# Protein misfolding, evolution and disease

The number of different types of proteins in the human body approaches 100 000. Proteins are engaged in promoting or controlling virtually every event on which our lives depend. In recent years we have become familiar with the intricate and elegant forms that these molecules adopt in their functional native states. Moreover, we have begun to learn a great deal about the manner in which these structures are attained through the complex process of protein folding, whether this takes place in the laboratory or within the natural environment of the living cell[1–3]. It is also becoming increasingly evident that biological systems have evolved elaborate procedures to ensure that proteins fold correctly or, if they do not, that they are detected and degraded before any serious harm can ensue to the host organism[4].

## Protein misfolding is linked to disease

Despite these controls, a range of debilitating human diseases is associated with protein misfolding events that result in the malfunctioning of the cellular machinery[5]. Cystic fibrosis is one example where mutations in the gene encoding a crucial transport protein result in the protein folding incorrectly and hence not being secreted in the quantity required for proper function. Other diseases, including some types of familial emphysema, result from mutations that result in improper trafficking of proteins to the sites where they are needed. Recently, however, most attention has been focused on a group of diseases where proteins or fragments of proteins convert from their normally soluble forms to insoluble fibrils or plaques, which accumulate in a variety of organs including the liver, spleen and brain[6–9] (Fig. 1). The final forms of these aggregates often have a well-defined fibrillar nature, and are known as amyloid (see Box 1), hence the use of the term amyloidosis to describe many of the clinical conditions with which they are associated.

This group of diseases, of which nearly 20 have been described, includes Alzheimer's and Parkinson's diseases, the spongiform encephalopathies such as Creutzfeldt–Jakob disease, type II diabetes and a range of less well-known but often equally serious conditions such as fatal familial insomnia[6]. These diseases can be sporadic, inherited or even infectious, and are often manifest only late in life. Each disease is associated with a particular protein and aggregates of these proteins are thought to be the direct or indirect origin of the pathological conditions associated with the disease in question. In some cases, the quantity of material involved is enormous, with several kilograms of protein being deposited in certain manifestations of systemic amyloidosis. Remarkably, despite the range of proteins involved in these diseases, including several well-known proteins such as lysozyme, transthyretin and the prions, all of which have unique and characteristic native folds, the fibrils in which they are found in the disease states are extremely similar in their overall appearance[10].

## Soluble proteins convert into aggregates under denaturing conditions

Studies of the mechanism of the conversion of the normally soluble proteins into amyloid fibrils have benefited from the fact that, in many cases, the structural transitions of the disease-associated molecules can be reproduced under laboratory conditions[7]. In order to achieve this, a common procedure has been to expose the folded proteins to mildly denaturing conditions, such as low pH or elevated temperatures. Those transitions that have been studied in detail have generally revealed the existence of intermediates prior to the formation of well-defined aggregates, and some studies have shown a clear kinetic lag phase before rapid development of fibrils occurs. This latter behaviour has analogies with the more familiar, extensively studied processes of crystal growth and polymer gelation, and is thought to be associated with the need to

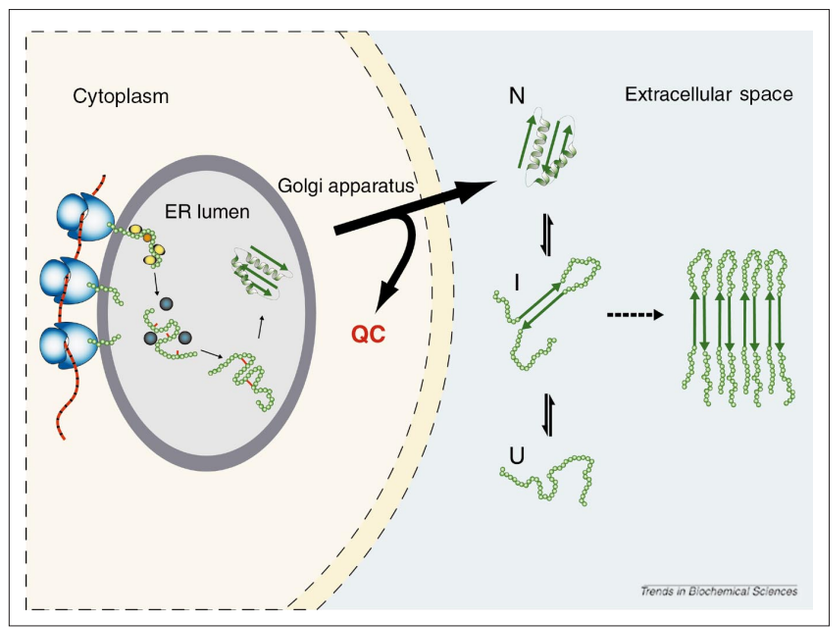


**Figure 1**
Schematic representation of the possible mechanism of formation of amyloid fibrils by a globular protein. After synthesis on the ribosome, the protein is assumed to fold in the endoplasmic reticulum (ER), aided by molecular chaperones that deter aggregation of incompletely folded species. The correctly folded protein is secreted from the cell and functions normally in its extracellular environment. Under certain conditions the protein unfolds, at least partially, and becomes prone to aggregation. This can result in the formation of fibrils and other aggregates that accumulate in tissue. It is likely that small aggregates, as well as the highly organized fibrils and plaques, can give rise to pathological conditions in at least some cases. N, I and U refer to native, partially folded (intermediate) and unfolded states of the protein, respectively. QC refers to the quality control mechanism that prevents incompletely folded proteins being secreted from the endoplasmic reticulum[35].

0968–0004/99/$ – See front matter © 1999, Elsevier Science Ltd. All rights reserved. PII: S0968-0004(99)01445-0



Petitioner Merck, Ex. 1140, p. 329
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00004

**Box 1. Amyloid**

The term 'amyloid' was originally used to describe proteinaceous aggregates associated with diseases of the type discussed in this article because some of their properties resembled those of starch (amylose). The traditional test for amyloid in tissue involves the observation of a red shift in the light absorption of the dye Congo red and of a characteristic green birefringence under polarized light. Both these effects are attributable to the interaction of the dye molecule with the regularly spaced protein chains. The detection of the characteristic 'cross-β' structure in X-ray diffraction and typical morphology in the EM micrographs are now generally considered to be requisites for protein aggregates to be properly described as amyloid.

develop 'nuclei' or 'seeds': small aggregates from which larger molecular assemblies can grow[11]. Indeed, the ability to enhance growth of fibrils by seeding is a possible mechanism for the rapid progression of many of these diseases following onset, and indeed for the infectivity of the spongiform encephalopathies associated with the prion proteins[12,13].

Despite the association of prions with spongiform encephalopathies, no recombinant prion protein has yet been shown to be able to induce infectivity in animals and hence provide the ultimate proof of the 'protein only' hypothesis of the prion diseases[14,15]. Moreover, production of the aggregated (scrapie) form from the soluble (cellular) form of the prion protein has proved to be difficult, although some conversion to protease-resistant material can be achieved by addition of extracts from infective tissue[16]. Recently, however, the formation of fibrillar material with the characteristic morphology of amyloid from a large recombinant fragment of the human prion protein has been reported[17]. This was achieved by exposing the soluble form of the protein to rather extreme conditions, including reduction of the single disulphide bond that stabilizes the native state. Interestingly, this conversion was associated with the presence of a soluble intermediate with a predominantly β structure, instead of the largely helical nature of the native protein. Amyloid fibrils have long been known to be rich in β structure, giving rise to the well-known α-helix-to-β-sheet transition characterizing the conversion of cellular to scrapie forms of the prion protein, and indeed of the conversion to aggregates of those other proteins whose native folds are largely helical[10,18].

## Amyloid is a generic structural form of proteins

The involvement of only a handful of proteins in amyloid diseases has commonly been thought to be associated with some particular conformational character of the protein sequences involved. But recently compelling evidence has accumulated that the ability to form amyloid is not a peculiarity of this small group of proteins. A clue came, fortuitously, from an NMR study of the SH3 module of PI3 kinase, which has no connection with any known disease[19]. At low pH, where the protein is at least partially unfolded, the solution turned into a viscous gel after several hours. When this gel was examined by electron microscopy and other techniques it was found to contain well-defined fibrils with all the characteristics of those associated with the amyloid diseases (Fig. 2). Formation of similar fibrils has also been observed when a domain of fibronectin was heated to temperatures above that required for unfolding[20] and, indeed, is now being reported for an increasing range of other proteins and fragments of proteins.

A recent study set out to test the hypothesis that amyloid could be a common state of all proteins, by designing solution conditions to see if fibrils could be deliberately formed from a soluble protein not associated with any disease[21]. The selected method was to add trifluoroethanol – a solvent known to denature proteins but, unlike traditional denaturants such as urea, one that generates partially unfolded states where hydrogen bonds between peptide groups are still stable. After several days under the chosen conditions, the protein involved, acylphosphatase, duly formed amyloid fibrils with all the characteristics of those associated with disease. Interestingly, there are a number of observations in the early literature of fibrous gels being found under a variety of non-physiological conditions for a range of different proteins[22]. Although the relationship between these fibrils and those found in disease was not recognized at the time, these findings reinforce the idea that the ability to form amyloid is a generic property of polypeptide chains. This ability can readily be explained by the fact that the intermolecular bonds that stabilize this material involve the peptide backbone, which is common to all proteins.

The degree of structural regularity of the fibrils grown under carefully controlled laboratory conditions can be greater than that of those commonly observed *in vivo*. Indeed, careful growth of the SH3 fibrils over a period of weeks resulted in material sufficiently well ordered to be studied in detail by cryo-electron microscopy[23]. Image reconstruction techniques devised by Saibil and colleagues[23] show clearly that the fibrils consist of protofilaments, containing specific segments of β sheet, wound around each other in a helical array. The resolution of these techniques is not yet sufficient to reveal directly the structure in molecular detail. X-ray fibre diffraction studies, however, show that the polypeptide chains adopt a 'cross-β' structure in which hydrogen bonds are formed between polypeptide chains in directions parallel to the fibre axis[10]. Using this information, a model was built[23] to suggest how the polypeptide chains are assembled in the protofilaments (Fig. 3). This arrangement of the polypeptide chain, albeit no doubt with variations on a theme for different sequences, might represent a first glimpse of the common structure of this

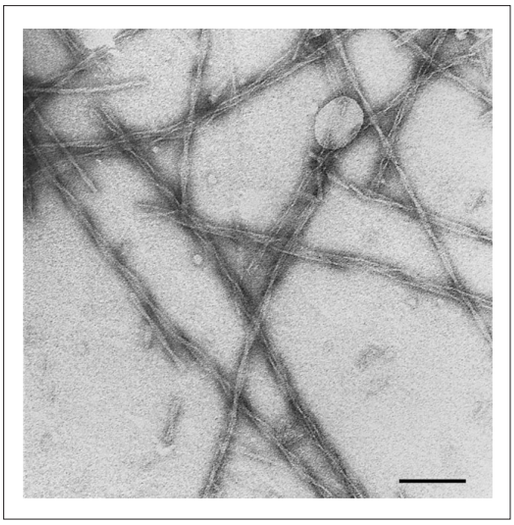

**Figure 2**
Electron micrograph of fibrils formed from an SH3 domain by incubation of a solution containing the protein at low pH (Ref. 19). Under these solution conditions the protein is partially unfolded and slowly aggregates to form a gel that contains the fibrils. Fibrils associated with the various amyloid diseases have a highly similar appearance to these fibrils formed under laboratory conditions[10]. The scale bar is 100 nm. Reproduced, with permission, from Ref. 19.



Petitioner Merck, Ex. 1140, p. 330
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00004

alternative, highly organized form of protein molecules.

### Living systems avoid forming amyloid

If amyloid fibril formation is a generic property of polypeptide chains, why is its occurrence in biology restricted to a very small number of proteins? What prevents the rest of our proteins forming this material in our bodies? And why are aggregates normally seen only when the usual amino-acid sequence of the disease-related proteins is altered by mutation, or following infection, or in old age? These questions are particularly pertinent because, once formed, amyloid fibrils are essentially indestructible under physiological conditions, probably because of the large number of hydrogen bonds that must be disrupted to rescue the polypeptide chain from the aggregated state. Undoubtedly, some amino-acid sequences are more prone to aggregation than others and some proteins are present *in vivo* at very much higher concentrations than others. But the fundamental answers to these questions must be that biology has found a way to avoid the formation of this unwanted material under normal physiological conditions. Many factors must be involved in this protective mechanism, but the selection of sequences during evolution that can fold efficiently to a globular form in which the polypeptide chain and the hydrophobic residues are hidden in the interior is likely to be particularly important[21]. Thus, except when the protein is exposed to denaturing conditions, the peptide backbone is not accessible to form the interchain hydrogen bonds associated with amyloid fibrils (Fig. 3).

One particularly interesting characteristic of proteins is that, under normal conditions, their native states are only marginally stable relative to denatured states. How, then, can the population of unfolded or partially folded protein species in a biological environment be sufficiently small to avoid aggregation? The answer is likely to be connected with the cooperativity of the protein-folding process[24]. The 'two-state' nature of protein denaturation that is commonly observed means that the populations of partially folded intermediate states are small and, indeed, that all segments of the polypeptide chain are firmly locked into the close-packed structure. Cooperativity of this type has long been associated with the existence of well-defined native proteins capable of carrying out their diverse functions, including resisting degradation by proteolysis. It now appears that, cooperativity may be equally crucial for the avoidance of aggregation[24]. Another very important requisite is that the cellular environment is such that denaturation of proteins does not normally occur under conditions where unfolded chains tend to aggregate. In particular, pH and temperature are generally carefully controlled, and molecular chaperones and degradation mechanisms are present to cope with the majority of aggregation-prone species.

The generic nature of amyloid formation therefore suggests that the existence of the diseases associated with this material results, at least in part, from the breakdown of the normal control and regulation mechanisms within a living organism. In support of this conclusion, conditions such as the low pH environments in endosomes associated with protein translocation or in lysosomes associated with protein degradation, have been implicated in these diseases[8]. Indeed, it has even been suggested that the prion proteins might experience environments in endosomes capable of reduction of the disulphide bond found to enhance the conversion to the fibrillar state[17]. Moreover, at least the majority of the mutations associated with the familial amyloid-related diseases are destabilizing, and many give rise to intermediates[7] that are an indication of the loss of the cooperativity associated with the wild-type protein. Aggregates in Alzheimer's disease are associated with peptide fragments of a precursor protein that result from the processing or partial degradation of this protein within the cellular environment[13]. Such peptides do not have a stable globular fold to protect them against aggregation.

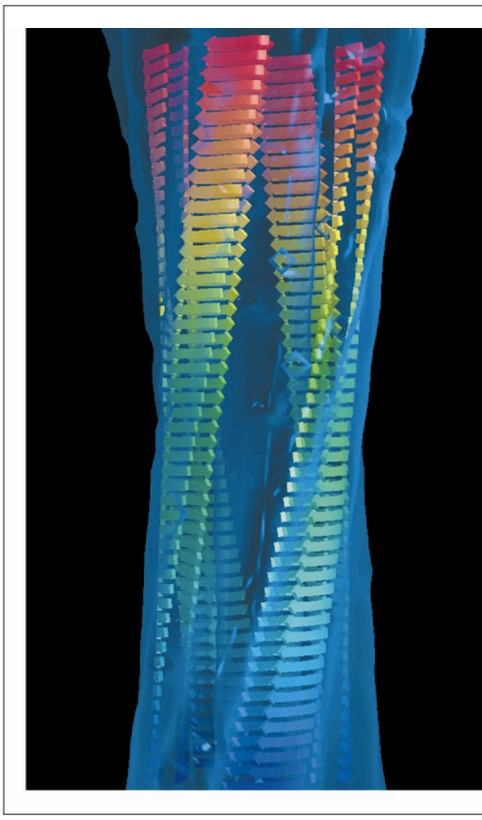

**Figure 3**
Molecular model of an amyloid fibril derived from cryo-EM analysis of fibrils grown from an SH3 domain. The fibril consists of four 'protofilaments' that twist around one another to form a hollow tube with a diameter of approximately 60 Å (Ref. 23). The model shown here represents one way in which regions of the polypeptide chain involved in β-sheet structure could be-assembled within the fibrils.

### New insights into evolutionary biology?

The conventional view of protein structures is that they have emerged under the pressure of the development of greater efficiency and new functionalities. Properties such as the cooperativity of their structures ensure that a single, well-defined structure exists in solution to allow efficient binding to other molecules and effective functional behaviour, for example, catalysis of very specific chemical reactions. It is interesting to speculate that avoidance of aggregation, particularly to highly insoluble amyloid fibrils, might be an equally important driving force in the evolutionary design of natural proteins. This is particularly significant in the context of the extraordinarily crowded nature of the cellular environment[25]. Interestingly, several proteins have been found to fold at rates close to the limits imposed by the physical principles that apply to all molecules[26]. This fits in with our preconceived notions that biology strives for maximum efficiency, in the same way that at least some enzymes appear to have reached perfection in terms of their catalytic efficiency[27]. The ability of proteins to fold rapidly is not only an effective way of generating biological



Petitioner Merck, Ex. 1140, p. 331
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00004

activity quickly, but it could, in at least some cases, confer an evolutionary advantage by minimizing the competing intermolecular processes of aggregation. It is interesting in this regard that the globular region of the prion protein has been found to fold rapidly in the absence of pathogenic mutations[28].

It appears, therefore, that biology has generally succeeded, through the coupled evolution of protein sequences and the environments in which they fold and function, in avoiding the formation of amyloid fibrils by the molecules that carry out the myriad functions of a living cell. But has the amyloid form of protein molecules evolved to be of any functional benefit in living systems? In general this seems unlikely, because of the intractable nature of the fibrils and the difficulty in controlling their growth once it is initiated. It is probable, however, that the major pathogenic species in at least some diseases are not the fibrils themselves but smaller aggregates that are precursors to their formation. If this is the case, the formation of fibrils could in fact be a means of sequestering such species in a form in which they are less harmful[8]. However, recent findings suggest that the conversion of soluble proteins to fibrillar structures could be involved in biological processes of a rather different type from those associated with mammalian diseases.

Studies of *Saccharomyces cerevisiae* have resulted in the discovery of proteins that have infective properties related to those of the mammalian prions[29,30]. These 'yeast prions' have been found to convert *in vitro* into ordered aggregates with all the characteristics of amyloid fibrils, and to display classic seeding behaviour[31,32]. A particularly remarkable finding of this work, however, is that the ability to convert into the amyloid form can be passed from a parent to its offspring through the process of cell division. The yeast prions have therefore been described as non-chromosomal genetic material as they are the basis for inherited traits within the organism. This concept is revolutionary enough but recently, proteins with similar characteristics have been found in filamentous fungi. In this case, however, there is evidence that these prions might be involved in the mediation of the normal cellular functions in these organisms and not simply linked to disease[29,30]. If this is correct, it is one of the more extraordinary examples of a generally emerging concept that the control of folding events is intimately linked with the control of biological activity.

The discovery of mammalian prions has led to the hypothesis that infection can occur without the need for bacteria or viruses. The properties of the yeast and fungal prions now suggest that transfer of genetic information is possible without nucleic acids. This neatly turns the tables on RNA research, where the discovery of ribozymes took the exclusivity of biological catalysis away from proteins[33]. It might also give food for thought about how the transfer of information could have taken place in early life forms, in much the same way as the properties of ribozymes stimulated ideas about the control of chemical reactions in primitive organisms[34]. A better understanding of amyloid fibrils and the way they form should result in improved knowledge of the pathological conditions that lead to many of the diseases that are emerging in the aging population of the modern world, and of opportunities for their prevention and treatment. And it might stimulate new ideas about the nature of the sequences that are emerging from the various genome projects, and perhaps lead to a deeper understanding of the fundamental driving forces behind biological evolution.

## Acknowledgements

I acknowledge very valuable discussions on this article with John Ellis and Carol Robinson. I am grateful to Jose Jiminez and Helen Saibil for providing Figure 3 and to Adam Rostom for producing Figure 1. This paper is a contribution from the Oxford Centre for Molecular Sciences, which is funded by the BBSRC, EPSRC and MRC. The research of C. M. D. is also supported by the Howard Hughes Medical Institute and the Wellcome Trust.

## References

1 Fersht, A. R. (1999) *Structure and Mechanism in Protein Science*, W. H. Freeman
2 Dobson, C. M. and Ptitsyn, O. B., eds (1999) *Curr. Opin. Struct. Biol.* 9, 89–141
3 Radford, S. E. and Dobson, C. M. (1999) *Cell* 97, 291–298
4 Plemper, R. K. and Wolf, D. H. (1999) *Trends Biochem. Sci.* 29, 266–270
5 Thomas, P. J., Qu, B-H. and Pederson, P. L. (1995) *Trends Biochem. Sci.* 20, 456–459
6 Tan, S. Y. and Pepys, M. B. (1994) *Histopathology* 25, 403–414
7 Kelly, J. W. (1998) *Curr. Opin. Struct. Biol.* 8, 101–106
8 Lansbury, P. T. (1999) *Proc. Natl. Acad. Sci. U. S. A.* 96, 3342–3344
9 Perutz, M. F. (1999) *Trends Biochem. Sci.* 24, 58–63
10 Sunde, M. and Blake, C. C. F. (1997) *Adv. Protein Chem.* 50, 123–159
11 Mullins, J. (1993) *Crystallization* (3rd edn), Butterworth–Heinemann
12 Harper, J. D. and Lansbury, P. T. (1997) *Annu. Rev. Biochem.* 66, 385–407
13 Horwich, A. L. and Weissman, J. S. (1997) *Cell* 89, 499–510
14 Prusiner, S. B. (1997) *Science* 278, 245–251
15 Weissman, C. (1999) *J. Biol. Chem.* 274, 3–6
16 Caughey, B. and Chesbro, B. (1997) *Trends Cell Biol.* 7, 56–62
17 Jackson, G. S. *et al.* (1999) *Science* 283, 1935–1937
18 Fink, A. (1998) *Fold. Des.* 3, R9–R23
19 Guijarro, J. I. *et al.* (1998) *Proc. Natl. Acad. Sci. U. S. A.* 95, 4224–4228
20 Litvinovich, S. V. *et al.* (1998) *J. Mol. Biol.* 280, 245–258
21 Chiti, F. *et al.* (1999) *Proc. Natl. Acad. Sci. U. S. A.* 96, 3590–3594
22 Clarke, A. H. *et al.* (1981) *Int. J. Pept. Protein Res.* 17, 380–392
23 Jiminez, J. L. *et al.* (1999) *EMBO J.* 18, 815–821
24 Dobson, C. M. and Karplus, M. (1999) *Curr. Opin. Struct. Biol.* 9, 92–101
25 Luby-Phelps, K. (1994) *Curr. Opin. Cell Biol.* 6, 3–9
26 Eaton, W. A. *et al.* (1998) *Acc. Chem. Res.* 31, 745–753
27 Shakhnovich, E., Abkevich, V. and Ptitsyn, O. B. (1996) *Nature* 379, 96–98
28 Wildegger, G., Liemann, S. and Glockshuber, R. (1999) *Nat. Struct. Biol.* 6, 550–553
29 Lindquist, S. (1997) *Cell* 89, 495–498
30 Wickner, R. B. *et al.* (1999) *J. Biol. Chem.* 274, 554–558
31 Glover, J. R. *et al.* (1997) *Cell* 89, 811–819
32 King, C-Y. *et al.* (1997) *Proc. Natl. Acad. Sci. U. S. A.* 94, 6618–6622
33 Cech, T. R. (1986) *Sci. Am.* 255, 64–75
34 Csermely, P. (1997) *Trends Biochem. Sci.* 22, 147–148
35 Trombetta, E. S. and Helenius, A. (1998) *Curr. Opin. Struct. Biol.* 8, 587–592

**CHRISTOPHER M. DOBSON**

Oxford Centre for Molecular Sciences, University of Oxford, New Chemistry Laboratory, South Parks Road, Oxford, UK OX1 3QT.

**Next month in *Trends in Cell Biology***

Septins: structural polymers or signalling GTPases?
*Christine M. Field and Douglas Kellogg*

Protein translocation in apoptosis
*Alan G. Porter*

Unravelling the kinetoplastid paraflagellar rod
*J.A. Maga and J.H. LeBowitz*

Bacterial Type II protein export and pilus biogenesis: more than just homologies?
*David Nunn*

Titin: a molecular-control freak
*John Trinick and Larissa Tskhovrebova*



Petitioner Merck, Ex. 1140, p. 332
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00004