Amyloid Fibrils

The amyloid state and its association with protein misfolding diseases

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

From a wide range of *in vitro* experiments on peptides and proteins we now know that the formation of amyloid structures is not a rare phenomenon associated with a small number of diseases but rather **that it reflects a well-defined structural form of the protein that is an alternative to the native state** — a form that may in principle be adopted by many, if not all, polypeptide sequences

*A selection of diseases associated with extracellular amyloid deposits or intracellular inclusions with amyloid-like characteristics. See REF. 5 for a more comprehensive list of the approximately 50 human protein misfolding diseases and their associated proteins.

How do proteins either adopt and maintain the specific states that are needed to carry out given functions or instead misfold and form potentially pathogenic aggregates such as amyloid fibrils? Importance of investigating the nature and properties of the various states in which these molecules can be found.

How is the conversion of proteins into the amyloid state generally avoided in living systems? Importance of define the specific mechanisms by which aggregation occurs and the manner in which it can induce pathogenic behaviour.

Progress from:

- chemical kinetics, nanotechnology and microfluidics.
- comparative studies of the behaviour of a wide range of peptides and proteins, also non pathogenic

The multiplicity of protein states

We discussed how the varied and intricate structures of globular proteins are encoded by their amino acid sequences, and that these molecules have an intrinsic ability to fold spontaneously.

From these studies we now know that the **folding process** can be described as a **diffusional search on a free energy surface**. A combination of experimental and theoretical studies has clarified how **given sequences define specific free energy surfaces that enable folding**.

However, **free energy surface depends on a very large number of relatively weak interactions**, including hydrogen bonds, electrostatic interactions, dispersion forces and **interactions with solvent molecules** that underlie the hydrophobic effect.

Moreover, in some cases the native state of a given peptide or protein may not be structured in a globular form but disordered.

The functional native state is likely to only reflect a local free energy minimum at physiological concentrations, as self-association into aggregated protein species may in many cases lower the global free energy. Indeed, the maintenance of **protein solubility** has emerged as a central aspect of the more general topic of protein homeostasis.

Figure 1 | A protein can exist in various different states. The populations of the different states and their interconversion rates are determined by their different thermodynamic stabilities, by the free energy barriers that are associated with the corresponding transitions as well as by the rates of synthesis and degradation, the propensity to interact with chaperones and to undergo post-translational and other chemical modifications. The amyloid state of a protein is a highly ordered form of aggregate in which the polypeptide chains adopt a fibrillar structure, which is capable of self-replication, for example through secondary processes. Amyloid fibrils are rich in β-sheet structure and typically form from unfolded or partially unfolded conformations of proteins and peptides, some of which are fragments of larger proteins. The amyloid state is 'generic' in that its characteristic architecture is not encoded by specific amino acid sequences.

Intrinsically disordered proteins:

are capable to interact in multiple ways with different partner (required for functional reasons, including trafficking to specific cellular locations that require translocation across mitochondrial and even nuclear membranes);

are not necessarily prone to aggregation, as their sequences have usually evolved to maintain the **level of solubility that is required for their optimal function**;

the existence of extensive regions that are highly abundant in charged and polar groups disfavour intermolecular association from a thermodynamic point of view.

Strong **kinetic barriers to aggregation** to overcome.

Misfolding diseases often involve IDP or disordered domains (in soluble form), which are particularly vulnerable to misfolding and aggregation from which they must be protected in living systems.

Figure 1 | A protein can exist in various different states. The populations of the different states and their interconversion rates are determined by their different thermodynamic stabilities, by the free energy barriers that are associated with the corresponding transitions as well as by the rates of synthesis and degradation, the propensity to interact with chaperones and to undergo post-translational and other chemical modifications. The amyloid state of a protein is a highly ordered form of aggregate in which the polypeptide chains adopt a fibrillar structure, which is capable of self-replication, for example through secondary processes. Amyloid fibrils are rich in β-sheet structure and typically form from unfolded or partially unfolded conformations of proteins and peptides, some of which are fragments of larger proteins. The amyloid state is 'generic' in that its characteristic architecture is not encoded by specific amino acid sequences.

Extreme cases:

proteins that normally function in globular states often adopt intermediate conformations (corresponding to local minima on the free energy surface) before becoming fully folded and incorporated into their biological environments (for example, in the cytosol or within membranes) following their synthesis on the ribosome, particularly those proteins that are large and have complex folds.

proteins that have folded correctly can subsequently unfold, at least locally, often simply as a consequence of the dynamical fluctuations that occur within protein molecules

Figure 1 | A protein can exist in various different states. The populations of the different states and their interconversion rates are determined by their different thermodynamic stabilities, by the free energy barriers that are associated with the corresponding transitions as well as by the rates of synthesis and degradation, the propensity to interact with chaperones and to undergo post-translational and other chemical modifications. The amyloid state of a protein is a highly ordered form of aggregate in which the polypeptide chains adopt a fibrillar structure, which is capable of self-replication, for example through secondary processes. Amyloid fibrils are rich in β-sheet structure and typically form from unfolded or partially unfolded conformations of proteins and peptides, some of which are fragments of larger proteins. The amyloid state is 'generic' in that its characteristic architecture is not encoded by specific amino acid sequences.

Although most common forms of proteins are soluble, some functional states are insoluble, for example the fibrillar assemblies that form the cellular cytoskeleton or spider silk (which are not amyloids!!!).

Proteins exhibit generic polymeric and colloidal patterns of behaviour, and many non-biological and synthetic polymers show similarly condensed phases, including, for example, filamentous and particulate gel states.

Aggregated forms of proteins can be generally amorphous on an ultrastructural level, consisting of more or less disordered assemblies of interacting chains of the same or of different sequences.

Of particular fascination, because of their remarkable structures and properties, are **highly ordered self-associated species of peptides and proteins, notably amyloid fibrils** and closely related prion-like states.

In the native state of a protein, most of the interactions between amino acid residues are intramolecular, whereas in the amyloid state intermolecular interactions generally dominate. The native state is thermodynamically stable relative to the amyloid state if the free energy (G) of the peptide or protein molecule is lower in the native state than in the amyloid state. As a protein will not spontaneously transition from a state of lower free energy to a state of higher free energy, the conversion into the amyloid state will only take place when its free energy is lower than that of the native state. As the stability of the amyloid state (ΔG) is dependent on the protein concentration, whereas that of the native state (unless it exists in a functional complex) is to a good approximation independent of the protein concentration, there is a concentration at which the stability of the amyloid state is the same as that of the native state; this is the critical concentration. At concentrations exceeding this critical value, a protein is more stable in the amyloid state than in its native state. In such situations, the native state can only persist if there are high free energy barriers that hinder the transition into the more stable amyloid state (see the figure, part a). Under such conditions, the native state is then said to be kinetically metastable^{41,42}. See the figure, part **b**, for a depiction of experimentally determined standard free energies per residue $(\Delta G/N)$, red dots) for a set of peptides and proteins as a function of the length, N, of the polypeptide chain⁴¹; for reference, the free energy differences between native and denatured (unfolded) states of proteins with sequence lengths in the same range are also shown (green crosses)⁴¹. Topological constraints associated with the packing of a long polypeptide chain decrease the free energy gain from the conversion to the amyloid form and for polypeptide chains longer than ~150 residues. The standard free energy associated with amyloid formation is on average comparable or more negative than that associated with protein folding. Part b adapted with permission from Baldwin, A. J. et al. Metastability of native proteins and the phenomenon of amyloid formation. J. Am. Chem. Soc. 133, 14160-14163 (2011). © (2011) American Chemical Society.

There has been increasing evidence, that the amyloid state might be **thermodynamically more stable than the functional native states** of many protein molecules even under physiological conditions. As amyloid involves formation of intermolecular contacts, the **thermodynamic stability is favoured at higher concentrations**. The critical concentration above which the stability of the amyloid state exceeds the soluble state may be lower than the physiological concentration of a given protein *in vivo*, which suggests the possibility that some, and perhaps many, proteins might routinely function at concentrations higher than for their thermodynamic solubility .

The amyloid state was first observed in the context of systemic amylodidosis more than 150 years ago, and indeed the name 'amyloid' means 'starch-like', as the deposits observed in the tissues and organs of patients who died from these conditions contained deposits that stained with iodine, which is used to detect starch

In the native state of a protein, most of the interactions between amino acid residues are intramolecular, whereas in the amyloid state intermolecular interactions generally dominate. The native state is thermodynamically stable relative to the amyloid state if the free energy (G) of the peptide or protein molecule is lower in the native state than in the amyloid state. As a protein will not spontaneously transition from a state of lower free energy to a state of higher free energy, the conversion into the amyloid state will only take place when its free energy is lower than that of the native state. As the stability of the amyloid state (ΔG) is dependent on the protein concentration, whereas that of the native state (unless it exists in a functional complex) is to a good approximation independent of the protein concentration, there is a concentration at which the stability of the amyloid state is the same as that of the native state; this is the critical concentration. At concentrations exceeding this critical value, a protein is more stable in the amyloid state than in its native state. In such situations, the native state can only persist if there are high free energy barriers that hinder the transition into the more stable amyloid state (see the figure, part a). Under such conditions, the native state is then said to be kinetically metastable^{41,42}. See the figure, part **b**, for a depiction of experimentally determined standard free energies per residue $(\Delta G/N)$, red dots) for a set of peptides and proteins as a function of the length, N, of the polypeptide chain⁴¹; for reference, the free energy differences between native and denatured (unfolded) states of proteins with sequence lengths in the same range are also shown (green crosses)⁴¹. Topological constraints associated with the packing of a long polypeptide chain decrease the free energy gain from the conversion to the amyloid form and for polypeptide chains longer than ~150 residues. The standard free energy associated with amyloid formation is on average comparable or more negative than that associated with protein folding. Part b adapted with permission from Baldwin, A. J. et al. Metastability of native proteins and the phenomenon of amyloid formation. J. Am. Chem. Soc. 133, 14160-14163 (2011). © (2011) American Chemical Society.

The amyloid state essential architecture is not encoded by the amino acid sequence, although the details of its structure and stability can be markedly sequencedependent.

The amyloid structure can in principle be adopted by any polypeptide chain

Such deposits are usually primarily composed of **one protein**, although they are typically associated *in vivo* with various other molecules.

Remarkably, there is **no evident similarity in the sequences**, native structures or functions of the group of disease-associated proteins.

Despite such differences, the corresponding **amyloid fibrils all contain a common 'cross-β' pattern in X-ray fibre diffraction studies that is indicative of the component β-strands being oriented perpendicularly to the fibril axis.**

Amyloid fibrils from different proteins seem to be remarkably similar at the nanometre length scale. Indeed, under electron microscopy (Cryo-EM) or atomic force microscopy (AFM), amyloid fibrils tend to appear as **unbranched filamentous structures only a few nanometers in diameter but often micrometres in length**. They are typically observed to consist of multiple protofilaments that twist around each other to form mature fibrils. X-ray fibre diffraction studies indicate **that the core of each protofilament adopts a cross-β structure, in which β-strands form effectively continuous hydrogen-bonded β-sheets that run along the length of the fibril**

Figure 2 | Generic features of the amyloid structure. Amyloid structures exhibit a range of specific features and possess common characteristics. Representation of the 'cross-ß' structure common to amyloid fibrils (part a). The spacing between polypeptide chains along the fibril axis is constant to a good approximation even for very different polypeptide sequences, a generic property arising from the common inter-main chain hydrogen bonding constraints (orange line in part b). By contrast, the spacing between the β -sheets in the direction perpendicular to the fibril axis is highly dependent on the nature of the side chains and originates from variable packing constraints. Correlation of the inter-sheet spacing (green circles; green line indicates specific properties) with the average van der Waals volumes of the side chains for a range of different amyloid fibrils (part b); filled green circles indicate fibrils that are formed by homopolymers (that is, polypeptide repeats of the same type of amino acid residue)³¹. Part **b** reprinted with permission from REF, 31, Wiley.

NMR and Cryo-EM studies confirm the generic nature of the overall structures of different fibrils, which can be attributed to the common properties of the polypeptide backbone that support the hydrogen bonding pattern in the fibril core and to variations that result from the manner in which the different sets of side chains are incorporated into the common fibrillar architecture.

The cross-β architecture provides very great stability to the fibrils, as it allows the formation of a **continuous array of hydrogen bonds**

Figure 3 | Structure of an amyloid fibril at atomic resolution. The structure shown is of one of several polymorphs of the amyloid fibrils that are formed from an 11-residue fragment of transthyretin⁷⁵. The combination of cryo-electron microscopy imaging (part a) with solid-state NMR analysis has enabled the determination of an atomic-level structure (part b). A more detailed view (part c) shows the hierarchical organization of this amyloid fibril in which the three filaments that form the mature fibril illustrated here are in turn formed by pairs of cross- β protofilaments, which are each composed of pairs of β -sheets⁷⁵. The fibril surfaces are shown as electron density maps, and the constituent β -sheets are shown in a ribbon representation; oxygen, carbon, and nitrogen atoms are shown in red, gray and blue, respectively. Adapted with permission from Fitzpatrick, A.W. P. et al. Atomic structure and hierarchical assembly of a cross- β

The common cross-β architecture of amyloid structures originates from the **universal propensity of polypeptide chains to form backbone hydrogen bonding** — a propensity **that**, in a condensed state, **can be most readily accommodated through extended intermolecular β-sheets**. The lateral packing of such β-sheets, however, relies on specific patterns of interactions between side chains that depend on the amino acid sequences of the component proteins.

Fibrillar structures have remarkable mechanical properties, including a high **Young's modulus** and tensile strength, which arise in large part from the hydrogen bonding network of the cross-β structure (**intermediate between carbon nanotubes, and actin- tubulin**)

Figure 3 | Structure of an amyloid fibril at atomic resolution. The structure shown is of one of several polymorphs of the amyloid fibrils that are formed from an 11-residue fragment of transthyretin⁷⁵. The combination of cryo-electron microscopy imaging (part a) with solid-state NMR analysis has enabled the determination of an atomic-level structure (part b). A more detailed view (part c) shows the hierarchical organization of this amyloid fibril in which the three filaments that form the mature fibril illustrated here are in turn formed by pairs of cross- β protofilaments, which are each composed of pairs of β -sheets⁷⁵. The fibril surfaces are shown as electron density maps, and the constituent β -sheets are shown in a ribbon representation; oxygen, carbon, and nitrogen atoms are shown in red, gray and blue, respectively. Adapted with permission from Fitzpatrick, A.W. P. et al. Atomic structure and hierarchical assembly of a cross- β

WHY this specific structure?

It is connected to the CHIRALITY of the central (alpha) carbon atom of each amino acid and to the polypeptide low persistence length.

The constant width comes from this competition: chirality enhances the twisting, flexibility penalty reduces it.

Remains true that amyloids have strong stability. Lower free energy of the native state of the components, and that this thermodynamic stability is favoured at high concentrations. Up to the limit of 150 residues.

Evolution, perhaps for this reason, has created proteins with about 300 residues, on average!!!

Figure 3 | Structure of an amyloid fibril at atomic resolution. The structure shown is of one of several polymorphs of the amyloid fibrils that are formed from an 11-residue fragment of transthyretin⁷⁵. The combination of cryo-electron microscopy imaging (part a) with solid-state NMR analysis has enabled the determination of an atomic-level structure (part b). A more detailed view (part c) shows the hierarchical organization of this amyloid fibril in which the three filaments that form the mature fibril illustrated here are in turn formed by pairs of cross- β protofilaments, which are each composed of pairs of β -sheets⁷⁵. The fibril surfaces are shown as electron density maps, and the constituent β -sheets are shown in a ribbon representation; oxygen, carbon, and nitrogen atoms are shown in red, gray and blue, respectively. Adapted with permission from Fitzpatrick, A.W. P. et al. Atomic structure and hierarchical assembly of a cross- β

Box 2 | Kinetics of amyloid formation

Whereas thermodynamics describes whether or not a transition from one state to another is spontaneous and can occur without an external driving force, kinetics addresses the question of how fast such a transformation will take place. Kinetic measurements have historically been one of the main tools for elucidating reaction mechanisms for both small-molecule reactions and for enzymology¹⁷⁴. As aggregate populations observed during amyloid formation are heterogeneous, a convenient tool for describing the kinetics of amyloid formation is the master equation formalism^{99,100}. This formalism connects the microscopic steps that underlie amyloid formation with their macroscopic manifestations by considering the different ways in which new aggregates can be formed: for example, from monomer through primary nucleation (see the figure, part a; top; monomers are shown in green); from existing fibrils through fragmentation (see the figure, part a; bottom; fibrils are shown in blue); or from a combination of both monomeric and aggregated species through secondary nucleation (see the figure, part a; centre). Integrated rate laws that are obtained for filament growth processes that occur under the action of these various microscopic processes commonly take the form of sigmoidal functions (see the figure, part b; blue line), which are characterized by a lag time $(\tau_{i_{\alpha}})$ and a maximal growth rate (r_{max} ; red line). Comparison of such integrated rate laws with experimental kinetic measurements enables the relative importance of specific microscopic processes to be tested. In this example (see the figure, part c) the aggregation kinetics for increasing concentrations (see the figure, part c; coloured lines) of the am yloid- β peptide are compared with integrated rate laws that contain primary nucleation (see the figure, part c; left), fragmentation (see the figure, part c; centre) and monomer-dependent secondary nucleation (see the figure, part c; right). The data show that secondary nucleation is the dominant process under these conditions (n_and n, indicate the sizes of the critical nuclei for primary and secondary nucleation, respectively; k and k indicate the rates of primary nucleation and fragmentation, respectively; and m and M indicate the concentrations of the monomers and the fibril mass, respectively), at time t. To avoid overfitting, global analysis approaches have emerged as powerful tools^{99,100}. In this strategy, a known parameter, most commonly the concentration of a monomeric protein or peptide at the beginning of the aggregation reaction is varied in a known manner, and then the entire data set in terms of both time and concentration is fitted to a single rate law. More details can be found in the literature^{99,100}. Parts a and c adapted with permission from Cohen, S. I. A. et al. Proliferation of amyloid- β 42 aggregates occurs through a secondary nucleation mechanism. Proc. Natl Acad. Sci. USA 110, 9758-9763 (2013). Part a © (2005) National Academy of Sciences, USA. Lührs, T. et al. 3D structure of Alzheimer's amyloid-β(1-42) fibrils. Proc. Natl Acad. Sci. USA 102, 17342-17347 (2005). Part b from Knowles, T. P. J. et al. An analytical solution to the kinetics of breakable filament assembly. Science 326, 1533-1537 (2009). Reprinted with permission from AAAS.

- IDPs flexibility allow interact with multiple partners
- to be functional, specific regions within IDPs undergo an induced folding that is triggered by either binding to their biological partner(s) or due to environmental changes.
- formation of complexes that possess a more ordered structure than the original components and lower free energy minima than the native state

Typical sigmoidal reaction time course.

This type of behaviour, in which a lag phase is observed before a rapid growth phase, is a feature of nucleated polymerization; in cases in which the total quantity of protein is limited, the growth phase is followed by a plateau phase in which the reaction rate declines as a result of the depletion of the soluble species that is being monitored as it converts into fibrils. When monomers add to the ends of the fibrils, they adopt the cross-β conformation to match that of the peptides already present in the aggregate that hence function as templates. Indeed, when pre-formed aggregates are added to a solution during its lag phase, rapid fibril formation can take place as a result of seeding, a process analogous to that familiar from studies of crystallization phenomena. Thus, the conversion of a protein molecule from its soluble state into the amyloid form can be triggered by nucleation, as well as by templating or seeding from existing aggregates.

According to the classical nucleation process, a **primary nucleation step** is necessary for the formation of aggregates, but in the case of amyloid fibrils several secondary steps can also be involved. In these secondary processes, the formation of the nuclei is catalyzed either by the fragmentation of the already formed fibrils or through a surface-catalyzed secondary nucleation mechanism, whereby the existing fibrils catalyze the nucleation of further nuclei at their surface.

The conventional "first-misfolding-thenaggregation" paradigm is the generally accepted process of amyloid formation, but several observations have shown that the misfolding process could take place after a first step in which native monomers aggregate, i.e.

"first-aggregation-then-misfolding"

These native oligomers undergo a structural misfolding to form the early cross sheetaggregate whereas the f inal amyloid fibrillar structures only form in a second step.

Despite the initial difference, both pathways are conceptually similar: a misfolded state, monomeric or oligomeric, is necessary to nucleate the formation of the universal amyloidogenic cross- sheet structure.

kinetic descriptions of the mechanistic details of amyloid formation at the microscopic level works, but the full elucidation of the process requires the identification of all the conformational and oligomeric states adopted by the protein and of its possible aggregates. Indeed, structural polymorphism can be encountered at all aggregation levels, and it originates as a consequence of the glassy, frustrated energy landscape that underlies misfolding and aggregation. Oligomeric and protofibrillar structures can form on- and off-pathway during the formation of mature fibrils.

Parkinsons' disease

Statistical characterization of alfa-synuclein amyloid aggregates cross-sectional dimensions. A) Monomeric and early oligomeric structures. B) Protofibrillar and fibrillar aggregates.

Box 3 | Strategies for therapeutic intervention

The maintenance of biological functionality. As proteins in their functional forms can be thermodynamically and chemically metastable, mechanisms have evolved to maintain their solubility *in vivo* for prolonged periods of time and to avoid their conversion into non-funtional amyloid states. Exceptions are cases in which the amyloid state is functional. Thus, for example, Pmel17 is a highly aggregation-prone protein that forms functional amyloid structures that are involved in melanosome biogenesis, and it has been found that certain peptide and protein hormones in secretory granules of the endocrine system are stored in cross-β conformations. Rather few such examples have, however, been observed, particularly in mammalian systems, and functional amyloid formation is undoubtedly very carefully regulated.

particularly important to reduce the population of oligomeric species (pre-fibrillar species) by disrupting the processes of their formation (see the figure; dashed arrows) or by promoting the pathways of their removal (dotted arrows). These strategies can be implemented in various ways¹⁶⁶; for example, through the modulation of production processes (targeting synthesis or proteolysis)¹⁷⁵, degradation (targeting the ubiquitinproteasome^{126,127} and the autophagy^{129,130} systems) or stability (usually targeting the native state)^{167,168}. For example, one can use the ability of antibodies^{176,177}, or artificiallygenerated analogues such as affibodies¹⁷⁸, to selectively bind to the native states of aggregation-prone proteins, as binding generally results in increased stability and hence to a reduction in aggregation propensity. In some cases, it might be possible to use antibody-based immunotherapy approaches to reduce the level of highly aggregation-prone species (such as amyloid-B) by stimulating their clearance^{179,180}. Antibodies and their analogues also offer other possibilities, one of which could be to mimic the action of natural molecular chaperones by targeting the aberrant misfolded species that give rise to cellular damage. If such 'artificial chaperones' can be developed and can be targeted to the appropriate location (for example, by enhancing their ability to cross the blood-brain barrier), then they could represent a highly effective therapy^{3,43}. Furthermore, it is also becoming evident that small molecules can also function to suppress the early stages of protein aggregation, for example by binding to specific amyloidogenic species and by reducing the risk of nucleation and proliferation of pathogenic agents¹⁶⁸⁻¹⁷².