin | 32 | Intrinsically disordered |
| Injection-localized amyloidosis[c] | Insulin | 21 + 30[h] | α-Helical |

From Dobson 2013; adapted, with permission. A more comprehensive list is given in Chiti and Dobson 2017.

Ig, Immunoglobulin; AL, amyloid light chain; AA, serum amyloid A protein; Apo A-I, apolipoprotein A-I; IAPP, islet amyloid polypeptide.

[a]Data do not refer to the number of amino acid residues of the precursor proteins, but to the lengths of the processed polypeptide chains that are present in the aggregates that are deposited in the disease states.

[b]This column reports the structural class and general native fold; both refer to the processed peptides or proteins that deposit into aggregates before aggregation and not to the precursor proteins.

[c]Predominantly sporadic, although in some of these diseases hereditary forms associated with specific mutations are well documented.

[d]Predominantly hereditary, although in some of these diseases sporadic cases are well documented.

[e]5% of cases are infectious (iatrogenic).

[f]Fragments of various lengths are generated and reported in ex vivo fibrils.

[g]Lengths refer to the normal sequences with nonpathogenic traits of polyglutamine (poly[Q]).

[h]Human insulin consists of two chains (A and B with 21 and 30 residues, respectively) covalently bonded by disulfide bridges.

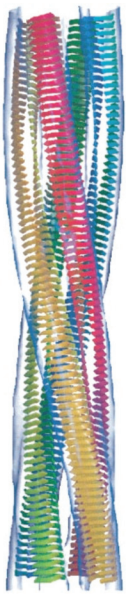

Figure 1. Model of one of the polymorphs of the amyloid fibrils formed from insulin as defined from cryogenic electron microscopy (cryo-EM) analysis. This particular fibril contains four protofilaments that twist around each other to form the mature fibril. Each of the protofilaments has a pair of nearly flat β-sheets, with the component strands oriented perpendicularly to the main fibril axis. (From Jiménez et al. 2002; reprinted, with permission, from the National Academy of Sciences © 2002.)

molecules is likely to disfavor their assembly into polymeric amyloid structures relative to folded native states (Baldwin et al. 2011).

More generally, proteins and indeed other biomolecules have coevolved with their biological environments, and factors ranging from localization in cellular compartments to the regulation of synthesis and degradation, and the existence of molecular chaperones, act to limit the probability of the formation and accumulation of misfolded and aggregated peptides and proteins (Morimoto 2008; Hartl et al. 2011; Vendruscolo et al. 2011). Indeed, proteins can adopt a multiplicity of states following their synthesis (Fig. 2) and such mechanisms ensure in a properly functional organism that the right state of a given protein is present at the right time and in the right place, much as enzymes control the nature, location, and populations of small molecules, for example in the regulation of metabolic processes (Knowles et al. 2014). It is increasingly clear, however, that as proteins begin to aggregate such protective mechanisms can become inadequate and ultimately result in a widespread loss of protein homeostasis (Balch et al. 2008; Morimoto 2008; Hartl et al. 2011; Vendruscolo et al. 2011; Labbadia and Morimoto 2015).

## THE KINETICS AND MECHANISM OF AMYLOID FORMATION

In the light of the importance of the rates at which specific proteins aggregate to form the amyloid state, a key step in understanding its occurrence in given environments is to define the mechanisms by which it is generated. Amyloid fibrils are typically composed of a number of protofilaments that twist around each other, the core of each adopting a cross-β structure in which β-strands are oriented perpendicularly to the fibril axis to create hydrogen-bonded β-sheets running along its length; such structural features make the structures extremely stable, both thermodynamically and kinetically, and also resistant to degradation by proteolytic enzymes or other mechanisms (Knowles et al. 2007, 2014).

Methodological developments in X-ray crystallography (Wei et al. 2017), solid-state nuclear magnetic resonance (NMR) spectroscopy (Petkova et al. 2002; Colvin et al. 2015), and cryogenic electron microscopy (cryo-EM) techniques (Sunde et al. 1997; Sachse et al. 2008; Fitzpatrick et al. 2017), have resulted in a steady increase in our detailed knowledge of the molecular structures of amyloid fibrils, including those of fibrillar deposits extracted from diseased human brain tissue (Fitzpatrick et al. 2017). The structures formed from different proteins, and under different conditions, differ both in the manner in which the various side chains are incorporated into the overall architecture of the fibrillar state and in the way specific regions of the polypeptide chain are brought together to form the cross-β cores that are characteristic of the amyloid state. Such differences, in particular in the case of prions, have been related to the various "strains" of disease, and it is likely that

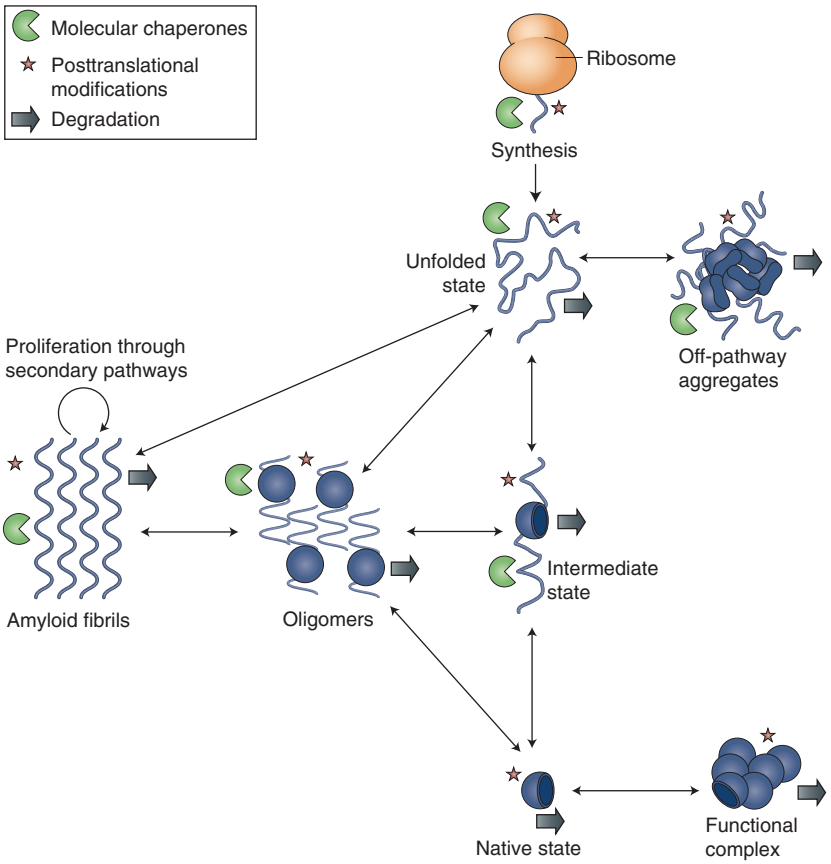


**Figure 2.** The different states of a protein and the process of their interconversion. Schematic description of a selection of the possible states that can be formed by proteins under different circumstances. The populations and their rates of interconversion are determined by their different thermodynamic stabilities and free energy barriers associated with the various transitions as well as by the rates of synthesis and degradation, the propensity to interact with molecular chaperones, and to undergo posttranslational and other chemical modifications such as proteolytic cleavage. (From Knowles et al. 2014; reprinted, with permission, from Springer Nature © 2014.)

other amyloid structures also have structures characteristic of specific manifestations of disease within individual patients (Safar et al. 1998; Petkova et al. 2005).

A wide variety of studies has shown that the process of conversion from a soluble, typically monomeric, state of a protein into the polymeric fibrillar state involves a series of well-defined molecular steps and a variety of intermediate species. Moreover, it is increasingly clear that these elementary steps operate together to give rise to a complex phenomenology and a range of associative, dissociative, and rearrangement events taking place during the course of the process that culminates in the formation of mature fibrils (Michaels et al. 2018). Our understanding of the nature of these intermediate species and the kinetics and mechanisms of their formation and interconversion has increased dramatically in recent years as a result of careful experimental studies coupled to the application of mathematical methods of analysis (Knowles et al. 2009; Cohen et al. 2013; Meisl et al. 2016; Michaels et al. 2018). These advances have allowed the framework of microscopic rate laws and chemical kinetics to be used quantitatively in the area of protein aggregation studies and to relate macroscopic measurements of aggregation behavior

to the fundamental molecular level processes and their rates. In particular, these methods have revealed that the macroscopic observations of the time dependence of fibril formation, measured, for example, by the intensity of the fluorescence of the amyloid-specific dye thioflavin T, are the result of a network of distinct microscopic steps, including primary and secondary nucleation and other processes such as fragmentation and elongation of the filamentous structures (Fig. 3; Knowles et al. 2009; Cohen et al. 2013; Meisl et al. 2016; Michaels et al. 2018).

These different microscopic processes depend in characteristic ways on variables such as concentration and the presence of specific quantities of preformed fibrils (seeding), and systematic variations in these conditions can enable the network of contributing microscopic processes to be analyzed to define their rate constants and relative significance (Knowles et al. 2009; Cohen et al. 2013; Meisl et al. 2016; Michaels et al. 2018). One of the most interesting results from such kinetic analysis has been the finding of the importance of secondary nucleation in the mechanism of the aggregation of the 40 and 42 residue Aβ-peptides that are linked to Alzheimer's disease (Cohen et al. 2013; Meisl et al. 2016; Törnquist et al. 2018). Secondary nucleation is a process generating new fibrils from monomeric precursor proteins that is catalyzed on the surfaces of aggregated species, a process that generates positive feedback as the number of fibrils in turn increases the available catalytic surface and hence results in their rapid proliferation (Cohen et al. 2013; Michaels et al. 2018; Törnquist et al. 2018). Other biological surfaces in addition to those of amyloid fibrils can also catalyze the nucleation step in amyloid formation. Indeed, studies of other systems, show that differences exist in the relative importance of specific steps; in the case of α-synuclein, whose aggregation is linked to Parkinson's disease, primary nucleation has been found to be a key process that is enhanced very significantly by the surfaces of the lipid bilayers that are the major components of cell membranes (Galvagnion et al. 2015). In other systems, notably both yeast and mammalian prions, fibril multiplication through fragmentation processes appears to be of particular significance in the proliferation of aggregates (Aguzzi and Calella 2009; Knowles et al. 2009).

Further information about the nature and properties of the species formed during the aggregation process has come from the concerted application of a series of experimental techniques, including the measurement of their sizes and structural characteristics by single-molecule optical methods (Cremades et al. 2012), and

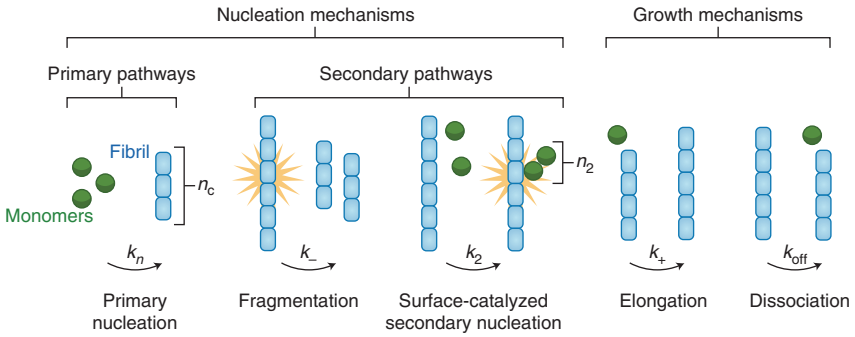


**Figure 3.** The microscopic steps that can contribute to the macroscopic conversion of soluble proteins into amyloid fibrils. The mechanisms of these microscopic steps can be divided into nucleation (i.e., fibril-forming) and growth processes. Events producing new fibrils are further classified as primary or secondary processes on the basis of their dependence (secondary) or lack of dependence (primary) on the population of aggregates. Here, $k_n$, $k_-$, $k_+$, and $k_{off}$ represent rate constants, and $n_c$ and $n_2$ represent the reaction orders of primary and secondary (surface-catalyzed) nucleation. (From Michaels et al. 2018; reprinted, with permission, from *Annual Reviews* © 2018.)

their diffusion properties by microfluidics (Knowles et al. 2011). The latter properties are of particular interest not just in understanding the fundamental mechanism of aggregation but also the ability of the component species to proliferate spatially and to promote further aggregation (Herling et al. 2018). Although such measurements have primarily been made under laboratory conditions, it is possible to relate them to the events occurring in living systems (Luheshi et al. 2008; Kundel et al. 2018), and some biophysical techniques, for example those involving fluorescence measurements, can be used to study the aggregation process directly within living systems and hence to compare the events occurring in different environments (Kaminski et al. 2011).

## THE BIOLOGICAL CONSEQUENCES OF AMYLOID FORMATION

The presence of amyloid deposits in a range of diseases has been known for many years and the link between this phenomenon and the various pathologies has been the subject of much debate (Caughey and Lansbury 2003; Haass and Selkoe 2007; Jack et al. 2018). For systemic amyloidosis, the accumulation of large quantities of fibrillar deposits in vital organs is likely to result in their inability to function normally. In the case of neurodegenerative diseases, however, the amyloid burden can be quite low, and does not always relate to the severity of the symptoms (Lue et al. 1999). There is now a great deal of evidence that the most highly pathogenic species associated with amyloid formation are not the fibrillar deposits themselves but the prefibrillar oligomeric species that are populated during the aggregation process (Lue et al. 1999; Bucciantini et al. 2002; Kayed et al. 2003; Tomic et al. 2009; Benilova et al. 2012). Indeed, cellular toxicity has been observed to arise from oligomeric species formed by the aggregation of proteins unconnected to any pathological condition as well as those known to be associated with specific misfolding diseases (Bucciantini et al. 2002; Chiti and Dobson 2017).

The fundamental origin of such toxicity is attributable to the fact that these oligomers are inherently misfolded and, therefore, are likely to interact inappropriately with many of the functional components of the highly complex and crowded environments with which the latter have coevolved; the high surface-to-volume ratios of these species relative to larger aggregates, and their greater ability to spread by diffusion or other means, are also likely to be important factors in promoting high levels of cellular damage (Stefani and Dobson 2003; Guo and Lee 2014; Herling et al. 2018). The evidence from mechanistic studies that a range of species is present in a dynamic network throughout the process of protein aggregation strongly suggests that there is no unique "toxic agent" but that many misfolded species have the potential to generate at least some degree of toxicity. Indeed, it is clear that oligomeric intermediates can differ very significantly in their properties and their ability to induce given types of cellular dysfunction (Campioni et al. 2010; Cremades et al. 2012).

Although much remains to be learned about their nature and properties, it seems probable that the oligomers formed initially during the aggregation process of cytosolic proteins, at least, are relatively disordered species whose characteristics are likely to be strongly influenced by the sequestration from the aqueous environment of hydrophobic residues. In some cases, at least, it then appears that one or more conformational changes occurs to generate less disordered species with a significant content of β-sheet structure and a more highly hydrophobic surface that has greater potential to interact with cellular components such as membranes (Campioni et al. 2010; Cremades et al. 2012; Fusco et al. 2017). As such species increase in size, and ultimately form mature fibrils and plaques, many such hydrophobic regions are likely to become buried, reducing their potential to generate pathological effects; in addition, the increasing content of highly ordered structure progressively reduces their susceptibility to proteolysis and other degradation mechanisms, resulting in their increasing ability to accumulate in organs and tissue.

The ability of oligomeric species to disrupt cellular membranes is increasingly evident from experiments with lipid bilayers, where aggregates known to be cytotoxic have been shown

to be able to insert into the bilayers (Fig. 4) and allow the ingress of $Ca^{2+}$ ions, events that are then coupled to increases in reactive oxidative species (ROS) and ultimately cell death (Emak and Davies 2002; Stefani and Dobson 2003; Fusco et al. 2017). There is also evidence that the differences in the toxicity of oligomeric species can be related to specific structural characteristics; studies of stabilized oligomers of α-synuclein have, for example, revealed the importance of the exposure of the amino-terminal region of the protein in such species for insertion into lipid bilayers (Fusco et al. 2017). It is also clear, however, that interactions of aggregates with specific receptors are also associated with the induction of cellular damage, and there is increasing recognition of the role of the inflammatory response in neurodegenerative diseases (Amor et al. 2010; Benilova et al. 2012).

The observation that many proteins can be thermodynamically unstable with respect to aggregation even under physiological conditions has been discussed above. Additional studies have provided yet further insight into this issue, as they have revealed that proteins appear generally to be present at concentrations that are at, or are close to, their solubility limits within the crowded cellular environments with which they have coevolved (Tartaglia et al. 2007). Such a situation reflects the balance between the need for proteins to remain soluble to perform their normal functions and the fact that random mutations tend to reduce protein solubility; the result is that even small changes, induced for example by mutations, posttranslational modifications or changes of concentration, can generate an increased risk of aggregation. Moreover, it has also emerged that some proteins are "supersaturated," in that they are present at levels in which they are highly metastable. It is particularly interesting that many of these proteins are found in pathways that are associated with neurodegenerative conditions, making them particularly vulnerable to aberrant behavior (Ciryam et al. 2015).

## AMYLOID FORMATION IN THE CONTEXT OF PROTEIN HOMEOSTASIS

We now know that the inherent risks of protein aggregation, and of their likely deleterious consequences, are countered by the existence of an array of mechanisms with the ability of maintaining protein homeostasis (Fig. 1). In the context of amyloid formation, as well as the overall regulation of protein concentrations, molecular chaperones play key roles in targeting misfolded and aggregation-prone species either to induce correct refolding or to target them for degradation. Indeed, pathogenic behavior occurs when the quantities of misfolded and aggregation-prone species reach levels in which the range of

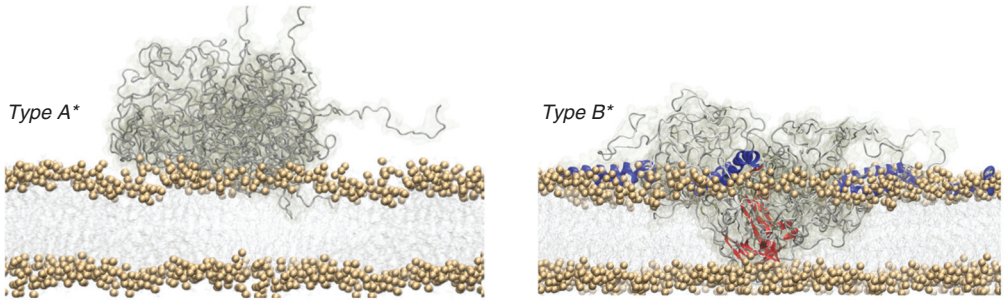


**Figure 4.** The interaction of different types of stabilized oligomers of α-synuclein with a lipid bilayer. Schematic representations of the binding of type A* (largely disordered) (*left*) and type B* (having both ordered [red] and disordered [gray] regions) (*right*) oligomers to a lipid bilayer. The amino-terminal regions of the type B* oligomers fold into amphipathic α-helices (blue) upon interaction with the bilayer, and the ordered regions, which are rich in β-sheet structure insert into the bilayer and disrupt its integrity. (From Fusco et al. 2017; reprinted, with permission, from the American Association for the Advancement of Science © 2017.)

regulatory mechanisms is unable to maintain protein solubility and homeostasis (Morimoto 2008; Douglas and Dillin 2010; Labbadia and Morimoto 2015). In the light of the key role that oligomeric aggregates appear to play in neurodegenerative conditions, much effort has been expended in exploring the way that the presence of such species is minimized, particularly under conditions of aging when the challenges in maintaining homeostasis increase (Morimoto 2008; Douglas and Dillin 2010; Labbadia and Morimoto 2015). Under such conditions, the initiation of the aggregation of even one particularly vulnerable protein may lead to a range of downstream processes involving many other proteins that then contribute substantially to the onset of disease.

Molecular chaperones have been shown in multiple studies to inhibit protein aggregation in vitro, and a wide range of biophysical studies have directly shown the interaction of molecular chaperones with oligomeric species. As an example, single-molecule fluorescence techniques have revealed the interaction of clusterin with the ensemble of oligomers observed during the aggregation of the Aβ-peptide (Narayan et al. 2011). Molecular chaperones have also been shown to reduce the toxicity associated with aggregation reactions both in cellular systems and in model organisms (Morimoto 2008; Hartl et al. 2011). In addition to their effects on the aggregation reactions themselves, in studies with stabilized and purified oligomeric species, the reduction of toxicity has been attributable to a variety of effects associated with the interactions of molecular chaperones with such species, including the inhibition of the binding of oligomers to cell membranes and the promotion of the further aggregation of the oligomers into larger and less damaging species (Mannini et al. 2012; Fusco et al. 2017).

Of particular interest are recent conclusions that have arisen from the extension of the methods of kinetic analysis discussed above in the context of the microscopic steps in the overall aggregation reaction. These studies have revealed that molecular chaperones have the ability to inhibit with remarkable efficacy very specific steps in the mechanisms that result in amyloid formation. One important example of this behavior has been the demonstration that the chaperone Brichos can inhibit extremely selectively the secondary nucleation processes associated with the aggregation of the Aβ-peptide (Fig. 5; Cohen et al. 2015). This inhibition has been shown to be associated with the binding of the molecular chaperone to the surfaces of the fibrils, thereby suppressing their ability to catalyze secondary nucleation processes. Such suppression is predicted from the kinetic analysis to reduce very significantly the population of oligomers, and indeed experiments confirm that the aggregation of the Aβ-peptide generates a much lower level of toxicity in the presence of Brichos (Cohen et al. 2015).

Additional studies reveal that other molecular chaperones are able to inhibit different steps in the aggregation process of the Aβ-peptides, and indeed of other proteins (Arosio et al. 2016). These observations suggest that the abundance of different molecular chaperones present within living organisms, and that are up-regulated in response to cellular stress, act in concert to inhibit the multitude of different aggregation pathways within the dynamic networks of intermediate species that are associated with the conversion of soluble proteins into amyloid fibrils.

More generally, increasing evidence indicates that certain cell and tissue types are more vulnerable than others to the effects of protein misfolding because of specific differences in the wide range of components of the protein homeostasis system responsible for the regulation of protein aggregation, including in particular trafficking and degradation pathways. Such cells and tissues are characterized by different transcriptional and proteomic signatures associated with the maintenance of aggregation-prone proteins, which are typically present for functional requirements, such as synaptic transmission and oxidative phosphorylation (Ciryam et al. 2016, 2017; Freer et al. 2016; Kundra et al. 2017; Fu et al. 2019). Further studies in this direction will increase our system-level understanding of misfolding disorders, and may provide new targets for therapeutic interventions as well as biomarkers for developing innovative diagnostic approaches.

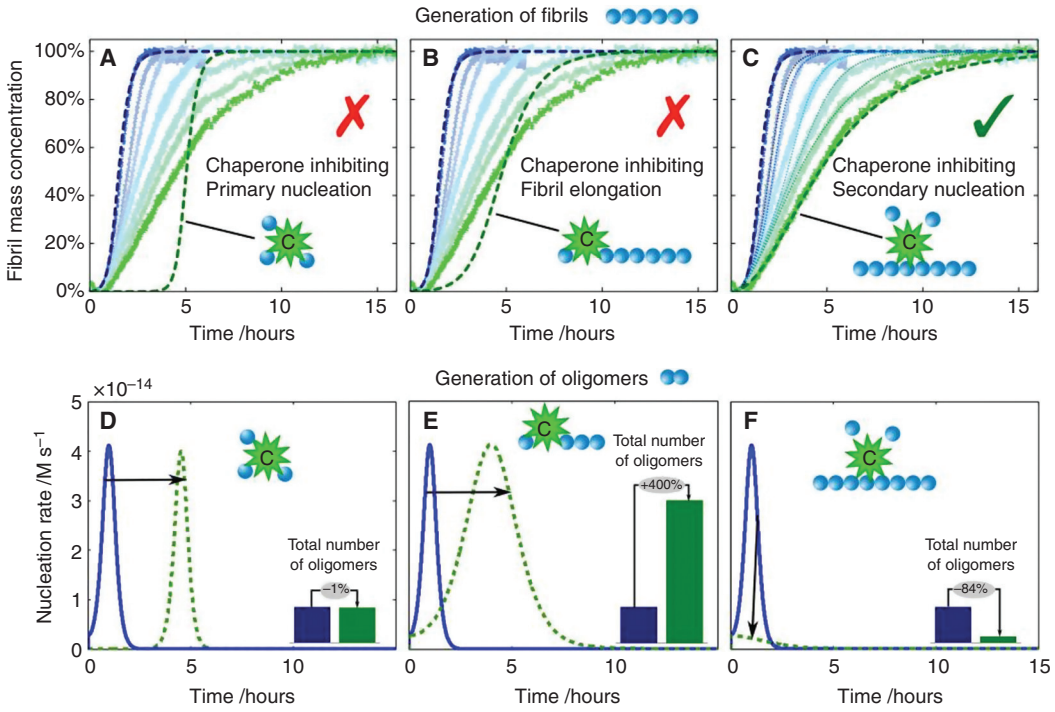


Figure 5. Kinetics of the 42 residue Aβ-peptide (Aβ42) aggregation in the presence of the molecular chaperone Brichos. (*A–C*) Reaction profiles from *left* (blue) to *right* (green) in the absence of Brichos and in the presence of increasing concentrations of Brichos up to a 1:1 stoichiometry. The blue dashed line is the integrated rate law for Aβ42 aggregation in the absence of Brichos using the rate constants determined previously. The green dashed lines show predictions of the resulting reaction profiles when each of (*A*) primary nucleation, (*B*) fibril elongation, and (*C*) secondary (surface-catalyzed) nucleation is inhibited by the chaperone. The thin dotted lines in (*C*) are theoretical predictions for the reaction profiles at the intermediate Brichos concentrations using the association and dissociation rate constants determined for its binding by means of surface plasmon resonance (SPR) measurements. (*D–F*) Time evolution of the nucleation rates calculated from the kinetic analysis. The blue line corresponds to the situation in the absence of Brichos and the green dashed lines show predictions for the cases when each of (*D*) primary nucleation, (*E*) fibril elongation, and (*F*) secondary nucleation is inhibited by the chaperone. The insets show the relative numbers of oligomers predicted to be generated during the aggregation reaction. (From Cohen et al. 2015; reprinted, with permission, from Nature Publishing Group © 2015 courtesy of the Open Access Licensing Policy.)



## STRATEGIES FOR THERAPEUTIC INTERVENTION

Therapeutic strategies for long-established diseases, such as bacterial and viral infections, cancer, and heart conditions, involve such processes as the prevention of infection by external agents and the selective targeting of specific biochemical pathways associated with disease. These strategies, which have been developed over decades or even centuries of study, have proved to be increasingly effective in reducing the incidence of disease and also in limiting their effects on the individuals concerned. It is largely the advances in the prevention and treatment of these conditions that have resulted in the dramatic increase in human life span in recent years, and hence to the increasing prevalence of age-related amyloid disorders, particularly neurodegenerative conditions such as Alzheimer's disease. Such conditions, however, require different therapeutic approaches as they result fundamentally from the failure of control and regulatory processes that normally prevent

our essential functional protein molecules converting into aberrant species with the potential to disrupt the normal processes on which living systems depend.

This rapid increase in the incidence of amyloid-related diseases has been such that until recently little detailed research had been performed to explore their molecular origins and means of progression. Early drug discovery efforts, at least partly for this reason, have led to very disappointing clinical trials, in particular those targeted at neurodegenerative conditions (Cummings et al. 2014, 2018). The progress in understanding the fundamental principles underlying misfolding and aggregation that has been made in recent years, however, now provides new opportunities to engage in rational therapeutic strategies (Fig. 6; Dobson 2004; Balch et al. 2008; Ong and Kelly 2011). In the following section of this review, this topic will be discussed in the context of the underlying molecular events associated with the aggregation process that have been discussed above, and of means of enhancing the ability of the natural protective mechanisms within living systems.

A major objective of any approach to the maintenance of human health is to try to prevent the onset of disease or to make its effects less deleterious. In the context of misfolding disorders, the initiation of the aggregation process can be followed by the rapid proliferation of aggregated proteins, making its suppression a particularly attractive therapeutic target. A pioneering demonstration of this approach is that designed for the treatment of amyloid-related conditions based on the aggregation of transthyretin (Johnson et al. 2005; Ong and Kelly 2011). This protein, whose function is to transport the thyroid hormone thyroxin, is tetrameric in its native state, but certain mutations result in its destabilization into monomers that are highly aggregation prone and give rise to both systemic and neurological conditions (Table 1). The native state of the protein can, however, be stabilized by binding to a substrate analog, hence reducing the risk of aggregation (Johnson et al.

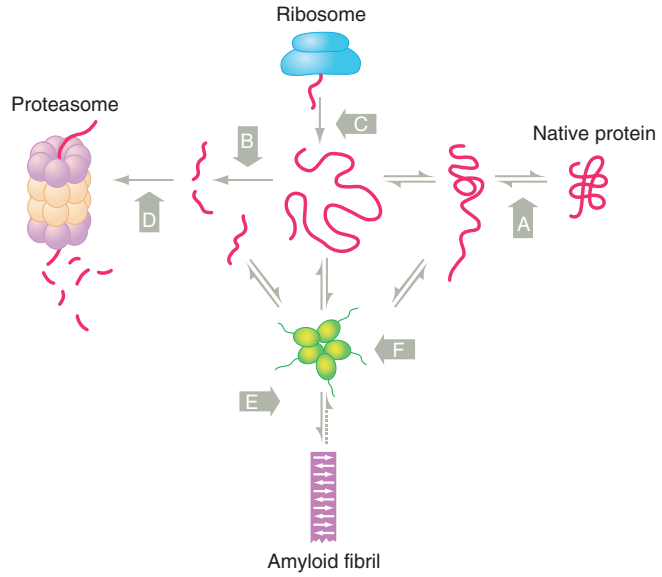


**Figure 6.** Schematic diagram indicating some of the processes following protein synthesis that can potentially be perturbed for therapeutic purposes to combat amyloid disorders. The various possible strategies include (*A*) stabilizing the native state; (*B*) inhibiting enzymes that process proteins into peptides with a higher propensity to aggregate; (*C*) inhibiting protein synthesis; (*D*) stimulating clearance of misfolded proteins, for example, by boosting proteasomal degradation; (*E*) perturbing the assembly of fibrils; and (*F*) suppressing the formation of toxic oligomeric fibril precursors. (Image based on data in Ciryam et al. 2017.)

2005). This strategy has been used very effectively to develop a drug, tafamidis, which is currently in clinical use to treat familial amyloid polyneuropathy (Said et al. 2012).

Stabilizing the native state, however, is problematic for the proteins whose aggregation is associated with most neurological conditions, including Alzheimer's and Parkinson's diseases, which are natively disordered and therefore lack specific tight binding sites for other molecules (Dyson and Wright 2005; Heller et al. 2018). An alternative strategy to suppress aggregation is therefore to inhibit one or more of the microscopic steps involved in the process of amyloid formation, thereby reducing the risk in particular of the accumulation of toxic oligomeric intermediates, in an analogous way to that observed to result from the presence of molecular chaperones. Recent studies of the Aβ-peptides and α-synuclein have resulted in the identification of a range of small molecules that are highly effective in vitro in reducing the rates of primary or secondary nucleation, or indeed both; such molecules have also been found to be effective in reducing very substantially the generation of amyloid deposits in vivo using *Caenorhabditis elegans* as a model system (Fig. 7; Habchi et al. 2016; Perni et al. 2017). With the development of high-throughput screening methods, it has been shown that the efficacy of such inhibitors can be increased very significantly by methods analogous to those used in conventional drug discovery techniques, except that they use the measurement of kinetic parameters rather than binding affinities (Chia et al. 2018).

Some of the molecules that are emerging from such studies are natural products of drugs already approved for other conditions, making repurposing a very real possibility. In addition to small molecules, antibodies are being widely explored in the context of potential therapeutic strategies to combat neurodegenerative conditions (Schenck 2002; Sevigny et al. 2016), and recent work has revealed their ability to inhibit individual steps in the aggregation of specific proteins. A series of single-domain antibodies has been designed by computational methods to bind to short stretches along the sequence of the Aβ-peptide; these molecules have been found to inhibit the microscopic steps in the aggregation process in distinctive ways, affecting primary and secondary nucleation to different extents (Fig. 8; Aprile et al. 2017). Taken together, these results suggest that it should be possible to reduce the level of toxic oligomeric species, by means of small molecules or macromolecules, to levels that enable the protective cellular mechanisms to maintain protein homeostasis for longer periods of time and hence postpone into older ages the onset of disease.

## LOOKING TO THE FUTURE

Our knowledge of the fundamental nature and significance of the amyloid states of proteins has developed very rapidly in recent years, enabling us to understand at a molecular level many aspects of their structures and properties, the mechanisms by which they can be formed from the functional states of proteins, and the manner in which such mechanisms are controlled and regulated in healthy living systems. As a result, it is becoming possible to gain new insights into the way that the properties of functional proteins have coevolved with their environments to enable protein homeostasis to be maintained under normal physiological conditions, but to be vulnerable to impairment under others, for example, as a consequence of mutations or of aging. Such insights also suggest opportunities to intervene therapeutically to reduce the risk of the initiation of aggregation, an approach that should be preventative against the onset of disease, or to reduce the generation of toxic oligomeric species during the course of aggregation, a process that should represent a potential means of treatment. In both situations, a key objective is to enhance by such means the natural protective mechanisms within cells and tissues to enable protein homeostasis to be maintained, particularly into older age.

Although a great deal has now been established about the physical nature of the amyloid phenomenon and its relationship to biology and medicine, there remains much to be done. Thus, for example, we need to determine in greater detail the structural characteristics of the oligomeric states that are an inherent feature of the process

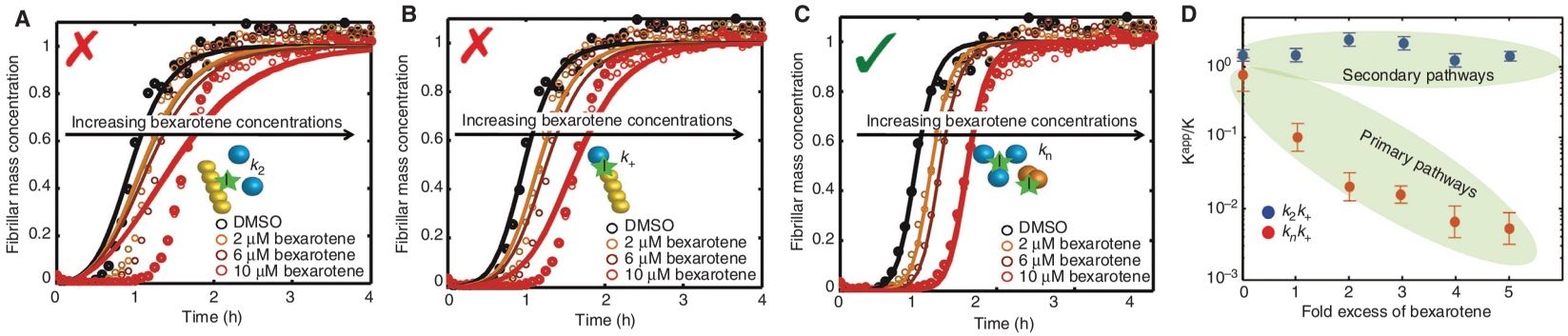


**Figure 7.** Inhibition of Aβ42 aggregation by a small molecule. (*A–C*). Kinetic profiles of the aggregation of Aβ42 in the absence or presence of bexarotene (a small- molecule drug developed for cancer chemotherapy) in concentration ratios from 1:1 to 5:1 Aβ42: bexarotene. The solid lines show predictions for the resulting reaction profiles when secondary nucleation (*A*), fibril elongation (*B*), or primary nucleation (*C*) is inhibited by bexarotene. Only the prediction for inhibition of primary nucleation closely fits the experimental data. (*D*) Evolution of the apparent reaction rate constants with increasing concentrations of bexarotene. The definitions of the rate constants are as in Figure 3, and $k$ represents in each case either $k_nk_+$ or $k_2k_+$. The data show the significant decrease in primary pathways when compared with secondary pathways as the concentration of bexarotene is increased. (From Habchi et al. 2016; reprinted, with permission, from The American Association for the Advancement of Science © 2016 under the terms of the Creative Commons Attribution-NonCommercial license.)

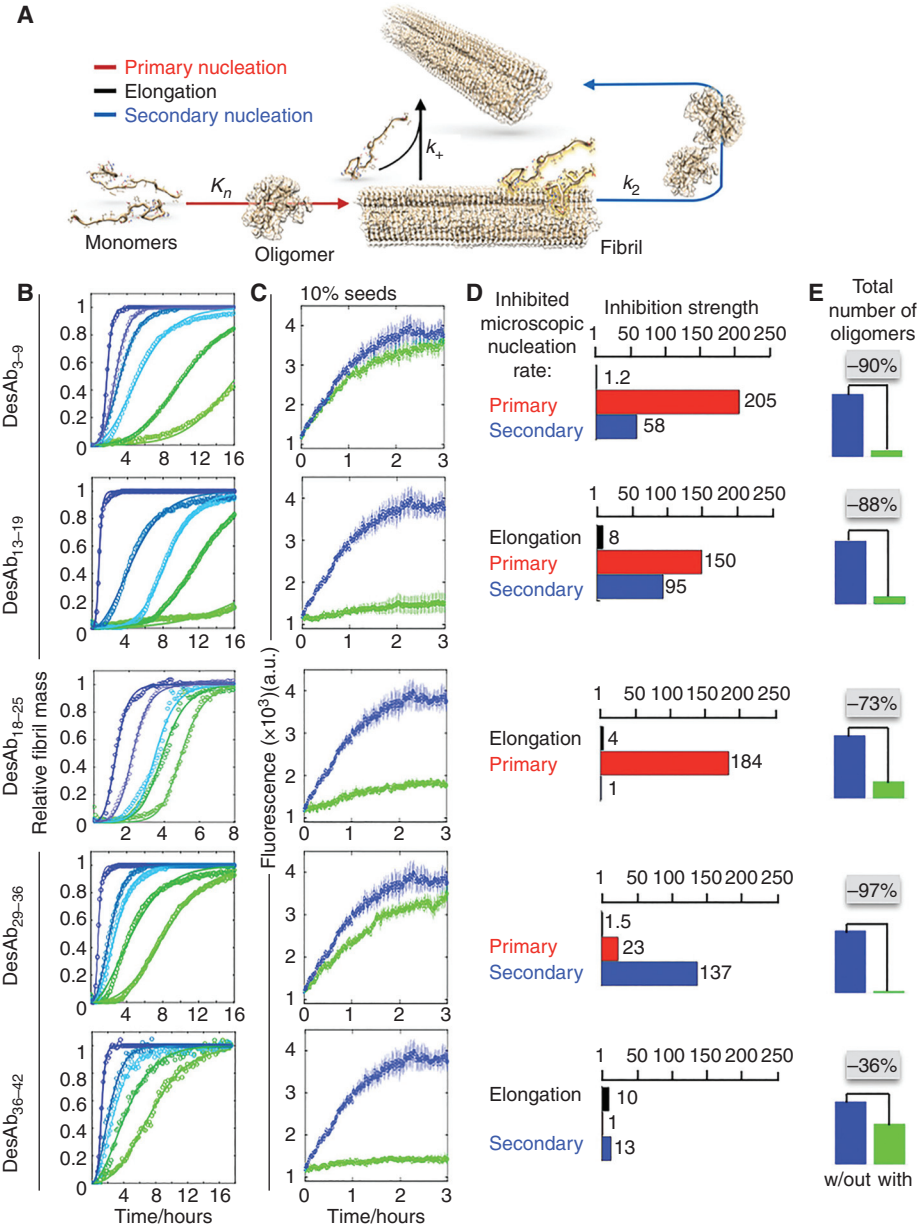


**Figure 8.** The antibody scanning method produces antibodies that affect different microscopic steps in the aggregation of Aβ42. (*A*) Schematic of Aβ42 aggregation showing the primary (red arrow) and secondary (blue arrow) nucleation of the aggregation process and the elongation of fibrils (black). (*B*) The designed antibodies (DesAbs) were generated to target five different epitopes in the Aβ42 sequence (indicated on the *left* side) and the kinetics of aggregation were monitored at different concentrations of each DesAb at increasing (blue to green) concentrations relative to that of Aβ42. (*C*) Seeded aggregation at a low (blue) or high (green) concentration of added fibrils. (*D*) Bar plots showing that each DesAb inhibits the microscopic steps in the aggregation process in a different way; the fold change of each of the rate constants is shown at the top of the corresponding bar. (*E*) The relative number of oligomers generated during the aggregation reaction with 1:2 DesAb:Aβ42 ratios. (From Aprile et al. 2017; reprinted, with permission, from the American Association for the Advancement of Science © 2016 under the terms of the Creative Commons Attribution-NonCommercial license.)

of protein aggregation, to define more fully the dynamic network of events involved in their interconversion, and to understand in greater depth the nature and consequences of their interactions with living systems. We need also to know more about the mechanisms by which homeostasis is preserved, and the manner in which it becomes impaired during the process of aging. In the context of human health, particularly for neurodegenerative conditions, we need enhanced diagnostic methods and improved biomarkers through which to monitor the effects of potential drugs more quantitatively, and at earlier stages of disease, than is currently possible. The increasing recognition of the impact of these conditions, and the greater interactions that are now developing in this area between the physical, biological, and medical sciences, suggest that we should be optimistic about rapid progress in this vital field of human health and welfare.

Finally, it is hard to overestimate the need for enhanced levels of research to generate rapid progress. A large fraction of the human race is now suffering from highly debilitating and incurable amyloid-related conditions that already place huge burdens on the financial resources and health care systems of the world (World Alzheimer Report 2018). The increase in longevity achieved in recent decades has already resulted in neurodegenerative conditions becoming the leading cause of death in many parts of the world, and suggests that the numbers of people afflicted by such disorders alone will more than double in the next 25 years, the majority of whom will be in low and middle-income countries (World Alzheimer Report 2018). The study of the nature of protein misfolding and aggregation, and its contribution to loss of protein homeostasis in aging, promises to be the crucial step in the development of the means to address one of the greatest challenges facing humanity in the modern age.

## ACKNOWLEDGMENTS

We acknowledge the major contributions to the work described in this article of the many graduate students, postdoctoral fellows, colleagues, and other collaborators whose discoveries and ideas are reflected in its content, many of whose names appear in the citations to published work. This article is an extensively revised and updated version of material discussed in Knowles et al. (2014) and Dobson (2013). We also express our gratitude to the many organizations that have funded our research over many years, including the Wellcome Trust, the Leverhulme Trust, the Royal Society, Alzheimer's Research UK, Parkinson's UK, the European Commission, the UK Research Councils, and the Cambridge Centre for Misfolding Diseases.

## REFERENCES

Aguzzi A, Calella AM. 2009. Prions: Protein aggregation and infectious diseases. *Physiol Rev* **89:** 1105–1152. doi:10.1152/physrev.00006.2009

Amor S, Puentas F, Baker D, van der Valk P. 2010. Inflammation in neurodegenerative diseases. *Immunology* **129:** 154–169. doi:10.1111/j.1365-2567.2009.03225.x

Aprile FA, Sormanni P, Perni M, Arosio P, Linse S, Knowles TPJ, Dobson CM, Vendruscolo M. 2017. Selective targeting of primary and secondary nucleation pathways in Aβ42 aggregation using a rational antibody scanning method. *Sci Adv* **3:** e1700488. doi:10.1126/sciadv.1700488.

Arosio P, Michaels TCT, Linse S, Månsson C, Emanuelsson C, Presto J, Johansson J, Vendruscolo M, Dobson CM, Knowles TPJ. 2016. Kinetic analysis reveals the diversity of microscopic mechanisms through which molecular chaperones suppress amyloid formation. *Nat Commun* **7:** 10948. doi:10.1038/ncomms10948

Balch WE, Morimoto RI, Dillin A, Kelly JW. 2008. Adapting proteostasis for disease intervention. *Science* **319:** 916–919. doi:10.1126/science.1141448

Baldwin AJ, Knowles TPJ, Tartaglia GG, Fitzpatrick AW, Devlin GL, Shammas SL, Waudby CA, Mossuto MF, Meehan S, Gras SL, et al. 2011. Metastability of native proteins and the phenomenon of amyloid formation. *J Am Chem Soc* **133:** 14160–14163. doi:10.1021/ja2017703

Benilova I, Karran E, De Stroopoer B. 2012. The toxic Aβ oligomer and Alzheimer's disease: An emperor in need of clothes. *Nat Neurosci* **15:** 349–357. doi:10.1038/nn.3028

Broome BM, Hecht MH. 2000. Nature disfavours sequences of alternative polar and non-polar amino acids: Implications for amyloidogenesis. *J Mol Biol* **296:** 961–968. doi:10.1006/jmbi.2000.3514

Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, Zurdo J, Taddei N, Ramponi G, Dobson CM, Stefani M. 2002. Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416:** 507–511. doi:10.1038/416507a

Campioni S, Mannini B, Zampagni M, Pensalfini A, Parrini C, Evangelisti E, Relini A, Stefani M, Dobson CM, Cecchi C, et al. 2010. A causative link between the structure of aberrant protein oligomers and their toxicity. *Nat Chem Biol* **6:** 140–147. doi:10.1038/nchembio.283

Caughey B, Lansbury PT. 2003. Protofibrils, pores, fibrils, and neurodegeneration: Separating the responsible protein aggregates from the innocent bystanders. *Annu Rev Neurosci* **26:** 267–298. doi:10.1146/annurev.neuro.26.010302.081142

Chia S, Habchi J, Michaels TCT, Cohen SIA, Linse S, Dobson CM, Knowles TPJ, Vendruscolo M. 2018. SAR by kinetics for drug discovery in protein misfolding diseases. *Proc Natl Acad Sci* **115:** 10245–10250. doi:10.1073/pnas.1807884115

Chiti F, Dobson CM. 2017. Protein misfolding, amyloid formation, and human disease: A summary of progress over the last decade. *Ann Rev Biochem* **86:** 27–68. doi:10.1146/annurev-biochem-061516-045115

Ciryam P, Kundra R, Morimoto RI, Dobson CM, Vendruscolo M. 2015. Supersaturation is a major driving force for protein aggregation in neurodegenerative diseases. *Trends Pharmacol Sci* **36:** 72–77. doi:10.1016/j.tips.2014.12.004

Ciryam P, Kundra R, Freer R, Morimoto RI, Dobson CM, Vendruscolo M. 2016. A transcriptional signature of Alzheimer's disease is associated with a metastable subproteome at risk of aggregation. *Proc Natl Acad Sci* **113:** 4753–4758. doi:10.1073/pnas.1516604113

Ciryam P, Lambert-Smity IA, Bean DM, Freer R, Cid F, Tartaglia GG, Saunders DN, Wilson MR, Oliver SG, Morimoto RI, et al. 2017. Spinal motor neuron protein supersaturation patterns are associated with inclusion body formation in ALS. *Proc Natl Acad Sci* **114:** E3935–E3943. doi:10.1073/pnas.1613854114

Cohen SIA, Linse S, Luheshi LM, Hellstrand E, White DA, Rajah L, Otzen DE, Vendruscolo M, Dobson CM, Knowles TPJ. 2013. Proliferation of amyloid-β42 aggregates occurs through a secondary nucleation mechanism. *Proc Natl Acad Sci* **110:** 9758–9763. doi:10.1073/pnas.1218402110

Cohen SIA, Arosio P, Presto J, Kurudenkandy FR, Biverstål H, Dolfe L, Dunning C, Yang X, Frohm B, Vendruscolo M, et al. 2015. A molecular chaperone breaks the catalytic cycle that generates toxic Aβ oligomers. *Nat Struct Mol Biol* **22:** 207–213. doi:10.1038/nsmb.2971

Colvin MT, Silvers R, Frohm B, Su Y, Linse S, Griffin RG. 2015. High resolution structural characterization of Aβ42 amyloid fibrils by magic angle spinning NMR. *J Am Chem Soc* **137:** 7509–7518. doi:10.1021/jacs.5b03997

Cremades N, Cohen SIA, Deas E, Abramov AY, Chen AY, Orte A, Sandal M, Clarke RW, Dunne P, Aprile FA, et al. 2012. Direct observation of the interconversion of normal and toxic forms of α-synuclein. *Cell* **149:** 1048–1059. doi:10.1016/j.cell.2012.03.037

Cummings JL, Morstorf T, Zhong K. 2014. Alzheimer's disease drug-development pipeline: Few candidates, frequent failures. *Alzheimers Res Ther* **6:** 37. doi:10.1186/alzrt269

Cummings J, Lee G, Ritter A, Zhong K. 2018. Alzheimer's disease drug development pipeline: 2018. *Alzheimers Dement* **4:** 195–214.

Dobson CM. 2003. Protein folding and misfolding. *Nature* **426:** 884–890. doi:10.1038/nature02261

Dobson CM. 2004. In the footsteps of alchemists. *Science* **304:** 1259–1262. doi:10.1126/science.1093078

Dobson CM. 2013. The amyloid phenomenon and its significance. In *Amyloid fibrils and prefibrillar aggregates: Molecular and biological properties* (ed. DE Otzen), pp. 1–19. Wiley, Hoboken, NJ.

Douglas PM, Dillin A. 2010. Protein homeostasis and aging in neurodegeneration. *J Cell Biol* **190:** 719–729. doi:10.1083/jcb.201005144

Dyson HJ, Wright PE. 2005. Intrinsically unstructured proteins and their functions. *Nat Rev Mol Cell Biol* **6:** 197–208. doi:10.1038/nrm1589

Eisenberg D, Jucker M. 2012. The amyloid state of proteins in human diseases. *Cell* **148:** 1188–1203. doi:10.1016/j.cell.2012.02.022

Emak G, Davies KJ. 2002. Calcium and oxidative stress: From cell signaling to cell death. *Mol Immunol* **38:** 713–721. doi:10.1016/S0161-5890(01)00108-0

Fändrich M, Dobson CM. 2002. The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *EMBO J* **21:** 5682–5690. doi:10.1093/emboj/cdf573

Fitzpatrick AWP, Falcon B, He S, Murzin AG, Murshudov G, Garringer HJ, Crowther RA, Ghetti B, Goedert M, Scheres SHW. 2017. Cryo-EM structures of tau filaments from Alzheimer's disease. *Nature* **547:** 185–190. doi:10.1038/nature23002

Fowler DM, Koulov AV, Balch WE, Kelly JW. 2007. Functional amyloid—From bacteria to humans. *Trends Biochem Sci* **32:** 217–224. doi:10.1016/j.tibs.2007.03.003

Freer R, Sormanni P, Vecchi G, Ciryam P, Dobson CM, Vendruscolo M. 2016. A protein homeostasis signature in healthy brains recapitulates tissue vulnerability to Alzheimer's disease. *Sci Adv* **2:** e1600947. doi:10.1126/sciadv.1600947

Fu H, Possenti A, Freer R, Nakano Y, Vallegas NCH, Tang M, Cauhy PVM, Lassus BA, Chen S, Fowler SL, et al. 2019. A tau homeostasis signature is linked with the cellular and regional vulnerability of excitatory neurons to tau pathology. *Nat Neurosci* **22:** 47–56. doi:10.1038/s41593-018-0298-7

Fusco G, Chen SW, Williamson PTF, Cascella R, Perni M, Jarvis JA, Cecchi C, Vendruscolo M, Chiti F, Cremades N, et al. 2017. Structural basis of membrane disruption and cellular toxicity by α-synuclein oligomers. *Science* **358:** 1440–1443. doi:10.1126/science.aan6160

Galvagnion C, Buell AK, Meisl G, Michaels TCT, Vendruscolo M, Knowles TPJ, Dobson CM. 2015. Lipid vesicles trigger α-synuclein aggregation by stimulating primary nucleation. *Nat Chem Biol* **11:** 229–234. doi:10.1038/nchembio.1750

Gazit E. 2002. The "correctly folded" state of proteins: Is it a metastable state? *Angew Chem Int Ed Engl* **41:** 257–259. doi:10.1002/1521-3773(20020118)41:2<257::AID-ANIE257>3.0.CO;2-M

Guo JL, Lee VM. 2014. Cell-to-cell transmission of pathogenic proteins in neurodegenerative diseases. *Nat Med* **20:** 130–138. doi:10.1038/nm.3457

Haass C, Selkoe DJ. 2007. Soluble protein oligomers in neurodegeneration: Lessons from the Alzheimer's amyloid β-peptide. *Nat Rev Mol Cell Biol* **8:** 101–112. doi:10.1038/nrm2101

Habchi J, Arosio P, Perni M, Costa AR, Yagi-Utsumi M, Joshi P, Chia S, Cohen SIA, Müller MBD, Linse S, et al. 2016. An anticancer drug suppresses the primary

nucleation reaction that initiates the production of the toxic Aβ42 aggregates linked with Alzheimer's disease. *Sci Adv* **2:** e1501244. doi:10.1126/sciadv.1501244

Hartl FU, Bracher A, Hayer-Hartl M. 2011. Molecular chaperones in protein folding and proteostasis. *Nature* **475:** 324–332. doi:10.1038/nature10317

Heller GT, Bonomi M, Vendruscolo M. 2018. Structural ensemble modulation upon small-molecule binding to disordered proteins. *J Mol Biol* **430:** 2288–2292. doi:10.1016/j.jmb.2018.03.015

Herling TW, Levin A, Saar KL, Dobson CM, Knowles TPJ. 2018. Microfluidic approaches for probing amyloid formation and behaviour. *Lab Chip* **18:** 999–1016. doi:10.1039/C7LC01241A

Jack CR Jr., Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, Holtzman DM, Jagust W, Jessen F, Karlawish J, et al. 2018. NIA-AA research framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement* **14:** 535–562. doi:10.1016/j.jalz.2018.02.018.

Jiménez JL, Nettleton EJ, Bouchard M, Robinson CV, Dobson CM, Saibil HR. 2002. The protofilament structure of insulin amyloid fibrils. *Proc Natl Acad Sci* **99:** 9196–9201. doi:10.1073/pnas.142459399

Johnson SM, Wiseman RL, Sekijima Y, Green NS, Adamski-Werner SL, Kelly JW. 2005. Native state kinetic stabilization as a strategy to ameliorate protein misfolding diseases: A focus on the transthyretin amyloidoses. *Acc Chem Res* **38:** 911–921. doi:10.1021/ar020073i

Kaminski Schierle GS, van de Linde S, Erdelyi M, Esbjörner EK, Klein T, Rees E, Bertoncini CW, Dobson CM, Sauer M, Kaminski CF. 2011. In situ measurements of the formation and morphology of intracellular β-amyloid fibrils by super-resolution fluorescence imaging. *J Am Chem Soc* **133:** 12902–12905. doi:10.1021/ja201651w

Kayed R, Head E, Thompson JL, McIntire TM, Milton SC, Cotman CW, Glabe CG. 2003. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science* **300:** 486–489. doi:10.1126/science.1079469

Knowles TPJ, Buehler MJ. 2011. Nanomechanics of functional and pathological amyloid materials. *Nat Nanotechnol* **6:** 469–479. doi:10.1038/nnano.2011.102

Knowles TPJ, Fitzpatrick AW, Meehan S, Mott HR, Vendruscolo M, Dobson CM, Welland ME. 2007. Role of intermolecular forces in defining material properties of protein nanofibrils. *Science* **318:** 1900–1903. doi:10.1126/science.1150057

Knowles TPJ, Waudby CA, Devlin GL, Cohen SIA, Aguzzi A, Vendruscolo M, Terentjev EM, Welland ME, Dobson CM. 2009. An analytical solution to the kinetics of breakable filament assembly. *Science* **326:** 1533–1537. doi:10.1126/science.1178250

Knowles TPJ, White DA, Abate AR, Agresti JJ, Cohen SIA, Sperling RA, De Genst EJ, Dobson CM, Weitz DA. 2011. Observation of spatial propagation of amyloid assembly from single nuclei. *Proc Natl Acad Sci* **108:** 14746–14751. doi:10.1073/pnas.1105555108

Knowles TPJ, Vendruscolo M, Dobson CM. 2014. The amyloid state and its association with protein misfolding diseases. *Nat Rev Mol Cell Biol* **15:** 384–396. doi:10.1038/nrm3810

Kundel F, Hong L, Falcon B, McEwan WA, Michaels TCT, Meisl G, Esteras N, Abramov AY, Knowles TPJ, Goedert M, et al. 2018. Measurement of tau filament fragmentation provides insights into prion-like spreading. *ACS Chem Neurosci* **9:** 1276–1282. doi:10.1021/acschemneuro.8b00094

Kundra R, Ciryam P, Morimoto RI, Dobson CM, Vendruscolo M. 2017. Protein homeostasis of a metastable subproteome associated with Alzheimer's disease. *Proc Natl Acad Sci* **114:** E5703–E5711. doi:10.1073/pnas.1618417114

Labbadia J, Morimoto RI. 2015. The biology of proteostasis in aging and disease. *Annu Rev Biochem* **84:** 435–464. doi:10.1146/annurev-biochem-060614-033955

Lue LF, Kuo YM, Roher AE, Brachova L, Shen Y, Sue L, Beach T, Kurth JH, Rydel RE, Rogers J. 1999. Soluble amyloid β peptide concentration as a predictor of synaptic change in Alzheimer's disease. *Am J Pathol* **155:** 853–862. doi:10.1016/S0002-9440(10)65184-X

Luheshi LM, Crowther DC, Dobson CM. 2008. Protein misfolding and disease: From the test tube to the organism. *Curr Opin Chem Biol* **12:** 25–31. doi:10.1016/j.cbpa.2008.02.011

Maji SK, Perrin MH, Sawaya MR, Jessberger S, Vadodaria K, Rissman RA, Singru PS, Nilsson KP, Simon R, Schubert D, et al. 2009. Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* **325:** 328–332. doi:10.1126/science.1173155

Mannini B, Cascella R, Zampagni M, van Waarde-Verhagen M, Meehan S, Roodveldt C, Campioni S, Boninsegna M, Penco A, Relini A, et al. 2012. Molecular mechanisms used by chaperones to reduce the toxicity of aberrant protein oligomers. *Proc Natl Acad Sci* **109:** 12479–12484. doi:10.1073/pnas.1117799109

Meisl G, Kirkegaard JB, Arosio P, Michaels TCT, Vendruscolo M, Dobson CM, Linse S, Knowles TPJ. 2016. Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nat Protoc* **11:** 252–272. doi:10.1038/nprot.2016.010

Michaels TCT, Šarić A, Habchi J, Chia S, Meisl G, Vendruscolo M, Dobson CM, Knowles TPJ. 2018. Chemical kinetics for bridging molecular mechanisms and macroscopic measurements of amyloid fibril formation. *Annu Rev Phys Chem* **69:** 273–298. doi:10.1146/annurev-physchem-050317-021322

Morimoto RI. 2008. Proteotoxic stress and inducible chaperone networks in neurodegenerative disease and aging. *Genes Dev* **22:** 1427–1438. doi:10.1101/gad.1657108

Narayan P, Orte A, Clarke RW, Bolognesi B, Hook S, Ganzinger KA, Meehan S, Wilson MR, Dobson CM, Klenerman D. 2011. The extracellular chaperone clusterin sequesters oligomeric forms of the amyloid-β(1–40) peptide. *Nat Struct Mol Biol* **19:** 79–83. doi:10.1038/nsmb.2191

Nelson R, Sawaya MR, Balbirnie M, Madsden AØ, Riekel C, Grothe R, Eisenberg D. 2005. Structure of the cross-β spine of amyloid-like fibrils. *Nature* **435:** 773–778. doi:10.1038/nature03680

Ong DS, Kelly JW. 2011. Chemical and/or biological therapeutic strategies to ameliorate protein misfolding diseases. *Curr Opin Cell Biol* **23:** 231–238. doi:10.1016/j.ceb.2010.11.002.

Cold Spring Harbor Perspectives in Biology
www.cshperspectives.org

Pawar AP, Dubay KF, Zurdo J, Chiti F, Vendruscolo M, Dobson CM. 2005. Prediction of "aggregation-prone" and "aggregation-susceptible" regions in proteins associated with neurodegenerative diseases. *J Mol Biol* **350:** 379–392. doi:10.1016/j.jmb.2005.04.016

Perni M, Galvagnion C, Maltsev A, Meisl G, Müller MB, Challa PK, Kirkegaard JB, Flagmeier P, Cohen SIA, Cascella R, et al. 2017. A natural product inhibits the initiation of α-synuclein aggregation and suppresses its toxicity. *Proc Natl Acad Sci* **114:** E1009–E1017. doi:10.1073/pnas.1610586114

Petkova AT, Ishii Y, Balbach JJ, Antzutkin ON, Leapman RD, Delaglio F, Tycko R. 2002. A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc Natl Acad Sci* **99:** 16742–16747. doi:10.1073/pnas.262663499

Petkova AT, Leapman RD, Guo Z, Yau WM, Mattson MP, Tycko R. 2005. Self-propagating, molecular-level polymorphism in Alzheimer's β-amyloid fibrils. *Science* **307:** 262–265. doi:10.1126/science.1105850

Robinson CV, Sali A, Baumeister W. 2007. The molecular sociology of the cell. *Nature* **450:** 973–982. doi:10.1038/nature06523

Sachse C, Fändrich M, Grigorieff N. 2008. Paired β-sheet structure of an Aβ(1–40) amyloid fibril revealed by electron microscopy. *Proc Natl Acad Sci* **105:** 7462–7466. doi:10.1073/pnas.0712290105

Safar J, Wille H, Itri V, Groth D, Serban H, Torchia M, Cohen FE, Prusiner SB. 1998. Eight prion strains have PrP(Sc) molecules with different conformations. *Nat Med* **4:** 1157–1165. doi:10.1038/2654

Said G, Grippon S, Kirkpatrick P. 2012. Tafamidis. *Nat Rev Drug Discov* **11:** 185–186. doi:10.1038/nrd3675.

Schenck D. 2002. Amyloid-β immunotherapy for Alzheimer's disease: The end of the beginning. *Nat Rev Neurosci* **3:** 824–828. doi:10.1038/nrn938

Sevigny J, Chiao P, Bussière T, Weinreb PH, Williams L, Maier M, Dunstan R, Salloway S, Chen T, Ling Y, et al. 2016. The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. *Nature* **537:** 50–56. doi:10.1038/nature19323.

Shimanovich U, Efimov I, Mason TO, Flagmeier P, Buell AK, Gedanken A, Linse S, Åkerfeldt KS, Dobson CM, Weitz DA, et al. 2015. Protein microgels from amyloid fibril networks. *ACS Nano* **9:** 43–51. doi:10.1021/nn504869d

Stefani M, Dobson CM. 2003. Protein aggregation and aggregate toxicity: New insights into protein folding, misfolding diseases and biological evolution. *J Mol Med* **81:** 678–699. doi:10.1007/s00109-003-0464-5

Sunde M, Serpell LC, Bartlam M, Fraser PE, Pepys MB, Blake CCF. 1997. Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J Mol Biol* **273:** 729–739. doi:10.1006/jmbi.1997.1348

Tartaglia GG, Pechmann S, Dobson CM, Vendruscolo M. 2007. Life on the edge: A link between gene expression levels and aggregation rates of human proteins. *Trends Biochem Sci* **32:** 204–206. doi:10.1016/j.tibs.2007.03.005

Tartaglia GG, Pawar AP, Campioni S, Dobson CM, Chiti F, Vendruscolo M. 2008. Prediction of aggregation-prone regions in structured proteins. *J Mol Biol* **380:** 425–436. doi:10.1016/j.jmb.2008.05.013

Tomic JL, Pensalfini A, Head E, Glabe CG. 2009. Soluble fibrillar oligomer levels are elevated in Alzheimer's disease brain and correlate with cognitive dysfunction. *Neurobiol Dis* **35:** 352–358. doi:10.1016/j.nbd.2009.05.024

Törnquist M, Michaels TCT, Sanagavarapu K, Yang X, Meisl G, Cohen SIA, Knowles TPJ, Linse S. 2018. Secondary nucleation in amyloid formation. *Chem Commun* **54:** 8667–8684. doi:10.1039/C8CC02204F

Vendruscolo M, Knowles TPJ, Dobson CM. 2011. Protein solubility and protein homeostasis: A generic view of protein misfolding disorders. *Cold Spring Harb Perspect Biol* **3:** a010454. doi:10.1101/cshperspect.a010454

Wei G, Su Z, Reynolds NP, Arosio P, Hamley IW, Gazit E, Mezzenga R. 2017. Self-assembling peptide and protein amyloids: From structure to tailored function in nanotechnology. *Chem Soc Rev* **46:** 4661–4708. doi:10.1039/C6CS00542J

World Alzheimer Report. 2018. Alzheimer's Disease International. www.alz.co.uk/worldreport2018.

Cold Spring Harbor Perspectives in Biology

# The Amyloid Phenomenon and Its Significance in Biology and Medicine

Christopher M. Dobson, Tuomas P.J. Knowles and Michele Vendruscolo

*Cold Spring Harb Perspect Biol* 2020; doi: 10.1101/cshperspect.a033878 originally published online April 1, 2019

**Subject Collection** Protein Homeostasis



For additional articles in this collection, see http://cshperspectives.cshlp.org/cgi/collection/

# Protein Misfolding, Functional Amyloid, and Human Disease

Fabrizio Chiti[1] and Christopher M. Dobson[2]

[1]Dipartimento di Scienze Biochimiche, Università degli Studi di Firenze, I-50134 Firenze, Italy; email: fabrizio.chiti@unifi.it

[2]Department of Chemistry, University of Cambridge, Cambridge, CB2 1EW, United Kingdom; email: cmd44@cam.ac.uk

Annu. Rev. Biochem.
2006. 75:333–66

The *Annual Review of Biochemistry* is online at
biochem.annualreviews.org

doi: 10.1146/
annurev.biochem.75.101304.123901



0066-4154/06/0707-
0333$20.00

## Key Words

aggregation mechanism, Alzheimer, Parkinson, prion, protein aggregation

## Abstract

Peptides or proteins convert under some conditions from their soluble forms into highly ordered fibrillar aggregates. Such transitions can give rise to pathological conditions ranging from neurodegenerative disorders to systemic amyloidoses. In this review, we identify the diseases known to be associated with formation of fibrillar aggregates and the specific peptides and proteins involved in each case. We describe, in addition, that living organisms can take advantage of the inherent ability of proteins to form such structures to generate novel and diverse biological functions. We review recent advances toward the elucidation of the structures of amyloid fibrils and the mechanisms of their formation at a molecular level. Finally, we discuss the relative importance of the common main-chain and side-chain interactions in determining the propensities of proteins to aggregate and describe some of the evidence that the oligomeric fibril precursors are the primary origins of pathological behavior.

## Contents



## INTRODUCTION

Writing a review focused on protein misfolding and the diseases with which it is related is both an exciting and a challenging activity. This is in part because recent interest in this topic has led to an explosion in the number of papers published across a broad spectrum of disciplines, and in part because many of the pathological features of the different diseases, and the characteristics of the proteins

with which they are associated, appear at first sight to be quite diverse. Despite this diversity, it is increasingly evident from the experimental data emerging from a wide range of studies that there are some, perhaps many, common features in the underlying physicochemical and biochemical origins of the various disorders and, indeed, of the cases in which similar processes contribute positively to biological function. It has been one of our primary objectives during the writing of this article to explore the extent to which such common features can provide the foundation on which to develop a deeper understanding of the various phenomena associated with protein misfolding and its consequences. Fortunately, within the past year or two, a variety of excellent reviews and books has appeared on the more specific features of many aspects of this complex subject, such as the two-volume book entitled *Protein Misfolding, Aggregation and Conformational Diseases* (1).

To provide a framework on which to build this article, we first describe the variety of human diseases that are now thought to arise from the misfolding of proteins, particularly those, perhaps the majority, in which misfolding results in the formation of highly organized and generally intractable thread-like aggregates termed amyloid fibrils. We point out, however, that in addition living organisms can take advantage of the inherent ability of proteins to form such structures to generate novel and diverse biological functions. Second, we describe the dramatic advances that have recently been made toward the elucidation of the structures of amyloid fibrils at a molecular level and emphasize that our knowledge of these structures is no longer limited to the notion of a fibrillar morphology and an ordered "cross-β" arrangement of the polypeptide chains of which they are composed. We then describe the progress that is being made toward understanding the mechanism of aggregation and toward identifying the nature of key intermediates in the aggregation process. Finally, we discuss some of the important ideas that are emerging about the pathogenesis of the various protein deposition diseases and show that, in at least some cases, the prefibrillar aggregates, rather than the mature and stable fibrils into which they convert, are the likely origins of pathological behavior.

From the evidence that emerges from such considerations, we have tried to pull together the various threads of this complex subject in an attempt to identify both the common features of the various disorders and the differences that lead to their individual identities. We also try to show that, in delving into the general phenomenon of protein misfolding, considerable light can be shed on the origins of some of the most debilitating and increasingly common diseases that affect humanity as well as on the strategies that are likely to be most effective for their prevention and treatment.

**Protein misfolding:** the conversion of a protein into a structure that differs from its native state

**Amyloid fibrils:** protein aggregates having a cross-β structure and other characteristics, e.g., specific dye-binding

**Protein deposition disease:** any pathological state associated with the formation of intracellular or extracellular protein deposits

## THE ROLE OF AMYLOID-LIKE STRUCTURES IN DISEASE AND IN NORMAL BIOLOGY

A broad range of human diseases arises from the failure of a specific peptide or protein to adopt, or remain in, its native functional conformational state. These pathological conditions are generally referred to as *protein misfolding* (*or protein conformational*) *diseases*. They include pathological states in which an impairment in the folding efficiency of a given protein results in a reduction in the quantity of the protein that is available to play its normal role. This reduction can arise as the result of one of several posttranslational processes, such as an increased probability of degradation via the quality control system of the endoplasmic reticulum, as occurs in cystic fibrosis (2), or the improper trafficking of a protein, as seen in early-onset emphysema (3). The largest group of misfolding diseases, however, is associated with the conversion of specific peptides or proteins from their soluble functional states ultimately into highly organized fibrillar aggregates. These structures are

generally described as amyloid fibrils or plaques when they accumulate extracellularly, whereas the term "intracellular inclusions" has been suggested as more appropriate when fibrils morphologically and structurally related to extracellular amyloid form inside the cell (4). For simplicity, however, we shall describe all such species as amyloid fibrils in this article. It is also becoming clear that fibrillar species with amyloid characteristics can serve a number of biological functions in living organisms, provided they form under controlled conditions. Perhaps the most fascinating of these functions lies in the ability of such structures to serve as transmissible genetic traits distinct from DNA genes.

**Amyloidosis:** any pathological state associated with the formation of extracellular amyloid deposits

**TEM:** transmission electron microscopy

**AFM:** atomic force microscopy

**ThT:** thioflavin T

**CR:** Congo red

**Protofilaments:** the constituent units of amyloid fibrils. They should not be confused with protofibrils

## Many Human Diseases Are Associated with Protein Aggregation

A list of known diseases that are associated with the formation of extracellular amyloid fibrils or intracellular inclusions with amyloid-like characteristics is given in **Table 1**, along with the specific proteins that in each case are the predominant components of the deposits. The diseases can be broadly grouped into neurodegenerative conditions, in which aggregation occurs in the brain, nonneuropathic localized amyloidoses, in which aggregation occurs in a single type of tissue other than the brain, and nonneuropathic systemic amyloidoses, in which aggregation occurs in multiple tissues (**Table 1**).

Some of these conditions, such as Alzheimer's and Parkinson's diseases, are predominantly sporadic (labeled [c] in **Table 1**), although hereditary forms are well documented. Other conditions, such as the lysozyme and fibrinogen amyloidoses, arise from specific mutations and are hereditary (labeled [d] in **Table 1**). In addition to sporadic (85%) and hereditary (10%) forms, spongiform encephalopathies can also be transmissible (5%) in humans as well as in other mammals. It has also been found that intravenous injection or oral administration of preformed fibrils from different sources can result in accelerated AA amyloidosis in mice subjected to an inflammatory stimulus (5, 6). It has therefore been postulated that an environment enriched with fibrillar material could act as a risk factor for amyloid diseases (6). Similarly, injection of the recombinant mouse prion protein in the form of amyloid-like fibrils has been reported to generate disease in mice that express the prion protein (7).

The extracellular proteinaceous deposits found in patients suffering from any of the amyloid diseases have a major protein component that forms the core and then additional associated species, including metal ions, glycosaminoglycans, the serum amyloid P component, apolipoprotein E, collagen, and many others (8, 9). Ex vivo fibrils, representing the amyloid core structures, can be isolated from patients, and closely similar fibrils can also be produced in vitro using natural or recombinant proteins; in this case, mildly denaturing conditions are generally required for their rapid formation, at least for proteins that normally adopt a well-defined folded structure (see below).

The fibrils can be imaged in vitro using transmission electron microscopy (TEM) or atomic force microscopy (AFM). These experiments reveal that the fibrils usually consist of a number (typically 2–6) of protofilaments, each about 2–5 nm in diameter (10). These protofilaments twist together to form rope-like fibrils that are typically 7–13 nm wide (10, 11) or associate laterally to form long ribbons that are 2–5 nm thick and up to 30 nm wide (12–14). X-ray fiber diffraction data have shown that in each individual protofilament the protein or peptide molecules are arranged so that the polypeptide chain forms β-strands that run perpendicular to the long axis of the fibril (11). The fibrils have the ability to bind specific dyes such as thioflavin T (ThT) and Congo red (CR) (15), although the specificity of binding of CR to amyloid fibrils and the resulting green birefringence under

**Table 1 Human diseases associated with formation of extracellular amyloid deposits or intracellular inclusions with amyloid-like characteristics**

| Disease | Aggregating protein or peptide | Number of residues[a] | Native structure of protein or peptide[b] |
|---|---|---|---|
| Neurodegenerative diseases | | | |
| Alzheimer's disease[c] | Amyloid β peptide | 40 or 42[f] | Natively unfolded |
| Spongiform encephalopathies[c,e] | Prion protein or fragments thereof | 253 | Natively unfolded (residues 1–120) and α-helical (residues 121–230) |
| Parkinson's disease[c] | α-Synuclein | 140 | Natively unfolded |
| Dementia with Lewy bodies[c] | α-Synuclein | 140 | Natively unfolded |
| Frontotemporal dementia with Parkinsonism[c] | Tau | 352–441[f] | Natively unfolded |
| Amyotrophic lateral sclerosis[c] | Superoxide dismutase 1 | 153 | All-β, Ig like |
| Huntington's disease[d] | Huntingtin with polyQ expansion | 3144[g] | Largely natively unfolded |
| Spinocerebellar ataxias[d] | Ataxins with polyQ expansion | 816[g,h] | All-β, AXH domain (residues 562–694); the rest are unknown |
| Spinocerebellar ataxia 17[d] | TATA box-binding protein with polyQ expansion | 339[g] | α+β, TBP like (residues 159–339); unknown (residues 1–158) |
| Spinal and bulbar muscular atrophy[d] | Androgen receptor with polyQ expansion | 919[g] | All-α, nuclear receptor ligand-binding domain (residues 669–919); the rest are unknown |
| Hereditary dentatorubral-pallidoluysian atrophy[d] | Atrophin-1 with polyQ expansion | 1185[g] | Unknown |
| Familial British dementia[d] | ABri | 23 | Natively unfolded |
| Familial Danish dementia[d] | ADan | 23 | Natively unfolded |
| Nonneuropathic systemic amyloidoses | | | |
| AL amyloidosis[c] | Immunoglobulin light chains or fragments | ~90[f] | All-β, Ig like |
| AA amyloidosis[c] | Fragments of serum amyloid A protein | 76–104[f] | All-α, unknown fold |
| Familial Mediterranean fever[c] | Fragments of serum amyloid A protein | 76–104[f] | All-α, unknown fold |
| Senile systemic amyloidosis[c] | Wild-type transthyretin | 127 | All-β, prealbumin like |
| Familial amyloidotic polyneuropathy[d] | Mutants of transthyretin | 127 | All-β, prealbumin like |
| Hemodialysis-related amyloidosis[c] | β2-microglobulin | 99 | All-β, Ig like |
| ApoAI amyloidosis[d] | N-terminal fragments of apolipoprotein AI | 80–93[f] | Natively unfolded |
| ApoAII amyloidosis[d] | N-terminal fragment of apolipoprotein AII | 98[i] | Unknown |
| ApoAIV amyloidosis[c] | N-terminal fragment of apolipoprotein AIV | ~70 | Unknown |
| Finnish hereditary amyloidosis[d] | Fragments of gelsolin mutants | 71 | Natively unfolded |
| Lysozyme amyloidosis[d] | Mutants of lysozyme | 130 | α+β, lysozyme fold |
| Fibrinogen amyloidosis[d] | Variants of fibrinogen α-chain | 27–81[f] | Unknown |
| Icelandic hereditary cerebral amyloid angiopathy[d] | Mutant of cystatin C | 120 | α+β, cystatin like |

(*Continued*)

**Table 1** (*Continued*)

| Disease | Aggregating protein or peptide | Number of residues[a] | Native structure of protein or peptide[b] |
|---|---|---|---|
| Nonneuropathic localized diseases | | | |
| Type II diabetes[c] | Amylin, also called islet amyloid polypeptide (IAPP) | 37 | Natively unfolded |
| Medullary carcinoma of the thyroid[c] | Calcitonin | 32 | Natively unfolded |
| Atrial amyloidosis[c] | Atrial natriuretic factor | 28 | Natively unfolded |
| Hereditary cerebral haemorrhage with amyloidosis[d] | Mutants of amyloid β peptide | 40 or 42[f] | Natively unfolded |
| Pituitary prolactinoma | Prolactin | 199 | All-α, 4-helical cytokines |
| Injection-localized amyloidosis[c] | Insulin | 21 + 30[j] | All-α, insulin like |
| Aortic medial amyloidosis[c] | Medin | 50[k] | Unknown |
| Hereditary lattice corneal dystrophy[d] | Mainly C-terminal fragments of kerato-epithelin | 50–200[f] | Unknown |
| Corneal amylodosis associated with trichiasis[c] | Lactoferrin | 692 | α+β, periplasmic-binding protein like II |
| Cataract[c] | γ-Crystallins | Variable | All-β, γ-crystallin like |
| Calcifying epithelial odontogenic tumors[c] | Unknown | ~46 | Unknown |
| Pulmonary alveolar proteinosis[d] | Lung surfactant protein C | 35 | Unknown |
| Inclusion-body myositis[c] | Amyloid β peptide | 40 or 42[f] | Natively unfolded |
| Cutaneous lichen amyloidosis[c] | Keratins | Variable | Unknown |

[a]Data refer to the number of residues of the processed polypeptide chains that deposit into aggregates, not of the precursor proteins.
[b]According to Structural Classification Of Proteins (SCOP), these are the structural class and fold of the native states of the processed peptides or proteins that deposit into aggregates prior to aggregation.
[c]Predominantly sporadic, although in some cases hereditary forms associated with specific mutations are well documented.
[d]Predominantly hereditary, although in some cases sporadic forms are documented.
[e]Five percent of the cases are transmitted (e.g., iatrogenic).
[f]Fragments of various lengths are generated and have been reported to be present in ex vivo fibrils.
[g]Lengths shown refer to the normal sequences with nonpathogenic traits of polyQ.
[h]Length shown is for ataxin-1.
[i]The pathogenic mutation converts the stop codon into a Gly codon, extending the 77-residue protein by 21 additional residues.
[j]Human insulin consists of two chains (A and B, with 21 and 30 residues, respectively) covalently linked by disulfide bridges.
[k]Medin is the 245–294 fragment of human lactadherin.



cross-polarized light has recently been questioned (16, 17).

The proteins found as intractable aggregates in pathological conditions do not share any obvious sequence identity or structural homology to each other. Considerable heterogeneity also exists as to secondary structure composition or chain length (**Table 1**). Interestingly, some amyloid deposits in vivo and fibrils generated in vitro have both been found to include higher-order assemblies, including highly organized species known as spherulites, which can be identified from a characteristic Maltese cross pattern when observed under cross-polarized light (18, 19). Such species are also observed in preparations of synthetic polymers, such as polyethylene, a finding consistent with the idea that amyloid fibrils have features analogous to those of classical polymers.

## Formation of Amyloid Fibrils Is Sometimes Exploited by Living Systems

An increasing number of proteins with no link to protein deposition diseases has been found to form, under some conditions in vitro, fibrillar aggregates that have the morphological, structural, and tinctorial properties that allow them to be classified as amyloid fibrils (20, 21). This finding has led to the idea that the ability to form the amyloid structure is an inherent or generic property of polypeptide chains, although, as we discuss below, the propensity to form such a structure can vary dramatically with sequence. This generic ability can increasingly be seen to have been exploited by living systems for specific purposes, as some organisms have been found to convert, during their normal physiological life cycle, one or more of their endogenous proteins into amyloid fibrils that have functional rather than disease-associated properties. A list of such proteins is reported in **Table 2**.

One particularly well-studied example of functional amyloid is that of the proteinaceous fibrils formed from the protein curlin that are used by *Escherichia coli* to colonize inert surfaces and mediate binding to host proteins. Consistent with the characteristics of other amyloid structures, these fibrils are 6–12 nm in diameter, possess extensive β-sheet structure, as revealed by circular dichroism (CD) spectroscopy, and bind to CR and ThT (22). A second example involves the filamentous bacterium *Streptomyces coelicolor* that produces aerial hyphae, which allow its spores to be dispersed efficiently; a class of secreted proteins called chaplins has been identified in the hyphae of this organism with the ability to form amyloid fibrils that act cooperatively to bring about aerial development (23). All these systems have extremely highly regulated assembly processes; generation of the bacterial *curli*, for example, involves several proteins, including one that nucleates a different protein to form fibrils.

As well as these examples from bacteria, the formation of functional amyloid-like structures has recently been observed in a mammalian system. The melanosomes, lysosome-related organelles that differentiate in melanocytes to allow the epidermal production of the melanin pigment, are characterized by intralumenal fibrous striations upon which melanin granules form. This fibrous material, sharing significant analogies with amyloid fibrils, is assembled from the intralumenal domain of the membrane protein Pmel17 that is proteolyzed by a proprotein convertase (24). This result is a direct indication that even in higher organisms amyloid formation can be physiologically useful for specific and specialized biological functions, provided it is regulated and allowed to take place under highly controlled conditions.

**Functional amyloid:** an amyloid structure found to have a beneficial function in living systems

**CD:** circular dichroism

## Amyloid Structures Can Serve as Nonchromosomal Genetic Elements

As we discussed in the previous paragraph, it is clear that living systems can utilize the amyloid structure as the functional state of some specific proteins. It is also clear, however, that nature has selected, or at the very least has not selected against, some proteins that can exist within normally functioning biological systems in both a soluble conformation and in an aggregated amyloid-like form. Remarkably, this phenomenon has resulted in the latter state being self-perpetuating, infectious, and inheritable as a non-Mendelian nonchromosomal genetic trait (25). Proteins with such behavior are called prions and are listed in **Table 2**. Although the only endogenous mammalian protein so far recognized to have such properties is associated with the group of invariably fatal and transmissible diseases, the heritable conformational changes of prion proteins from some other organisms have, in some cases, been found beneficial.

The prion proteins from *Saccharomyces cerevisiae*, including Ure2p and Sup35p, give rise to distinct phenotypes when adopting either one or the other forms of the

**Table 2 Proteins forming naturally nonpathological amyloid-like fibrils with specific functional roles**

| Protein | Organism | Function of the resulting amyloid-like fibrils | References |
|---|---|---|---|
| Curlin | *Escherichia coli* (bacterium) | To colonize inert surfaces and mediate binding to host proteins | 22 |
| Chaplins | *Streptomyces coelicolor* (bacterium) | To lower the water surface tension and allow the development of aerial hyphae | 23 |
| Hydrophobin[a] EAS | *Neurospora crassa* (fungus) | To lower the water surface tension and allow the development of aerial hyphae | 23a |
| Proteins of the chorion of the eggshell[b] | *Bombyx mori* (silkworm) | To protect the oocyte and the developing embryo from a wide range of environmental hazards | 23b |
| Spidroin | *Nephila edulis* (spider) | To form the silk fibers of the web | 23c |
| Intralumenal domain of Pmel17 | *Homo sapiens* | To form, inside melanosomes, fibrous striations upon which melanin granules form | 24 |
| Ure2p (prion) | *Saccharomyces cerevisiae* (yeast) | To promote the uptake of poor nitrogen sources ([URE3]) | 25 |
| Sup35p (prion) | *Saccharomyces cerevisiae* (yeast) | To confer new phenotypes ([PSI+]) by facilitating the readthrough of stop codons on mRNA | 26–28 |
| Rnq1p (prion) | *Saccharomyces cerevisiae* (yeast) | Not well understood ([RNQ+], also known as [PIN+], phenotype) | 28a |
| HET-s (prion) | *Podospora anserina* (fungus) | To trigger a complex programmed cell death phenomenon (heterokaryon incompatibility) | 31, 32 |
| Neuron-specific isoform of CPEB (prion) | *Aplisia californica* (marine snail) | To promote long-term maintenance of synaptic changes associated with memory storage | 30 |

[a]Other proteins from this class, collectively called hydrophobins, have been found to play similar roles in other species of filamentous fungi.
[b]Suggested to form amyloid-like fibrils in vivo, although amyloid formation has only been observed in vitro.

protein (soluble or fibrillar). These proteins are not related to each other, although they do have some characteristics in common, such as the presence of a globular domain and an unstructured portion of the sequence and the high occurrence of glutamine and asparagine residues in the unstructured domain. The polymerization-mediated inactivation of Sup35p, a protein involved in the termination of mRNA translation, confers a wide variety of novel phenotypes ([PSI+]) by facilitating the readthrough of stop codons (26–28). The aggregation of Ure2p destroys its ability to bind and sequester the transcription factor Gln3p; this results in the activation of a series of genes involved in the uptake of poor nitrogen sources (25). The resulting yeast cells [URE3] can grow on media that, for example, lack uracil but contain its precursor ureidosuccinate (25). Although the low natural occurrence of [URE3] and [PSI+] strains suggest that the corresponding phenotypes are not generally beneficial (29), they can still be advantageous under particular environmental circumstances.

In the marine snail *Aplysia californica*, a neuron-specific isoform of cytoplasmic polyadenylation element-binding protein (CPEB) has also been found to exist in a soluble and a self-perpetuating prion form (30). The prion form was found to be more active than the soluble form in stimulating translation of CPEB-regulated mRNA. From this finding, the suggestion was made that the polymerization of the protein could be essential for the long-term maintenance of synaptic changes associated with memory storage (30). Finally, the polymerization of the HET-s protein from *Podospora anserina* is involved in a controlled programmed cell death phenomenon termed heterokaryon incompatibility (31, 32).

It is evident, even from the relatively few examples that have been studied in detail so far, that the aggregation of proteins into amyloid-like structures can generate a number of extremely diverse biological functions. The presence of many other sequences in the genomes of different organisms with the characteristics of prions suggests that there may yet be surprises in store for us when their properties are investigated.

## THE STRUCTURES OF AMYLOID FIBRILS

For many years the only structural information about amyloid fibrils came from imaging techniques such as TEM, and more recently AFM, and from X-ray fiber diffraction (10, 11, 33). Despite the structural insight given by these techniques, as outlined above, one of the most common statements in the introductory sections of papers in this field until about three years ago was to the effect that "amyloid fibrils cannot be characterized in detail at the molecular level because they are not crystalline yet they are too large to be studied by solution NMR spectroscopy." The situation has changed dramatically recently as a result of major progress in the application of solid-state NMR (SSNMR) spectroscopy to preparations of amyloid fibrils (34–36) and of successes in growing nano- or microcrystals of small peptide fragments that have characteristics of amyloid fibrils yet are amenable to single crystal X-ray diffraction analysis (37, 38).

### High-Resolution Structural Studies Using Solid-State NMR

Using SSNMR, in conjunction with computational energy minimization procedures, Tycko and coworkers (34, 39, 40) have put forward a structure of the amyloid fibrils formed from the 40-residue form of the amyloid β peptide ($A\beta_{1-40}$) at pH 7.4 and 24°C under quiescent conditions. In this structure, each $A\beta_{1-40}$ molecule contributes a pair of β-strands, spanning approximately residues 12–24 and 30–40, to the core region of the fibrils (**Figure 1*a***). These strands, connected by the loop 25–29, are not part of the same β-sheet, however, but participate in the formation of two distinct β-sheets within the same protofilament (**Figure 1*a***). The different Aβ molecules are stacked on to each other, in a parallel arrangement and in register, at least from residue 9 to 39 (39, 40). By invoking additional experimental constraints, such as the diameter of the protofilaments observed using TEM, and the mass per unit length, measured by means of scanning transmission electron microscopy (STEM) (34, 41), it has been suggested that a single protofilament is composed of four β-sheets separated by distances of ~10 Å (**Figure 1*a***).

Support for key elements of this proposed structure comes from experiments of site-directed spin labeling coupled to electron paramagnetic resonance (SDSL-EPR) (42). The values of the inverse central line width in the EPR spectra for a series of labeled residues indicate that the segments of the $A\beta_{1-42}$ molecule corresponding to residues 13–21 and 30–39 are highly structured in the fibrils, parallel and in register. High flexibility and exposure to the solvent of the N-terminal region, in contrast to considerable structural rigidity detected for the remainder of the sequence, are also suggested by experimental strategies that use hydrogen-deuterium exchange methods in conjunction with mass spectrometry (43), limited proteolysis (44), and proline-scanning mutagenesis (45).

SSNMR, in conjunction with site-directed fluorescence labeling and an ingenious hydrogen/deuterium exchange protocol applied previously to probe the regions of β2-microglobulin fibrils that are involved in persistent structure (46), has led to identification of the regions of the C-terminal fragment of HET-s that are involved in the core of the fibril (36). In the proposed structure, each molecule contributes four β-strands, with strands one and three forming the same parallel β-sheet and with strands two and four

**SSNMR:** solid-state nuclear magnetic resonance

**STEM:** scanning transmission electron microscopy

**SDSL-EPR:** site-directed spin labeling coupled to electron paramagnetic resonance

**Aβ:** amyloid β peptide

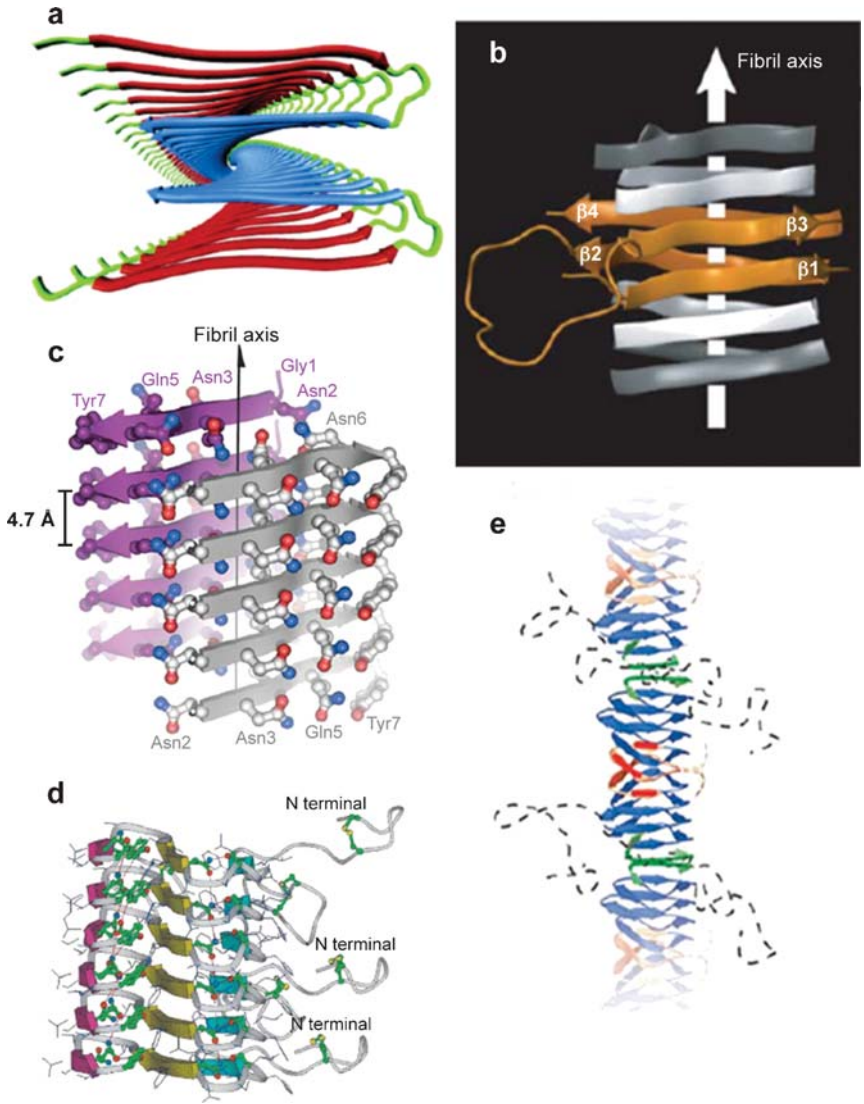


**Figure 1**

Recent three-dimensional structural models of fibrillar aggregates from different sources. (*a*) The protofilament of Aβ viewed down the long axis of the fibril. Reprinted with permission (177), copyright (2003) American Chemical Society. The segments 12–24 (*red*) and 30–40 (*blue*) are shown. (*b*) The fibril from the C-terminal domain 218–289 of the fungal prion protein HET-s [reproduced with permission (36)]. The ribbon diagram shows the four β-strands (*orange*) (residues 226–234, 237–245, 262–270, and 273–282) and the long loop between β2 and β3 from one molecule. Flanking molecules along the fibril axis (*gray*) are shown. (*c*) Atomic structure of the microcrystals assembled from the GNNQQNY peptide [reproduced with permission (38)]. Each β-strand is a peptide molecule. (*d*) The protofilament from amylin [reprinted with permission from Elsevier (51)]. Green, yellow, and pink β-strands indicate residues 12–17, 22–27, and 31–37, respectively. The unstructured N-terminal tail is shown on the right of the panel along with the disulfide bridge between Cys2 and Cys7. (*e*) The fibril from the NM region of Sup35p [reproduced with permission (52)]. The colored ribbons indicate residues 25–38 (*red*), 39–90 (*blue*), and 91–106 (*green*). The unstructured regions 1–20 (*red dashed lines*) and 158–250 (*black dashed lines*) are shown.

forming another parallel β-sheet ~10 Å away (**Figure 1*b***).

Advances in SSNMR techniques that enable specific internuclear distances and torsion angles to be measured have also allowed the structure of a 11-residue fragment of transthyretin within an amyloid-like fibril to be defined in atomic detail (35, 47). This study shows that the peptide adopts an extended β-strand within the fibrils. Most importantly, however, this pioneering study reveals that the molecules within the fibrils possess a degree of uniformity, even at the level of the side-chain torsion angles, that has previously only been associated with crystalline materials. Because this regularity is reflected in the very narrow resonance lines in the SSNMR spectra, we can anticipate that complete atomic-level structures will soon begin to emerge for a range of systems, transforming our understanding of this facet of the amyloid phenomenon.

## High-Resolution Structural Studies Using X-ray Crystallography

The remarkable achievement of inducing a peptide derived from Sup35p (GNNQQNY) and another with sequence KFFEAAAKKFFE, to form three-dimensional crystals that possess key characteristics of amyloid fibrils, has allowed both the structure of the peptides and the way the molecules could be packed together to be determined with unprecedented resolution (37, 38). In the case of the Sup35p fragment, the crystal consists of pairs of parallel β-sheets in which each individual peptide molecule contributes a single β-strand (**Figure 1*c***). The stacked β-strands are parallel and in register in both sheets. The two sheets interact with each other through the side chains of Asn2, Gln4, and Asn6 to such a degree that water is excluded from the region between them. The remaining side chains on the outer faces of the sheets are hydrated and more distant from the next pair of β-sheets, suggesting that this less intimate interaction could represent a crystal contact rather than a feature of the fibrillar state.

A particularly significant aspect of these structures determined with X-ray or SSNMR is that they are strikingly similar to proposals from cryo-electron microscopy (EM) analysis of the amyloid fibrils formed from an SH3 domain and from insulin, in which the electron density maps were interpreted as arising from pairs of relatively flat untwisted β-sheets (48, 49). Such similarities suggest that many amyloid fibrils could have core structures that have very similar features, which are primarily dictated by the intrinsic conformational preferences of polypeptide chains. The specific nature of the side-chain packing, including such characteristics as the alignment of adjacent strands and the separation of the sheets (50), however, provides an explanation for the occurrence of variations in the details of the structures for specific types of fibril. Hopefully, these pioneering X-ray and SSNMR studies may represent the first entries in a new database of structures similar to the current Protein Data Bank.

## Other Approaches to Defining the Structural Properties of Amyloid Fibrils

As discussed above, SSNMR and X-ray crystallography have recently made major contributions to our knowledge of the structures of amyloid fibrils. Considerable progress in this quest has also come from other approaches, typically involving the combination of data from a number of different biophysical experiments (13, 51, 52). One example is the three-dimensional structure proposed for amyloid fibrils from amylin (**Figure 1*d***) (51). The polypeptide chains were configured within the fibrils on the basis of a cross-β structure, deduced from X-ray diffraction data along with measurements of the protofilament diameter and mass per unit length, determined using TEM and STEM, respectively (53, 54). Additional constraints were provided by evidence of a parallel and in

register arrangement of the β-strands formed by adjacent molecules from SDSL-EPR data (55) and by evidence of the high propensity of various amylin segments to form fibrils when dissected from the rest of the sequence (51). In the resulting model, the N-terminal "tail" (residues 1–11) is unstructured, and residues 12–17, 22–27, and 31–37 form β-strands in a serpentine arrangement, contributing to different β-ribbons in the protofilament (**Figure1*d***).

In another particularly elegant example, detailed structural information on the fibrils, formed from the NM region (residues 1–250) of the yeast prion protein Sup35p, was also obtained by combining a variety of experimental strategies (52). Carefully chosen residues spaced along the fragment of the protein were mutated so as to generate 37 variants, each having a single cysteine residue at a desired position in the molecule. The variants were then labeled with fluorescent probes. The wavelength maximum and total emission intensity of the fluorescent probes were then used to provide information about the degree of burial from solvent of the various residues and about the distances between probes attached to different molecules within the fibrils. Dimeric constructs were also generated for each variant by covalently linking the free thiol group of one molecule to the same group in a second molecule, either directly by a disulfide bridge or by the insertion of a linker. The ability, or lack of ability, of such dimers to form fibrils was used to estimate the distances between corresponding regions of the sequence from adjacent molecules in the fibrils.

Taken together, these complementary sets of data allowed a model to be defined that describes the molecular structure of the filaments (52). In this structure (**Figure 1*e***), two segments of the N domain, corresponding to residues 25–38 and 91–106 (colored green and red in **Figure 1*e***, respectively), interact with the corresponding regions in other molecules to form a "head-to-head" and "tail-to-tail" arrangement. The large central region of the sequence between these two segments (blue in **Figure 1*e***) is folded in such a way that it forms only intramolecular interactions. The C-terminal region of the N domain and the proximal portion of the M domain (residues 107–157) are also structured within the fibrils, whereas the N-terminal region (residues 1–20) and the distal end of the M domain (residues 158–250) appear to be structurally heterogeneous and solvent exposed (dashed lines in **Figure 1*e***).

Finally, although detailed structural models have not yet been proposed, much has been learned about the characteristics of other types of fibrils through similar approaches. This has led, for example, to the identification of regions of the polypeptide chain that are associated with an ordered structure in α-synuclein and tau fibrils using SDSL-EPR (56, 57). It was also possible to determine the most structured regions in α-synuclein as well as in both straight and curly fibrils from β2-microglobulin using hydrogen-deuterium exchange (46, 58, 58a), limited proteolysis (59, 59a), and SSNMR (59b). In addition, from X-ray fiber diffraction studies a cylindrical β-sheet model for fibrils from a poly-Gln peptide and the exon-1 peptide of huntingtin has been proposed (60). The polyglutamine fibrils are of particular interest because of the possibility that the additional array of hydrogen-bonding interactions involving the side chains results in a structure significantly different from that of the classical amyloid fibrils. Evidence that this situation can arise comes from the absence of the 10 Å reflection in the X-ray fiber diffraction patterns of these systems.

## Similarities and Differences in Fibrillar Structures from Various Systems

Comparison of the information about the structural properties of various fibrillar systems, discussed in the previous three paragraphs, allows us to draw a number of tentative conclusions about their similarities and differences. Different fibrils clearly have many properties in common, including the

canonical cross-β structure and the frequent presence of repetitive hydrophobic or polar interactions along the fibrillar axis. The ubiquitous presence of a cross-β structure strongly supports the view that the physicochemical properties of the polypeptide chain are the major determinants of the fibrillar structure in each case. Moreover, several of the proposed structures, despite very different sequences of their component polypeptides, suggest that the core region is composed of two to four sheets that interact closely with each other. An interesting feature of these sheets is that they appear to be much less twisted than expected from the analysis of the short arrays of β-strands that form β-sheets in globular protein structures. This feature was first proposed from cryo-EM and has been supported by Fourier transform infrared (FTIR) analyses (48, 61).

Nevertheless, it is clear that there are significant differences in detail attributable to the influence of the side chains on the structures adopted by the various systems. These appear to include the lengths of the β-strands and whether they are arranged in a parallel or antiparallel arrangement within each sheet; the lengths and conformational properties of the loops, turns, and other regions that are not included within the core structure; and the number of β-sheets in the protofilament. It is clear that the fraction of the residues of a polypeptide chain that are incorporated in the core structure can vary substantially (e.g., from all the residues of the 7-mer peptide to only about 13% of the residues in the full-length HET-s) and that the exact spacing between the β-sheets varies with factors such as the steric bulk of the side chains that are packed together in the core (50). In addition, the presence of disulfide bonds in proteins such as insulin may perturb the way in which the sheets can stack together (49). In cases such as the polyglutamine sequences, other interactions between the side chains may generate larger perturbations of the structure to generate such motifs as β-helices (60), which are also seen under similar circumstances in the structures of globular proteins.

The structure that will normally be adopted in the fibrils will be the lowest in free energy and/or the most kinetically accessible. What is clear, therefore, is that the interactions of the various side chains with each other and with solvent are crucial in determining the variations in the fibrillar architecture even though the main-chain interactions determine the overall framework within which these variations can occur. In other words, the interactions and conditions (see below) involving the side chains in a given sequence can tip the balance between the alternative "variations on a common theme" arrangements of a polypeptide "polymer" chain in its fibrillar structure. Such a situation contrasts with that pertaining to the native structures of the highly selected protein molecules, which are able to fold to unique structures that are significantly more stable for a given sequence than any alternatives.

**FTIR:** Fourier transform infrared

## The Polymorphism of Amyloid Fibrils

Even before the molecular structures of amyloid fibrils began to emerge, it was clear that significant morphological variation can exist between different fibrils formed from the same peptide or protein (12, 48, 49, 54). Evidence is now accumulating that such variations in morphology is linked to heterogeneity in molecular structure, i.e., in the structural positioning of the polypeptide chains within the fibrils. One example of such heterogeneity involves the peptide hormone glucagon, wherein fibrils formed at different temperatures (25°C or 50°C) are morphologically distinct; measurements of CD and FTIR spectroscopy reveal differences in the secondary structure adopted by the constituent peptide molecules (14). A particularly important study in this regard addresses the origin of the marked differences in the morphology of $A\beta_{1-40}$ fibrils that can be observed

in TEM studies of samples prepared under agitation or quiescent conditions; differences in the SSNMR spectra recorded from the different preparations provide clear evidence that this polymorphism is linked to differences in molecular structure (62).

Another example of conformational variability involves fibrils formed from the yeast prion protein Ure2p, where two independent studies came to somewhat different conclusions about the fibril structure. Both studies find that the globular C-terminal domain maintains a largely native-like structure. However, in one case, it appears that the fibrils possess a cross-β core involving only the N-terminal domains, each arranged in a serpentine fashion and forming a series of consecutive strands and loops (63–65). The parallel and in register stacking of serpentines from different molecules then forms the cross-β core with the C-terminal globular units decorating it (64). In the other study, the C- and N-terminal domains of the protein appear to interact with each other, and these fibrils do not have the characteristic 4.7-Å reflection typical of a cross-β structure (17, 66, 67). These apparently conflicting reports are likely to reflect structural differences in the fibrils, probably caused by the slightly different conditions used to prepare them.

Conformational polymorphism has also been found in other yeast prion proteins and is of particular significance because of the light it sheds on the existence of "strains" of mammalian prions and on the nature of the crucial barriers to infectivity that limit transmissibility between species (68). Efficiency of interspecies prion transmission decreases as the sequences of the infectious prions diverge, probably because each prion sequence can give rise to a limited number of misfolded conformations, which have low cross-seeding efficiency. However, a strain conformation of Sup35p has recently been identified that allows transmission from *S. cerevisiae* to the highly divergent *Candida albicans* (68). Similarly, mammalian $PrP_{23-144}$ fibrils from different species vary in morphology and secondary structure, and these differences appear to be controlled by one or two residues in a critical region of the polypeptide sequence (69).

In all of these cases, preformed seeds can propagate their morphology and structure as well as overcome sequence- or condition-based structural preferences, resulting in fibrils that inherit the characteristics of the template (14, 62, 68, 69). These results show that each protein sequence can form a spectrum of structurally distinct fibrillar aggregates and that kinetic factors can dictate which of these alternatives is dominant under given circumstances. Of the many possible conformations that could be present in the amyloid core for a given protein, the specific ones that play this role will depend simply on the thermodynamic and, in many cases, the kinetic factors that are dominant under those circumstances. By contrast, natural globular proteins have been selected by evolution to fold into one specific three-dimensional structure, and the complex free-energy landscapes associated with their sequences have a single and well-defined minimum, under physiological conditions, corresponding to the native state.

## MECHANISMS OF AMYLOID FIBRIL FORMATION

The full elucidation of the aggregation process of a protein requires the identification of all the conformational states and oligomeric structures adopted by the polypeptide chain during the process and the determination of the thermodynamics and kinetics of all the conformational changes that link these different species. It also implies characterizing each of the transitions in molecular detail and identifying the residues or regions of the sequence that promote the various aggregation steps. The identification and characterization of oligomers preceding the formation of well-defined fibrils is of particular interest because of an increasing awareness that these species are likely to play a critical role in the pathogenesis of protein deposition diseases.

## Amyloid Formation Occurs via a Nucleated Growth Mechanism

It is widely established that amyloid fibril formation has many characteristics of a "nucleated growth" mechanism. The time course of the conversion of a peptide or protein into its fibrillar form (measured by ThT fluorescence, light scattering, or other techniques) typically includes a lag phase that is followed by a rapid exponential growth phase (70–73). The lag phase is assumed to be the time required for "nuclei" to form. Once a nucleus is formed, fibril growth is thought to proceed rapidly by further association of either monomers or oligomers with the nucleus.

Such a nucleated growth mechanism has been well studied both experimentally and theoretically in many other contexts, most notably for the process of crystallization of both large and small molecules (74). As with many other processes dependent on a nucleation step, including crystallization, addition of preformed fibrillar species to a sample of a protein under aggregation conditions ("seeding") causes the lag phase to be shortened and ultimately abolished when the rate of the aggregation process is no longer limited by the need for nucleation (70, 71). It has been shown also that changes in experimental conditions, or certain types of mutations, can also reduce or eliminate the length of the lag phase, again assumed to result from a situation wherein nucleation is no longer rate limiting (72, 73, 75). The absence of a lag phase, therefore, does not necessarily imply that a nucleated growth mechanism is not operating, but it may simply be that the time required for fibril growth is sufficiently slow relative to the nucleation process and that the latter is no longer the slowest step in the conversion of a soluble protein into the amyloid state. Although fibrils do not appear to a significant extent during the lag phase, it is increasingly clear that this stage in fibril formation is an important event in which a variety of oligomers form, including β-sheet-rich species that provide nuclei for the formation of mature fibrils.

The efficiency of preformed fibrils to promote further aggregation through a seeding mechanism decreases dramatically as the sequences diverge (68, 76, 76a). Using a number of immunoglobulin domains sharing different degrees of sequence identity, it was shown that coaggregation between different types of domain is not detectable if the sequence identity is lower than ~30% to 40% (76). A bioinformatics analysis of consecutive homologous domains in large multimodular proteins shows that such domains almost exclusively have sequence identities of less than 40%, suggesting that such low sequence identities could play a crucial role in safeguarding proteins against aggregation (76).

**Oligomers:** clusters of small numbers of protein or peptide molecules without a fibrillar appearance

**Protofibrils:** protein aggregates of isolated or clustered spherical beads 2–5 nm in diameter with β-sheet structure

## Oligomers Preceding Amyloid Fibril Formation: Structured Protofibrils

The past decade has seen very substantial efforts directed toward identifying, isolating, and characterizing the oligomeric species that are present in solution prior to the appearance of fibrils, both because of their likely role in the mechanism of fibril formation and because of their implication as the toxic species involved in neurodegenerative disorders. We focus initially on amyloid formation by the Aβ peptide because this has been widely studied owing to its links with Alzheimer's disease. Aggregation of this peptide is preceded by the formation of a series of metastable, nonfibrillar species that can be visualized using AFM and TEM (33, 77–79). Some appear to be spherical beads of 2–5 nm in diameter. Others appear to be beaded chains with the individual beads again having a diameter of 2–5 nm and seeming to assemble in linear and curly chains. Yet others appear as annular structures, apparently formed by the circularization of the beaded chains. All of these aggregates, which have been termed protofibrils by the authors who first observed them (33, 77–79), should not be confused with the

protofilaments that are the constituent units of mature fibrils. Protofibrils from Aβ can bind CR and ThT (79), contain an extensive β-sheet structure (79), and, in the form of the smaller spherical species, are made up of ~20 molecules (80). A first exciting attempt to determine the structure of Aβ protofibrils was published using proline-scanning mutagenesis (81).

Analogous spherical and chain-like protofibrillar structures have been observed for many other systems, including α-synuclein (82), amylin (80), the immunoglobulin light chain (83), transthyretin (84), polyQ-containing proteins (80), β2-microglobulin (85), equine lysozyme (86), the *Sulfolobus solfataricus* acylphosphatase (Sso AcP) (87), and an SH3 domain (87a). These species are generally characterized by extensive β-structure and sufficient structural regularity to bind ThT and CR. The exciting finding that a specific antibody can bind to protofibrillar species from different sources, but not to their corresponding monomeric or fibrillar states, suggests that such soluble amyloid oligomers have some important common structural elements (88).

Data have been reported showing that in some cases protofibrils can be on-pathway to fibrils (33, 71). In other cases, they appear to be off-pathway (85, 89). It has been reported that the transition from the protofibrillar to the fibrillar state of the peptide 109–122 of the Syrian hamster prion protein occurs concomitantly with the alignment of β-strands within sheets in which the strands are initially misaligned (89a). Such an aligment involves detachment and re-annealing of the strands, but may also occur through an internal structural reorganization within the sheets, depending on conditions (89b). Regardless of the precise role played by protofibrils in the overall process of fibril formation, the elucidation of their mechanism of formation and of their structures is extremely important, not least because these species could be the primary toxic agents involved in neurodegenerative disorders.

## Oligomers Preceding Fibril and Protofibril Formation: Unstructured Aggregates

Following the isolation and characterization of protofibrils, studies based on photo-induced cross-linking of unmodified proteins (PICUP) began to identify other oligomeric species that appeared to precede their formation (90, 91). Both the 40 and 42 residue forms of Aβ have been shown to exist as soluble oligomers in rapid equilibrium with the corresponding monomeric forms. These oligomers appear to be composed of 2–4 and 5–6 molecules for $A\beta_{1-40}$ and $A\beta_{1-42}$, respectively, and CD measurements suggest that they are relatively disorganized (91). Interest in these low-molecular-weight oligomers has been particularly intense as species of this type have also been detected in the brains of Alzheimer's disease patients (92) and in the lysates and conditioned media of cultured cells expressing the amyloid β protein precursor (93, 94).

The NM region of the yeast prion Sup35p has been shown to form "structurally fluid" oligomers rapidly, and these oligomers only later convert to species with extensive β-structures that are capable of nucleating fibril formation (71). Such a conversion has been found to be facilitated by the covalent dimerization of NM molecules when residues in the "head" region of N (residues 25–38) are cross-linked (52). Moreover, if the fluid oligomers are maintained under oxidizing conditions, intermolecular disulfide bridges are found to form more easily for variants in which cysteine residues are introduced into the head region of N rather than elsewhere. These results indicate that the interaction of the head regions of two N molecules nucleates the formation of an amyloid-like structure within the aggregates (52).

Similar behavior has been observed for the aggregation of denatured yeast phosphoglycerate kinase at low pH using dynamic light scattering and far-UV CD spectroscopy (95). β-sheet structure is increasingly stabilized as

the aggregates grow in size. When a critical mass is reached, the oligomers associate with each other to form short, curly protofibrils that are similar in appearance to those observed with Aβ and α-synuclein (95). Moreover, unfolding of the SH3 domain from the bovine phosphatidylinositol 3′ kinase at pH 3.6 results in the rapid formation of a broad distribution of unstructured oligomers that subsequently convert into thin, curly, ThT-binding protofibrils (87a). All these experimental results, along with computer simulations carried out using simple polyalanine peptides (96), suggest that structured protofibrillar species can form from the reorganization or assembly of small and relatively disorganized oligomers that are formed rapidly after the initiation of the aggregation process.

## Aggregation of Globular Proteins Can Occur via Partial Unfolding

So far we have discussed systems that are largely unstructured prior to the aggregation process. It is generally believed that globular proteins need to unfold, at least partially, to aggregate into amyloid fibrils (21, 97, 98). Evidence supporting this hypothesis comes from a large body of experimental data. It is clear, for example, that globular proteins have an increased propensity to aggregate under conditions that promote their partial unfolding, such as high temperature, high pressure, low pH, or moderate concentrations of organic solvents (85, 99–102). In addition, for some familial forms of disease in which the proteins involved in aggregation normally adopt folded conformations (see **Table 1**), there is clear evidence that a destabilization of the native structure, resulting in an increase in the population of nonnative states, is the primary mechanism through which natural mutations mediate their pathogenicity (103–105).

A strong correlation between a decreased conformational stability of the native state and an increased propensity to aggregate into amyloid-like structures has also been shown in vitro for nondisease-associated proteins (100, 106). Remarkably, aggregation of human lysozyme and HypF-N can be initiated by a population of less than 1% of a partially folded state that is in equilibrium with the native conformation (104, 107). Conversely, the binding of ligands and other species, such as antibodies, that stabilize the native state can decrease dramatically the propensity of proteins to aggregate (108–111). Such observations have inspired an extensive search of potential pharmaceutical compounds for the treatment of the diseases associated with transthyretin through specific binding to the tetrameric native state of the protein (109).

## Aggregation of Globular Proteins Can Occur via Formation of Native-Like Oligomers

Although the "conformational change hypothesis" is undoubtedly the most appropriate way to describe the formation of amyloid fibrils by many globular proteins, recent observations have suggested that in some cases the major conformational change associated with amyloid aggregation may not take place until after the initial aggregation step. Formation of amyloid fibrils by insulin at low pH, for example, is preceded by an oligomerization step in which a native-like content of α-helical structure is almost completely retained, and aggregates with a morphology reminiscent of amyloid protofibrils and with a high content of β-structure appear only later in the process (112). In addition, within a group of variants of the protein S6 from *Thermus thermophilus*, no significant correlation was found between the rate of fibril formation under conditions in which a quasi-native state was populated prior to aggregation and the unfolding rate or conformational stability (73). Similarly, the native state of the pathogenic variant of ataxin-3, the protein associated with spinocerebellar ataxia type-3, does not appear to be significantly destabilized, leading to the proposal that the pathway for fibril formation can be distinct from that of unfolding (113).

Details of the manner in which aggregation under these conditions can take place has come from studies of the aggregation of Sso AcP. These studies have shown that unfolding of the protein can be two orders of magnitude slower than the formation of amyloid protofibrils when the protein is placed under conditions in which the native state is thermodynamically more stable than the dominant partially unfolded state (87). The first event in the aggregation of Sso AcP under these conditions is the formation of oligomers that do not bind to ThT or CR and, remarkably, not only have a native-like topology but also retain enzymatic activity (114). These native-like oligomers then undergo structural reorganization to form amyloid protofibrils that have extensive β-structure, bind ThT and CR, but are not enzymatically active. The fact that protofibril formation is also faster than the rate of disaggregation of the initially formed oligomers shows that dissolution of the latter followed by renucleation cannot be the dominant process giving rise to the structural conversion.

In the case of Ure2p, a mechanism of the type observed for Sso AcP appears to give rise to a situation wherein a native-like conformation is even retained in the fibrils themselves under some conditions (17, 66, 67). The significant propensity of native or native-like structures to aggregate is not surprising if we consider that there is a multitude of conformers even in the native ensemble of a globular protein (115). Some of these conformers will be only transiently populated but could be significant for aggregation just as they are for the hydrogen exchange of their main-chain amide groups.

Finally in this section, despite their apparent differences, there are in fact substantial similarities between the fundamental mechanism of aggregation described here for folded proteins and that of natively unfolded systems, such as Aβ and Sup35p NM. In both cases, the polypeptide molecules assemble first into species that can have characteristics far from those of the final aggregates but similar to those of the precursor structures, whether natively unfolded or natively folded. The initial aggregates then transform into species that are not yet fibrillar in their morphologies but have other properties characteristic of amyloid-like structures, notably β-sheet structure and binding to CR and ThT. Clearly, fully or partially unfolded states of globular proteins are generally more susceptible to aggregation than the native states. Nevertheless, in some situations, particularly those close to physiological, the much higher populations of the latter can result in their playing an important role in initiating an aggregation process that could be significant on the very slow timescales of the amyloid disorders.

## A Multitude of Conformational States Is Accessible to Polypeptide Chains

The differing features of the aggregation processes, described in the previous paragraphs, reveal that polypeptide chains can adopt a multitude of conformational states and interconvert between them on a wide range of timescales. The network of equilibria, which link some of the most important of such states both inside and outside the cell, is schematically illustrated in **Figure 2**. Following biosynthesis on a ribosome, a polypeptide chain is initially unfolded. It can then populate a wide distribution of conformations, each of which contains little persistent structure, as in the case of natively unfolded proteins, or fold to a unique compact structure, often through one or more partly folded intermediates. In such a conformational state, the protein can remain as a monomer or associate to form oligomers or higher aggregates, some of which are functional with characteristics far from those of amyloid structures, such as in actin, myosin, and microtubules. Sooner or later, the vast majority of proteins will be degraded, usually under very carefully

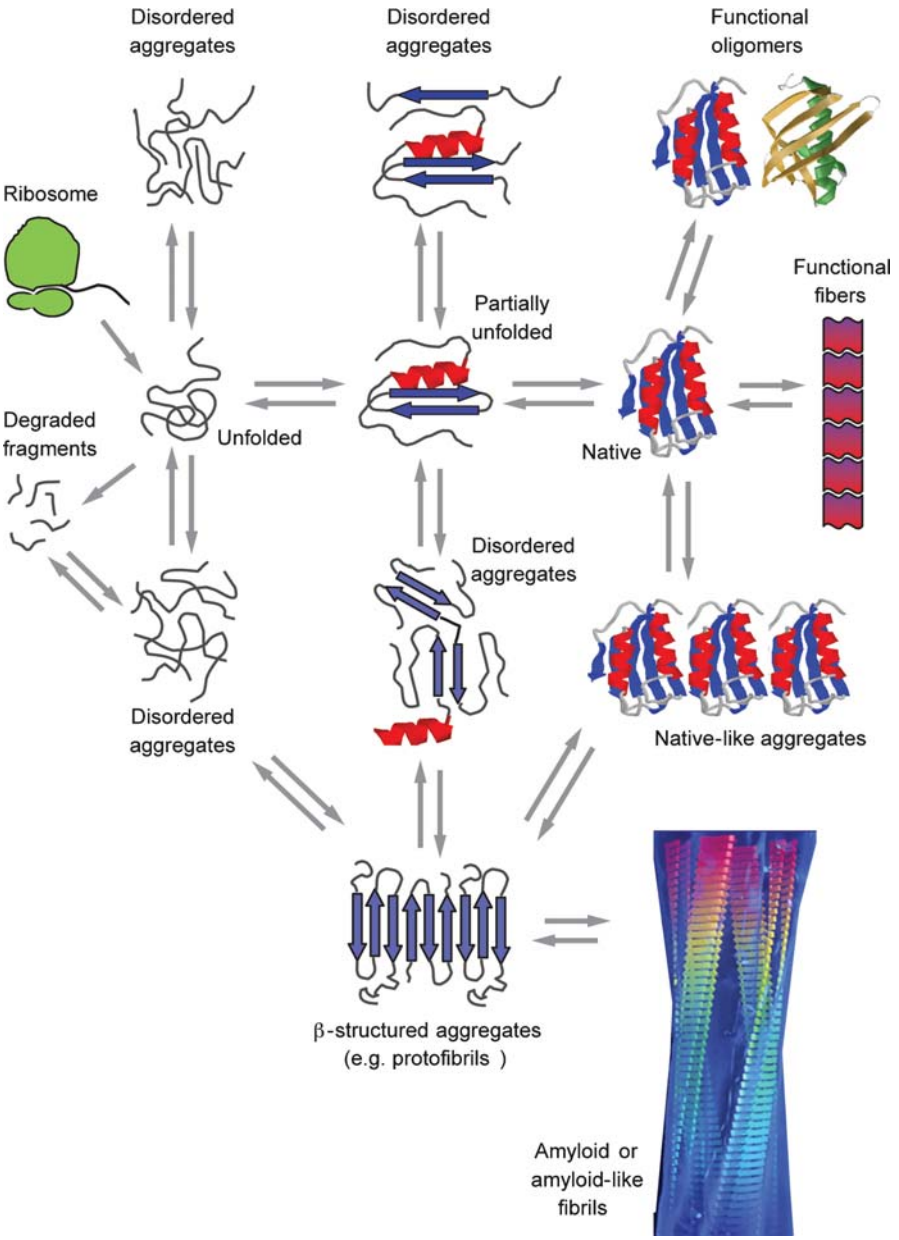


**Figure 2**

A schematic representation of some of the many conformational states that can be adopted by polypeptide chains and of the means by which they can be interconverted. The transition from β-structured aggregates to amyloid fibrils can occur by addition of either monomers or protofibrils (depending on protein) to preformed β-aggregates. All of these different conformational states and their interconversions are carefully regulated in the biological environment, much as enzymes regulate all the chemistry in cells, by using machinery such as molecular chaperones, degradatory systems, and quality control processes. Many of the various states of proteins are utilized functionally by biology, including unfolded proteins and amyloid fibrils, but conformational diseases will occur when such regulatory systems fail, just as metabolic diseases occur when the regulation of chemical processes becomes impaired.

controlled conditions and as a part of normal biochemical processes, with their amino acids often being recycled.

This description of normal functional behavior, honed by millions of years of evolution, is, however, only part of the story. Fully or partially unfolded ensembles on the pathways to their functional states (or generated as the result of stress, chemical modification, or genetic mutation) are particularly vulnerable to aggregation (**Figure 2**). Peptides and proteins that are natively unfolded, as well as fragments of proteins generated by proteolysis and unable to fold in the absence of the remainder of the polypeptide chain, can also aggregate under some circumstances, for example, if their concentrations become elevated. Some of the initial amorphous aggregates simply dissociate again, but others may reorganize to form oligomers with the germ of an amyloid structure, including the spherical, chain-like, and annular amyloid protofibrils observed for many systems. In order to generate long-range order in such structures, a critical number of molecules must be present such that the favorably enthaplic terms associated with their regular stacking can most effectively offset the accompanying loss of configurational entropy.

The structured polypeptide aggregates can then sometimes grow into mature fibrils by further self-association or through the repetitive addition of monomers. Proteins that adopt a folded structure under physiological conditions can also aggregate under some circumstances. This latter type of protein can either unfold, fully or partially, and aggregate through the mechanism described above or they can oligomerize prior to such a substantial conformational change. In the latter process, a structural reorganization to give amyloid-like assemblies occurs later and may in some cases be promoted by the existence of intermolecular contacts within native-like aggregates.

Every state of a polypeptide molecule, except the unique native state of globular proteins wherein the side chains pack together in a unique manner, is a broad ensemble of often diverse conformations. It is not surprising, therefore, that even the fibrillar end products of aggregation processes are characterized by morphological and structural diversity, representing variations on a common theme. Under most conditions in living systems, misfolding and aggregation of proteins are intrinsic side effects of the conformational transitions essential to the functioning of the organism. Formation of aggregates is normally inhibited by molecular chaperones and degradation processes as well as being disfavored by the amino acid sequences that are carefully selected by evolution to inhibit aggregation. But under some circumstances, as we discuss below, these aggregation processes can escape from the host of natural defenses and then give rise to pathogenic behavior.

## THE INFLUENCE OF SEQUENCE ON AMYLOID FORMATION

We have stressed that amyloid formation results primarily from the properties of the polypeptide chain that are common to all peptides and proteins. We have seen, however, that the sequence influences the relative stabilities of all the conformational states accessible to a given molecule, most notably the native state, and will thereby contribute to the susceptibility of a given polypeptide chain to convert into amyloid fibrils. Moreover, it is clear that polypeptide chains with different sequences can form amyloid fibrils at very different rates, even when these processes occur from fully or partially unfolded states. We start the exploration of this topic with a description of the determinants of the aggregation of those unfolded polypeptide chains that can broadly be described as unstructured, i.e., having no significant elements of persistent or cooperative structure. By considering these systems, we can examine how the properties of the sequence influence its intrinsic aggregation behavior rather than affect the stability of a given protein fold.

## Hydrophobicity, Charge, and Secondary Structure Propensities Strongly Influence Amyloid Formation

One important determinant of the aggregation of an unfolded polypeptide chain is the hydrophobicity of the side chains. Amino acid substitutions within regions of the sequence that play a crucial role (e.g., if they are in the region that nucleates aggregation) in the behavior of the whole sequence can reduce (or increase) the aggregation propensity of a sequence when they decrease (or increase) the hydrophobicity at the site of mutation (116–118). Moreover, there is evidence that protein sequences have evolved to avoid clusters of hydrophobic residues; for example, groups of three or more consecutive hydrophobic residues are less frequent in natural protein sequences than would be expected in the absence of evolutionary selection (119).

Another property likely to be a key factor in protein aggregation is charge, as a high net charge either globally or locally may hinder self-association (120, 121). For example, the effects of single amino acid substitutions were investigated on the propensity of AcP denatured in trifluoroethanol to aggregate (120). Although mutations decreasing the positive net charge of the protein resulted in an accelerated formation of β-sheet containing aggregates able to bind CR and ThT, mutations increasing the net charge resulted in the opposite effect. Further indications of the importance of charge in protein aggregation come from observations that aggregation of polypeptide chains can be facilitated by interactions with macromolecules, which exhibit a high compensatory charge (50, 122–125).

Comparison of large data sets of natively unfolded and natively folded proteins has shown that the former have a lower content of hydrophobic residues and a higher net charge than the latter (126). These properties undoubtedly contribute to maintaining the aggregation propensity of natively unfolded proteins sufficiently low to avoid the formation of aggregates under normal physiological conditions despite the fact that all, or at least the very large majority, of the side chains are accessible for intermolecular interactions.

In addition to charge and hydrophobicity, a low propensity to form α-helical structure and a high propensity to form β-sheet structure are also likely to be important factors encouraging amyloid formation (45, 50, 101, 117, 127, 128). Patterns of alternating hydrophilic and hydrophobic residues have been shown to be less frequent in natural proteins than expected on a random basis, suggesting that evolutionary selection has reduced the probability of such sequence patterns that favor β-sheet formation (127). Furthermore, it has been suggested that the high conservation of proline residues in a fibronectin type III superfamily and of glycine residues in AcPs can be rationalized on the grounds that such residues have a low propensity to form β-structure and hence inhibit aggregation (129, 130).

## The Amino Acid Sequence Affects Fibril Structure and Aggregation Rate

The demonstration that the various physicochemical factors described in the previous paragraph are important determinants of the formation of amyloid structure by unfolded polypeptide chains has proved to be of great value in understanding the mechanism of aggregation at a molecular level. For example, changes in the rate of aggregation of unfolded AcP following a series of mutations were used to generate a phenomenological equation, based on physicochemical principles, that is able to rationalize these rates in a robust manner (131). This expression was, remarkably, found to rationalize just as well similar data for a whole series of other unstructured peptides and proteins (**Figure 3*a***). This finding also provides compelling evidence for the close similarity of the principles underlying the aggregation behavior of different polypeptide molecules.

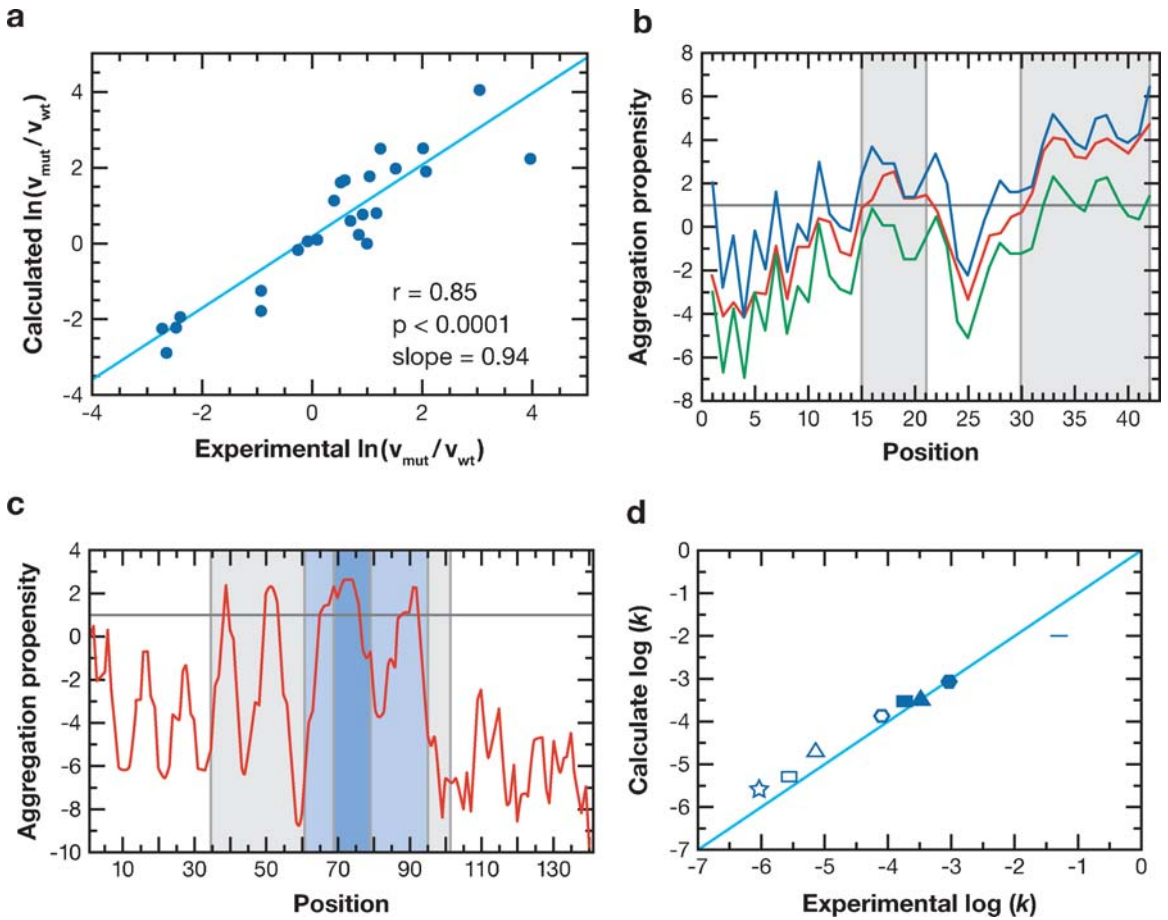


**Figure 3**

(*a*) Predicted versus experimental changes of aggregation rate following mutation. The mutations are in a variety of different unstructured polypeptide chains, such as Aβ, α-synuclein, amylin, and tau. Reproduced with permission (131). (*b*) Aggregation propensity profile (*red line*) for Aβ. The gray areas indicate regions of the sequence found experimentally to form and stabilize the amyloid cross-β core structure. A horizontal line at a propensity of 1 is drawn to highlight the aggregation-promoting regions that have values above this line. (*c*) Aggregation propensity profile (*red line*) for α-synuclein. The large region of the protein thought to be structured in the fibrils (*pale gray*) (56) is shown and includes all the peaks in the profile. The highly amyloidogenic NAC region (*light blue*) (178) and the 69–79 region (*dark blue*), found to be a particularly amyloidogenic segment within the NAC region (179) and containing the most prominent peak in the profile, are shown. The figures shown in panels *b* and *c* are reprinted with permission from Elsevier (134). (*d*) Predicted versus experimental aggregation rates (*k*) for a number of unstructured systems. The data points refer to $A\beta_{1-40}$ (Θ), ABri (□), denatured transthyretin (Δ), amylin (○), AChE (■), unfolded PrP (σ), unfolded human AcP (λ), and unfolded *E. coli* HypF-N (–). The straight line has a slope of 1 and indicates the ideal correlation between theory and experiment. Reprinted with permission from Elsevier (136).

A related approach, which considers additional factors such as the changes in the number of aromatic side chains, exposed surface area, and dipole moment upon mutation, has also been shown to predict the effect of mutations on the aggregation rates of a wide variety of polypeptide chains (132). The success of these rather straightforward

relationships between the rates of aggregation of unfolded polypeptide chains and simple physicochemical factors is strong support for the idea that such aggregation reflects the situation wherein a polypeptide chain behaves as a simple "polymer." Such behavior contrasts with the process of folding a globular protein for which the rates of folding are closely coupled to the specific structures of the highly evolved native states associated with individual sequences.

As described above, only a fraction of the residues of even the most highly amyloidogenic proteins are found in the core structure of the fibrils. In addition, conservative mutations have an effect on amyloid formation only when they are located in specific regions of the sequence (45, 117). Our increasing knowledge of such effects is beginning to lead to an understanding of the factors that cause specific segments of the sequence, rather than others, to form the characteristic cross-β structure. By extension of the analysis that led to the recognition of the links between aggregation propensities and the physicochemical characteristics of the constituent amino acid residues of a polypeptide chain, new algorithms have been developed to identify the regions of the sequence that are likely to promote aggregation within an unstructured polypeptide chain (133, 134). The outcome of both these approaches is a plot of aggregation propensity as a function of residue number, similar to the hydropathy profiles introduced to predict the regions of sequences that span the lipid bilayer in membrane proteins (135). The success of this type of approach is illustrated particularly well by the very good agreement between the regions of the sequence predicted to promote the aggregation of the Aβ peptide and α-synuclein and the regions found experimentally to form and stabilize the fibril core and/or to play a primary role in fibril formation (**Figure 3*b*,*c***) (134).

In a similar type of approach, multiple regression analysis has generated an equation that includes in the aggregation predictions the effects of extrinsic factors, such as the concentration of protein as well as the pH and ionic strength of the solution in which it is located, in addition to the intrinsic factors associated with the amino acid sequence (136). This equation reproduces the experimentally determined aggregation rates, which span five orders of magnitude, from the unstructured states of a set of nonhomologous protein sequences (**Figure 3*d***). In an exciting development, two reports have appeared recently that point to the validity of these concepts for in vivo situations, at least in bacteria (137, 138). The expenditure of effort and ingenuity in devising new methods to make quantitative analyses of such aggregation behavior in living organisms is therefore likely to bring rich rewards.

## Unfolded Regions Play Critical Roles in Promoting the Aggregation of Partially Folded States

Although the key regions of the sequence that promote fibril formation by an unfolded peptide or protein can now be broadly identified using relatively simple physicochemical parameters, the aggregation of polypeptides that contain significant levels of persistent secondary structure and long-range interactions will be influenced by additional factors. For example, the yeast prion proteins Sup35p, Ure2p, and HET-s all have unstructured and globular domains in their soluble states. In each of these three cases, the region forming the cross-β core and responsible for the prion properties has been found to be the domain that is unstructured in the soluble form of the protein (63, 139–141).

Further insights into this issue have again come from studies of proteins other than those found to form functional or pathogenic amyloid structures in vivo. An approach using limited proteolysis has shown that in the partially unfolded state adopted by AcP in the presence of moderate concentrations of trifluoroethanol, the regions of the sequence found to promote amyloid aggregation are flexible and/or solvent exposed in addition to having

an intrinsically high propensity to aggregate (142). Other regions that are not involved in aggregation despite having high propensities to aggregate were found to be at least partially buried in residual structure, whereas other solvent-exposed regions not involved in the aggregation process possess a low propensity to form amyloid fibrils. Similarly, the ease with which apomyoglobin converts to fibrils under different solution conditions correlates with the degree of denaturation, suggesting that fibrils assemble by association of unfolded polypeptide segments rather than by the docking of preformed structured elements (143).

Even α-synuclein, the protein associated with Parkinson's disease and assumed to have no significant structural preferences, has recently been shown to possess some significant long-range interactions between the negatively charged C-terminal region and the central amyloidogenic NAC region (144, 145). Structural perturbations that destabilize the interactions between these two portions of the protein molecule appear to increase the exposure of the amyloidogenic NAC region. Such perturbations include the presence of positively charged ions able to interact with the C terminus, a decrease in pH that reduces the net charge of the C-terminal region, and deletion of the C terminus; all result in a more rapid aggregation reaction (125, 146). Although it is clear that partial neutralization of the negative charge of α-synuclein will stimulate aggregation on a purely electrostatic argument, pairs of variants with a similar net charge but opposite signs (for example +3 and –3) aggregate more rapidly when the NAC region is unprotected (146).

## Variations in Fibrillar Structure Can Be Reconciled by Common Determinants of the Aggregation Process

Our ability to rationalize, and particularly to predict, important features of the process of amyloid assembly emphasizes in a dramatic manner that common traits are dominant in the aggregation behavior of different peptides and proteins. Although the structural analysis of fibrils at the level of specific residues (described above) highlights differences in the details of the manner in which individual molecules are incorporated into the fibrils, the fact remains that the generic cross-β structure and the frequent presence of stabilizing rows of hydrophobic interactions that run along the fibril axis (apart from important exceptions such as the fibrillar species associated with polyQ traits where additional side-chain hydrogen-bonding interactions are undoubtedly important) indicate the presence of common features in the aggregation of polypeptide chains. This commonality explains our ability to predict, often with a high degree of success, the regions involved in the formation of the amyloid core and the effect of mutations in this process.

Unlike the extreme dependence of the evolved native fold on protein sequence, it is unlikely that a single arrangement of a given chain in the amyloid core structure provides unique stability relative to all other arrangements. As noted above, this conclusion also means that the specific regions of a sequence found in such structures can vary with solution conditions, that there can be subtle differences in the manner in which a given polypeptide sequence is arranged in a cross-β core structure even under essentially identical conditions, and that the details of the resulting structures may be determined by kinetic rather than thermodynamic factors. This lack of a single unique structure, coupled with the extremely high degree of repetitive order within individual fibrils, may be the origin of the strain phenomena observed in both yeast and mammalian prions. Another important facet of this topic is that chemical modifications, for example those induced by physiologically formed metabolites (147, 148), or interactions with small molecules or metal ions (149) may play a much more important role in the aggregation process than might be imagined, e.g., by perturbing the thermodynamics of kinetics

sufficiently to alter the details of the resulting amyloid structure.

## THE PATHOGENESIS OF PROTEIN DEPOSITION DISEASES

The presence of highly organized and stable fibrillar deposits in the organs of patients suffering from protein deposition diseases led initially to the reasonable postulate that this material is the causative agent of the various disorders. This view was later reinforced by a number of observations; for example, amyloid fibrils formed from the Aβ peptide were found to be toxic to cultured neuronal cells (150, 151) and to cause both membrane depolarization and alterations in the frequency of their action potentials (152). Moreover, Aβ fibrils were shown to cause neuronal loss and microglial activation when injected into the cerebral cortex of aged rhesus monkeys (153). However, more recent findings have raised the possibility that precursors to amyloid fibrils, such as low-molecular-weight oligomers and/or structured protofibrils, are the real pathogenic species, at least in neuropathic diseases. Here we describe some of the most compelling evidence supporting this view, starting again from the well documented Aβ case.

### The Search Is on for the Causative Agents of Protein Aggregation Diseases

The severity of cognitive impairment in Alzheimer's disease correlates with the levels of low-molecular-weight species of Aβ, including small oligomers, rather than with the amyloid burden (154–156). In addition, transgenic mice show deficits in cognitive impairment, cell function, and synaptic plasticity well before the accumulation of significant quantities of amyloid plaques (157, 158). Similarly, phenotypic changes reminiscent of Alzheimer's disease precede amyloid plaque formation, or occur in their absence, in transgenic *Drosophila* expressing $A\beta_{1-42}$ and $A\beta_{1-40}$ (159, 160).

Further evidence comes from the finding that a single injection of a monoclonal anti-Aβ antibody does not reduce amyloid deposits in the brains of transgenic mice expressing $A\beta_{1-42}$, but it does reverse the associated memory loss, perhaps as a result of enhanced peripheral clearance and/or sequestration of soluble forms of the Aβ peptide (161). Genetic evidence also supports the theory that the precursor aggregates, as opposed to mature fibrils, are the pathogenic species: The aggressive "Arctic" (E693G) mutation of the amyloid β precursor protein, associated with a heritable early-onset manifestation of Alzheimer's disease, has been found in vitro to enhance protofibril, but not fibril, formation (162).

A similar scenario concerning the toxicity of early aggregates also holds for Parkinson's disease, a neurodegenerative condition associated with the formation of intracellular fibrillar deposits, notably Lewy bodies, in the dopaminergic neurons of the *substantia nigra*. In this disease, those dopaminergic neurons that survive, whether or not they contain Lewy bodies, show no quantifiable differences in viability (163, 164). Furthermore, mutations associated with juvenile Parkinson's disease or early-onset forms of Parkinsonism give rise to early neuronal degeneration in the absence of the accumulation of Lewy bodies (165). Overexpression of α-synuclein in transgenic flies or rats does not result in neuronal loss concomitant with the formation of detectable intracellular deposits (166, 167). By contrast, transgenic mice with nonfibrillar deposits of α-synuclein in various regions of the brain are characterized by substantial motor deficiencies and losses of dopaminergic neurons (168).

It is increasingly evident that prefibrillar aggregates from peptides and proteins other than Aβ and α-synuclein can either be toxic to cells or perturb their function. Early, nonfibrillar aggregates of transthyretin have been found toxic to neuronal cells under conditions

in which the native tetramer and the mature fibrils are not (169). Consistent with this finding, symptoms of familial amyloid polyneuropathy appear when transthyretin is deposited in an aggregated but nonfibrillar form that does not stain with CR (169). Reixach and coworkers (170) have found that such toxicity originates from low-molecular-weight oligomers of transthyretin of up to ~100 kDa in size.

The likelihood that such behavior is much more general is suggested by the finding that prefibrillar forms of the nondisease-related HypF-N from *E. coli*, the SH3 domain from bovine phosphatidylinositol 3′ kinase, lysozyme from horse, and apomyoglobin from sperm whale are also highly toxic to cultured fibroblasts and neurons, whereas the monomeric native states and the amyloid-like fibrils (all formed in vitro) displayed very little, if any, toxicity (171, 171a, 171b). Interestingly in this context, the most highly infective form of the mammalian prion protein has been identified as an oligomer of about 20 molecules, indicating that such small aggregates are the most effective initiators of transmissible spongiform encephalopathies (172).

## The Toxicity of Prefibrillar Aggregates Results from their Misfolded Nature

The reason why prefibrillar aggregates are toxic to cells, and hence appear to be the most likely culprits for the origins of at least some of the protein deposition diseases, is now at the front line of research in this field. A wide variety of biochemical, cytological, and physiological perturbations has been identified following the exposure of neurons to such species, both in vivo and in vitro. A detailed description of all of the reported effects is beyond the scope of the present review, and indeed, it is still too early to draw definitive conclusions about the similarities or differences of the effects of particular types of aggregates in different diseases.

Despite differences in the specific mechanisms of pathogenic behavior giving rise to distinct diseases, it is clear that the conversion of a protein from its soluble state into oligomeric forms will invariably generate a wide distribution of nonnative species, the populations of which will vary with sequence, time, and conditions. It seems likely that all of these inherently "misfolded" species will be toxic to some degree because they will inevitably expose on their surfaces an array of groups that are normally buried in globular proteins or dispersed in highly unfolded peptides or proteins. Small aggregates have a higher proportion of residues on their surfaces than larger aggregates, including mature amyloid fibrils, and therefore are likely in general to have a higher relative toxicity. In the crowded and highly organized environment of a living organism, the nonnative character of misfolded oligomers is particularly likely to trigger aberrant events resulting from their inappropriate interactions with cellular components, such as membranes, small metabolites, proteins, or other macromolecules. Such events will, in some situations, lead to the malfunctioning of crucial aspects of the cellular machinery, whether it is axonal transport, oxidative stress, ion balance, sequestration of essential proteins, or a combination of disparate factors, ultimately leading to apoptosis or other forms of cell death.

Although the natural defenses against misfolded proteins will act to sequester and neutralize such species, and/or inhibit their formation, it is inevitable that these mechanisms will sometimes be overwhelmed (98, 173). Such situations include mutations that dramatically increase aggregation rates, as in familial diseases; ingestion of preformed aggregates that are able to seed more extensive aggregation, as in prion diseases; or the age-related decline of chaperone and ubiquitin/proteasome responses, as in sporadic forms of diseases. Oligomer-mediated cytotoxicity is a key issue in neuropathic protein deposition diseases, although the question arises as to whether a similar mechanism is

central in the pathogenesis of nonneuropathic diseases. Systemic amyloidoses are often associated with accumulation of large quantities (even kilograms in some cases) of amyloid deposits in the affected tissues and organs (174). Undoubtedly, the impairment and disruption of tissue architecture, caused by these deposits in vital organs, are major features of these diseases and could well be the most important factors in the pathogenesis of at least some of these nonneuropathic degenerative conditions (174). Patients can have mechanical problems in carrying out even routine everyday tasks. Examples include difficulties in swallowing when amyloid accumulation occurs in the tongue and in moving because of extreme pain when amyloid accumulation occurs in joints. However, suggestions that early oligomeric species could have a more important role than fibril accumulation in the pathogenesis of nonneuropathic amyloidoses have been put forward (169, 174a). The elucidation of the mechanism of tissue damage by amyloid fibril proteins is undoubtedly an important issue in therapeutic approaches, although the optimum strategy must be to prevent aggregation or even production of the amyloidogenic protein before it can generate any potential damaging deposits.

## PERSPECTIVES

Despite the complexity of the protein aggregation process, the findings described above show that dramatic progress in its elucidation has been made in recent years. This progress relates particularly to our understanding of the nature and significance of amyloid formation and to how this process relates to the normal and aberrant behavior of living organisms. Increasingly sophisticated techniques are now being applied to elucidate the "amyloid phenomenon" in ever greater detail. Of special significance is the manner in which a wide variety of ideas from across the breadth of the biological, physical, and medical sciences is being brought together to probe important unifying principles. Much, of course, still remains to be discovered, but we are personally optimistic that the investigation of an increasing number of proteins both in vitro and in vivo will shed new light on the relationships between protein folding and misfolding as well as on the manner in which the multitude of different states accessible to proteins are regulated and interact with each other and with other cellular components. In addition, even our present understanding of the mechanism of amyloid formation is leading to more reliable methods of early diagnosis and to more rational therapeutic strategies that are either in clinical trials or approaching such trials (175, 176). Thus, despite the rapidity with which diseases of the type discussed here are increasingly afflicting the human populations of the modern world, there are grounds for optimism that present progress in understanding their nature and origins will lead, in the not too distant future, to the beginnings of widely applicable and effective means to combat their spread and their debilitating consequences.

### SUMMARY POINTS

1. A variety of human diseases is now thought to be associated with the formation of highly organized and generally intractable thread-like aggregates termed amyloid or amyloid-like fibrils.
2. Living organisms can take advantage of the inherent ability of proteins to form such structures to generate novel and diverse biological functions.
3. Dramatic advances have recently been made toward the elucidation of the structures of amyloid fibrils at a molecular level.

4. Amyloid fibril formation is preceded by formation of a wide range of aggregates such as unstructured oligomers and structured protofibrils.
5. We are now able to rationalize some of the issues regarding the molecular mechanism of amyloid formation, e.g., identify the regions of the sequence that form and stabilize the fibril core and/or play a primary role in fibril formation.
6. At least in some cases, prefibrillar aggregates, rather than the mature fibrils into which they convert, are the likely origins of pathological behavior. Despite obvious differences in detail, the pathogenic nature of these species lies in the exposure of groups that are normally buried in a folded protein or dispersed in an unfolded ensemble.

## FUTURE ISSUES TO BE RESOLVED

1. Although considerable progress has been made in the elucidation of amyloid fibril properties at a molecular level, very little is yet known about the structure of the amyloid protofibrils and unstructured aggregates that precede their formation and are likely to play a key role in the pathogenesis of protein deposition diseases.
2. Present research has been remarkably successful in providing a framework for understanding the fundamental nature of protein aggregation. The challenge now is to explore in more detail the links between these largely structural principles and the cellular and animal environments in which aggregation takes place.
3. The precise origin of the pathogenic nature of the amyloid deposits and their precursors remains elusive in each pathological condition associated with formation of these species.
4. The rational design of successful therapeutic strategies requires further characterization of the processes of amyloid formation occuring in vivo and of the interaction of the resulting aggregates with the various components of living organisms.

## LITERATURE CITED


1. Uversky VN, Fink AL. 2006. *Protein Misfolding, Aggregation and Conformational Diseases*, Vols. 1, 2. Kluwer Academic/Plenum
2. Amaral MD. 2004. *J. Mol. Neurosci.* 23:41–48
3. Lomas DA, Carrell RW. 2002. *Nat. Rev. Genet.* 3:759–68
4. Westermark P, Benson MD, Buxbaum JN, Cohen AS, Frangione B, et al. 2005. *Amyloid* 12:1–4
5. Lundmark K, Westermark GT, Nystrom S, Murphy CL, Solomon A, Westermark P. 2002. *Proc. Natl. Acad. Sci. USA* 99:6979–84
6. Lundmark K, Westermark GT, Olsen A, Westermark P. 2005. *Proc. Natl. Acad. Sci. USA* 102:6098–102
7. Legname G, Baskakov IV, Nguyen HO, Riesner D, Cohen FE, et al. 2004. *Science* 305:673–76
8. Hirschfield GM, Hawkins PN. 2003. *Int. J. Biochem. Cell. Biol.* 35:1608–13
9. Alexandrescu AT. 2005. *Protein Sci.* 14:1–12

10. Serpell LC, Sunde M, Benson MD, Tennent GA, Pepys MB, Fraser PE. 2000. *J. Mol. Biol.* 300:1033–39
11. Sunde M, Blake C. 1997. *Adv. Protein Chem.* 50:123–59
12. Bauer HH, Aebi U, Haner M, Hermann R, Muller M, Merkle HP. 1995. *J. Struct. Biol.* 115:1–15
13. Saiki M, Honda S, Kawasaki K, Zhou D, Kaito A, et al. 2005. *J. Mol. Biol.* 348:983–98
14. Pedersen JS, Dikov D, Flink JL, Hjuler HA, Christiansen G, Otzen D. 2005. *J. Mol. Biol.* 355:501–23
15. Nilsson MR. 2004. *Methods* 34:151–60
16. Khurana R, Uversky VN, Nielsen L, Fink AL. 2001. *J. Biol. Chem.* 276:22715–21
17. Bousset L, Redeker V, Decottignies P, Dubois S, Le Marechal P, Melki R. 2004. *Biochemistry* 43:5022–32
18. Jin LW, Claborn KA, Kurimoto M, Geday MA, Maezawa I, et al. 2003. *Proc. Natl. Acad. Sci. USA* 100:15294–98
19. Krebs MRH, MacPhee CE, Miller AF, Dunlop LE, Dobson CM, Donald AM. 2004. *Proc. Natl. Acad. Sci. USA* 101:14420–24
20. Stefani M, Dobson CM. 2003. *J. Mol. Med.* 81:678–99
21. Uversky VN, Fink AL. 2004. *Biochim. Biophys. Acta* 1698:131–53
22. Chapman MR, Robinson LS, Pinkner JS, Roth R, Heuser J, et al. 2002. *Science* 295:851–55
23. Claessen D, Rink R, de Jong W, Siebring J, de Vreugd P, et al. 2003. *Genes Dev.* 17:1714–26
23a. Mackay JP, Matthews JM, Winefield RD, Mackay LG, Haverkamp RG, et al. 2001. *Structure* 9:83–91
23b. Iconomidou VA, Vriend G, Hamodrakas SJ. 2000. *FEBS Lett.* 479:141–45
23c. Kenney JM, Knight D, Wise MJ, Vollrath F. 2002. *Eur. J. Biochem.* 269:4159–63
24. Berson JF, Theos AC, Harper DC, Tenza D, Raposo G, Marks MS. 2003. *J. Cell Biol.* 161:521–33
25. Chien P, Weissman JS, DePace AH. 2004. *Annu. Rev. Biochem.* 73:617–56
26. Eaglestone SS, Cox BS, Tuite MF. 1999. *EMBO J.* 18:1974–81
27. True HL, Lindquist SL. 2000. *Nature* 407:477–83
28. True HL, Berlin I, Lindquist SL. 2004. *Nature* 431:184–87
28a. Giaever G, Chu AM, Ni L, Connelly C, Riles L, et al. 2002. *Nature* 418:387–91
29. Nakayashiki T, Kurtzman CP, Edskes HK, Wickner RB. 2005. *Proc. Natl. Acad. Sci. USA* 102:10575–80
30. Si K, Lindquist S, Kandel ER. 2003. *Cell* 115:879–91
31. Coustou V, Deleu C, Saupe S, Begueret J. 1997. *Proc. Natl. Acad. Sci. USA* 94:9773–78
32. Saupe SJ. 2000. *Microbiol. Mol. Biol. Rev.* 64:489–502
33. Harper JD, Lieber CM, Lansbury PT Jr. 1997. *Chem. Biol.* 4:951–59
34. Petkova AT, Ishii Y, Balbach JJ, Antzutkin ON, Leapman RD, et al. 2002. *Proc. Natl. Acad. Sci. USA* 99:16742–47
35. Jaroniec CP, MacPhee CE, Astrof NS, Dobson CM, Griffin RG. 2002. *Proc. Natl. Acad. Sci. USA* 99:16748–53
36. Ritter C, Maddelein ML, Siemer AB, Luhrs T, Ernst M, et al. 2005. *Nature* 435:844–48
37. Makin OS, Atkins E, Sikorski P, Johansson J, Serpell LC. 2005. *Proc. Natl. Acad. Sci USA* 102:315–20
38. Nelson R, Sawaya MR, Balbirnie M, Madsen AO, Riekel C, et al. 2005. *Nature* 435:773–78

39. Antzutkin ON, Balbach JJ, Leapman RD, Rizzo NW, Reed J, Tycko R. 2000. *Proc. Natl. Acad. Sci. USA* 97:13045–50
40. Balbach JJ, Petkova AT, Oyler NA, Antzutkin ON, Gordon DJ, et al. 2002. *Biophys. J.* 83:1205–16
41. Goldsbury CS, Wirtz S, Muller SA, Sunderji S, Wicki P, et al. 2000. *J. Struct. Biol.* 130:217–31
42. Torok M, Milton S, Kayed R, Wu P, McIntire T, et al. 2000. *J. Biol. Chem.* 277:40810–15
43. Kheterpal I, Zhou S, Cook KD, Wetzel R. 2000. *Proc. Natl. Acad. Sci. USA* 97:13597–601
44. Kheterpal I, Williams A, Murphy C, Bledsoe B, Wetzel R. 2001. *Biochemistry* 40:11757–67
45. Williams AD, Portelius E, Kheterpal I, Guo JT, Cook KD, et al. 2004. *J. Mol. Biol.* 335:833–42
46. Hoshino M, Katou H, Hagihara Y, Hasegawa K, Naiki H, Goto Y. 2002. *Nat. Struct. Biol.* 9:332–36
47. Jaroniec CP, MacPhee CE, Bajaj VS, McMahon MT, Dobson CM, Griffin RG. 2004. *Proc. Natl. Acad. Sci. USA* 101:711–16
48. Jimenez JL, Guijarro JI, Orlova E, Zurdo J, Dobson CM, et al. 1999. *EMBO J.* 18:815–21
49. Jimenez JL, Nettleton EJ, Bouchard M, Robinson CV, Dobson CM, Saibil HR. 2002. *Proc. Natl. Acad. Sci. USA* 99:9196–201
50. Fandrich M, Dobson CM. 2002. *EMBO J.* 21:5682–90
51. Kajava AV, Aebi U, Steven AC. 2005. *J. Mol. Biol.* 348:247–52
52. Krishnan R, Lindquist SL. 2005. *Nature* 435:765–72
53. Makin O, Serpell LC. 2004. *J. Mol. Biol.* 335:1279–88
54. Goldsbury CS, Cooper GJ, Goldie KN, Muller SA, Saafi EL, et al. 1997. *J. Struct. Biol.* 119:17–27
55. Jayasinghe SA, Langen R. 2004. *J. Biol. Chem.* 279:48420–25
56. Der-Sarkissian A, Jao CC, Chen J, Langen R. 2003. *J. Biol. Chem.* 278:37530–35
57. Margittai M, Langen R. 2004. *Proc. Natl. Acad. Sci. USA* 101:10278–83
58. Yamaguchi K, Katou H, Hoshino M, Hasegawa K, Naiki H, Goto Y. 2004. *J. Mol. Biol.* 338:559–71
58a. Del Mar C, Greenbaum EA, Mayne L, Englander SW, Woods VL Jr. 2005. *Proc. Natl. Acad. Sci. USA* 102:15477–82
59. Monti M, Principe S, Giorgetti S, Mangione P, Merlini G, et al. 2002. *Protein Sci.* 11:2362–69
59a. Miake H, Mizusawa H, Iwatsubo T, Hasegawa M. 2002. *J. Biol. Chem.* 2002. 277:19213–19
59b. Heise H, Hoyer W, Becker S, Andronesi OC, Riedel D, Baldus M. 2005. *Proc. Natl. Acad. Sci. USA* 102:15871–76
60. Perutz MF, Finch JT, Berriman J, Lesk A. 2002. *Proc. Natl. Acad. Sci. USA* 99:5591–95
61. Zandomeneghi G, Krebs MR, McCammon MG, Fandrich M. 2004. *Protein Sci.* 13:3314–21
62. Petkova AT, Leapman RD, Guo Z, Yau WM, Mattson MP, Tycko R. 2005. *Science* 307:262–65
63. Baxa U, Taylor KL, Wall JS, Simon MN, Cheng NQ, et al. 2003. *J. Biol. Chem.* 278:43717–27
64. Kajava AV, Baxa U, Wickner RB, Steven AC. 2004. *Proc. Natl. Acad. Sci. USA* 101:7885–90
65. Baxa U, Cheng NQ, Winkler DC, Chiu TK, Davies DR, et al. 2005. *J. Struct. Biol.* 150:170–79

66. Bousset L, Thomson NH, Radford SE, Melki R. 2002. *EMBO J.* 17:2903–11
67. Bousset L, Briki F, Doucet J, Melki R. 2003. *J. Struct. Biol.* 141:132–42
68. Tanaka M, Chien P, Yonekura K, Weissman JS. 2005. *Cell* 121:49–62
69. Jones EM, Surewicz WK. 2005. *Cell* 121:63–72
70. Naiki H, Hashimoto N, Suzuki S, Kimura H, Nakakuki K, Gejyo F. 1997. *Amyloid* 4:223–32
71. Serio TR, Cashikar AG, Kowal AS, Sawicki GJ, Moslehi JJ, et al. 2000. *Science* 289:1317–21
72. Uversky VN, Li J, Souillac P, Millett IS, Doniach S, et al. 2002. *J. Biol. Chem.* 277:11970–78
73. Pedersen JS, Christensen G, Otzen DE. 2004. *J. Mol. Biol.* 341:575–88
74. Chayen NE. 2005. *Prog. Biophys. Mol. Biol.* 88:329–37
75. Fezoui Y, Teplow DB. 2002. *J. Biol. Chem.* 277:36948–54
76. Wright CF, Teichmann SA, Clarke J, Dobson CM. 2005. *Nature* 438:878–81
76a. Krebs MR, Morozova-Roche LA, Daniel K, Robinson CV, Dobson CM. 2004. *Protein Sci.* 13:1933–38
77. Harper JD, Wong SS, Lieber CM, Lansbury PT Jr. 1997. *Chem. Biol.* 4:119–25
78. Walsh DM, Lomakin A, Benedek GB, Condron MM, Teplow DB. 1997. *J. Biol. Chem.* 272:22364–72
79. Walsh DM, Hartley DM, Kusumoto Y, Fezoui Y, Condron MM, et al. 1999. *J. Biol. Chem.* 274:25945–52
80. Kayed R, Sokolov Y, Edmonds B, McIntire TM, Milton SC, et al. 2004. *J. Biol. Chem.* 279:46363–66
81. Williams AD, Sega M, Chen ML, Kheterpal I, Geva M, et al. 2005. *Proc. Natl. Acad. Sci. USA* 102:7115–20
82. Conway KA, Harper JD, Lansbury PT Jr. 2000. *Biochemistry* 39:2552–63
83. Ionescu–Zanetti C, Khurana R, Gillespie JR, Petrick JS, Trabachino LC, et al. 1999. *Proc. Natl. Acad. Sci. USA* 96:13175–79
84. Quintas A, Vaz DC, Cardoso I, Saraiva MJM, Brito RMM. 2001. *J. Biol. Chem.* 276:27207–13
85. Gosal WS, Morten IJ, Hewitt EW, Smith DA, Thomson NH, Radford SE. 2005. *J. Mol. Biol.* 351:850–64
86. Malisauskas M, Zamotin V, Jass J, Noppe W, Dobson CM, Morozova-Roche LA. 2003. *J. Mol. Biol.* 330:879–90
87. Plakoutsi G, Taddei N, Stefani M, Chiti F. 2004. *J. Biol. Chem.* 279:14111–19
87a. Bader R, Bamford R, Zurdo J, Luisi BF, Dobson CM. 2006. *J. Mol. Biol.* 356:189–208
88. Kayed R, Head E, Thompson JL, McIntire TM, Milton SC, et al. 2003. *Science* 300:486–89
89. Morozova-Roche LA, Zamotin V, Malisauskas M, Ohman A, Chertkova R, et al. 2004. *Biochemistry* 43:9610–19
89a. Petty SA, Adalsteinsson T, Decatur SM. 2005. *Biochemistry* 44:4720–26
89b. Petty SA, Decatur SM. 2005. *Proc. Natl. Acad. Sci. USA* 102:14272–77
90. Bitan G, Lomakin A, Teplow DB. 2001. *J. Biol. Chem.* 276:35176–84
91. Bitan G, Kirkitadze MD, Lomakin A, Vollers SS, Benedek GB, Teplow DB. 2003. *Proc. Natl. Acad. Sci. USA* 100:330–35
92. Roher AE, Chaney MO, Kuo YM, Webster SD, Stine WB, et al. 1996. *J. Biol. Chem.* 271:20631–35
93. Podlisny MB, Ostaszewski BL, Squazzo SL, Koo EH, Rydell RE, et al. 1995. *J. Biol. Chem.* 270:9564–70

94. Walsh DM, Tseng BP, Rydel RE, Podlisny MB, Selkoe DJ. 2000. *Biochemistry* 39:10831–39
95. Modler AJ, Gast K, Lutsch G, Damaschun G. 2003. *J. Mol. Biol.* 325:135–48
96. Nguyen HD, Hall CK. 2004. *Proc. Natl. Acad. Sci. USA* 101:16180–85
97. Kelly JW. 1998. *Curr. Opin. Struct. Biol.* 8:101–6
98. Dobson CM. 1999. *Trends Biochem. Sci.* 24:329–32
99. Guijarro JI, Sunde M, Jones JA, Campbell ID, Dobson CM. 1998. *Proc. Natl. Acad. Sci. USA* 95:4224–28
100. Chiti F, Taddei N, Bucciantini M, White P, Ramponi G, Dobson CM. 2000. *EMBO J.* 19:1441–49
101. Villegas V, Zurdo J, Filimonov VV, Aviles FX, Dobson CM, Serrano L. 2000. *Protein Sci.* 9:1700–8
102. Ferrao-Gonzales AD, Souto SO, Silva JL, Foguel D. 2000. *Proc. Natl. Acad. Sci. USA* 97:6445–50
103. Raffen R, Dieckman LJ, Szpunar M, Wunschl C, Pokkuluri PR, et al. 1999. *Protein Sci.* 8:509–17
104. Canet D, Last AM, Tito P, Sunde M, Spencer A, et al. 2002. *Nat. Struct. Biol.* 9:308–15
105. Hammarstrom P, Jiang X, Hurshman AR, Powers ET, Kelly JW. 2002. *Proc. Natl. Acad. Sci. USA* 99:16427–32
106. Ramirez-Alvarado M, Merkel JS, Regan L. 2000. *Proc. Natl. Acad. Sci. USA* 97:8979–84
107. Marcon G, Plakoutsi G, Canale C, Relini A, Taddei N, et al. 2005. *J. Mol. Biol.* 347:323–35
108. Chiti F, Taddei N, Stefani M, Dobson CM, Ramponi G. 2001. *Protein Sci.* 10:879–86
109. Sacchettini JC, Kelly JW. 2002. *Nat. Rev. Drug Discov.* 1:267–75
110. Dumoulin M, Last AM, Desmyter A, Decanniere K, Canet D, et al. 2003. *Nature* 424:783–88
111. Ray SS, Nowak RJ, Brown RH Jr, Lansbury PT Jr. 2005. *Proc. Natl. Acad. Sci. USA* 102:3639–44
112. Bouchard M, Zurdo J, Nettleton EJ, Dobson CM, Robinson CV. 2000. *Protein Sci.* 9:960–67
113. Chow MK, Ellisdon AM, Cabrita LD, Bottomley SP. 2005. *J. Biol. Chem.* 279:47643–51
114. Plakoutsi G, Bemporad F, Calamai M, Taddei N, Dobson CM, Chiti F. 2005. *J. Mol. Biol.* 351:910–22
115. Lindorff-Larsen K, Best RB, Depristo MA, Dobson CM, Vendruscolo M. 2005. *Nature* 433:128–32
116. Otzen DE, Kristensen O, Oliveberg M. 2000. *Proc. Natl. Acad. Sci. USA* 97:9907–12
117. Chiti F, Taddei N, Baroni F, Capanni C, Stefani M, et al. 2002. *Nat. Struct. Biol.* 9:137–43
118. Wurth C, Guimard NK, Hecht MH. 2002. *J. Mol. Biol.* 319:1279–90
119. Schwartz R, Istrail S, King J. 2001. *Protein Sci.* 10:1023–31
120. Chiti F, Calamai M, Taddei N, Stefani M, Ramponi G, Dobson CM. 2002. *Proc. Natl. Acad. Sci. USA* 99:16419–26
121. Schmittschmitt JP, Scholtz JM. 2003. *Protein Sci.* 12:2374–78
122. Konno T. 2001. *Biochemistry* 40:2148–54
123. Goers J, Uversky VN, Fink AL. 2003. *Protein Sci.* 12:702–7
124. Giasson BI, Forman MS, Higuchi M, Golbe LI, Graves CL, et al. 2003. *Science* 300:636–40
125. Fernandez CO, Hoyer W, Zweckstetter M, Jares-Erijman EA, Subramaniam V, et al. 2004. *EMBO J.* 23:2039–46

126. Uversky VN. 2002. *Protein Sci*. 11:739–56
127. Broome BM, Hecht MH. 2000. *J. Mol. Biol.* 296:961–68
128. Kallberg Y, Gustafsson M, Persson B, Thyberg J, Johansson J. 2001. *J. Biol. Chem.* 276:12945–50
129. Steward A, Adhya S, Clarke J. 2002. *J. Mol. Biol.* 318:935–40
130. Parrini C, Taddei N, Ramazzotti M, Degl'Innocenti D, Ramponi G, et al. 2005. *Structure* 13:1143–51
131. Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM. 2003. *Nature* 424:805–8
132. Tartaglia GG, Cavalli A, Pellarin R, Caflisch A. 2004. *Protein Sci*. 13:1939–41
133. Fernandez-Escamilla AM, Rousseau F, Schymkowitz J, Serrano L. 2004. *Nat. Biotechnol*. 22:1302–6
134. Pawar AP, DuBay KF, Zurdo J, Chiti F, Vendruscolo M, Dobson CM. 2005. *J. Mol. Biol.* 350:379–92
135. Kyte J, Doolittle RF. 1982. *J. Mol. Biol.* 157:105–32
136. DuBay KF, Pawar AP, Chiti F, Zurdo J, Dobson CM, Vendruscolo M. 2004. *J. Mol. Biol*. 341:1317–26
137. Ignatova Z, Gierasch LM. 2004. *Proc. Natl. Acad. Sci. USA* 101:523–28
138. Calloni G, Zoffoli S, Stefani M, Dobson CM, Chiti F. 2005. *J. Biol. Chem*. 280:10607–13
139. Paushkin SV, Kushnirov VV, Smirnov VN, Ter-Avanesyan MD. 1996. *EMBO J.* 15:3127–34
140. Masison DC, Maddelein ML, Wickner RB. 1997. *Proc. Natl. Acad. Sci. USA* 94:12503–8
141. Balguerie A, Dos Reis S, Ritter C, Chaignepain S, Coulary-Salin B, et al. 2003. *EMBO J.* 22:2071–81
142. Monti M, di Bard BLG, Calloni G, Chiti F, Amoresano A, et al. 2004. *J. Mol. Biol.* 336:253–62
143. Fandrich M, Forge V, Buder K, Kittler M, Dobson CM, Diekmann S. 2003. *Proc. Natl. Acad. Sci. USA* 100:15463–68
144. Bertoncini CW, Jung YS, Fernandez CO, Hoyer W, Griesinger C, et al. 2005. *Proc. Natl. Acad. Sci. USA* 102:1430–35
145. Dedmon MM, Lindorff-Larsen K, Christodoulou J, Vendruscolo M, Dobson CM. 2005. *J. Am. Chem. Soc.* 127:476–77
146. Hoyer W, Cherny D, Subramaniam V, Jovin TM. 2004. *Biochemistry* 43:16233–42
147. Nilsson MR, Driscoll M, Raleigh DP. 2002. *Protein Sci.* 11:342–49
148. Zhang Q, Powers ET, Nieva J, Huff ME, Dendle MA. 2004. *Proc. Natl. Acad. Sci. USA* 101:4752–57
149. Bush AI. 2003. *Trends Neurosci*. 26:207–14
150. Pike CJ, Walencewicz AJ, Glabe CG, Cotman CW. 1991. *Brain Res.* 563:311–14
151. Lorenzo A, Yankner BA. 1994. *Proc. Natl. Acad. Sci. USA* 91:12243–47
152. Hartley DM, Walsh DM, Ye CP, Diehl T, Vasquez S, et al. 1999. *J. Neurosci*. 19:8876–84
153. Geula C, Wu CK, Saroff D, Lorenzo A, Yuan M, Yankner BA. 1998. *Nat. Med*. 4:827–31
154. Lue LF, Kuo YM, Roher AE, Brachova L, Shen Y, et al. 1999. *Am. J. Pathol.* 155:853–62
155. McLean CA, Cherny RA, Fraser FW, Fuller SJ, Smith MJ, et al. 1999. *Ann. Neurol.* 46:860–66
156. Wang J, Dickson DW, Trojanowski JQ, Lee VM. 1999. *Exp. Neurol.* 158:328–37
157. Moechars D, Dewachter I, Lorent K, Reverse D, Baekelandt V, et al. 1999. *J. Biol. Chem.* 274:6483–92
158. Larson J, Lynch G, Games D, Seubert P. 1999. *Brain Res.* 840:23–35
159. Iijima K, Liu HP, Chiang AS, Hearn SA, Konsolaki M, Zhong Y. 2001. *Proc. Natl. Acad. Sci. USA* 101:6623–28

160. Crowther DC, Kinghorn KJ, Miranda E, Page R, Curry JA, et al. 2005. *Neuroscience* 132:123–35
161. Dodart JC, Bales KR, Gannon KS, Greene SJ, DeMattos RB, et al. 2002. *Nat. Neurosci.* 5:452–57
162. Nilsberth C, Westlind-Danielsson A, Eckman CB, Condron MM, Axelman K, et al. 2001. *Nat. Neurosci.* 4:887–93
163. Gertz HJ, Siegers A, Kuchinke J. 1994. *Brain Res.* 637:339–41
164. Bergeron C, Petrunka C, Weyer L. 1996. *Am. J. Pathol.* 148:273–79
165. Kitada T, Asakawa S, Hattori N, Matsumine H, Yamamura Y, et al. 1998. *Nature* 392:605–8
166. Auluck PK, Chan HY, Trojanowski JQ, Lee VM, Bonini NM. 2002. *Science* 295:865–68
167. Lo Bianco C, Ridet JL, Schneider BL, Deglon N, Aebischer P. 2002. *Proc. Natl. Acad. Sci. USA* 99:10813–18
168. Masliah E, Rockenstein E, Veinbergs I, Mallory M, Hashimoto M, et al. 2000. *Science* 287:1265–69
169. Sousa MM, Cardoso I, Fernandes R, Guimaraes A, Saraiva MJ. 2001. *Am. J. Pathol.* 159:1993–2000
170. Reixach N, Deechongkit S, Jiang X, Kelly JW, Buxbaum JN. 2004. *Proc. Natl. Acad. Sci. USA* 101:2817–22
171. Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, et al. 2002. *Nature* 416:507–11
171a. Sirangelo I, Malmo C, Iannuzzi C, Mezzogiorno A, Bianco MR, et al. 2004. *J. Biol. Chem.* 279:13183–89
171b. Malisauskas M, Ostman J, Darinskas A, Zamotin V, Liutkevicius E, et al. 2005. *J. Biol. Chem.* 280:6269–75
172. Silveira JR, Raymond GJ, Hughson AG, Race RE, Sim VL, et al. 2005. *Nature* 437:257–61
173. Soti C, Pal C, Papp B, Csermely P. 2005. *Curr. Opin. Cell. Biol.* 17:210–15
174. Pepys MB. 2001. *Philos. Trans. R. Soc. London Ser. B* 356:203–10
174a. Merlini G, Bellotti V. 2003. *N. Engl. J. Med.* 349:583–96
175. Cohen FE, Kelly JW. 2003. *Nature* 426:905–9
176. Merlini G, Westermark P. 2004. *J. Intern. Med.* 255:159–78
177. Tycko R. 2002. *Biochemistry* 42:3151–59
178. Han HY, Weinreb PH, Lansbury PT Jr. 1995. *Chem. Biol.* 2:163–69
179. El-Agnaf OMA, Irvine GB. 2002. *Biochem. Soc. Trans.* 30:559–65

Annual Review
of Biochemistry

Volume 75, 2006

# Contents

BMC Genomics

BioMed Central

Research



# Unfoldomics of human diseases: linking protein intrinsic disorder with diseases

Vladimir N Uversky*[1,2,3], Christopher J Oldfield[1], Uros Midic[4], Hongbo Xie[4], Bin Xue[1,2], Slobodan Vucetic[4], Lilia M Iakoucheva[5], Zoran Obradovic[4] and A Keith Dunker[1]

Address: [1]Center for Computational Biology and Bioinformatics, Indiana University School of Medicine, Indianapolis, IN 46202, USA, [2]Institute for Intrinsically Disordered Protein Research, Indiana University School of Medicine, Indianapolis, IN 46202, USA, [3]Institute for Biological Instrumentation, Russian Academy of Sciences, 142290 Pushchino, Moscow Region, Russia, [4]Center for Information Science and Technology, Temple University, Philadelphia, PA 19122, USA and [5]Laboratory of Statistical Genetics, The Rockefeller University, New York, NY 10065 USA

Email: Vladimir N Uversky* - vuversky@iupui.edu; Christopher J Oldfield - cjoldfie@iupui.edu; Uros Midic - uros@ist.temple.edu; Hongbo Xie - michaelxie@ist.temple.edu; Bin Xue - binxue@iupui.edu; Slobodan Vucetic - vucetic@ist.temple.edu; Lilia M Iakoucheva - lilia@rockefeller.edu; Zoran Obradovic - zoran@ist.temple.edu; A Keith Dunker - kedunker@iupui.edu

* Corresponding author

*from* The 2008 International Conference on Bioinformatics & Computational Biology (BIOCOMP'08)
Las Vegas, NV, USA. 14–17 July 2008

Published: 7 July 2009

*BMC Genomics* 2009, **10**(Suppl 1):S7 doi:10.1186/1471-2164-10-S1-S7

This article is available from: http://www.biomedcentral.com/1471-2164/10/S1/S7



## Abstract

**Background:** Intrinsically disordered proteins (IDPs) and intrinsically disordered regions (IDRs) lack stable tertiary and/or secondary structure yet fulfills key biological functions. The recent recognition of IDPs and IDRs is leading to an entire field aimed at their systematic structural characterization and at determination of their mechanisms of action. Bioinformatics studies showed that IDPs and IDRs are highly abundant in different proteomes and carry out mostly regulatory functions related to molecular recognition and signal transduction. These activities complement the functions of structured proteins. IDPs and IDRs were shown to participate in both one-to-many and many-to-one signaling. Alternative splicing and posttranslational modifications are frequently used to tune the IDP functionality. Several individual IDPs were shown to be associated with human diseases, such as cancer, cardiovascular disease, amyloidoses, diabetes, neurodegenerative diseases, and others. This raises questions regarding the involvement of IDPs and IDRs in various diseases.

**Results:** IDPs and IDRs were shown to be highly abundant in proteins associated with various human maladies. As the number of IDPs related to various diseases was found to be very large, the concepts of the disease-related unfoldome and unfoldomics were introduced. Novel bioinformatics tools were proposed to populate and characterize the disease-associated unfoldome. Structural characterization of the members of the disease-related unfoldome requires specialized experimental approaches. IDPs possess a number of unique structural and functional features that determine their broad involvement into the pathogenesis of various diseases.

**Conclusion:** Proteins associated with various human diseases are enriched in intrinsic disorder. These disease-associated IDPs and IDRs are real, abundant, diversified, vital, and dynamic. These

proteins and regions comprise the disease-related unfoldome, which covers a significant part of the human proteome. Profound association between intrinsic disorder and various human diseases is determined by a set of unique structural and functional characteristics of IDPs and IDRs. Unfoldomics of human diseases utilizes unrivaled bioinformatics and experimental techniques, paves the road for better understanding of human diseases, their pathogenesis and molecular mechanisms, and helps develop new strategies for the analysis of disease-related proteins.

## Background

### *Introducing intrinsically disordered proteins*

Proteins are the major components of the living cell. They play crucial roles in the maintenance of life. Protein dysfunctions may cause development of various pathological conditions For more than 75 years it has been believed that the specific functionality of a given protein is predetermined by its unique 3-D structure [1,2]. For these structured proteins, the sequence → structure → function paradigm has become paramount. According to this view, a protein's function depends on its prior folding into a unique three-dimensional structure. In such cases, the amino acid sequence determines the protein's unique 3-D structure.

Evidence is rapidly accumulating that many protein regions and even entire proteins lack stable tertiary and/or secondary structure in solution yet possess crucial biological functions [3-25]. These naturally flexible proteins and regions are known by different names, including intrinsically disordered [8], natively denatured [26], natively unfolded [27], intrinsically unstructured [4], natively disordered [21], and inherently disordered [25,28,29]. In this article, the terms "intrinsically disordered proteins" and "intrinsically disordered regions" (IDPs and IDRs, respectively) are used to describe such proteins and regions generally and "natively denatured" or "intrinsically unstructured" are used for collapsed and extended random coils that lack significant amounts of stable secondary structure (see below).

IDPs and IDRs can contain collapsed-disorder, semi-collapsed disorder, or extended-disorder under physiological conditions *in vitro* [6,12,19]. Collapsed-disorder consists mainly of molten globules, which are formed by hydrophobic collapse, which have stable but dynamic secondary structure, and which have flexible and dynamic side chains [14,30-38]. Semi-collapsed structures arise because water is a poor solvent for the peptide backbone and include, for example, polyglutamine regions [28], other polar sequences [29], and pre-molten globules [14,39-41]. Pre-molten globules may contain regions with transient secondary structure or small amounts of localized, fairly stable secondary structure. Extended-disorder arises from chains having repulsion arising from a net charge, and these proteins and regions resemble the more classical idealized random coil. Because of the lack of a hydrophobic core and the presence of only the marginal levels of residual secondary structure, native coils and native pre-molten globules are grouped together in a class of natively unfolded or intrinsically unstructured proteins [11,12].

In contrast to the long history regarding structured proteins, the study of the IDP phenomenon is emerging only very recently (Figure 1). This transition is occurring mostly due to the efforts of four research groups, which almost simultaneously and completely independently came to the important conclusion that naturally flexible proteins, instead of being just rare exceptions, represent a new and very broad class of proteins [1,2,4,11]. This important conclusion was reached from different starting points using very different experimental approaches, including: bioinformatics (Dr. A.K. Dunker's group), NMR spectroscopy (Dr. P.E. Wright's group), multiparametric protein folding/misfolding studies (Dr. V.N. Uversky's group), and protein structural characterization (Dr. P. Tompa's group). The bioinformatics approach has played an especially crucial role in shaping this field, bringing coherence and recognition to proteins that were previously viewed individually as outliers from the main stream [42]. After publication of key studies and reviews describing this new concept, the literature on IDPs and IDRs is virtually exploding (see Figure 1).

Figure 2 represents the modern understanding of the fate of a polypeptide chain inside a cell and schematically shows the three types of intrinsic disorder mentioned above, native coil, native pre-molten globule, and native molten globule. According to this hypothesis, newly synthesized proteins can either fold to gain a unique structure necessary for the catalytic and transport activities, can stay substantially non-folded, or can misfold under some circumstances to form amyloid-like fibrils. Importantly, both folded and non-folded polypeptide chains have specific biological functions. The three endpoints are further interlinked and some changes in the environment, interaction with specific binding partners, or mutations may bring subsequent structural rearrangements. As a result, an intrinsically disordered polypeptide can partially or completely fold or misfold and form amyloid-like fibrils,

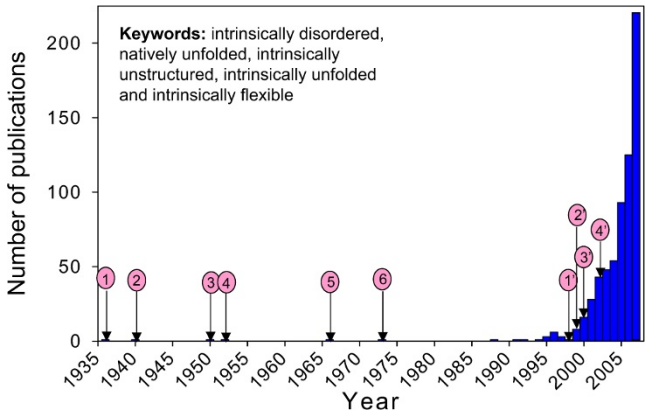


**Figure 1**
**Increase in the number of publications dealing with the IDPs**. Circles 1–6 correspond to some key IDP-related publications in the pre-bioinformatics era. They are: 1) Landsteiner, 1936 [110]; 2) Pauling, 1940 [111]; 3) Karush, 1950 [112]; 4) McMeekin, 1962 [113]; 5) Jirgenesons, 1966 [114]; 6) Doolittle, 1973 [115]. Circles 1'-4' correspond to key research bioinformatics articles and reviews that created and shaped the IDP field: 1') Romero et al., 1998 [3]; 2') Wright & Dyson, 1999 [4]; 3') Uversky et al., 2000 [6]; 4') Tompa, 2002 [13].

whereas ordered protein can likewise misfold and assemble into a pathogenic fibrillar form (see Figure 2) [14].

Just as the amino acid sequence of an ordered protein contains the information for a highly specific folding, the amino acid sequence of an IDP codes for lack of structure or disorder. The validity of this hypothesis is supported by the development of various disorder predictors [29-32] all based on well-defined biases in the IDP sequences and amino acid compositions [3,6,7,42-45]. For example, natively unfolded proteins were shown to be specifically localized within a unique region of charge-hydropathy phase space characterized by a combination of low overall hydropathy and high net charge [6]. More specifically, in comparison with structured proteins, IDPs and IDRs are generally depleted in the structure-promoting residues (including cysteine, tryptophan, tyrosine, isoleucine, phenylalanine, valine, leucine, histidine, threonine, and asparagine) and noticeably enriched in the disorder-promoting residues (aspartic acid, methionine, lysine, arginine, serine, glutamine, proline, and glutamic acid) [5,29,33,34].

IDPs and IDRs are highly abundant in nature. This follows from the results of disorder prediction for many whole proteomes. The fraction of proteins with substantial amounts of disorder is found to be proportional to the complexity of the organisms. IDPs/IDRs are more abundant in eukaryotes than in archaea and prokaryotes. Furthermore, multicellular eukaryotes were shown to have much more predicted disorder than unicellular eukaryotes [5,46,47]. In general, for mammals, ~75% of their signaling proteins are predicted to contain long IDRs (> 30 residues), about half of their total proteins are predicted to contain such long IDRs, and ~25% of their proteins are predicted to be fully disordered [25].

IDPs and IDRs carry out pivotal biological functions, participating in recognition and in various signaling and regulatory pathways, via specific protein-protein, protein-nucleic acid and protein-ligand interactions [22,48-50]. Sites of various post-translational modifications (PTMs) and sites of proteolytic attack are frequently associated with regions of intrinsic disorder [50]. The capability of non-folding proteins and regions to interact with collections of partners is utilized in organizing complex protein-protein interaction networks. In fact, hub proteins have been shown to have multiple interactions, either being intrinsically disordered and serving as an anchor, or acting as a stable globular anchor that interacts with intrinsically disordered regions of its targets [21,51-56].

Summarizing, whole proteins or protein regions are intrinsically disordered if they fail to fold into 3-D structures, remaining as floppy ensembles with specific biological functions. In our view, IDPs include molten globules,

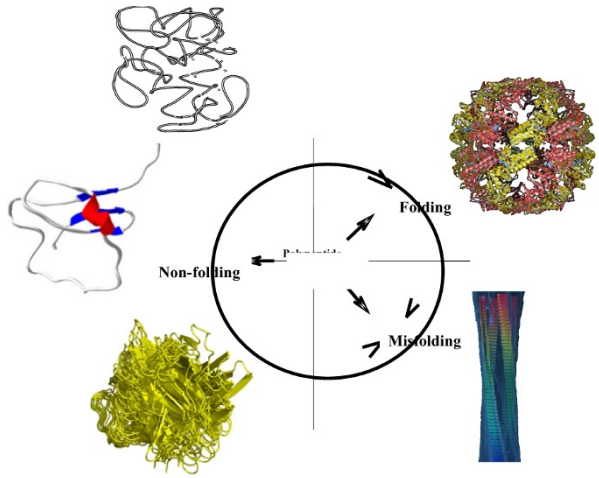


**Figure 2**
**The modern understanding of the fate of a polypeptide chain inside a cell**. Three types of IDPs, native coil, native pre-molten globule, and native molten globule are schematically shown together with the structure of an ordered protein and an amyloid fibril. Here, newly synthesized polypeptide chain can either undergo specific folding to gain a unique structure necessary for the catalytic and transport activities, or stay substantially non-folded or misfold and form amyloid-like fibrils. Both folded and non-folded proteins have specific biological functions.

pre-molten globules, random coils and transiently structured forms. IDPs are highly abundant and carry out numerous vital functions. IDPs and IDRs can be predicted by a variety of algorithms. Experimentally, they can be identified using various biophysical techniques, including NMR (especially $^{1}H$-$^{15}N$ NOEs), X-ray crystallography (especially missing density regions), circular dichroism, protease sensitivity, and many others [57].

## IDPs in human diseases: illustrative case studies

Proteins are involved in the maintenance of all stages of the life cycle. The fact that protein dysfunction can cause development of various pathological conditions was known for a very long time. Currently, a broad range of human diseases is linked to the failure of a specific peptide or protein to adopt its functional conformational state; i.e., to protein misfolding, loss of normal function, gain of toxic function, and/or protein aggregation. Although each of such diseases originates from the misfunction of a particular protein, they all are grouped together as protein-conformation or protein-misfolding diseases to emphasize the common molecular mechanisms of their origin. Triggers for misfolding vary for different proteins. Some disease-related proteins have an intrinsic propensity to form pathologic conformation(s). For other proteins, interactions or impaired interactions with chaperones, intracellular or extracellular matrixes, other proteins, small molecules and other endogenous factors can induce conformational changes and increase propensity to misfold. Often, misfolding and misfunction originate from point mutation(s) or result from an exposure to internal or external toxins, impaired posttranslational modifications (phosphorylation, advanced glycation, deamidation, racemization, etc.), an increased probability of degradation, impaired trafficking, lost binding partners or oxidative damage. All these factors can act independently, additively or synergistically.

Protein-conformation diseases can affect a single organ or be spread through multiple tissues. For example, numerous amyloidoses and various neurodegenerative disorders originate from the conversion of specific proteins from their soluble functional states into stable, highly ordered amyloid fibrils, and from the deposition of these aggregates in the variety of organs and tissues. Although protein aggregation is the most visible and the best studied consequence of protein misfolding, pathogenesis of many human diseases might depend on more subtle structural changes that lead to misfunction. Many of the proteins associated with the various conformational diseases are involved in recognition, regulation and cell signaling and a great number of these proteins are IDPs. This review is an attempt to develop an overall understanding of the roles of IDPs in various human diseases. We will start with a couple of illustrative examples where well-characterized IDPs were shown to be associated with the pathogenesis of specific diseases. We will consider here α-synuclein, p53 and HPV proteins. Additional illustrative examples can be found in our recent review [58]. The abundance of intrinsic disorder in various disease-associated proteins will be revealed using specific bioinformatics and computational tools. Then, we will attempt to answer the question why IDPs are so frequently associated with human diseases. The overall goal of this review is to introduce a concept of the disease-related unfoldome and to describe a set of bioinformatics approaches that serve as specific unfoldomics tools.

### *α*-Synuclein, Parkinson's Diseases and other synucleinpathies

α-Synuclein is one of the most intensively studied IDPs [59-61]. This is because of its association with a group of neurodegenerative disorders, synucleinopathies, characterized by the fibrillar α-synuclein aggregates in the cytoplasm of selective populations of neurons and glia [62-65] and by a chronic and progressive decline in motor, cognitive, behavioral, and autonomic functions, with the disease phenotype depending on the distribution of the lesions. Some of the most common synucleinopathies are Parkinson's disease (PD), dementia with Lewy bodies (DLB), Alzheimer's disease (AD), Down's syndrome, multiple system atrophy (MSA), and neurodegeneration with brain iron accumulation type 1 (NBIA1). A more complete list of synucleinopathies is shown in Additional file 1.

Depending on the type of pathology, α-synuclein inclusions are present in neurons (both dopaminergic and non-dopaminargic), where they can be deposited in perikarya or in axonal processes of neurons, and in glia. At least five morphologically different α-synuclein-containing inclusions have been described: Lewy bodies, Lewy neurites (dystrophic neurites), glial cytoplasmic inclusions, neuronal cytoplasmic inclusions and axonal spheroids [60,61].

The protein that links various synucleinopathies is α-synuclein, which is a typical IDP with low level of ordered structure under the physiological conditions *in vitro* [59]. According to the detailed conformational studies, the structure of α-synuclein is extremely sensitive to the environment, and this protein is known to adopt a variety of structurally unrelated conformations. The list includes a natively unfolded (mostly disordered) state, an amyloidogenic partially folded conformation, and different α-helical or β-structural species folded to a different degree, both monomeric and oligomeric [59]. It might also form aggregates with different morphology, oligomers (spherical or annular), amorphous aggregates, and amyloid-like fibrils [59]. Finally, similar to other fibrillating proteins

[66], α-synuclein was shown to assemble into the annular aggregates able to form ion-conducting, transmembrane channels [67-69]. As α-synuclein has a high intrinsic propensity to aggregate, it represents a unique model for the structural and mechanistic analysis of amyloidogenic IDPs.

To describe the structural malleability of α-synuclein, the concept of a protein-chameleon was proposed, according to which the structure of α-synuclein depends on its environment and the choice between various conformations is determined by the peculiarities of the protein's surroundings [59]. This conformational plasticity is determined by a specific folding-energy landscape of an IDP, which in contrast to that of an ordered protein, is characterized by numerous local energy minima, leading to a highly frustrated system without any stable well-folded conformation [58]. Such an energy landscape can explain the conformational plasticity of an IDP and show how such a protein can specifically interact with many ligands of different nature and respond differently to various environmental challenges. The interaction with a particular binding partner (or other changes in the environment) affects the IDP folding landscape making some energy minima deeper and some energy barriers higher, therefore determining the ability of such a protein to fold in a template-dependent manner [58].

#### *p53 and cancer*

The p53 protein is a transcription factor located at the center of a large signaling network. It regulates expression of over 150 genes, including *p21*, *GADD45*, *MDM2*, *IGFBP3*, and *BAX* [70]. Some of the genes induced or inhibited by p53 are involved in such cellular processes as cell cycle progression, apoptosis induction, DNA repair, response to cellular stress, among other functions [71]. When p53 function is lost, either directly through mutation or indirectly through several other mechanisms, the cell often undergoes cancerous transformation [72]. For this reason, a loss of p53 function is considered as a major factor in cancer development [72].

To carry out its numerous signal transduction functions, p53 interacts physically with a large number of other proteins. Many of these interactors are transcription factors, and many more are activators or inhibitors of p53 transactivation activities. The p53-Mdm2 interaction is of special interest due to its direct relation to the oncogenesis. The Mdm2 protein inactivates p53 by binding to its transcription activation domain [73]. This interaction prevents p53 from activating its target genes in three ways [74]: (*i*) It directly blocks p53 from binding to various transcription factors; (*ii*) Mdm2 acts as a ubiquitin ligase, targeting p53 for destruction; (*iii*) Mdm2 contains a nuclear export signal, so the p53-Mdm2 complex tends to be exported from the nucleus, thereby preventing p53 from activating genes.

Several interactions have been mapped to the N-terminal domain (i.e., the transactivation domain), the C-terminal domain (i.e., the regulatory domain), and the DNA binding domain (DBD) of p53 [58,71]. These domains have also been characterized in terms of their intrinsic disorder-order state, where the DNA binding domain is intrinsically structured and the terminal domains are intrinsically disordered [51,58,75,76]. Additionally, many sites of various posttranslational modifications have been identified in p53. Overall, ~70% of the interactions are mediated by IDRs in p53 [51]. A bias toward intrinsic disorder is even more pronounced in the sites of posttranslational modifications, with 86%, 90%, and 100% of observed acetylation, phosphorylation, and protein conjugation sites, respectively, found in IDRs [51,58]. Clearly, p53 extensively uses disordered regions to mediate and modulate interactions with other proteins. This is illustrated by Figure 3, which represents a set of complexes of various p53 fragments or domains with numerous binding partners.

#### *Intrinsic disorder, HPV proteins and cervical cancer*

There are more than 100 different types of human papillomaviruses (HPVs), which are the causative agents of benign papillomas/warts, and cofactors in the development of carcinomas of the genital tract, head and neck and epidermis. With respect to their association with cancer, HPVs are grouped into two classes, known as low- (e.g., HPV-6 and HPV-11) and high-risk (e.g., HPV-16 and HPV-18) types. The entire proteome of HPV includes two structural proteins, L1 and L2, and six nonstructural proteins: E1, E2, E4, E5, E6 and E7. The last two, E6 and E7, are known to function as oncoproteins in the high-risk HPVs. The correlation between the amount of ID and the ability of human papillomaviruses to cause the carcinoma development has been recently evaluated [77]. To this end, a detailed bioinformatics analysis of proteomes of high-risk and low-risk HPVs with the major focus on E6 and E7 oncoproteins was performed. The results of this analysis were consistent with the conclusion that high-risk HPVs were characterized by the increased amount of intrinsic disorder in transforming proteins E6 and E7 [77].

## IDPs in human diseases: from individual cases to general picture

Although the illustrative examples given above demonstrate the involvement of IDPs and IDRs in various diseases, and despite several more cases that are scattered in literature, all of these examples together are not sufficient to determine the extent of IDP involvement in the pathogenesis of human diseases. Simply put, how generally do IDPs and IDRs play important roles in human disease? To answer this big question, appropriate analytical tools are

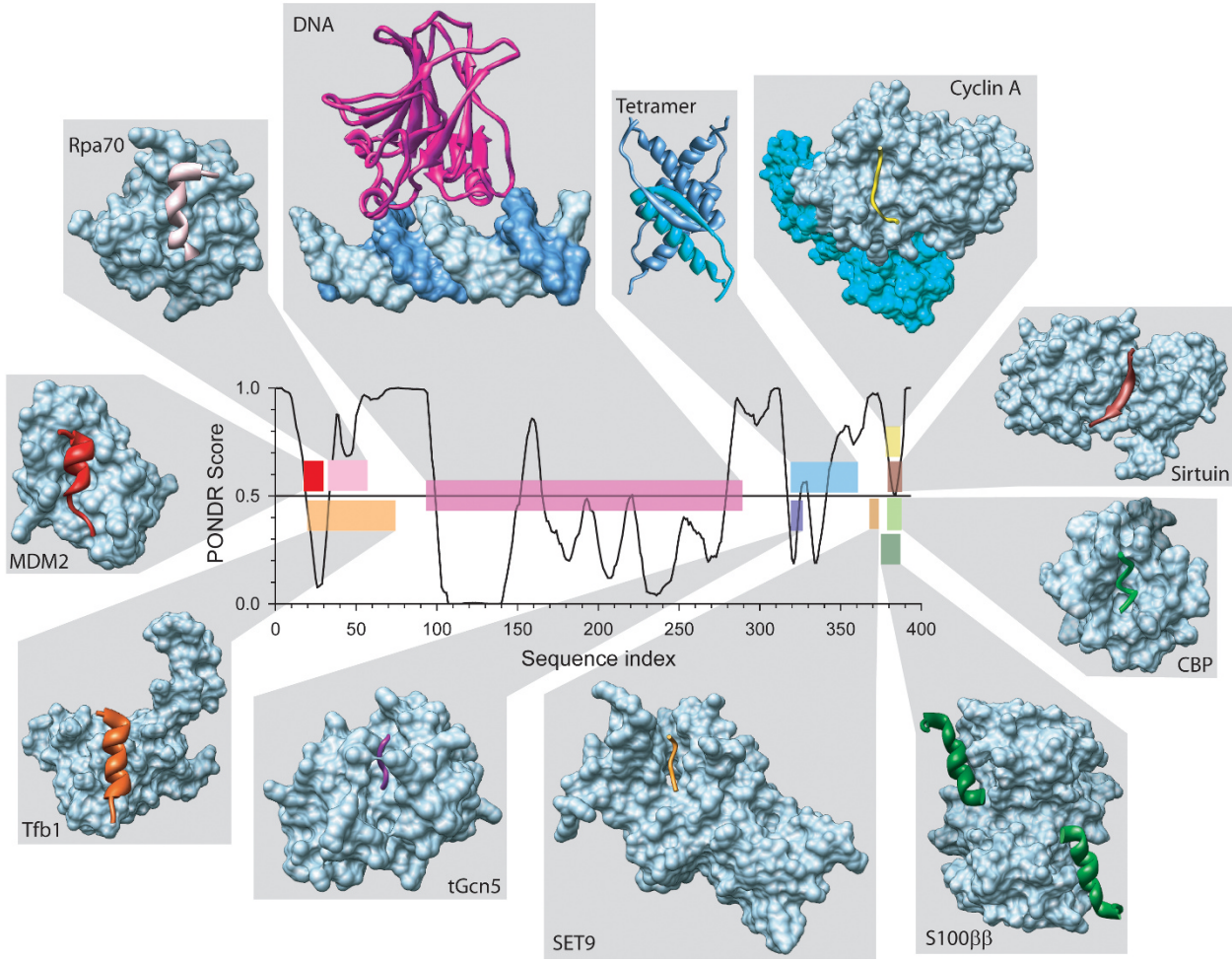


**Figure 3**

**Disorder profile and functionality of p53**. Intrinsic disorder was predicted by the PONDR® VLXT. Segments with scores above 0.5 correspond to disordered regions, while those below 0.5 correspond to ordered regions/binding sites. p53 is at the center of a large signaling network, regulating expression of genes involved in a variety of cellular processes and interacting with a large number of other proteins. The interaction sites are signaled by downward spikes in the plot of the predicted disorder. The structures of the complexes containing various p53 binding regions are displayed around the predicted disorder pattern. In complexes, the structures of p53 segments bound to their partners are shown in different colors. These color codes are also used for bars in the PONDR® VLXT plot to indicate the positions of the regions of known structure in the context of the intrinsic disorder predictions. The Protein Data Bank IDs and partner names for the structures (from upper left, clockwise) are as follows: (1tsr DNA), (1gzh 53BP1), (1q2d gcn5), (3sak p53 (tet dom)), (1xqh set9), (1h26 cyclinA), (1ma3 sirtuin), (1jsp CBP bromo domain), (1dt7 s100bb), (2h11 sv40 Large T antigen), (1ycs 53BP2), (2gs0 PH), (1ycr MDM2), and (2b3g rpa70).

needed. In a fashion similar to the history of the development of the IDP concept itself, bioinformatics is being used to determine the extent and generality of the involvement of IDPs and IDRs in human disease.

### *Unfoldomics of human diseases: tools to establish and analyze disease-related unfoldome*

#### *Unfoldome and unfoldomics*

Since IDPs are highly abundant in various diseases (see below), the "disorder in disorders" or $D^2$ concept was introduced to summarize work in this area [58]. As the number of IDPs related to various diseases is very large, it makes sense to develop the disease-related unfoldome and unfoldomics concepts.

The use of the suffix '-ome' has a long history while '-omics' is much more recent. The Oxford English Dictionary (OED) attributes 'genome' to Hans Winkler from his 1920 work [78]. While the OED suggests that 'genome' arose as a portmanteau of 'gene' and 'chromosome', this

does not seem to be supported by the literature. Instead, Lederberg and McCray suggest that, as a botanist, Winkler must have been familiar with terms such as biome (a biological community), rhizome (a root system), and phyllome (the leaves covering a tree) among others, all of which were in use well before 1920 and all of which signify the collectivity of the units involved [79]. Thus, 'ome' implies the complete set of the objects in question, with genome signifying the set of genes of an organism. By changing the 'e' in '-ome' to '-ics', the new word is created that indicates the scientific study of the '-ome' in question. For genome, the change to 'genomics' did not occur until 1987 when a journal by this name was founded by Victor McKusick and Frank Ruddle [79].

Many additional conversions from -ome to -omics have subsequently occurred and a large number of "-omes" have been accepted in biology, including but not limited to the following: genome, proteome, interactome, metabolome, transcriptome, diseasome, toxicogenome, nutrigenome, cytome, oncoproteome, epitome, and glycome, etc. For a more complete list, the reader is directed to http://omics.org. Interestingly, some of the -ome words at this website cannot be found in PubMed searches, whereas similar words can be found. For example, 'foldome' and 'foldomics' are both listed on this website, but a search of these words in PubMed yields no hits for either word, while the similar word, 'foldeomics,' yields one hit, which leads to a database containing information about protein folding. The suffixes -ome and -omics imply a new layer of knowledge, especially when a scientist is dealing with the data produced by the large-scale studies, including the high throughput experiments and the computational/bioinformatics analyses of the large datasets.

The unfoldome and unfoldomics concepts are built on the ideas given above. Unfoldomics is the field that focuses on the unfoldome. The unfoldome is the set of IDPs, which are also known as natively unfolded proteins, hence the unfoldome. We are also using unfoldome to cover segments or regions of proteins that remain unfolded in the functional state. Unfoldomics considers not only the identities of the set of proteins and protein regions in the unfoldome of a given organism, but also their functions, structures, interactions, evolution, etc. Because IDPs and IDRs are highly abundant in nature (~50% eukaryotic proteins are either entirely disordered or contain long disordered regions), have amazing structural variability and possess a very wide variety of functions, we thought it appropriate to name this realm of proteins the unfoldome, with unfoldomics reflecting the totality of the phenomena associated with IDPs and IDRs.

*Computational tools for the unfoldome analysis*

Obviously, when the scale of analysis increases from one protein to many, new analytical tools are required. The set of computational tools utilized in the bioinformatics studies on disease-related unfoldomes is briefly introduced below. This set includes compositional profiling, disorder prediction, evaluation of the number of potential binding sites, analysis of alternative splicing, and determination of posttranslational modifications.

*Compositional profiling*

A specific feature of a probable IDR is its amino-acid compositional bias characterized by a low content of so-called order-promoting residues such as Cys, Trp, Phe, Tyr, Val, Leu, and Ile, and a high content of so-called disorder-promoting residues, Glu, Lys, Arg, Asp, Gln, Ser, Pro, and Thr [7,45,80,81]. This bias can be visualized by plotting the fractional difference in composition between a given set of proteins and a set of ordered proteins [7,81]. These fractional differences in composition between the studied set and a set of ordered proteins are calculated for each amino acid residue as $(C_x\text{-}C_{order})/C_{order}$, where $C_x$ is the content of a given amino acid in the set of interest, and $C_{order}$ is the corresponding content in a set of ordered proteins. The analysis can be performed using a web Composition Profiler tool http://www.cprofiler.org/.

*Disorder predictions*

Predictions of the intrinsic disorder propensity can be performed using a set of per-residue Predictors Of Natural disordered regions (PONDR®) algorithms, PONDR® VLXT, VL3 and VSL1/2 or a set of binary predictors that predict disorder on the level of whole proteins, charge-hydrophathy plot (CH-plot) and cumulative distribution function (CDF) analysis. Many research groups have developed a number of different predictors of disorder in addition to the examples listed above. Links to many of these predictors can be found at http://www.disprot.org.

PONDR®-VLXT combines three neural networks, one for internal sequences and one for each terminus of the sequence. The internal predictor was trained on disordered sequences from only 15 proteins whose disorder was characterized by either X-ray or NMR studies [80]. The terminal predictors were trained on short regions of X-ray characterized disorder from the N- and C-terminus [82]. The merger was accomplished by performing overlapping predictions, followed by averaging the outputs. The VLXT training set included disordered segments of 40 or more amino acid residues as characterized by X-ray and NMR for the predictor of the internal regions, and segments of five or more amino acid residues for the predictors of the two terminal regions. VLXT most likely underestimates the occurrence of long disordered regions in proteins.

However, this algorithm is very important for finding potential binding sites (see below).

PONDR®-VL3 combines the predictions of 30 neural networks for the entire protein sequence and was trained using disordered regions from more than 150 proteins characterized by the methods mentioned above plus circular dichroism, limited proteolysis and other physical approaches [83]. This is one of the most accurate predictors of long disordered regions.

PONDR®-VSL1/2 is a recently developed Various Short-Long, version 1/2 (PONDR® VSL1/2) algorithm, which is an ensemble of logistic regression models that predict per-residue order-disorder [84,85]. Two models predict either long or short disordered regions – greater or less than 30 residues – based on features similar to those used by VLXT. The algorithm calculates a weighted average of these predictions, where the weights are determined by a meta-predictor that approximates the likelihood of a long disordered region within its 61-residue window. Predictor inputs include PSI-BLAST [86] profiles and PHD [87], and PSI-PRED secondary structure predictions [88].

#### *CDF analysis*

Originally, cumulative distribution function (CDF) analysis summarized the per-residue disorder predictions by plotting PONDR® VLXT scores [80,82,89] against their cumulative frequency, which allows ordered and disordered proteins to be distinguished based on the distribution of prediction scores [47]. At any given point on the CDF curve, the ordinate gives the proportion of residues with a PONDR® score less than or equal to the abscissa. The optimal boundary that provided the most accurate order-disorder classification was shown to represent seven points located in the $12^{th}$ through $18^{th}$ bin [47]. Thus, for CDF analysis, order-disorder classification is based on whether a CDF curve is above or below a majority of boundary points. Recently, CDF analysis was extended to include several other per-residue predictors of intrinsic disorder [90].

#### *CH-plot analysis*

Ordered and intrinsically unstructured proteins occupy non-overlapping regions in the charge-hydropathy plots (CH-plots), with natively unfolded proteins being specifically localized within a particular region of charge-hydropathy phase space, satisfying the following relationship [6,47]:

$$\langle H\rangle \le \langle H\rangle_b = \frac{\langle R\rangle + 1.151}{2.785},$$

where $\langle H\rangle$ and $\langle R\rangle$ are the mean hydropathy and the mean net charge of the given protein, respectively, whereas $\langle H\rangle_b$ is the "boundary" mean hydropathy value, below which a polypeptide chain with a given $\langle R\rangle$ will be most probably unfolded. The mean hydropathy, $\langle H\rangle$, is defined as the sum of the normalized hydropathy of all residues divided by the number of residues in the polypeptide. The mean net charge $\langle R\rangle$ is defined as the net charge at pH 7.0, divided by the total number of residues [6,47].

#### *α-MoRF predicitions*

The order/disorder tendencies of IDPs as revealed by PONDR® VLXT could be used to find disordered region(s) involved in interaction with specific binding partners. In fact, often IDPs have a peculiar and well-recognizable pattern, where short region of predicted order is surrounded by extended regions predicted disorder. This specific pattern was used to develop a unique bioinformatics tool dedicated to the identification of potential protein-protein interaction sites in IDPs, namely the identifier of α-helix forming Molecular Recognition Features, α-MoRF, which is focused on short binding regions within long regions of disorder that are likely to form helical structure upon binding [19,91]. The predictor utilizes a stacked architecture, where PONDR® VLXT is used to identify short predictions of order within long predictions of disorder and then a second level predictor determines whether the order prediction is likely to be a binding site based on attributes of both the predicted ordered region and the predicted surrounding disordered region. An α-MoRF prediction indicates the presence of a relatively short (20 residues), loosely structured helical region within a largely disordered sequence [19,91]. Such regions gain functionality upon a disorder-to-order transition induced by binding to partner.

#### *Alternative splicing analysis*

Alternative splicing (AS) is a process responsible for the production of multiple, mature mRNAs from a single precursor pre-mRNA by the inclusion and omission of different segments [92]. Therefore, the AS regions are defined as exons or parts of exons that are expressed in some, but not in all protein sequences transcribed from a given gene. AS is prevalent in multicellular eukaryotes [93], and it is estimated that 40 – 60% of human genes yield multiple proteins via this process [94]. These observations suggest that AS provides an important mechanism for enhancing the diversity of the proteome in multicellular eukaryotes [95]. As AS impacts many protein functions such as ligand binding, enzymatic activity, and protein-protein interactions, not surprisingly, abnormal AS has been associated with various human diseases, including myotonic dystrophy [96], axoospermia [97], Alzheimer's disease [98], cancer [99,100] and many others.

In the disease-related unfoldome, the sequence alignments of genes with multiple isoforms provide informa-

tion on the AS regions. Similarly as for a whole protein, disorder content for an AS region is estimated as the fraction of its residues that are predicted to be disordered.

*PTM analysis*

Posttranslational modifications (PTMs) are widely employed by cells to modulate the functionalities of many of their proteins. Some proteins require different types of posttranslational modifications for their function. PTMs are classified according to the mechanisms that are involved: addition of functional groups (e.g., acylation, alkylation, phosphorylation, glucosylation, etc.); attachement of other proteins and peptides (e.g., ubiquitination, SUMOylation, etc.); changing of the chemical nature of amino acids (deamidation, deimidation, oxidation, etc.); and dissection of the backbone by proteolytic cleavage. Additionally, according to the conformational state of the potential PTM site, PTMs can be grouped into two major classes. The first class involves modifications that are associated primarily with structured proteins and regions, whereas the second class combines modifications that are associated primarily with IDPs and IDRs [50]. The first class of PTMs is crucial for providing moieties for catalytic functions, for modifying enzyme activities or for stabilizing protein structure. This includes formylation, protein splicing, oxidation and covalent attachment of quinones and organic radicals [50]. The abundance of IDRs among the primary targets for the second class PTMs is likely determined by the need of the modifying enzymes to bind to their corresponding substrates via high specificity/low affinity interactions; such characteristics are typical of signaling interactions and typically involve disorder-to-order transitions of at least one of the partners [50]. Among the second class of PTMs are phosphorylation, acetylation, acylation, adenylylation, ADP ribosylation, amidation, carboxylation, formylation, glycosylation, methylation, sulfation, prenylation, ubiquitination, and Ubl-conjugation (i.e., covalent attachment of ubiquitin-like proteins, including SUMO, ISG15, Nedd8, and Atg8) [50].

As amino acid compositions, sequence complexity, hydrophobicity, charge and other sequence attributes of regions adjacent to phosphorylation sites were found to be very similar to those of IDPs and IDRs, a specific web-based tool for the prediction of protein phosphorylation sites, DISPHOS (DISorder-enhanced PHOSphorylation predictor, http://www.ist.temple.edu/DISPHOS) was elaborated [101]. Recent studies further support the view that phosphorylation occurs much more often in IDPs and IDRs as compared to structured proteins and regions [102,103]. A predictive tool similar to DISPHOS is also available for protein methylation [104]. These tools can be utilized to evaluate the abundance of PTMs in the disease-related unfoldome.

*Establishing and analyzing the disease-related unfoldomes*

Three approaches were elaborated to estimate the abundance of IDPs in various pathological conditions. The first approach is based on the assembly of specific datasets of proteins associated with a given disease and the computational analysis of these datasets using a number of disorder predictors [9,58,77,105]. In essence, this is an analysis of individual proteins extended to a set of independent proteins. A second approach utilized network of genetic diseases where the related proteins are interlinked within one disease and between different diseases [106]. A third approach is based on the evaluation of the association between a particular protein function (including the disease-specific functional keywords) with the level of intrinsic disorder in a set of proteins known to carry out this function [48-50]. These three approaches are briefly described below, whereas the results of their application are presented in the subsequent section.

*Simple dataset analysis*

The simplest analysis of the abundance of intrinsic disorder in a given disease is based on the two-stage protocol, where a set of related proteins is first assembled by searching various databases and then the collected group of proteins is analyzed for intrinsic disorder. The depth of this analysis is based on the breadth of the search for the disease-related proteins and on the number of different computational tools utilized to find disordered proteins/regions. For example, a dataset of human cancer-associated proteins (HCAP) extracted from SWISS-PROT http://www.expasy.ch/sprot using keywords Anti-oncogene; Oncogene; Proto-oncogene; tumor in the description field and "human" in the organism field contained 231 proteins [9]. Whereas 487 proteins associated with cardiovascular disease (CVD) were found in SWISS-PROT using an exhaustive list of CVD-related keywords: Aneurysm; Angina Pectoris; Angioneurotic Edema; Aortic Valve Stenosis; Arrhythmia; Arrhythmogenic; Arteriosclerosis; Arteriovenous Malformations; Atrial Fibrillation; Behcet Syndrome; Bradycardia; Cardiac Tamponade; Cardiomegaly; Cardiomyopathy; Cardiovascular Disease; Carotid Stenosis; Cerebral Hemorrhage; Churg-Strauss Syndrome; Ebstein's Anomaly; Eisenmenger Complex; Embolism; Cholesterol; Endocarditis; Fibromuscular Dysplasia; Heart Block; Heart Defects; Heart Disease; Heart Failure; Heart Valve Diseases; Hematoma; Hippel Lindau Disease; Hyperemia; Hypertension; Hypertrophy; Hypoplastic Left Heart Syndrome; Hypotension; Intermittent Claudication; Klippel-Trenaunay-Weber Syndrome; Lateral Medullary Syndrome; Long QT Syndrome; Microvascular AnginaOR Mitral Valve Prolapse; Moyamoya Disease; Mucocutaneous Lymph Node Syndrome; Myocardial Infarction; Myocardial Ischemia; Myocarditis; Pericarditis; Peripheral Vascular Diseases; Phlebitis; Polyarteritis Nodosa; Pulmonary Atresia; Raynaud Disease; Sneddon

Syndrome; Superior Vena Cava Syndrome; Tachycardia; Takayasu's Arteritis; Telangiectasia; Telangiectasis; Temporal Arteritis; Tetralogy of Fallot; Thromboangiitis Obliterans; Thrombosis; Tricuspid Atresia; Varicose Veins; Vascular Disease; Vasculitis; Vasospasm; Ventricular Fibrillation; Williams Syndrome; Wolff-Parkinson-White Syndrome; Heart disease; Stroke; Thromb; Cardio-vascular disease; Blood coagulation; Heart muscle; Cardiovascular disease; Plasma; Vascular disease in the description field and "human" in the organism field [105]. The intrinsic disorder analysis in the assembled datasets of disease-related proteins includes various computational tools described in a previous section.

#### *Functional keyword analysis*

A computational tool for the evaluation of a correlation between the functional annotations in the SWISSPROT database and the predicted intrinsic disorder was elaborated [48-50]. First, functional keywords associated with 20 or more proteins in SWISSPROT were determined and corresponding protein datasets were assembled. Then, for each keyword-associated set, a length-matching set of random proteins was drawn from the SWISSPROT. Order-disorder predictions were carried out for the keyword-associated sets and for the random sets. If a function described by a given keyword were carried out by a long region of disordered protein, one would expect the keyword-associated set to have a greater amount of predicted disorder compared to the random set. The keyword-associated set would have less prediction of disorder compared to the random set if the keyword-associated function were carried out by structured protein. Given the two sets of predictions for the pairs of sets, it is possible to calculate the p-values, where a p-value > 0.95 suggests a disorder-associated function, a p-value < 0.05 suggests an order-associated function, and intermediate p-values are ambiguous [48-50].

#### *Genetic diseasenNetwork analysis*

To estimate whether human genetic diseases and the corresponding disease genes are related to each other at a higher level of cellular and organism organization, a bipartite graph was utilized in a dual way: to represent a network of genetic diseases, the "human disease network", HDN, where two diseases are directly linked if there is a gene that is directly related to both of them, and a network of disease genes, the "disease gene network", DGN, where two genes are directly linked if there is a disease to which they are both directly related [107]. This framework, called the human diseasome, systematically linked the human disease phenome (which includes all the human genetic diseases) with the human disease genome (which contains all the disease-related genes). This diseasome opened a new avenue for the analysis and understanding of human genetic diseases, moving from single gene-single disease viewpoint to a framework-based approach [107].

Using this approach various diseases were classified into 20 types, some diseases were unclassified, and several diseases were annotated as belonging to multiple classes. Similarly, genes were clustered into classes via their associations with specific diseases [107]. The analysis of these networks revealed that of 1,284 genetic diseases, 867 had at least one link to other diseases, and 516 diseases formed a giant component, suggesting that the genetic origins of most diseases, to some extent, were shared with other diseases. Similalrly, in the DGN, 1,377 of 1,777 disease genes were shown to be connected to other disease genes, and 903 genes belonged to a giant cluster [107]. The vast majority of genes associated with genetic diseases was non-essential and showed no tendency to encode hub proteins. In fact, many of the disease-related genes were shown to be localized in the functional periphery of the network [107]. The large-scale analysis of the abundance of intrinsic disorder in transcripts of the various disease-related genes was performed using a set of computational tools described in a previous section [106]. The results of this analysis suggest that IDPs are broadly involved in human diseases (see below).

### *IDPs in cancer, CVD, neurodegenerative diseases and diabetes*

For the first time, the dataset analysis approach was used in 2002 [9], when significant fractions of cancer-associated and cell-signaling proteins were found to contain predicted IDRs of 30 residues or longer (see Figure 4). This was in a sharp contrast to a set of structured (ordered) proteins with well-defined 3-D structures, which was shown to contain only 13% of the proteins with predicted IDRs ≥ 30 residues. Following a similar analytical model, a dataset of 487 proteins related to cardiovascular disease (CVD) was collected and analyzed [105]. On average, CVD-related proteins were found to be highly disordered. They were depleted in major order-promoting residues (Trp, Phe, Tyr, Ile, and Val) and enriched in some disorder-promoting residues (Arg, Gln, Ser, Pro, and Glu). High level of intrinsic disorder and a substantial number of potential interaction sites were also found using a set of computational tools. The percentage of proteins with 30 or more consecutive disordered residues was ~60% for CVD-associated proteins (see Figure 4). Many proteins were predicted to be wholly disordered, with 101 proteins from the CVD dataset predicted to have a total of almost 200 specific disorder-based binding motifs (thus about 2 binding sites per protein). These binding sites are called α-helical molecular recognition features, α-MoRFs, and have been well studied from protein complexes taken from PDB [105]. All of this clearly suggested that IDPs might play key roles in CVD.

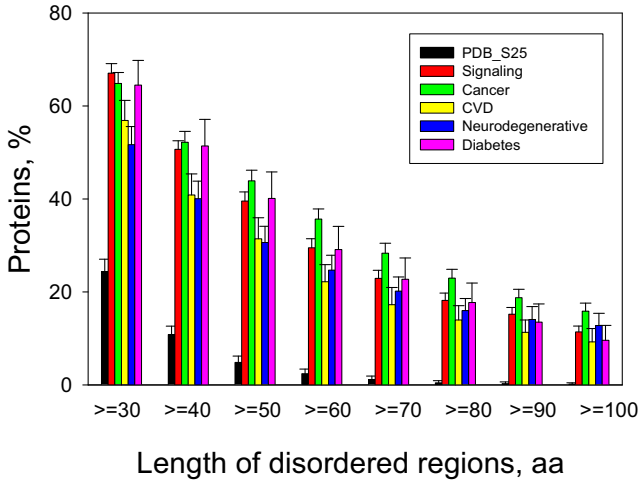


**Figure 4**
**Abundance of intrinsic disorder in disease-associated proteins**. Percentages of disease associated proteins with ≥ 30 to ≥ 100 consecutive residues predicted to be disordered. The error bars represent 95% confidence intervals and were calculated using 1,000 bootstrap re-sampling. Corresponding data for signaling and ordered proteins are shown for the comparison. Analyzed sets of disease-related proteins included 1786, 487, 689, and 285 proteins for cancer, CVD, neurodegenerative disease and diabetes, respectively.

In addition to being abundant in cancer- and CVD-related proteins, intrinsic disorder was commonly found in such maladies as neurodegenerative diseases and diabetes. Figure 4 represents this as the percentages of proteins with ≥ 30 consecutive residues predicted to be disordered in datasets of proteins associated with all four diseases. This figure shows that intrinsic disorder is highly prevalent in proteins associated with all of the studied diseases, being comparable with that of signaling proteins and significantly exceeding the levels of intrinsic disorder in eukaryotic and in non-homologous, structured proteins [58].

### *Functional anthology of intrinsic disorder and human diseases*

The application of the functional keyword analysis tool revealed that out of 710 SWISSPROT keywords each being assigned to at least 20 proteins, 310 had p-values < 0.05, suggesting order-associated functions, 238 had p-values > 0.95, suggesting disorder-associated functions, and the remainder, 162, gave intermediate p-values, yielding ambiguity in the likely function-structure associations [48-50].

When the functional keywords were partitioned into eleven functional categories (Biological processes, cellular components, developmental stage, etc.) order-associated keywords were found for seven of the categories, but disorder-associated keywords were found for all eleven categories [48]. Figure 5 represents the results of this analysis and show that many diseases were strongly correlated with proteins predicted to be disordered. Contrary to this, we did not find disease-associated proteins to be strongly correlated with absence of disorder [50]. Among disease-related Swiss-Prot keywords strongly associated with intrinsic disorder were oncoproteins, malaria, trypanosomiasis, human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS), deafness, obesity, cardiovascular disease, diabetes mellitus, albinism, and prion [50]. In agreement with this bioinformatics analysis, we were able to find at least one illustrative, experimentally validated example of functional disorder or order for the vast majority of functional keywords related to diseases [50].

### *Intrinsic disorder in proteins from the genetic disease network*

The dual Human Disease Network/Disease Gene Network (HDN/DGN) consists of two types of nodes that represent human genes (1,777) and diseases (1,284), and links that connect diseases with related genes [107]. A set of disease genes from DGN with human genes with known protein sequences was used to collect protein sequences for all human genes from NCBI Gene database [106]. All model proteins obtained solely with automated genome annotation processing were excluded from the consideration. After this exclusion, the diseasome included 1,751 human disease related genes. The transcripts of the genetic dis-

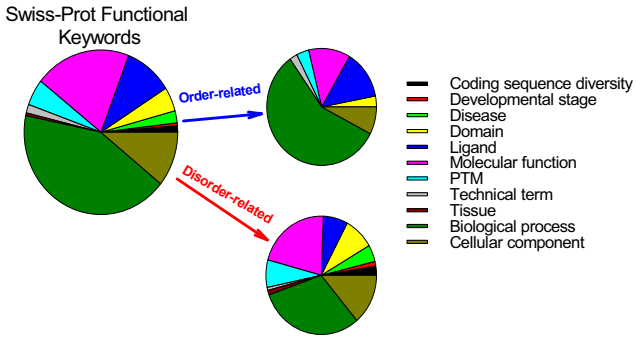


**Figure 5**
**Functional anthology of IDPs**. More than 200,000 proteins and 710 SWISSPROT functional keywords each associated with at least 20 different proteins were analyzed [48-50]. Based on the bioinformatics analysis, 238 keywords were associated with the predicted intrinsic disorder. These keywords covered various functions and included almost all disease-related keywords. This is in a strict contrast to 302 keywords which were associated with the predicted order. Functionally, the vast majority of these keywords were various "ases". They contained almost no disease-related keywords.

ease-associated genes were compared with proteins encoded by 16,358 other human genes with known protein sequences [106].

The abundance of intrinsic disorder in these diseasome network proteins was evaluated by means of several prediction algorithms, including PONDR® VSL2, CDF-analysis, CH-plot [106]. The functional repertoires of these proteins were analyzed based on prior studies relating disorder to function [48-50]. These analyses uncovered an unfoldome associated with human genetic diseases and revealed several interesting peculiarities [106]:

(i) Intrinsic disorder is common in proteins associated with many human genetic diseases;

(ii) Different disease classes vary significantly in the IDP contents of their associated proteins;

(iii) Molecular recognition features, which are relatively short loosely structured protein regions within mostly disordered sequences and which gain structure upon binding to partners, are common in the diseasome, and their abundance correlates with the intrinsic disorder level;

(iv) Some disease classes have a significantly higher fraction of genes affected by alternative splicing, and the alternatively spliced regions in the corresponding proteins are predicted to be highly disordered and in some disease classes contain a significantly higher number of MoRFs;

(v) Correlations were found among the various diseasome graph-related properties and intrinsic disorder. In agreement with earlier studies, hub proteins were shown to be more disordered.

### *Why the unfoldome and why IDPs?*

All the data presented above provide evidence that IDPs are very common in various diseases and therefore comprise a disease-related unfoldome. The introduction of the unfoldome and unfoldomics concepts pave the way for a better understanding of the molecular aspects of human diseases, including a better understanding of their pathogenesis and molecular mechanisms. This concept is also important for the development of the appropriate strategies dedicated to the targeted analysis of the disease-related proteins. As many of these proteins are either completely disordered or contain long disordered regions, it would be a clear mistake to analyze them using only the experimental tools developed for the characterization of structured proteins. The appropriate conformational analyses should utilize the fact that IDPs and IDRs possess a range of structural properties that are quite different from those of ordered proteins [6,7,11,12,14,21]. The techniques used for such analysis were described in a recent review [57]. Some of these techniques are briefly considered below.

(i) Although X-ray crystallography is traditionally used to characterize structure of ordered proteins, it repeatedly defines missing electron density in many protein structures, which may correspond to disordered region(s). The increased flexibility of atoms in the IDR leads to the non-coherent X-ray scattering, making them unobserved or at least smearing out their electron densities. Missing regions of structure can be structured but wobbly domains rather than disordered regions, and so further studies on X-ray identified IDRs using other methods is very important.

(ii) A solution-based counterpart of X-ray crystallography is heteronuclear multidimensional NMR. This is an extremely powerful technique for protein 3D-structure determination in solution and for the characterization of protein dynamics. Recent advances in this technology have allowed the complete assignment of resonances for several unfolded and partially folded proteins, as well as for the several IDPs and IDRs.

(iii) Circular dichroism (CD) is another powerful technique for the evaluation of the overall tertiary structure of a protein. CD spectra in the near UV region (250–350 nm) reflect the asymmetry of the environment of aromatic amino acid residues and, consequently, are characteristic of protein tertiary structure. IDPs may be detected by their display of simplified near-UV CD spectra.

(iv) Decreased content of ordered secondary structure in IDPs may be detected by several spectroscopic techniques including far-UV CD, optical rotary dispersion (ORD), Fourier transform infrared spectroscopy (FTIR), Raman optical activity and deep UV Raman spectroscopy.

(v) Hydrodynamic parameters obtained from techniques such as gel-filtration, viscometry, small angle X-ray or neutron scattering (SAXS or SANS, respectively), sedimentation, dynamic and static light scattering may help in determining the degree of a polypeptide chain compaction.

(vi) Another very important structural parameter is the degree of globularity, which reflects the presence or absence of a tightly packed core in a protein molecule. This information may be extracted from the analysis of SAXS data in form of a Kratky plot, the shape of which is sensitive to the conformational state of the scatter-

ing protein molecules. The Kratky plot of a globular molecule (ordered or molten globular) has a characteristic maximum, which is absent from the Kratky plot of a coil-like or pre-molten globule-like IDP.

(vii) Different fluorescence characteristics provide a wealth of knowledge on the intramolecular mobility and compactness of a protein. This includes FRET, shape and position of the intrinsic fluorescence spectrum, fluorescence anisotropy and lifetime, accessibility of the chromophore groups to external quenchers, and steady state and time-resolved parameters of the fluorescent dyes.

(viii) Increased proteolytic degradation *in vitro* of IDPs and IDRs indirectly confirms their increased flexibility.

(ix) Protein disorder may also be evaluated by immunochemical methods or via the interaction with molecular chaperones.

(x) Finally, IDPs may be detected by their response to the environmental changes or via the analysis of protein conformational stability.

(xi) Aberrant mobility during the SDS-PAGE gel electrophoresis may be suggestive of intrinsic disorder since disordered proteins usually migrate slower than their calculated molecular weight.

As discussed above, IDPs and IDRs can be characterized by a variety of biophysical and biochemical methods. As a result, a very large number of disease-associated proteins have been experimentally shown to be IDPs or to contain IDRs as indicate by the illustrative examples at the beginning of this article. This leads naturally to the following question: from a biological perspective, why have such proteins been so heavily linked to human diseases? To answer this question, some specific features of IDPs that potentially make them key players in the development of pathological conditions need to be considered. Many of these features are linked to the function of IDPs in signaling, regulation and control. The list of these features includes [24]:

(i) Decoupled specificity and strength of binding leading to high-specificity-low-affinity interactions;

(ii) Increased speed of interaction due to greater capture radius and the ability to spatially search interaction space;

(iii) Flexible encounter complexes (less stringent spatial orientation requirements);

(iv) Controlled regulation via high sensitivity to proteolytic degradation when in the free state;

(v) Increased interaction (surface) area per residue;

(vi) A one-to-many binding mode and binding promiscuity by which a single IDP/IDR binds to multiple structurally diverse partners. This is accomplished by plasticity, by which a given IDR folds into distinctive conformations to accommodate the diverse binding sites of its different partners

(vii) A many-to-one binding mode, by which many different IDPs/IDRs bind to one site on a single ordered partner. Again this is accomplished by plasticity, by which different IDRs fold into similar conformations that all fit into a single binding site on one partner.

(viii) Induced folding where an IDR folds as it binds to a specific partner;

(ix) Low steric restrictions allowing the elongation or contraction of a given binding area;

(x) Ease of regulation or reorganization of signaling networks by posttranslational modification;

(xi) Ease of regulation or reorganization of signaling networks by alternative splicing;

(xii) Overlapping of binding sites due to use of extended linear conformations for association;

(xiii) High evolutionary rates leading to rapid adaptability and easy modification of signaling networks;

(xiv) Flexibility that allows masking (or not) of interaction sites or that allows multiple interactions between bound partners.

***Induced folding, binding promiscuity, and binding plasticity***

Protein-protein and protein-nucleic acid interactions are central to many processes in molecular biology. They often involve coupled folding and binding of at least one of the partners [4,6-8,10,13,17,22,108,109]. Among the list of structural features that make IDPs especially useful for their signaling and regulation functions include induced folding, binding promiscuity, and binding plasticity. The p53 protein molecule represents an especially dramatic example for which intrinsic disorder is heavily utilized for function via induced folding, binding promiscuity (i.e., the ability of a given IDP to bind interact with several binding partners), and binding plasticity (which is determined as the ability of a given IDR to gain different

folds to accommodate diverse binding sites of different partners). As it has been already mentioned, p53 regulates expression of over 150 genes and binds to over 100 proteins [51,70,71]. These many interactions represent an illustrative example of the one-to-many biding mode [51]. The 3-D structures of several complexes between the various p53 regions and unique binding partners have been determined (see Figure 3). The interactions with 10 of these partners are mediated by region experimentally characterized as IDRs. Figure 3 shows that PONDR® VLXT is able to detect the majority of these binding regions as short predictions of order within a longer prediction of disorder. These structures are complexes between p53 and: cyclin A, sirtuin, CBP, S100ββ, set9, tGcn5, Rpa70, Mdm2, Tfb1, and itself. The remaining 4 interactions are mediated by the structured DBD, between p53 and: DNA, 53BP1, sv40 Large T antigen, and 53BP2 [51].

Of special interest is the C-terminal regulatory domain, which is involved in the formation of multiple complexes. Figure 3 shows that a single IDR of p53 derived from the C-terminal regulatory domain (residues 374–388) was observed to form all three major secondary structure types in the bound state: a helix when associating with S100ββ, a sheet with sirtuin, an irregular structure with CBP, and an irregular structure with a completely different trajectory with cyclin A2. The set of residues involved in these interactions exhibit a very high extent of overlap along the sequence [51]. Based on the fact that the secondary structures adopted by this IDR in different complexes were very distinct, it seemed reasonable to expect that p53 utilizes different residues for the interactions with these four different binding partners. This hypothesis is supported via the quantification of the buried surface area for each residue in each interaction by calculating their ΔASA [51]. In fact, the ΔASA-based binding profiles for the single IDR of p53 bound to four different partners were completely different, indicating that the same residues were used to different extents in the four interfaces, suggesting that the same IDR sequence is "read" by the different partners in entirely different ways [51].

This intriguing p53 example demonstrates the roles of IDRs in determining multiple specificities associated with the one-to-many binding mode, where remarkable conformational changes enable very distinct surfaces to be formed for binding to different partners. The mentioned interactions of the C-terminal regulatory domain of p53 with various binding partners are used for the activation or inhibition of its primary role as a transcription regulator. Therefore, it is possible that the disordered binding regions may play a passive regulatory role by providing a specific binding site, where IDRs serve as the identification sites of the protein to be regulated [13,51].

Because p53 is so heavily studied, we have learned about the use of IDPs and IDRs for its functions, especially as providing sites for protein-protein interactions, before we have gained such knowledge for other signaling proteins. However, sites of protein-protein interactions that are located within IDPs and IDRs and that are very similar to those observed for p53 are predicted to be extremely common in the proteins of mammalian proteomes [91]. Thus, what we have presented above for p53 likely provides a blueprint for the use of IDPs and IDRs for a very large number of proteins in the cell.

Overall, there are intriguing interconnections among intrinsic disorder, cell signaling and human diseases, suggesting that protein conformational diseases may result not only from protein misfolding, but also from misidentification and missignaling.

## Concluding remarks

Intrinsic disorder is highly abundant among proteins associated with various human diseases. This conclusion is based on the detailed analysis of several well-characterized disease-related IDPs and on the results of the extensive bioinformatics studies. As the number of disease-related IDPs is very large and as many of these proteins are interlinked, the concepts of the disease-related unfoldome and unfoldomics were introduced. Here, the disease-related unfoldome is attributed to a significant part of human proteome, which includes malady-associated IDPs, their functions, structures, interactions, evolution, etc. We believe that the unfoldomics concept helps lead to better understanding of various human diseases, their pathogenesis and molecular mechanisms. This concept might also help in the development of specialized strategies for the targeted analysis of functional and structural properties of disease-related proteins. The high degree of association between intrinsic disorder and many proteins implicated in various maladies is due to structural and functional peculiarities of IDPs and IDRs, which are typically involved in cellular regulation, recognition and signal transduction. One of the promising future developments in the field of the disease-related unfoldome and unfoldomics is the evaluation of IDP/IDR abundance in the framework of disease ontology. However, since the corresponding resources are not ready yet, such an analysis might be difficult at the current point.

## Competing interests

The authors declare that they have no competing interests.

## Authors' contributions

VNU was involved in design and planning of all the experiments, drafted the manuscript, revised the final version and headed the project. CJO, BX, UM, HX, SV and LMI performed the computational analysis, designed figures

and contributed to the manuscript writing. ZO and AKD were involved in design and planning of all the experiments and contributed to the manuscript writing. All authors have read and approved the final manuscript.

## Additional material

### Additional file 1

*Human neurodegenerative disorders characterized by the presence of the α-synuclein deposits.*
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2164-10-S1-S7-S1.doc]

## Acknowledgements

This work was supported in part by the grants R01 LM007688-01A1 (to A.K.D and V.N.U.) and GM071714-01A2 (to A.K.D and V.N.U.) from the National Institutes of Health and the Program of the Russian Academy of Sciences for the "Molecular and cellular biology" (to V. N. U.). L.M.I. was supported by NIH NCI 1R21CA113711 and NSF 0444818. We gratefully acknowledge the support of the IUPUI Signature Centers Initiative.

This article has been published as part of *BMC Genomics* Volume 10 Supplement 1, 2009: The 2008 International Conference on Bioinformatics & Computational Biology (BIOCOMP'08). The full contents of the supplement are available online at http://www.biomedcentral.com/1471-2164/10?issue=S1.

## References

1. Wu H: **Studies on denaturation of proteins XIII A theory of denaturation.** *Chin J Physiol* 1931, **1:**219-234.
2. Mirsky AE, Pauling L: **On the structure of native, denatured and coagulated proteins.** *Proc Natl Acad Sci USA* 1936, **22:**439-447.
3. Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Garner E, Guilliot S, Dunker AK: **Thousands of proteins likely to have long disordered regions.** *Pac Symp Biocomput* 1998:437-448.
4. Wright PE, Dyson HJ: **Intrinsically unstructured proteins: reassessing the protein structure-function paradigm.** *J Mol Biol* 1999, **293(2):**321-331.
5. Dunker AK, Obradovic Z, Romero P, Garner EC, Brown CJ: **Intrinsic protein disorder in complete genomes.** *Genome Inform Ser Workshop Genome Inform* 2000, **11:**161-171.
6. Uversky VN, Gillespie JR, Fink AL: **Why are "natively unfolded" proteins unstructured under physiologic conditions?** *Proteins* 2000, **41(3):**415-427.
7. Dunker AK, Lawson JD, Brown CJ, Williams RM, Romero P, Oh JS, Oldfield CJ, Campen AM, Ratliff CM, Hipps KW, *et al.*: **Intrinsically disordered protein.** *J Mol Graph Model* 2001, **19(1):**26-59.
8. Dunker AK, Obradovic Z: **The protein trinity – linking function and disorder.** *Nat Biotechnol* 2001, **19(9):**805-806.
9. Iakoucheva LM, Brown CJ, Lawson JD, Obradovic Z, Dunker AK: **Intrinsic disorder in cell-signaling and cancer-associated proteins.** *J Mol Biol* 2002, **323(3):**573-584.
10. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z: **Intrinsic disorder and protein function.** *Biochemistry* 2002, **41(21):**6573-6582.
11. Uversky VN: **Natively unfolded proteins: a point where biology waits for physics.** *Protein Sci* 2002, **11(4):**739-756.
12. Uversky VN: **What does it mean to be natively unfolded?** *Eur J Biochem* 2002, **269(1):**2-12.
13. Tompa P: **Intrinsically unstructured proteins.** *Trends Biochem Sci* 2002, **27(10):**527-533.
14. Uversky VN: **Protein folding revisited. A polypeptide chain at the folding-misfolding-nonfolding cross-roads: which way to go?** *Cell Mol Life Sci* 2003, **60(9):**1852-1871.
15. Tompa P: **Intrinsically unstructured proteins evolve by repeat expansion.** *Bioessays* 2003, **25(9):**847-855.
16. Tompa P, Csermely P: **The role of structural disorder in the function of RNA and protein chaperones.** *Faseb J* 2004, **18(11):**1169-1175.
17. Dyson HJ, Wright PE: **Intrinsically unstructured proteins and their functions.** *Nat Rev Mol Cell Biol* 2005, **6(3):**197-208.
18. Uversky VN, Oldfield CJ, Dunker AK: **Showing your ID: intrinsic disorder as an ID for recognition, regulation and cell signaling.** *J Mol Recognit* 2005, **18(5):**343-384.
19. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK: **Coupled folding and binding with alpha-helix-forming molecular recognition elements.** *Biochemistry* 2005, **44(37):**12454-12470.
20. Dunker AK, Cortese MS, Romero P, Iakoucheva LM, Uversky VN: **Flexible nets: The roles of intrinsic disorder in protein interaction networks.** *FEBS Journal* 2005, **272(20):**5129-5148.
21. Daughdrill GW, Pielak GJ, Uversky VN, Cortese MS, Dunker AK: **Natively disordered proteins.** In *Handbook of Protein Folding* Edited by: *Buchner J, Kiefhaber T. Weinheim, Germany: Wiley-VCH, Verlag GmbH & Co*; 2005:271-353.
22. Tompa P: **The interplay between structure and function in intrinsically unstructured proteins.** *FEBS Lett* 2005, **579(15):**3346-3354.
23. Fink AL: **Natively unfolded proteins.** *Curr Opin Struct Biol* 2005, **15(1):**35-41.
24. Cortese MS, Uversky VN, Keith Dunker A: **Intrinsic disorder in scaffold proteins: getting more from less.** *Prog Biophys Mol Biol* 2008, **98(1):**85-106.
25. Dunker AK, Silman I, Uversky VN, Sussman JL: **Function and structure of inherently disordered proteins.** *Curr Opin Struct Biol* 2008.
26. Schweers O, Schonbrunn-Hanebeck E, Marx A, Mandelkow E: **Structural studies of tau protein and Alzheimer paired helical filaments show no evidence for beta-structure.** *J Biol Chem* 1994, **269(39):**24290-24297.
27. Weinreb PH, Zhen W, Poon AW, Conway KA, Lansbury PT Jr: **NACP, a protein implicated in Alzheimer's disease and learning, is natively unfolded.** *Biochemistry* 1996, **35(43):**13709-13715.
28. Crick SL, Jayaraman M, Frieden C, Wetzel R, Pappu RV: **Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions.** *Proc Natl Acad Sci USA* 2006, **103(45):**16764-16769.
29. Tran HT, Mao A, Pappu RV: **Role of backbone-solvent interactions in determining conformational equilibria of intrinsically disordered proteins.** *J Am Chem Soc* 2008, **130(23):**7380-7392.
30. Dolgikh DA, Gilmanshin RI, Brazhnikov EV, Bychkova VE, Semisotnov GV, Venyaminov S, Ptitsyn OB: **Alpha-Lactalbumin: compact state with fluctuating tertiary structure?** *FEBS Lett* 1981, **136(2):**311-315.
31. Bychkova VE, Ptitsyn OB: **The molten globule in vitro and in vivo.** *Chemtracts Biochem Molec Biol* 1993, **4:**133-163.
32. Ptitsyn OB: **Kinetic and equilibrium intermediates in protein folding.** *Protein Eng* 1994, **7(5):**593-596.
33. Ptitsyn OB: **Molten globule and protein folding.** *Adv Protein Chem* 1995, **47:**83-229.
34. Ptitsyn OB, Bychkova VE, Uversky VN: **Kinetic and equilibrium folding intermediates.** *Philos Trans R Soc Lond B Biol Sci* 1995, **348(1323):**35-41.
35. Kuwajima K: **The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure.** *Proteins* 1989, **6(2):**87-103.
36. Kuwajima K: **The molten globule state of alpha-lactalbumin.** *Faseb J* 1996, **10(1):**102-109.
37. Christensen H, Pain RH: **Molten globule intermediates and protein folding.** *Eur Biophys J* 1991, **19(5):**221-229.
38. Fink AL: **Compact intermediate states in protein folding.** *Annu Rev Biophys Biomol Struct* 1995, **24:**495-522.
39. Uversky VN, Ptitsyn OB: **"Partly folded" state, a new equilibrium state of protein molecules: four-state guanidinium chloride-induced unfolding of beta-lactamase at low temperature.** *Biochemistry* 1994, **33(10):**2782-2791.
40. Uversky VN, Ptitsyn OB: **Further evidence on the equilibrium "pre-molten globule state": four-state guanidinium chloride-**

**induced unfolding of carbonic anhydrase B at low temperature.** *J Mol Biol* 1996, **255(1):**215-228.
41. Uversky VN, Karnoup AS, Segel DJ, Seshadri S, Doniach S, Fink AL: **Anion-induced folding of Staphylococcal nuclease: characterization of multiple equilibrium partially folded intermediates.** *J Mol Biol* 1998, **278(4):**879-894.
42. Radivojac P, Iakoucheva LM, Oldfield CJ, Obradovic Z, Uversky VN, Dunker AK: **Intrinsic disorder and functional proteomics.** *Biophys J* 2007, **92(5):**1439-1456.
43. Romero P, Obradovic Z, Kissinger CR, Villafranca JE, Dunker AK: **Identifying disordered regions in proteins from amino acid sequences.** *IEEE Int Conf Neural Netw* 1997, **1:**90-95.
44. Williams RM, Obradovi Z, Mathura V, Braun W, Garner EC, Young J, Takayama S, Brown CJ, Dunker AK: **The protein non-folding problem: amino acid determinants of intrinsic order and disorder.** *Pac Symp Biocomput* 2001:89-100.
45. Campen A, Williams RM, Brown CJ, Uversky VN, Dunker AK: **TOP-IDP-Scale: A new amino acid scale measuring propensity for intrinsic disorder.** *Protein and Peptide Letters* 2008, **15(9):**956-963.
46. Ward JJ, Sodhi JS, McGuffin LJ, Buxton BF, Jones DT: **Prediction and functional analysis of native disorder in proteins from the three kingdoms of life.** *J Mol Biol* 2004, **337(3):**635-645.
47. Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, Dunker AK: **Comparing and combining predictors of mostly disordered proteins.** *Biochemistry* 2005, **44(6):**1989-2000.
48. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Uversky VN, Obradovic Z: **Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions.** *J Proteome Res* 2007, **6(5):**1882-1898.
49. Vucetic S, Xie H, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN: **Functional anthology of intrinsic disorder. 2. Cellular components, domains, technical terms, developmental processes, and coding sequence diversities correlated with long disordered regions.** *J Proteome Res* 2007, **6(5):**1899-1916.
50. Xie H, Vucetic S, Iakoucheva LM, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN: **Functional anthology of intrinsic disorder. 3. Ligands, post-translational modifications, and diseases associated with intrinsically disordered proteins.** *J Proteome Res* 2007, **6(5):**1917-1932.
51. Oldfield CJ, Meng J, Yang JY, Yang MQ, Uversky VN, Dunker AK: **Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners.** *BMC Genomics* 2008, **9(Suppl 1):**S1.
52. Patil A, Nakamura H: **Disordered domains and high surface charge confer hubs with the ability to interact with multiple proteins in interaction networks.** *FEBS Lett* 2006, **580(8):**2041-2045.
53. Haynes C, Oldfield CJ, Ji F, Klitgord N, Cusick ME, Radivojac P, Uversky VN, Vidal M, Iakoucheva LM: **Intrinsic disorder is a common feature of hub proteins from four eukaryotic interactomes.** *PLoS Comput Biol* 2006, **2(8):**e100.
54. Dosztanyi Z, Chen J, Dunker AK, Simon I, Tompa P: **Disorder and sequence repeats in hub proteins and their implications for network evolution.** *J Proteome Res* 2006, **5(11):**2985-2995.
55. Ekman D, Light S, Bjorklund AK, Elofsson A: **What properties characterize the hub proteins of the protein-protein interaction network of Saccharomyces cerevisiae?** *Genome Biol* 2006, **7(6):**R45.
56. Singh GP, Ganapathi M, Dash D: **Role of intrinsic disorder in transient interactions of hub proteins.** *Proteins* 2007, **66(4):**761-765.
57. Receveur-Brechot V, Bourhis JM, Uversky VN, Canard B, Longhi S: **Assessing protein disorder and induced folding.** *Proteins* 2006, **62(1):**24-45.
58. Uversky VN, Oldfield CJ, Dunker AK: **Intrinsically disordered proteins in human diseases: introducing the D2 concept.** *Annu Rev Biophys* 2008, **37:**215-246.
59. Uversky VN: **A protein-chameleon: conformational plasticity of alpha-synuclein, a disordered protein involved in neurodegenerative disorders.** *J Biomol Struct Dyn* 2003, **21(2):**211-234.
60. Uversky VN: **Neuropathology, biochemistry, and biophysics of alpha-synuclein aggregation.** *J Neurochem* 2007, **103(1):**17-37.
61. Uversky VN: **Alpha-synuclein misfolding and neurodegenerative diseases.** *Curr Protein Pept Sci* 2008, **9(5):**507-540.
62. Goedert M: **Filamentous nerve cell inclusions in neurodegenerative diseases: tauopathies and alpha-synucleinopathies.** *Philos Trans R Soc Lond B Biol Sci* 1999, **354(1386):**1101-1118.
63. Spillantini MG, Goedert M: **The alpha-synucleinopathies: Parkinson's disease, dementia with Lewy bodies, and multiple system atrophy.** *Ann N Y Acad Sci* 2000, **920:**16-27.
64. Galvin JE, Lee VM, Trojanowski JQ: **Synucleinopathies: clinical and pathological implications.** *Arch Neurol* 2001, **58(2):**186-190.
65. Trojanowski JQ, Lee VM: **Parkinson's disease and related alpha-synucleinopathies are brain amyloidoses.** *Ann N Y Acad Sci* 2003, **991:**107-110.
66. Quist A, Doudevski I, Lin H, Azimova R, Ng D, Frangione B, Kagan B, Ghiso J, Lal R: **Amyloid ion channels: a common structural link for protein-misfolding disease.** *Proc Natl Acad Sci USA* 2005, **102(30):**10427-10432.
67. Ding TT, Lee SJ, Rochet JC, Lansbury PT Jr: **Annular alpha-synuclein protofibrils are produced when spherical protofibrils are incubated in solution or bound to brain-derived membranes.** *Biochemistry* 2002, **41(32):**10209-10217.
68. Volles MJ, Lee SJ, Rochet JC, Shtilerman MD, Ding TT, Kessler JC, Lansbury PT Jr: **Vesicle permeabilization by protofibrillar alpha-synuclein: implications for the pathogenesis and treatment of Parkinson's disease.** *Biochemistry* 2001, **40(26):**7812-7819.
69. Zakharov SD, Hulleman JD, Dutseva EA, Antonenko YN, Rochet JC, Cramer WA: **Helical alpha-synuclein forms highly conductive ion channels.** *Biochemistry* 2007, **46(50):**14369-14379.
70. Zhao R, Gish K, Murphy M, Yin Y, Notterman D, Hoffman WH, Tom E, Mack DH, Levine AJ: **Analysis of p53-regulated gene expression patterns using oligonucleotide arrays.** *Genes Dev* 2000, **14(8):**981-993.
71. Anderson CW, Appella E: **Signaling to the p53 tumor suppressor through pathways activated by genotoxic and nongenotoxic stress.** In *Handbook of Cell Signaling* Edited by: *Bradshaw RA, Dennis EA. New York: Academic Press*; 2004:237-247.
72. Hollstein M, Sidransky D, Vogelstein B, Harris CC: **p53 mutations in human cancers.** *Science* 1991, **253(5015):**49-53.
73. Oliner JD, Pietenpol JA, Thiagalingam S, Gyuris J, Kinzler KW, Vogelstein B: **Oncoprotein MDM2 conceals the activation domain of tumour suppressor p53.** *Nature* 1993, **362(6423):**857-860.
74. Chene P: **Targeting p53 in cancer.** *Curr Med Chem Anticancer Agents* 2001, **1(2):**151-161.
75. Dawson R, Muller L, Dehner A, Klein C, Kessler H, Buchner J: **The N-terminal domain of p53 is natively unfolded.** *J Mol Biol* 2003, **332(5):**1131-1141.
76. Lee H, Mok KH, Muhandiram R, Park KH, Suk JE, Kim DH, Chang J, Sung YC, Choi KY, Han KH: **Local structural elements in the mostly unstructured transcriptional activation domain of human p53.** *J Biol Chem* 2000, **275(38):**29426-29432.
77. Uversky VN, Roman A, Oldfield CJ, Dunker AK: **Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs.** *J Proteome Res* 2006, **5(8):**1829-1842.
78. Winkler H: **Verbreitung und Ursache der Parthenogenesis im Pflanzed- und Teirreiche.** *Jena: Verlag Fischer*; 1920.
79. Lederberg J, McCray AT: **'Ome' sweet 'omics' – a genealogical treasury of words.** *The Scientist* 2001, **15(7):**8.
80. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK: **Sequence complexity of disordered protein.** *Proteins* 2001, **42(1):**38-48.
81. Vacic V, Uversky VN, Dunker AK, Lonardi S: **Composition Profiler: a tool for discovery and visualization of amino acid composition differences.** *BMC Bioinformatics* 2007, **8:**211.
82. Li X, Romero P, Rani M, Dunker AK, Obradovic Z: **Predicting Protein Disorder for N-, C-, and Internal Regions.** *Genome Inform Ser Workshop Genome Inform* 1999, **10:**30-40.
83. Obradovic Z, Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK: **Predicting intrinsic disorder from amino acid sequence.** *Proteins* 2003, **53(Suppl 6):**566-572.
84. Peng K, Vucetic S, Radivojac P, Brown CJ, Dunker AK, Obradovic Z: **Optimizing long intrinsic disorder predictors with protein evolutionary information.** *J Bioinform Comput Biol* 2005, **3(1):**35-60.
85. Obradovic Z, Peng K, Vucetic S, Radivojac P, Dunker AK: **Exploiting heterogeneous sequence properties improves prediction of disorder.** *Proteins* 2005, **61(Suppl 7):**176-182.

86. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res* 1997, **25(17):**3389-3402.
87. Rost B, Sander C, Schneider R: **PHD – an automatic mail server for protein secondary structure prediction.** *Comput Appl Biosci* 1994, **10(1):**53-60.
88. McGuffin LJ, Bryson K, Jones DT: **The PSIPRED protein structure prediction server.** *Bioinformatics* 2000, **16(4):**404-405.
89. Li X, Obradovic Z, Brown CJ, Garner EC, Dunker AK: **Comparing predictors of disordered protein.** *Genome Inform Ser Workshop Genome Inform* 2000, **11:**172-184.
90. Xue B, Oldfield CJ, Dunker AK, Uversky VN: **CDF it all: Consensus prediction of intrinsically disordered proteins based on various cumulative distribution functions.** *BMC Genomics* 2009.
91. Cheng Y, Oldfield CJ, Meng J, Romero P, Uversky VN, Dunker AK: **Mining alpha-helix-forming molecular recognition features with cross species sequence alignments.** *Biochemistry* 2007, **46(47):**13468-13477.
92. Black DL: **Mechanisms of alternative pre-messenger RNA splicing.** *Annu Rev Biochem* 2003, **72:**291-336.
93. Ast G: **How did alternative splicing evolve?** *Nat Rev Genet* 2004, **5(10):**773-782.
94. Stamm S, Ben-Ari S, Rafalska I, Tang Y, Zhang Z, Toiber D, Thanaraj TA, Soreq H: **Function of alternative splicing.** *Gene* 2005, **344:**1-20.
95. Graveley BR: **Alternative splicing: increasing diversity in the proteomic world.** *Trends Genet* 2001, **17(2):**100-107.
96. Roberts R, Timchenko NA, Miller JW, Reddy S, Caskey CT, Swanson MS, Timchenko LT: **Altered phosphorylation and intracellular distribution of a (CUG)n triplet repeat RNA-binding protein in patients with myotonic dystrophy and in myotonin protein kinase knockout mice.** *Proc Natl Acad Sci USA* 1997, **94(24):**13221-13226.
97. Ma K, Inglis JD, Sharkey A, Bickmore WA, Hill RE, Prosser EJ, Speed RM, Thomson EJ, Jobling M, Taylor K, *et al.*: **A Y chromosome gene family with RNA-binding protein homology: candidates for the azoospermia factor AZF controlling human spermatogenesis.** *Cell* 1993, **75(7):**1287-1295.
98. Kar A, Kuo D, He R, Zhou J, Wu JY: **Tau alternative splicing and frontotemporal dementia.** *Alzheimer Dis Assoc Disord* 2005, **19(Suppl 1):**S29-36.
99. Skotheim RI, Nees M: **Alternative splicing in cancer: noise, functional, or systematic?** *Int J Biochem Cell Biol* 2007, **39(7–8):**1432-1449.
100. Venables JP: **Aberrant and alternative splicing in cancer.** *Cancer Res* 2004, **64(21):**7647-7654.
101. Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, Dunker AK: **The importance of intrinsic disorder for protein phosphorylation.** *Nucleic Acids Res* 2004, **32(3):**1037-1049.
102. Li T, Li F, Zhang X: **Prediction of kinase-specific phosphorylation sites with sequence features by a log-odds ratio approach.** *Proteins* 2008, **70(2):**404-414.
103. Collins MO, Yu L, Campuzano I, Grant SG, Choudhary JS: **Phosphoproteomic analysis of the mouse brain cytosol reveals a predominance of protein phosphorylation in regions of intrinsic sequence disorder.** *Mol Cell Proteomics* 2008, **7(7):**1331-1348.
104. Daily KM, Radivojac P, Dunker AK: **Intrinsic disorder and protein modifications: building an SVM predictor for methylation.** *IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology. San Diego, California, U.S.A* 2005.
105. Cheng Y, LeGall T, Oldfield CJ, Dunker AK, Uversky VN: **Abundance of intrinsic disorder in protein associated with cardiovascular disease.** *Biochemistry* 2006, **45(35):**10448-10460.
106. Midic U, Oldfield CJ, Dunker AK, Obradovic Z, Uversky VN: **Protein disorder in the human diseasome: Unfoldomics of human genetic diseases.** *BMC Genomics* 2009, **10(Suppl 1):**S12.
107. Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL: **The human disease network.** *Proc Natl Acad Sci USA* 2007, **104(21):**8685-8690.
108. Dyson HJ, Wright PE: **Coupling of folding and binding for unstructured proteins.** *Curr Opin Struct Biol* 2002, **12(1):**54-60.
109. Dunker AK, Brown CJ, Obradovic Z: **Identification and functions of usefully disordered proteins.** *Adv Protein Chem* 2002, **62:**25-49.
110. Landsteiner K: **The specificity of serological reactions.** *New York: Dover*; 1936.
111. Pauling L: **A theory of the structure and process of formation of antibodies.** *J Am Chem Soc* 1940, **62:**2643-2657.
112. Karush F: **Heterogeneity of the binding sites of bovine serum albumin.** *J Am Chem Soc* 1950, **72:**2705-2713.
113. McMeekin TL: **Milk proteins.** *J Milk Food Technol* 1952, **15:**57-63.
114. Jirgenesons B: **Classification of proteins according to conformation.** *Makromolekulare Chemie* 1966, **91:**74-86.
115. Doolittle RF: **Structural aspects of the fibrinogen to fibrin conversion.** *Adv Protein Chem* 1973, **27:**1-109.