

1-1-2008

## Unraveling the mysteries of protein folding and misfolding

Heath Ecroyd

*University of Wollongong*, [heathe@uow.edu.au](mailto:heathe@uow.edu.au)

John A. Carver

Follow this and additional works at: <https://ro.uow.edu.au/scipapers>



Part of the [Life Sciences Commons](#), [Physical Sciences and Mathematics Commons](#), and the [Social and Behavioral Sciences Commons](#)

---

### Recommended Citation

Ecroyd, Heath and Carver, John A.: Unraveling the mysteries of protein folding and misfolding 2008, 769-774.

<https://ro.uow.edu.au/scipapers/929>

---

## Unraveling the mysteries of protein folding and misfolding

### Abstract

This mini-review focuses on the processes and consequences of protein folding and misfolding. The latter process often leads to protein aggregation and precipitation with the aggregates adopting either highly ordered (amyloid fibril) or disordered (amorphous) forms. In particular, the amyloid fibril is discussed because this form has gained considerable notoriety due to its close links to a variety of debilitating diseases including Alzheimer's, Parkinson's, Huntington's, and Creutzfeldt-Jakob diseases, and type-II diabetes. In each of these diseases a different protein forms fibrils, yet the fibrils formed have a very similar structure. The mechanism by which fibrils form, fibril structure, and the cytotoxicity associated with fibril formation are discussed. The generic nature of amyloid fibril structure suggests that a common target may be accessible to treat amyloid fibril-associated diseases. As such, the ability of some molecules, for example, the small heat-shock family of molecular chaperone proteins, to inhibit fibril formation is of interest due to their therapeutic potential.

### Keywords

protein, unraveling, misfolding, mysteries, folding, CMMB

### Disciplines

Life Sciences | Physical Sciences and Mathematics | Social and Behavioral Sciences

### Publication Details

Ecroyd, H. & Carver, J. A. (2008). Unraveling the mysteries of protein folding and misfolding. *IUBMB Life*, 60 (12), 769-774.

## Unravelling the Mysteries of Protein Folding and Misfolding

Heath Ecroyd\* and John Carver

School of Chemistry and Physics, University of Adelaide, SA 5005

\*Corresponding author: heath.ecroyd@adelaide.edu.au

The dogma of protein folding, based largely on the work of Christian Anfinsen some 50 years ago (1), is that all the information required for a protein to fold into its proper three-dimensional structure (and hence functional form) is contained within its amino acid sequence. However, even if, following translation, a protein successfully attains its biologically active state, this often does not herald the end-point of its folding/unfolding life. Many proteins go through cycles of unfolding and refolding due to a variety of factors that include transport across a membrane, cellular secretion, or exposure to stress conditions (e.g. changes in pH, temperature). As a result, the chance for protein misfolding is relatively high and so the whole process must be tightly regulated to ensure that it functions smoothly. The failure of a protein to fold correctly can have serious consequences; it is now recognised that protein misfolding lies at the very heart of a number of our most debilitating diseases (Table 1). This review will focus on the process of protein misfolding, highlighting its consequences (i.e. protein aggregation and precipitation) and the forms that these misfolded protein aggregates may take (ordered vs disordered). In particular, the highly ordered protein aggregate known as an amyloid fibril will be discussed since this form has gained considerable notoriety due to its links to most of the aforementioned diseases.

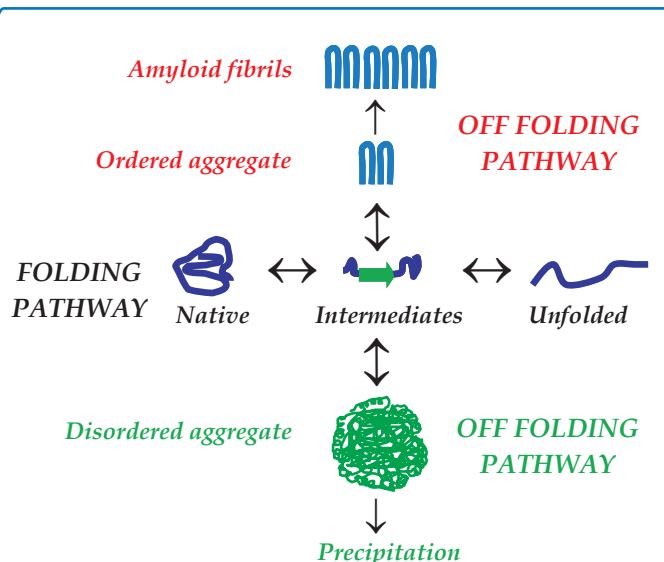
### Proteins Can Aggregate Through Two Distinct Mechanisms

During and immediately following its translation on the ribosome, the newly formed protein meets the first major hurdle of its life: to fold into the conformation it requires in order to fulfil its *raison d'être*. This, in itself, is not a trivial task since the number of theoretical interactions between each of its amino acid side chains far exceeds the total number of protein molecules within the cell and establishing the correct interactions is vital if the protein is to fold correctly. In addition, the protein must fold within the crowded environment of the cell, in which the intracellular concentration of protein can be as high as 350 mg/mL (2), and so the chance of it making inappropriate contacts with other proteins is very high. Yet, the driving force that pushes the protein to attain its lowest free energy state (i.e. its native conformation in the majority of cases) ensures that most proteins fold spontaneously and rapidly (in the order of micro- to milliseconds) and, more often than not, folding occurs without problems (3). Interestingly, some proteins never attain a defined conformation, and instead, in their biologically active state, remain intrinsically disordered, i.e. they have ill-defined secondary and tertiary structures in their native state.

For other proteins, folding does not occur unassisted and instead the whole folding process is overseen by a number of auxiliary proteins, such as catalysts and molecular chaperones, which ensure a high degree of folding fidelity. Whilst occurring quickly, the folding pathway of a protein typically does not occur in one step but instead proceeds through a number of intermediately folded states (each with lower energy than the unfolded protein) in which a few key initial contacts are established that are crucial in directing the correct protein structure. Subsequent hydrogen bonding and hydrophobic interactions enable the protein to attain its fully folded form (Fig. 1). The folding pathway is reversible. The folded protein, when required to or when subjected to stress (that causes the disruptions to hydrogen bonding and hydrophobic interactions between some side chains), partially unfolds to its intermediately folded state(s).

**Table 1. Some known diseases associated with amyloid fibril formation and the main protein component of the aggregates formed.**

Disease	Main protein
Alzheimer's disease	A $\beta$ peptides, Tau
Frontal-temporal dementias	Tau
Parkinson's disease	$\alpha$ -Synuclein
Dementia with Lewy bodies	$\alpha$ -Synuclein
Transmissible spongiform encephalopathies	Prion
Huntington's disease	Huntingtin
Type 2 diabetes	Amylin
Senile systemic amyloidosis	Transthyretin
Familial amyloid polyneuropathy I	Transthyretin
Familial amyloid polyneuropathy III	Apolipoprotein AI
Haemodialysis-related amyloidosis	$\beta_2$ -Microglobulin
Injection-localised amyloidosis	Insulin
Hereditary non-neuropathic systemic amyloidosis	Lysozyme
Spinocerebellar ataxias	Ataxins
Spinocerebellar ataxia 17	TATA-box binding protein
Primary systemic amyloidosis	Ig light chains
Secondary systemic amyloidosis	Serum amyloid A
Amyotrophic lateral sclerosis	Superoxide dismutase I
Medullary carcinoma of the thyroid	Calcitonin



**Fig. 1.** The protein on-folding pathway and the off-folding pathways that lead to protein aggregation. An unfolded protein folds to its native state via the formation of partially folded intermediates. This process is fast and reversible. However, under conditions in which the partially folded intermediates persist (e.g. during times of cellular stress or due to mutation), they can mutually associate via exposed hydrophobic regions that are normally buried into the core of the protein in its native state. When this occurs, the intermediates aggregate via either a disordered or ordered mechanism, leading to the formation of amorphous precipitates or amyloid fibrils, respectively.

Despite the number of checkpoints that exist to ensure proper folding of proteins, problems can arise due to undesirable interactions during folding. The main cause of this is the persistence of intermediately folded states of the protein on the folding pathway, a process that can be exacerbated by mutation and/or cellular stress. These intermediate states, which expose increased hydrophobicity to solution, are prone to self-association, leading to subsequent aggregation and precipitation. When this occurs, the protein leaves the folding pathway and enters the protein off-folding pathway, which is relatively slow (in the order of seconds) and driven primarily by the hydrophobic interactions between intermediately folded states (**Fig. 1**). The off-folding pathway comprises two distinct routes (into either disordered or ordered aggregates) by which aggregation of the protein may proceed. Determining which off-folding pathway predominates is often governed by the speed at which protein unfolding and aggregation occurs, its amino acid sequence and the nature of the intermediates that are formed.

A disordered aggregation mechanism results from the rapid unfolding and subsequent aggregation of intermediately folded proteins, in which individual monomers add to the growing clump of aggregated protein through a random process. This leads to the formation of amorphous aggregates which eventually

become so large that they form an insoluble precipitate. This type of aggregation is most often the bane of protein researchers as it is the underlying mechanism behind inclusion body formation in bacterial cells during recombinant protein expression and is also responsible for proteins 'falling out' of solution when changing buffer conditions. With regards to inclusion body formation, the huge amount of protein formed overwhelms the cell's ability to properly fold the newly expressed protein and so the misfolded protein aggregates and precipitates. However, under normal circumstances in the cell, amorphous aggregation is often not of major concern since it is well equipped to detect their formation and dispose of them into the proteasomal 'dustbin' before they precipitate.

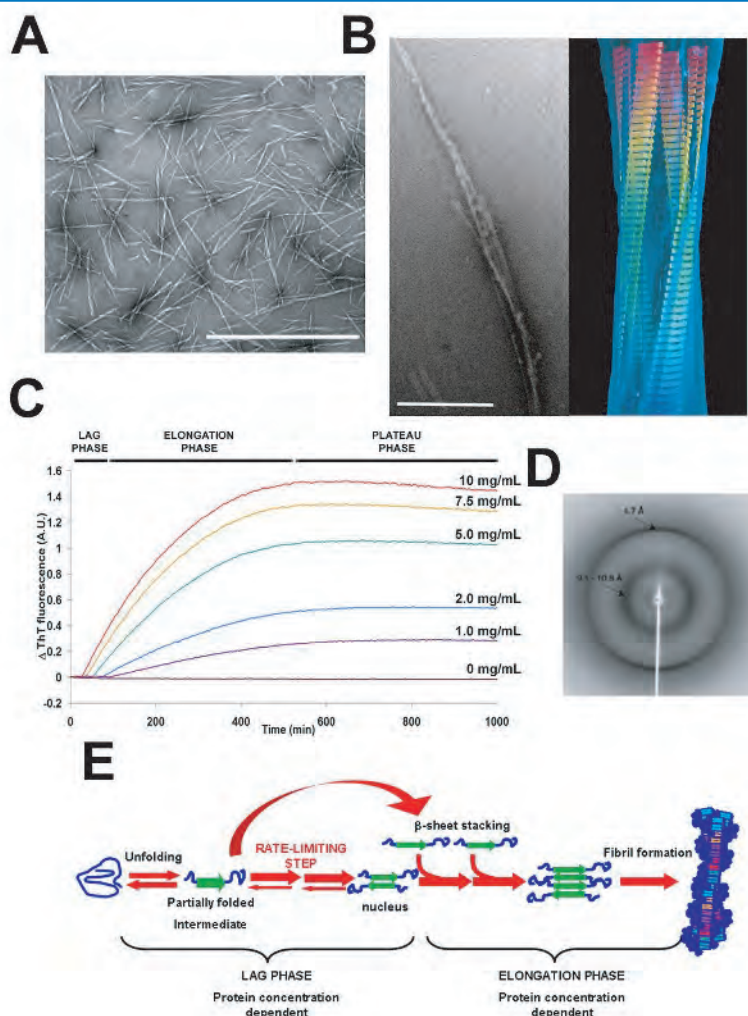
### Amyloid Fibrils are Formed Through an Ordered Aggregation Mechanism

In contrast to the formation of amorphous (disordered) clumps of protein, aggregation may occur more slowly through a highly ordered nucleation-dependent mechanism in which partially folded forms of the protein associate together to form a stable nucleus (the rate determining step). This nucleus then acts as a template to sequester other intermediates to add to the growing thread of aggregated protein (protofibril). The sequential addition of partially folded intermediates to the ends of the chain leads to the formation of a highly structured, insoluble form of protein known as an amyloid fibril (**Fig. 2A,B**). Such a mechanism explains the observed kinetics of fibril formation as monitored using amyloid-binding dyes such as thioflavin T or Congo red (**Fig. 2C**). Both the length of the lag phase and the rate of elongation are highly dependent on the protein concentration through their reliance on the concentration of partially folded intermediates present at any given time (4).

Amyloid fibril formation is often associated with disease and is believed to be causative, or at least linked, to disease onset and progression (5-7). However, the disease related proteins found as fibrillar aggregates *in vivo* share no obvious sequence or structural similarities in their native state (**Table 1**). Moreover, the amyloid fibril conformation has been found to be accessible to a diverse range of proteins, such that it is now thought to be a generic structural form that all proteins can adopt given appropriate conditions (8).

### The Generic Structure of Amyloid Fibrils

The characterisation of amyloid fibril formation by proteins *in vitro* has, to date, largely focused on biophysical studies to determine the structure of the fibril, and biochemical studies into the mechanism and kinetics of the process. Through such techniques as X-ray fibre diffraction, cryo-electron microscopy and solid state NMR spectroscopy we now have a detailed understanding of the core architecture of individual fibrils. All fibrils share a characteristic 'cross  $\beta$ -sheet array', so called since individual fibrils are made up of sheets of  $\beta$ -strands which lie perpendicular to the core axis of the fibril and which stack together to form an



**Fig. 2. Monitoring the formation of amyloid fibrils and their generic core architecture.**

**A.** The typical structure of amyloid fibrils as viewed by transmission electron microscopy, showing them as long, unbranched, rope-like fibres. Scale bar is 1  $\mu\text{m}$ .

**B.** In the left panel, a magnified view of a fibril highlighting its internal protofilament substructure (scale bar is 200 nm). In the right panel, a schematic view of an amyloid fibril formed from insulin. This model shows the core structure of each filament, i.e. the typical cross  $\beta$ -sheet array formed from sheets of  $\beta$ -strands lying perpendicular to the axis of the fibril and the aligning of these  $\beta$ -sheets into individual filaments (reproduced from (26)).

**C.** Monitoring amyloid fibril formation via the change in fluorescence of the amyloid-binding dye thioflavin T upon its binding to the fibril. The kinetics of fibril formation include a lag phase, elongation phase and plateau phase and the rate of fibril formation increases with increasing concentrations of protein.

**D.** X-ray fibre diffraction of amyloid fibrils showing the diagnostic meridional and equatorial reflections which form the 'cross  $\beta$ -sheet' pattern.

**E.** The nucleation-dependent model of amyloid fibril formation. Fibril formation commences with the unfolding of a native protein, forming a pool of partially folded intermediates, a process that is reversible. The partially folded intermediates are able to associate with each other until they reach a critical size/mass at which a stable nucleus is formed. The formation of this nucleus from the partially folded intermediates is slow and rate-limiting in the overall process of fibril formation (lag phase). Fibril elongation then proceeds via the addition of intermediates to the growing nucleus. The mechanism also explains how seeding the reaction increases the reaction rate and decreases the lag phase since addition of preformed fibrils overcomes the time required to form nuclei.

individual filament (**Fig. 2B**). This results in a characteristic cross formed by the meridional ( $\sim 4.7$  Å) and equatorial ( $\sim 9$ – $11$  Å) reflections in X-ray diffraction studies (**Fig. 2D**), which represent the hydrogen bonding distance between adjacent  $\beta$ -strands that make up a  $\beta$ -sheet and the distance between  $\beta$ -sheets, respectively. The presence of this 'cross  $\beta$ -sheet array' as the underlying architecture of fibrils observed by techniques such as transmission electron microscopy and atomic force microscopy is now seen as the diagnostic test for the presence of amyloid fibrils. Mature fibrils are commonly composed of two to six protofilaments that plait together into rope-like fibres, 5–10 nm in diameter and up to a few microns in length. The fibrils formed are often unbranched, extremely stable and resistant to degradation by proteases and denaturants. These properties are thought to be responsible for the difficulty the cell has in eliminating fibrils once they have been formed.

The overall stability of the fibril is achieved by intermolecular hydrogen bonds involving backbone atoms between the amide and carbonyl groups of the polypeptide main chain; the peptide backbone being common to all proteins is therefore thought to dictate why all fibrils share a common morphology. It also explains why very structurally diverse proteins are able to adopt the amyloid-fibril conformation, including those which are predominately  $\alpha$ -helical in their native state (9–11). However, the propensity for a given peptide or protein to form fibrils can vary dramatically with sequence and some regions of a protein may be more prone to aggregation than others. Thus, whilst it may be true that all proteins are capable of forming fibrils, the composition and amino acid sequence of a protein profoundly affects its propensity to adopt such structures. Moreover, the tendency of a polypeptide chain to aggregate, rather than fold correctly, depends on a number of intrinsic factors, including the propensity for it to form  $\beta$ -strands, its hydrophobicity and its overall net charge (12,13). The specific link between these physiochemical properties of constituent amino acid residues and their aggregation propensities has led to the development of a number of predictive algorithms for amyloidogenic regions of proteins that are based solely on their amino acid sequences (14,15).

## Toxic Protein Aggregation and its Prevention

In each of the amyloid diseases, and there have now been at least 20 which have been identified, the fibrils that are formed are primarily associated with one protein or protein fragment, e.g. the amyloid- $\beta$  peptides  $A\beta_{1-40}$  and  $A\beta_{1-42}$  in Alzheimer's disease,  $\alpha$ -synuclein in Parkinson's disease and the prion protein in the transmissible spongiform encephalopathies (see **Table 1**). In many of these diseases, the fibrils then self assemble into tangled plaques, the hallmark of most neurodegenerative conditions and the site at which the toxic effect of fibril formation is most evident. However, the cytotoxicity associated with amyloid formation is not restricted to disease-related proteins; fibrils and their precursors formed from non-disease related proteins, such as the SH3 domain from bovine phosphatidylinositol 3' kinase and the N-terminal domain of *Escherichia coli* HypF protein (HypF-N), show similar levels of cytotoxicity (16). There remains considerable debate as to the species responsible for the cytotoxicity of amyloid fibrils. Although there are obvious negative effects of extracellular amyloid plaque deposition, recent studies have suggested that it is primarily the soluble, pre-fibrillar dimers, trimers or other such oligomers, which are formed during the early stages of fibril formation, that are responsible for cell toxicity in neurodegenerative diseases such as Alzheimer's and Parkinson's disease (5,7). Others have indicated that the mature fibril can also be toxic (17,18) and, in fact, the cytotoxic species may vary depending on the precursor fibril-forming protein.

A wide variety of biochemical changes have been reported following exposure of neuronal cells in culture to amyloid fibrils or their precursors. It is not clear how the species formed during amyloid fibril assembly cause cell death, and indeed whether the mechanism behind the toxicity is the same for all amyloid fibril-forming proteins. A number of hypotheses have been proposed, for example, that a soluble precursor forms a pore-like structure in cell membranes (the amyloid channel or amyloid pore) and this culminates in neuronal death by unregulated membrane permeabilisation (19,20). Others have suggested that the toxicity of prefibrillar amyloid species is due to the production of reactive oxygen species (e.g. hydrogen peroxide) by the aggregating target protein itself, which are generated as a consequence of the fibril-forming process (21). In support of this, cells can be protected against amyloid aggregate toxicity by treatment with antioxidants such as tocopherol, lipoic acid and reduced glutathione. An advantage of some of these anti-oxidant compounds is that they also are able to inhibit the process of fibril formation (22), most likely due to a direct effect on the hydrophobic association steps required for nuclei formation. Thus, these anti-oxidants or derivatives thereof are promising drug targets due to their combined anti-oxidant and anti-amyloidogenic activities.

Other well-described inhibitors of protein aggregation, and therefore amyloid fibril formation, are molecular chaperone proteins, in particular, intracellular small heat shock proteins (sHsps) and extracellular clusterin (23). The sHsps are a ubiquitous group of proteins that are the first line of defence a cell has against physiological stress

conditions that promote protein aggregation. The chaperone action of sHsps does not require ATP hydrolysis and therefore they can be utilised by the cell under conditions in which energy levels are low, e.g. during cellular stress. sHsps seem to employ two distinct mechanisms to prevent protein aggregation. In some instances, they bind to long-lived, partially structured intermediates, primarily through hydrophobic interactions, to form a stable, soluble, chaperone-target protein complex (i.e. a 'reservoir of intermediates'). Neither sHsps or clusterin have the capability of refolding the target protein but instead act to maintain its solubility until cellular conditions allow them to be picked up and acted upon by other chaperones, such as Hsp70 and Hsp60, that use ATP hydrolysis to refold the protein. In other cases, sHsps may only transiently interact with the target to stabilise it and allow it to refold back to its native state upon release. We have found the latter to be the mechanism utilised by the sHsp,  $\alpha$ -crystallin, against fibril formation by apolipoprotein C-II (24).

No matter which mechanism is utilised, sHsps and clusterin are ideally suited to prevent amyloid fibril formation since they act very efficiently against slowly aggregating target proteins, a process that governs the ordered aggregation mechanism that leads to fibril formation. Interestingly, the expression of sHsps and clusterin is upregulated in many amyloid neurodegenerative diseases and they are found in high amounts in amyloid plaques (25), presumably as a result of their attempts to prevent protein aggregation. Studies in which these chaperone proteins are overexpressed in cellular models of amyloidoses will enable an assessment of their therapeutic potential in the treatment of such diseases.

## Conclusions

Whilst often overlooked, the folding and unfolding process faced by proteins during their life cycle is not a trivial one. Like all biological processes, the folding pathway is tightly regulated to ensure proteins reach their correct, functional form. However, problems can occur and the number of protein conformational diseases that are now recognised is an indication of the importance of proteins achieving and maintaining their correct fold. That the amyloid fibril conformation is potentially accessible to all proteins, no matter what their native state, indicates that for many this form of toxic protein aggregation is a constant threat. Significantly, diseases associated with amyloid fibril formation represent some of the western world's most debilitating conditions and, since many are associated with old age, will become more prevalent over the coming decades as the population ages. As such, a greater understanding of the mechanism by which fibrils are formed and strategies to prevent it are required.

## References

1. □ Anfinsen, C.B. (1973) *Science* **181**, 223-230
2. □ Ellis, R.J. (2001) *Curr. Opin. Struct. Biol.* **11**, 114-119
3. □ Dobson, C.M. (2003) *Nature* **426**, 884-890

references continued on page 11


**References continued from page 7**

4. □ Harper, J.D., and Lansbury, P.T. Jr. (1997) *Annu. Rev. Biochem.* **66**, 385-4075.
5. □ Chiti, F., and Dobson, C.M. (2006) *Annu. Rev. Biochem.* **75**, 333-366
6. □ Dobson, C.M. (2001) *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **356**, 133-145
7. □ Stefani, M., and Dobson, C.M. (2003) *J. Mol. Med.* **81**, 678-699
8. □ Dobson, C.M. (1999) *Trends Biochem. Sci.* **24**, 329-332
9. □ Guijarro, J.I., Sunde, M., Jones, J.A., Campbell, I.D., and Dobson, C.M. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 4224-4228
10. □ Fandrich, M., Fletcher, M.A., and Dobson, C.M. (2001) *Nature* **410**, 165-166
11. □ Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G., and Dobson, C.M. (1999) *Proc. Natl. Acad. Sci. USA* **96**, 3590-3594
12. □ Chiti, F., Stefani, M., Taddei, N., Ramponi, G., and Dobson, C.M. (2003) *Nature* **424**, 805-808
13. □ DuBay, K.F., Pawar, A.P., Chiti, F., Zurdo, J., Dobson, C.M., and Vendruscolo, M. (2004) *J. Mol. Biol.* **341**, 1317-1326
14. □ Fernandez-Escamilla, A.M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004) *Nat. Biotechnol.* **22**, 1302-1306
15. □ Pawar, A.P., Dubay, K.F., Zurdo, J., Chiti, F., Vendruscolo, M., and Dobson, C.M. (2005) *J. Mol. Biol.* **350**, 379-392
16. □ Bucciattini, M., Giannoni, E., Chiti, F., *et al.* (2002) *Nature* **416**, 507-511
17. □ Novitskaya, V., Bocharova, O.V., Bronstein, I., and Baskakov, I.V. (2006) *J. Biol. Chem.* **281**, 13828-13836
18. □ Ward, R.V., Jennings, K.H., Jepras, R., Neville, W., Owen, D.E., Hawkins, J., Christie, G., Davis, J.B., George, A., Karran, E.H., and Howlett, D.R. (2000) *Biochem. J.* **348**, 137-144
19. □ Caughey, B., and Lansbury, P.T. (2003) *Annu. Rev. Neurosci.* **26**, 267-298
20. □ Lashuel, H.A., and Lansbury, P.T. Jr. (2006) *Q. Rev. Biophys.* **39**, 167-201
21. □ Tabner, B.J., El-Agnaf, O.M., German, M.J., Fullwood, N.J., and Allsop, D. (2005) *Biochem. Soc. Trans.* **33**, 1082-1086
22. □ Ono, K., Hamaguchi, T., Naiki, H., and Yamada, M. (2006) *Biochim. Biophys. Acta.* **1762**, 575-586
23. □ Carver, J.A., Rekas, A., Thorn, D.C., and Wilson, M.R. (2003) *IUBMB Life* **55**, 661-668
24. □ Hatters, D.M., Lindner, R.A., Carver, J.A., and Howlett, G.J. (2001) *J. Biol. Chem.* **276**, 33755-33761
25. □ Sun, Y., and MacRae, T.H. (2005) *FEBS J.* **272**, 2613-2627
26. □ Jimenez, J.L., Nettleton, E.J., Bouchard, M., Robinson, C.V., Dobson, C.M., and Saibil, H.R. (2002) *Proc. Natl. Acad. Sci. USA* **99**, 9196-9201