Beilstein-Institut Molecular Informatics: Confronting Complexity, May 13th - 16th 2002, Bozen, Italy

PROTEIN MISFOLDING AND ITS LINKS WITH HUMAN DISEASE

CHRISTOPHER M. DOBSON

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK
E-Mail: cmd44@cam.ac.uk

Received: 18th June 2002 / Published: 15th May 2003

ABSTRACT

The ability of proteins to fold to their functional states following synthesis on the ribosome is one of the most remarkable features of biology. The sequences of natural proteins have emerged through evolutionary processes such that their unique native states can be found very efficiently even in the complex environment inside a living cell. But under some conditions proteins fail to fold correctly, or to remain correctly folded, in living systems, and this failure can result in a wide range of diseases. One group of diseases, known as amyloidoses, which includes Alzheimer's and the transmissible spongiform encephalopathies, involves deposition of aggregated proteins in a variety of tissues. These diseases are particularly intriguing because evidence is accumulating that the formation of the highly organized amyloid aggregates is a generic property of polypeptides, and not simply a feature of the few proteins associated with recognized pathological conditions. Moreover, such aggregates appear to posses inherent toxicity. That aggregates of this type are not normally found in properly functional biological systems is a further testament to the efficiency of biological evolution, in this case resulting in the emergence of a variety of mechanisms inhibiting their formation. Understanding the nature of such protective mechanisms is a crucial step in the development of strategies to prevent and treat these debilitating diseases.

PROTEIN FOLDING AND MISFOLDING

A living organism may contain as many as 50,000 different types of protein. Following synthesis on the ribosome, each protein molecule must fold into the specific conformational state that is encoded in its sequence in order to be able to carry out its biological function. How

this process happens is one of the most fascinating and challenging problems in structural biology (1, 2). In the cell, folding takes place in a complex and highly crowded environment, and the folding process is aided by a range of auxiliary proteins (3, 4). These proteins include molecular chaperones, whose main role is to protect the incompletely folded polypeptide chain from non-productive interactions, particularly those that result in aggregation, and folding catalysts, whose job is to speed up potentially slow steps in the folding process such as those associated with the isomerization of peptidylprolyl bonds and the formation of disulphide linkages. It is evident, however, that the code for folding is contained within the amino-acid sequence of the protein itself because it has been shown that proteins can reach their correct folded structure *in vitro* in the absence of any auxiliary factors, providing that appropriate conditions can be found (5). The questions of how the fold is encoded in the sequence, and how the process of folding takes place, are at last beginning to be answered in a credible manner. Progress in this area has come about as a consequence of novel experimental strategies to probe the structural transitions that take place during folding *in vitro,* and of innovative theoretical studies designed to simulate these transitions (6, 7, 8). Perhaps of greatest importance has been the fact that these approaches have been brought together in a synergistic manner to advance our fundamental understanding of this highly complex process (7, 8, 9).

Our present understanding of the folding *in vitro* of small proteins, typically those of less than about 100 residues, is that the rate of folding is limited primarily by the time required to find the crucial interactions that are needed to permit rapid progression to the native structure. For larger proteins, however, the folding process is typically slower and more complex, and is usually associated with the population of one or more partially folded intermediate states. In addition, events that may be termed misfolding may take place during the search for the stable native-like contacts between residues (7). That such complexities are seen even in the benign environment of a dilute solution of a pure protein suggests that they are even more likely to occur in the crowded environment of the cell. Undoubtedly, molecular chaperones are able to mitigate some of the consequences of this complex behaviour and provide some protection for the incompletely folded chain (4). But the idea that proteins can misfold, or fold to intermediates that may undergo undesirable side reactions such as aggregation, provides insight into potential problems that can arise during folding even in the best designed environments. Folding and unfolding are also now known to be coupled to many of the key events in the functioning of a biological system, including translocation of proteins across membranes, protein trafficking,

secretion of extracellular proteins, and the control and regulation of the cell cycle (10). Thus, the failure of proteins to fold, or to remain folded under physiological conditions, is likely to cause malfunctions and hence disease. Indeed, an increasing number of diseases is now linked to phenomena that can loosely be described as misfolding; a selection of these is given in table 1.

Table 1. Representative protein folding diseases. (ER, endoplasmic reticulum. Data from (11), (14), (15), (16)).

Disease	Protein	Site of folding
hypercholesterolaemia	low-density lipoprotein receptor	ER
cystic fibrosis	cystic fibrosis transmembrane regulator	ER
phenylketonuria	phenylalanine hydroxylase	cytosol
Huntington's disease	huntingtin	cytosol
Marfan syndrome	fibrillin	ER
osteogenesis imperfecta	procollagen	ER
sickle cell amaemia	haemoglobin	cytosol
α1-antitrypsin deficiency	α-1-antitrypsin	ER
Tay-Sachs disease	β-hexosaminidase	ER
scurvy	collagen	ER
Alzheimer's disease	β-amyloid/presenilin	ER
Parkinson's disease	α-synuclein	cytosol
scrapie/Creutzfeldt- Jakob disease	prion protein	ER
familial amyloidoses	transthyretin/lysozyme	ER
retinitis pigmentosa	rhodopsin	ER
cataracts	crystallins	cytosol
cancer	P53	cytosol

PROTEIN AGGREGATION AND AMYLOID DISEASES

Among the diseases listed in table 1 are those that are associated with the deposition of proteinaceous aggregates in a variety of organs such as the liver, heart and brain (14, 15, 16, 17). Many of these diseases are described as "amyloidoses", because the aggregated material stains with dyes such as Congo red in a manner similar to starch (amylose), and the typical fibrous structures as "amyloid fibrils". A list of known amyloid diseases is given in table 2, along with the protein component that is associated with the extracellular aggregates formed in each case (18).

It is evident that these diseases include many of the most debilitating conditions in modern society, particularly those associated with ageing such as type II diabetes and Alzheimer's disease. Some are familial, some are associated with medical treatment (e.g. haemodialysis) or infection (the prion diseases), and some are sporadic (e.g. most forms of Alzheimer's). Many of

the diseases (such as the amyloidoses associated with the protein transthyretin) can be found in both sporadic and familial forms. In addition to these diseases there are others, notably Parkinson's and Huntington's diseases (16, 19), that appear to involve very similar aggregates but which are intracellular not extracellular and are not therefore included in the strict definition of amyloidoses.

Table 2. Fibril protein components and precursors in amyloid diseases. (Data from (21)).

clinical syndrome	fibril component
Alzheimer's disease	Aβ peptide, 1-42, 1-43
spongiform encephalopathies	full length prion or fragments
primary systemic amyloidosis	intact light chain or fragments
secondary systemic amyloidosis	76-residue fragment of amyloid A protein
familial amyloidotic polyneuropathy I	transthyretin variants and fragments
senile systemic amyloidosis	wild-type transthyretin and fragments
hereditary cerebral amyloid angiopathy	fragment of cystatin-C
haemodialysis-related amyloidosis	β2-microglobulin
familial amyloidotic polyneuropahty II	fragments of apolipoprotein AI
Finnish hereditary amyloidosis	71-residue fragment of gelsolin
type II diabetes	fragment of islet-associafed polypeptide
medullary carcinoma of the thyroid	fragments of calcitonin
atrial amyloidosis	atrial natriuretic factor
lysozyme amyloidosis	full length lysozyme variants
insulin-related amyloid	full length insulin
fibrinogen α-chain amyloidosis	fibrinogen α-chain variants

Among the proteins linked with amyloidosis is lysozyme, a protein whose folding we have studied in particular depth (7, 10). Our studies led to the idea that this protein would be an exciting one to choose to try to understand at the molecular level the nature of the "misfolding" transition that converts the protein from a soluble to a fibrillar structure (20). One of the striking characteristics of the amyloid diseases is that the fibrils associated with all of them are very similar in their overall properties and appearance (18). The fibrils are typically long (often several microns), unbranched and ca. 10 nm in diameter. They have a variety of tinctorial properties, notably staining with Congo red and exhibiting a green birefringence under polarized light. A range of experiments, particularly X-ray fibre diffraction, indicates that the fibrils have extensive b-sheet character, and that these sheets run perpendicular to the fibril axis to generate what is described as a cross-b structure (18). This observation is remarkable in view of the fact that the soluble native forms of the proteins associated with these diseases vary considerably in their size and secondary structure. Moreover, some of the proteins are intact in

the fibrous form whilst others are at least partially degraded. This similarity of the fibrillar forms of the proteins prompted the proposal that there are strong similarities in the inherent structure of the amyloid fibrils and in the mechanism by which they are formed (12, 18). Thus the study in depth of the relationship between the folding and "misfolding" of one system could have very general value in understanding this whole class of diseases.

One of the very important observations in this regard is that the fibrillar forms of many of the disease-related proteins can be generated *in vitro* from the normal soluble forms. In the case of fibrils formed from peptides (including fragments of larger proteins) that are largely unstructured in solution, such fibrils typically form under a wide range of solution conditions. In the case of fibrils formed from intact globular proteins, however, the fibrils typically form under conditions under which the native state is significantly destabilized (15, 27). Thus in the case of the two known disease-related human lysozyme variants, fibrils form most readily at low pH or at elevated temperatures (20, 23) (see Figure 1).

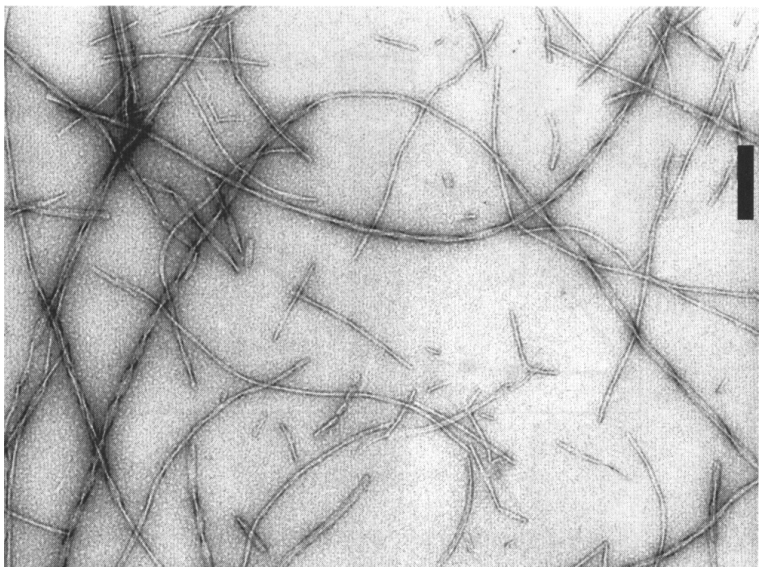

Figure 1. Amyloid fibrils from the Ile56Thr variant of human lysozyme produced by transmission electron microscopy. Scale bar, 200nm. (From (23)).

Experiments to examine the nature of the amyloidogenic variants (Ile56Thr and Asp67His) show that the structures of the proteins in their soluble native states are similar to that of the wild-type protein and have no obvious perturbations that could explain their tendency to aggregate (20). But experiments reveal that the two variants are destabilized relative to the wild-type protein to similar extent although the origin of this instability is different (24). Thus the Ile56Thr variant is destabilized largely because its folding rate is reduced, whilst the Asp67His

variant is destabilized largely because it unfolds more rapidly. It therefore appears that the decreased protein stability rather than the altered folding kinetics is a common feature of these two variants. In further experiments it has been demonstrated that the lower stability of the native state results in the population of a partially folded state that is very similar to the major (a-domain) intermediate populated on the folding pathway of the wild-type protein (24, 25). This finding can be rationalized because the mutations destabilizing the native fold are located in the b-domain of the protein, the region that is not highly structured in the predominant intermediate.

This observation suggests a mechanism for the formation of amyloid fibrils from the variant lysozymes, in which the partially folded intermediates aggregate as the first step in the formation of the ordered structures found in the fibrils (20, 23, 24). Calculations based on hydrogen exchange protection measurements suggest that the population of partially folded proteins under physiological conditions is about 1000 times greater in the variants than in wild-type lysozyme (24, 25). This conclusion allows one to speculate that the amyloidogenic variants have sufficient stability to fold efficiently so as to escape the quality control mechanisms in the endoplasmic reticulum and to be secreted into extracellular space (12). However, unlike the wild-type protein, they have insufficient stability to remain in their native states under all conditions to which they are exposed. Moreover, it has been speculated that endosomal compartments where the pH is reduced might be important in the formation of amyloid deposits. Under low pH conditions *in vitro* conversion to amyloid fibrils has been found to be particularly facile (23). In addition, *in vitro* experiments have shown that fibril formation is accelerated substantially when solutions are seeded with preformed fibrils. Such a mechanism has been suggested as being responsible for the rapid onset of some amyloidoses, and indeed of the infectivity of the prion diseases (16).

THE GENERIC NATURE OF THE AMYLOID STRUCTURE

In studies of the conformation of the SH3 domain from bovine PI3 kinase at low pH, when the protein is in a largely unfolded state, it was found that the protein readily formed a viscous gel (26). Examination of the gel using electron microscopy revealed the presence of large numbers of fibrils that closely resemble those formed from the proteins associated with amyloid diseases. Moreover, the aggregates showed all the other characteristics of amyloid fibrils, and were to all intents and purposes identical to these other structures. This observation prompted us to explore

the possibility that similar fibrils could be formed from other proteins by placing them under mildly denaturing conditions that do not immediately result in visible precipitation, and examining the solutions over often prolonged periods of time (27). For a range of representative proteins with no known connection with any disease we have been able to find conditions under which conversion occurs into fibrils very similar to those associated with amyloid disease. We shall refer to these types of structures as "amyloid fibrils" in future, regardless of whether or not they are associated with disease. The proteins studied included wild-type human lysozyme, which forms fibrils under similar but more destabilizing conditions than the amyloidogenic intermediates (23), and the archetypal globular protein, myoglobin, which readily forms fibrils when the heme group is removed (13). For myoglobin it is particularly evident that the protein has undergone a substantial conversion from its soluble a-helical form to the aggregated b-sheet conformation found in the fibrils. Such findings prompted us to conclude that the ability to form amyloid fibrils is not a characteristic associated wholly or primarily with those proteins found to be associated with amyloidoses, but a property that could be common to many or indeed all proteins under appropriate conditions (12, 26, 27, 28).

Models of the structure of amyloid fibrils indicate that the core regions of the protofilaments are based on hydrogen bonds between the polypeptide main chain (18, 29). As the main chain is common to all polypeptides, it explains how the fibrils from different proteins appear so similar, regardless of the length and sequence of the polypeptide involved. In contrast to the situation in native proteins, we suggest that the side chains are not a strong influence on the basic structure of the protofilaments (30). Nevertheless, the manner in which the protofilaments pack together to form mature fibrils may well depend significantly on those parts of the polypeptide chain that are not involved directly in the close-packed β-strands (31). Thus, the fibrils from different peptides and proteins are variations on a common theme. The ability of natural proteins to form amyloid structures does not violate the crucial hypothesis that a protein sequence codes for a single fold (5). The nature of the amyloid core structure is that it is not coded for by the sequence, as it is formed as a consequence of interactions involving the common polypeptide backbone of all proteins. Its rate and ease of formation will of course depend on the sequence, both as a consequence of the readiness for different side-chains to pack together within the structure, and as a consequence of the solubility and stability of the sequence in solution (32). It is the side-chains, however, that code for the specific fold of globular proteins by their ability to pack together in a unique manner to form compact globular structures.

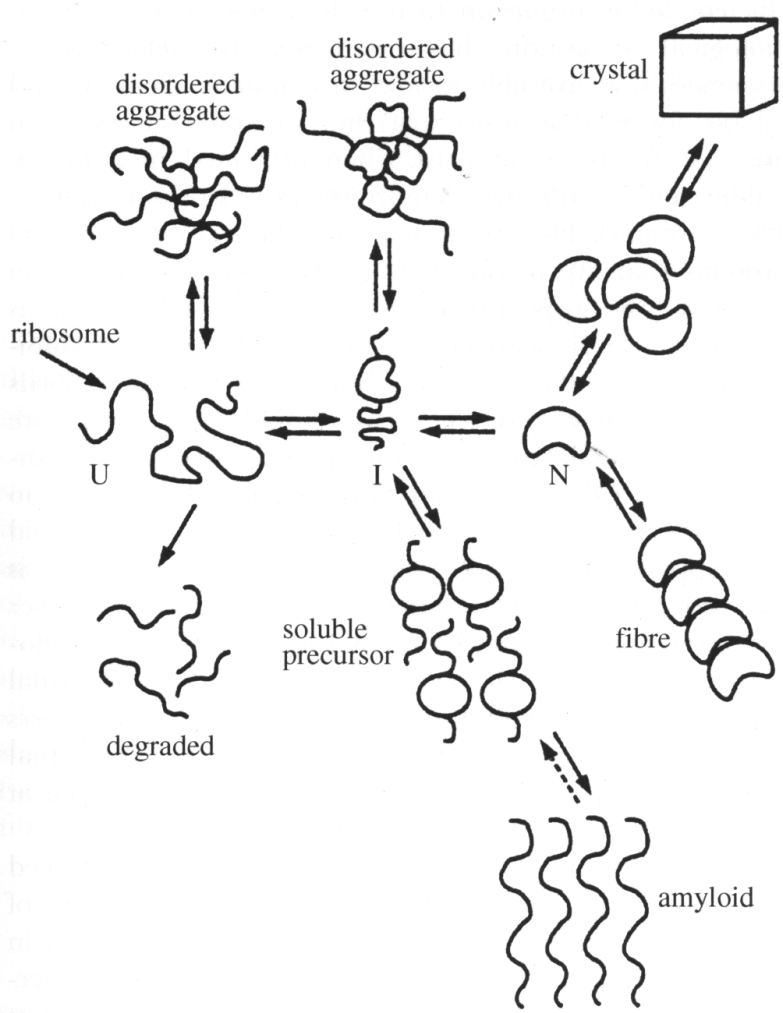

Figure 2. Schematic of some of the states accessible to a polypeptide chain following its biosynthesis. In its monomeric state, the protein is assumed to fold from it highly disordered unfolded state (U) through a partially structured intermediate (I) to a globular native state (N). The native state can form aggregated species, the most ordered of which is a three-dimensional crystal, whilst preserving its overall structure. The unfolded and partially folded states can form aggregated species that are frequently disordered, but highly ordered amyloid fibrils can form through a nucleation and growth mechanism. (From (33)).

The proposal that amyloid fibrils represent a generic structure of polypeptide chains has stimulated us to suggest that the conformational properties of all proteins should be considered in terms of the multiple states that are accessible to such structures (6, 33). This suggestion is illustrated in Figure 2 in a schematic manner. This diagram suggests that the various fates awaiting a polypeptide chain once it has been synthesized in the cell will depend on the kinetics and thermodynamics of the various equilibria between different possible states. Thus, the normal folding process may pass through partially folded states on the route to the fully native state, but the aggregation of these species will be minimized by the presence of molecular chaperones.

In addition, if the protein is able to fold rapidly, any partially folded species will have a short lifetime, reducing the probability of intermolecular interactions occurring. Moreover, once folded, the native state is generally a highly compact structure that conceals the polypeptide main chain within its interior. Such a state is protected from aggregation except through the relatively weak interactions of surface side chains and is unable to form the strong intermolecular hydrogen bonds associated with the polypeptide backbone. Provided that the native state is maintained under conditions where it remains folded, aggregation to amyloid fibrils will be resisted by the kinetic barrier associated with unfolding, even if the aggregated state is thermodynamically more stable. Importantly, the cooperative nature of protein structures means that virtually none of the polypeptide chain in individual molecules is locally unfolded, and that virtually no molecules in an ensemble are globally unfolded, even though native proteins are only marginally stable relative to denatured ones under normal physiological conditions (6, 33).

A Common Origin of Amyloid Diseases

This picture of the various structures accessible to polypeptide chains enables us to speculate on the origins of the amyloid diseases from the point of view of the physicochemical properties of protein molecules. If the stability or cooperativity of the native state of a protein is reduced, for example by a mutation, the population of non-native states will increase, as discussed above for the amyloidogenic variants of lysozyme. This rise will increase the probability of aggregation, as the concentration of polypeptide chains with at least partial exposure to the external environment will be greater. Whether or not aggregation does occur will depend on the concentration of protein molecules, the intrinsic propensity for a given sequence to aggregate when unfolded, and on the rate of the aggregation process. The fact that formation of ordered amyloid fibrils can be seeded, like the well-studied processes of crystallization and gelation, means that once the aggregation process is initiated it often proceeds very rapidly (23, 34, 35). In the absence of seeding there can be long “lag” phases before aggregation occurs. This lag can be thought of as arising because the growth of a fibril cannot occur until a “nucleus” of a small number of aggregated molecules is formed. Such a nucleus can be formed by the local fluctuations in concentration that occur in solution as a result of random molecular motion. When such fluctuations result in a local concentration of molecules above a critical value, the molecules associate with one other to form a species that is sufficiently large to have intrinsic

stability, and hence to grow in size by interacting with other molecules in the solution. The act of seeding provides such nuclei to the solution and hence reduces or abolishes the lag phase.

On this view of the aggregation process, the critical first step for globular proteins is the partial or complete unfolding of the native structure. In the case of most proteins, except perhaps the smallest ones, unfolding under physiological conditions will not generate the type of highly unfolded states seen in high concentrations of denaturant. Instead the denatured protein will be more stable in a partially collapsed state that may well resemble intermediates observed in the normal folding process (36). The generic nature of the structure and mechanism of amyloid formation suggests that the nature of the residual structure in such intermediates has little direct importance in dictating the structure of the resulting aggregates. It may, however, it may indicate which regions of the protein are most likely to be incorporated in the b-sheet segments of the fibrils. It is, however, the enhanced ability of some mutated proteins to access partially or completely unfolded states that is the underlying origin of many of the familial amyloid diseases such as those involving lysozyme that are discussed above (20, 25). Moreover, in some cases mutational changes can also enhance the propensity of unfolded or partially folded states to aggregate, providing an additional mechanism for the amyloidogenity of some disease-related variants (37, 38).

This general view of amyloid formation can readily be extended to include the existence of sporadic as well as familial and infectious diseases. Sporadic diseases could arise from the loss of the normal control and regulation processes that enable proteins to be maintained in their required states under all conditions in living organisms organism (12, 27). It is perhaps particularly likely that such control is lost in ageing, and the majority of the cases of sporadic diseases such as Alzheimer's or type II diabetes are associated with old age. It is significant that in a high proportion of elderly people even wild-type transthyretin, which in its mutant forms is associated with familial amyloidosis, is found as amyloid structures in organs such as the heart (15). The exact manner in which this happens is unclear, but it could be as the result of statistical factors (comparable with those observed in lag phases *in vitro*) or due to the effects of changes in the cellular environment, or the failure of the normal degradation mechanisms for proteins. Interestingly, many of the diseases in this category involve the deposition of peptide fragments, and the process of degradation in compartments such as lysozomes involves conditions such as low pH that serve to unfold proteins prior to the action of proteases. Such mildly denaturing conditions are particularly favourable for the nucleation and growth of amyloid structures.

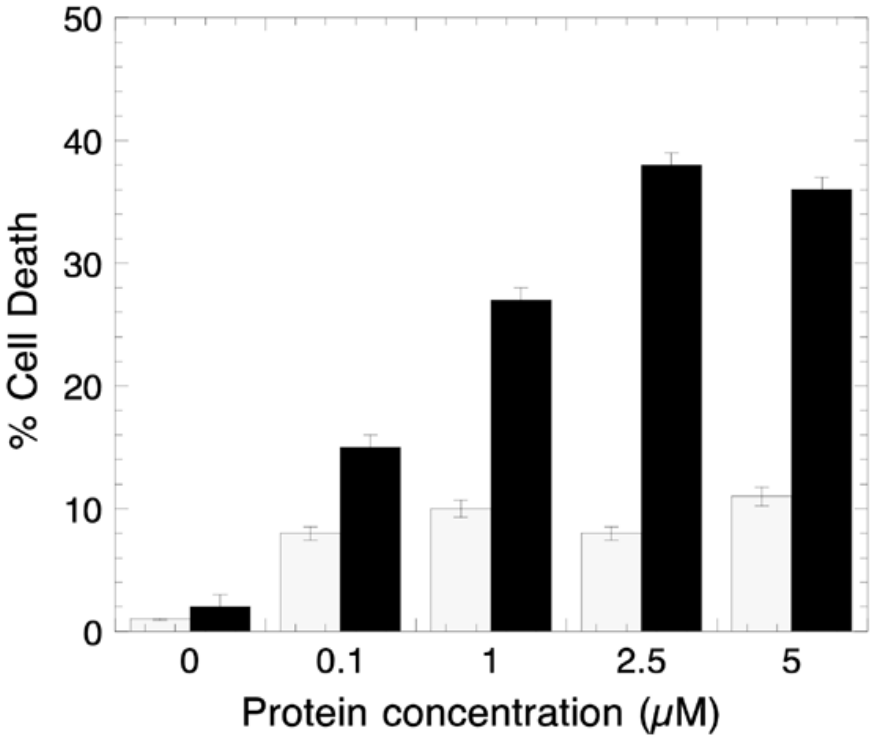

Figure 3. Percentage of cell deaths induced by 48-h-aged Hyp-F-N aggregates at different protein concentrations. Solid bars refert to aggregated protein, dotted bars to control experiments performed in the presence of soluble protein.

CONCLUDING REMARKS

It seems likely that biological evolution has resulted in the selection of polypeptide sequences that are able to fold to compact, globular and soluble forms that resist aggregation and conversion to fibrillar structures, at least when protected in a highly stable and controlled environment (12). These remarkable structures are the native states of proteins that are involved in every process occurring in the cell. The appearance of amyloid deposits in living systems may therefore be associated with mutations that destabilize proteins sufficiently for them to convert into fibrils when the wild-type protein would not, but leave them sufficiently stable to evade the quality control mechanisms in the cell and to function sufficiently normally to allow the organism to develop and reproduce. Amyloid deposits also appear in old age where evolutionary pressure is reduced after the reproductive life span, and in other conditions such as kuru or bovine spongiform encephalopathy (BSE), which are connected with abnormal practices such as ingestion of tissue from other members of the same species (17).

The generic picture of amyloid structure and the mechanism of its formation therefore provides a conceptual framework for linking together the various pathological conditions associated with

deposition of this material. This hypothesis has recently been extended by the observation that the species formed *in vitro* during the early stages of aggregation of proteins not linked to disease, like the early aggregates of disease associated proteins, can be highly toxic when added to cells in culture (39). This finding suggests that the control and regulation of aggregation can be even more crucial to the viability of the organism than was previously thought if misfolded proteins are not just non-functional but also potentially toxic. Moreover, it suggests that there may be at least common features in the pathology of the various misfolding diseases. Our knowledge of the mechanism of fibril growth and degradation should therefore help enormously to design effective strategies for the prevention or treatment of the various members of the family of amyloid diseases.

Moreover, the fundamental knowledge that is emerging from our ability to probe the aggregation properties of a range of sequences, rather than just those identified in association with recognized diseases, should enable the role of particular interactions and specific structural motifs in these processes to be explored (22, 27, 38). This knowledge could result in future approaches to drug design that are significantly more general and potentially more effective than those currently being explored. Given that many of the amyloid diseases are associated with old age, and modern medical and agricultural practices, the need for novel approaches to the avoidance or treatment of these conditions will be increasingly important in the future. One can speculate that as the human life span increases, and the complexity of our societies increases, the number of cases of the known diseases is likely to increase substantially, and indeed that novel diseases associated with the aggregation of proteins not so far linked to clinical symptoms might become highly significant in future years.

This article is an edited and updated version of a paper first published in Phil. Trans. R. Soc. Lond. B, 356, 133-145 (2001). The research of C.M.D is supported in part by a programme grant from the Wellcome Trust.

REFERENCES

[1] Dobson, C. M. & Fersht, A. R. (eds.) (1995). Protein folding. *Phil. Trans. R. Soc. Lond. B* **348**:1-119.

[2] Pain, R. H. (ed.) (2000). Mechanisms of Protein Folding, Second Edition Oxford.

[3] Gething, M.-J. & Sambrook, J. (1992). Protein folding in the cell. *Nature* **355**:33-45.

[4] Ellis, R. J. & Hartl, F. U. (1999). Principles of protein folding in the cellular environment. *Curr. Opin. Struct. Biol.* **9**: 102-110.

[5] Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science* **181**:223-230.

[6] Dobson, C. M. & Karplus, M. (1999). The fundamentals of protein folding: bringing together theory and experiment. *Curr. Opin. Struct. Biol.* **9**:92-101.

[7] Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M., Karplus, M. (2000). Understanding protein folding via free energy surfaces from theory and experiment. *Trends Biochem. Sci.* **25**:331-339.

[8] Daggett. V. & Fersht. A. R. (2002). *Cell*.

[9] Vendruscolo, M., Paci, E., Dobson, C. M., Karplus, M. (2001). Three Key Residues Form a Critical Contact Network in a Transition State for Protein Folding, *Nature* **409**: 641-646.

[10] Radford, S. E. & Dobson, C. M. (1999). From computer simulations to human disease: emerging themes in protein folding. *Cell* **97**:291-298.

[11] Thomas, P. J., Qu, B. H., Pederson, P. L. (1995). Defective protein folding as a basis of human disease. *Trends Biochem. Sci.* **20**:456-459.

[12] Dobson, C. M. (1999). Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24**:329-332.

[13] Fandrich, M., Fletcher, M. A., Dobson, C. M. (2001). Formation of amyloid fibrils from myoglobin. (Submitted).

[14] Tan, S. Y. & Pepys, M. B. (1994). Amyloidosis. *Histopathology* **25**:403-414.

[15] Kelly, J. W. (1998). The alternative conformations of amyloid proteins and their multi-step assembly pathways. *Curr. Opin. Struct. Biol.* **8**:101-106.

[16] Lansbury, P. T. (1999). Evolution of amyloid: what normal protein folding may tell us about fibrillogenesis and disease. *Proc. Natl. Acad. Sci. USA* **96**:3342-3344.

[17] Prusiner, S. (1997). Prion diseases and the BSE crisis. *Science* **278**:245-251.

[18] Sunde, M. & Blake, C. C. F. (1997). The structure of amyloid fibrils by electron microscopy and X-ray diraction. *Adv. Protein Chem.* **50**:123-159.

[19] Perutz, M. F. (1999). Glutamine repeats and neuro-degenerative disease: molecular aspects. *Trends Biochem. Sci.* **24**:58-63.

[20] Booth, D. R. (and 10 others) (1997). Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. *Nature* **385**:787-793.

[21] Sunde, M., Serpell, L. C., Bartlam, M., Fraser, P. E., Pepys, M. B., Blake, C. C. F. (1997). Common core structure of amyloid fibrils by synchrotron X-ray diffractions. *J. Mol. Biol.* **273**:729-739.

[22] Chiti, F., Taddei, N., White, P. M., Bucciantini, M., Magherini, F., Stefani, M., Dobson, C. M. (1999). Mutational analysis of acylphosphatase suggests the importance of topology and contact order in protein folding. *Nature Struct. Biol.* **6**:1005-1009.

[23] Morozova-Roche, L. A., Zurdo, J., Spencer, A., Noppe, W., Receveur, V., Archer, D. B., Joniau, M., Dobson, C. M. (2000). Amyloid fibril formation and seeding by wild type human lysozyme and its disease related mutational variants. *J. Struct. Biol.* **130**:339-351.

[24] Canet, D., Sunde, M., Last, A. M., Miranker, A., Spencer, A., Robinson, C. V., Dobson, C. M. (1999). Mechanistic studies of the folding of human lysozyme and the origin of amyloido-genic behavior in its disease related variants. *Biochemistry* **38**:6419-6427.

[25] Canet, D., Last, A. M., Tito, P., Sunde, M., Spencer, A., Archer, D. B., Redfield, C., Robinson, C. V., Dobson, C. M. (2002). Local Cooperativity in the Unfolding of the Asp67His Variant of Human Lysozyme is a Key Aspect of its Amyloidogenic Behaviour. *Nature Struct. Biol.* **9**:308-315.

[26] Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D., Dobson, C. M. (1998). Amyloid fibril formation by an SH3 domain. *Proc. Natl. Acad. Sci. USA* **95**:4224-4228.

[27] Chiti, F., Webster, P., Taddei, N., Clark, A., Stefani, M., Ramponi, G., Dobson, C. M. (1999). Designing conditions for in vitro formation of amyloid protolaments and fibrils. *Proc. Natl. Acad. Sci. USA* **96**:3590-3594.

[28] Dobson, C. M. (2001). Protein folding and human disease *Phil. Trans. R. Soc. Lond. B*.

[29] Minton, A. P. (2000). Implications of macromolecular crowding for protein assembly. *Curr. Opin. Struct. Biol.* **10**:34-39.

[30] Jimënez, J. L., Guijarro, J. I., Orlova, E., Zurdo, J., Dobson, C. M., Sunde, M., Saibil, H. R. (1999). Cryo-electron micro-scopy structure of an SH3 amyloid fibril and model of the molecular packing. *EMBOJ*. **18**:815-821.

[31] MacPhee, C. E. & Dobson, C. M. (2000). Formation of mixed fibrils reveals the generic nature and potential application of protein amyloid structures. *J. Am. Chem. Soc*. **122**:12707-12713.

[32] Chamberlain, A. K., MacPhee, C. E., Zurdo, J., Morozova-Roche, L. A., Hill, H. A. O., Dobson, C. M., Davis, J. J. (2000). The ultrastructural organisation of amyloid brils by atomic force microscopy. *Biophys. J.* **79**:3282-3293.

[33] Chiti, F., Capanni, C., Taddei, N., Stefani, M., Ramponi, G., Dobson, C. M. (2000). Specific regions of a protein determine the kinetics of aggregation and amyloid formation (in preparation).

[34] Dobson, C. M. (1999). How do we explore the energy landscape for folding? In: Simplicity and complexity in proteins and nucleic acids (ed. H. Fraunfelder, J. Deisenhofer & P. G. Wolynes), pp. 15-37. Berlin: Dahlem University Press.

[35] Harper, J. D. & Lansbury, P. T. (1997). Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *A. Rev. Biochem*. **66**, 385-407.

[36] Krebs, M. R. H., Wilkins, D. K., Chung, E.W., Pitkeathly, M. C., Chamberlain, A., Zurdo, J., Robinson, C. V., Dobson, C. M. (2000). Formation and seeding of amyloid fibrils from wild-type hen lysozyme and a peptide fragment from the b-domain. *J. Mol. Biol.* **300**:541-549.

[37] Dobson, C. M., Sali, A., Karplus, M. (1998). Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Ed. Eng*. **37**:868-893.

[38] Villegas, V., Zurdo, J., Filimonov, V. V., Aviles, F. X., Dobson, C. M., Serrano, L. (1999). Protein Engineering as a Strategy to Avoid Formation of Amyloid Fibrils, *Protein Sci*. **9**:1700-1708.

[39] Chiti, F. Taddei, N. Baroni, F. Capanni, C. Stefani, M. Ramponi G., Dobson, C. M. (2002). Kinetic Partitioning of Protein folding and Aggregation, *Nature Struct. Biol.* **9**:137-143.

[40] Bucciantini, M., E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C. M. Dobson, Stefani, M. (2002). Inherent Cytotoxicity of Aggregates Implies a Common Origin for Protein Misfolding Diseases. *Nature* **416**:507-511.