Mechanisms of amyloid fibril formation by proteins

Santosh Kumar and Jayant B. Udgaonkar*
National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bangalore 560 065, India

Understanding the structural heterogeneity inherent in the process of amyloid fibril formation is an important goal of protein aggregation studies. Structural heterogeneity in amyloid fibrils formed by a protein manifests itself in fibrils varying in internal structure and external appearance, and may originate from molecular level variations in the internal structure of the cross-β motif. Amyloid fibril formation commences from partially structured conformations of a protein, and in many cases, proceeds via pre-fibrillar aggregates (spherical oligomers and/or protofibrils). It now appears that structural heterogeneity is prevalent in the partially structured conformations as well as in the pre-fibrillar aggregates of proteins. Amyloid fibril formation may therefore potentially commence from many precursor states, and amyloid fibril polymorphism might be the consequence of the utilization of distinct nucleation and elongation mechanisms. This review examines the current understanding of the structural heterogeneity seen in amyloid fibril formation reactions, and describes how an understanding of the initial and intermediate stages of amyloid fibril formation reactions can provide an insight into the structural heterogeneity seen in mature fibrils.

Keywords: Alternative pathways, amyloid fibrils, amyloid protofibrils, spherical oligomers, structural heterogeneity.

The process of protein aggregation is a widely observed phenomenon in biology. A well-studied example is the aggregation of cytoskeletal proteins into filaments, which are vital for many cellular processes. Protein aggregation is also seen in disruptive contexts, where it affects the folding or normal functioning of proteins. In vitro studies of the refolding or unfolding of proteins at high concentrations are often hindered by the transient accumulation of protein aggregates, and avoiding aggregation can be a challenge during the industrial production of therapeutic proteins. While many such protein aggregates are disordered, protein aggregates can also be highly ordered. One example of ordered protein aggregates possessing a remarkably high internal order is the amyloid fibril.

Understanding the principles of amyloid fibril formation is an important problem in modern biology. Many human diseases, including Alzheimer’s disease, Parkinson’s disease, Huntington’s disease as well as the prion diseases, are associated with the formation of amyloid fibrils. In amyloidoses, amyloid fibrils accumulate in the brain, or in one or more other tissues. Amyloid fibrils are, however, not always harmful. It is now increasingly being seen that living organisms, ranging from prokaryotes to humans, exploit amyloid fibrils formed by their endogenous proteins for carrying out normal physiological functions. From the biotechnology perspective, amyloid fibrils also appear promising as macromolecular assembly based nanomaterials.

The term ‘amyloid’ was first used by Rudolf Virchow to describe a structured mass in human tissues, which was considered to be a cellulose-containing substance on the basis of its ability to be stained by iodine. Later, direct chemical analysis showed that the main component of amyloids is protein. Now, amyloid fibrils refer to elongated protein aggregates characterized by their long and relatively straight morphologies, cross-β diffraction patterns, specific dye binding properties and rigid core structures. They show a characteristic X-ray diffraction pattern with 4.7–4.8 Å meridional reflections and 10 Å equatorial reflections. They bind to and alter the spectroscopic characteristics of congo-red and thioflavin dyes. Hydrogen exchange experiments coupled with mass spectrometry (HIX-MS) and with NMR (HIX-NMR) have suggested that amyloid fibrils possess extensively hydrogen-bonded β-sheet core structures, which confer to them remarkable stability and resistance to protease cleavage.

Amyloid fibril formation involves a structural rearrangement of the native state into a β-sheet rich fibrillar conformation. β-sheets seem to provide a scaffold that is favourable for protein assembly: the edge strands of β-sheet structures are unstable, and the sheet can grow by interacting with any other β-strands it encounters. Natural β-sheet proteins are seen to utilize a number of mechanisms to avoid the edge-to-edge aggregation of their β-sheets. It now appears that all proteins can potentially assemble into amyloid fibrils.

Amyloid fibril formation is an extremely complex reaction. A protein can assemble into multiple structurally distinct fibrils. Structural heterogeneity also appears...
to be prevalent in the assembly intermediates formed at initial times of the reaction\(^\text{37-39}\). This review critically examines current knowledge and understanding of the mechanisms of amyloid fibril formation, the structural heterogeneity inherent in the process, as well as the role of structural heterogeneity in determining how fibrils form. The current molecular level understanding of the structural heterogeneity in amyloid fibrils is also discussed.

**Structure of amyloid fibrils**

Amyloid fibrils are \(\sim\)10 nm in their diameters, and are composed typically of 2–6 protofilaments. Amyloid fibrils of all proteins possess the same structural motif, the cross-\(\beta\) motif, wherein the \(\beta\)-strands are oriented perpendicular to, and the \(\beta\)-sheets parallel to the fibril axis\(^\text{20,21,40}\). In cross-\(\beta\) motifs, the separation between hydrogen-bonded \(\beta\)-strands is \(\sim\)0.48 nm, and that between \(\beta\)-sheet layers is \(\sim\)1.0–1.1 nm (ref. 41).

Understanding the molecular details of amyloid fibril structures has been a challenge owing to the large size, the low solubility and the noncrystalline nature of fibrils. Recently, however, the use of solid-state NMR\(^\text{42}\) has contributed to the considerable progress being made in the understanding of amyloid fibril structure. An elegant example is the structural model of the amyloid-\(\beta\) \(\text{40}\) protofilament (Figure 1a), which has been proposed on the basis of constraints from solid-state NMR studies, combined with measurements of fibril dimensions and of the mass-per-length (MPL) from electron microscopy images\(^\text{43,44}\). In this model, the first 10 residues of amyloid-\(\beta\) \(\text{40}\) molecule are in a disordered conformation. Residues 12–24 and 30–40 form the core region of the fibrils, and exist in a \(\beta\)-strand conformation. The two \(\beta\)-strands of each amyloid-\(\beta\) \(\text{40}\) molecule are connected via a bend region containing residues 25–29, and are parts of two distinct in-register, parallel \(\beta\)-sheets interacting through their side chains in the same protofilament. This suggests that a single cross-\(\beta\) unit consists of a double-layered \(\beta\)-sheet structure. A single amyloid-\(\beta\) \(\text{40}\) protofilament appears to comprise two cross-\(\beta\) motifs, i.e. four \(\beta\)-sheets with an intersheet distance of \(\sim\)1 nm. This structural model of amyloid-\(\beta\) \(\text{40}\) amyloid fibrils is consistent
with studies by other methods, such as X-ray fibre diffraction, electron paramagnetic resonance, hydrogen-exchange and proteolysis. Solid-state NMR and electron microscopy experiments have suggested that the fibrils formed by amyloid-β have similar supramolecular structures.

Our understanding of amyloid fibril structure has improved greatly by recent X-ray structure determinations of microcrystals of the amyloid-forming segments of 10 different amyloidogenic proteins. As suggested earlier, these studies indicate that the cross-β motifs in amyloid fibrils formed by these amyloidogenic segments consist of a pair of β-sheets. Three levels of organization are apparent. The first level of organization represents a β-sheet formed by the alignment of the peptide fragments. In the second level of organization, two such β-sheets self-complement to form a pair of sheet structures, in which the side chains protruding from the two sheets intercalate to form a dry 'steric zipper' (Figure 1b). In the third level of organization, interactions between the pairs of sheet structures lead to the formation of amyloid fibrils.

In the case of polyglutamine fibrils, it has been proposed that β-helices, structures significantly different from the classical amyloid fibrils, are generated by the involvement of additional hydrogen bonds between the side chains. These structures could be cylindrical β-sheets of 3.1 nm diameter with 20 residues per helical turn. In this cylinder, the neighbouring turns are linked by hydrogen bonds between backbone amides as well as by those between side-chain amides, and the side chains point alternatively in and out of the cylinder (Figure 1c).

In contrast to our knowledge of mature amyloid fibrils, very little is known about the internal structures of amyloid protofibrils. These are curvy and elongated nanostructures, which sometimes appear to circularize into annular protofibrils (see below), and which are seen to form at initial times of fibril formation by many proteins. Several studies utilizing fluorescence spectroscopy, Fourier-transform infrared (FTIR) spectroscopy or HX-MS have all suggested an increase in internal order from soluble oligomers to protofibrils to fibrils. The thioflavin T binding ability as well as the β-sheet content of protofibrils is seen to be less than that of mature fibrils. In the case of the amyloid-β protein, HX-MS has suggested that protofibrils possess β-sheet elements that extend to adjacent residues in mature fibrils. The internal organization of the β-sheets in protofibrils remains to be investigated by higher resolution structural probes.

**Mechanisms of protein polymerization**

The mechanism of polymerization of cytoskeletal proteins and sickle cell haemoglobin has been studied in great detail, and has been described in terms of two basic models, namely, nucleation-dependent polymerization and isodesmic (linear) polymerization. Since the data on protein aggregation reactions leading to protofibril and fibril formation are often evaluated in terms of these models, we first describe them briefly.

**Nucleation-dependent polymerization**

In a nucleation-dependent polymerization (NDP) reaction, the initial steps are slower than the later ones. A complete mathematical description of the kinetics of an NDP reaction requires forward and reverse rate constants for each step. A simplifying strategy for analysis considers the initial steps to be close to equilibrium, and thus reduces the kinetic problem to an equilibrium one. From a thermodynamic viewpoint, an NDP reaction (Figure 2a–c) can be described as follows. The initial steps (nucleation) consist of a number of unfavourable equilibria (Figure 2a), that makes the initiation (nucleation) of polymerization difficult, and the system can be viewed as climbing an energy barrier which must be crossed for the polymerization to proceed (Figure 2b). The peak of the free energy curve corresponds to a species (A2 in Figure 2a) which marks a turning point in the polymerization reaction, after which downstream steps (elongation) become thermodynamically favourable. This high energy and thus very scarce species is the nucleus, and it constitutes a bottle neck in the polymerization reaction.

The slope of the free energy barrier (Figure 2b) at any value of aggregate size is determined by the product of the concentration and the ratio of the association to dissociation rate constants. In the nucleation phase (Figure 2a and b), the dissociation rate constants are greater than the association rate constants. Once the nucleus is formed, the slope of the energy curve (Figure 2b) reverses its direction, and for all the subsequent steps, the association rate constants become greater than the dissociation rate constants. Thus, in terms of the reaction kinetics, the nucleus represents the smallest protein aggregate for which the rate constant of association is greater than that of dissociation.

**Characteristics of an NDP reaction**

An NDP reaction has the following characteristics. (1) The kinetics of polymer formation shows a lag phase. The lag time represents the weak initiation phase of the kinetics, and appears to be describable by a t^2 function (Figure 2c inset). The lag time in the kinetics of an NDP reaction arises because the dissociation rate constant is greater than the association rate constant in the initial part of the reaction (nucleation phase). The duration of the lag phase is proportional to the steepness of the energy curve in the initial part (Figure 2b), and depends on protein concentration.
lag time on protein concentration is controlled by the values of the association and dissociation rate constants as well as by the size of the nucleus (i.e. the number of monomers in the nucleus)\textsuperscript{52}. (2) There is a critical concentration for the formation of polymer. The lag phase of an NDP reaction shows a strong dependence on protein concentration; the lag time increases with a decrease in protein concentration. This implies that at a sufficiently low monomer concentration, which would vary from protein to protein, no polymer will form. This characteristic monomer concentration is referred to as the critical concentration. At equilibrium, a finite amount of the monomer would exist in equilibrium with the polymer\textsuperscript{52}. The critical concentration is usually determined from a plot of the rate of polymer formation (or amount of polymer) versus protein concentration (Figure 2d). (3) The lag phase is abolished if a small amount of pre-formed nuclei (seed) is provided at the beginning of the reaction (Figure 2e). This phenomenon is referred to as seeding.

**Nucleation-dependent polymerization with secondary pathways**

The theory of the NDP reaction successfully describes the kinetics of polymerization of many proteins. But in a few cases, the kinetics of the increase in the amount of polymerized material is much more abrupt than that predicted by a \( t^\text{\*} \) dependence, and is better described as an exponential time dependence. To explain this exponential time dependence of polymerization kinetics, the theory of NDP reaction was extended to include secondary mechanisms of polymer formation\textsuperscript{56,61}, such as fragmentation\textsuperscript{64,66}, branching and heterogeneous nucleation\textsuperscript{66}.

**Isodesmic (linear) polymerization**

In an isodesmic (linear) polymerization reaction\textsuperscript{56} (Figure 2f), there is no separate nucleation and elongation phase\textsuperscript{59,67}. Rather, polymerization can commence from any of the monomeric subunits. Each association step involves an identical bond, i.e. the rate constants are independent of the size of the polymer. Thus, an isodesmic polymerization reaction can be considered to be similar to the elongation phase of the NDP model. In the kinetics of an isodesmic polymerization reaction, no lag phase is seen, and the rate is fastest at the start of the reaction where the concentration of monomers is the highest; thereafter the rate decreases as the reaction proceeds towards equilibrium. There exists no critical concentration barrier.
Establishing polymerization mechanisms by kinetic analysis

A polymerization reaction is considered to be nucleation-dependent if it shows all the three characteristic features of the NDP mechanism (see above). The features of the NDP reaction are prominent at lower protein concentrations. At very high protein concentrations, nucleation may, however, become relatively favourable, and the lag phase and the dependence on protein concentration of the kinetics may disappear.

Generally, an isodesmic polymerization reaction does not display any of the three characteristic features of an NDP reaction. But it is not always straightforward to distinguish between the two polymerization mechanisms, because the distinction between them is subtle, and rests solely on the nucleus size and the rate constants for dissociation and association. Under some circumstances, an isodesmic polymerization mechanism can readily mimic the features of the NDP mechanism. For a polymerization reaction to be considered as an NDP reaction, it needs to show all three characteristic features (see above), because an isodesmic polymerization reaction can show at least two of the three features. Finally, it is important to realize that the NDP and isodesmic mechanisms represent two extreme cases of polymerization, and a given polymerization reaction may involve both the mechanisms at the same time.

Mechanism of amyloid fibril formation

Onset of amyloid fibril formation

The process of amyloid fibril formation seems to commence from partially structured conformers of proteins (Figure 3). The partial (un)folding of proteins seems to facilitate specific intermolecular interactions, such as hydrophobic and electrostatic interactions, which are required to drive the polymerization of protein molecules into amyloid fibrils. But direct structural information on monomeric partially unfolded conformers competent to form amyloid fibrils, is available in only a few cases, because it is not easy to trap such partially unfolded conformers. Amyloid fibril formation by tetrameric transthyretin commences only after its dissociation into monomers, and the propensity to fibrillate is related inversely to the stability of the tetramer. An HX-NMR study of monomeric transthyretin under amyloidogenic conditions suggested that the formation of the aggregation-competent intermediate is associated with the desta-