ELSEVIER

Seminars in Cell & Developmental Biology 15 (2004) 3–16

seminars in
CELL & DEVELOPMENTAL
BIOLOGY

www.elsevier.com/locate/semcdb

Principles of protein folding, misfolding and aggregation

Christopher M. Dobson*

Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

Abstract

This review summarises our current understanding of the underlying and universal mechanism by which newly synthesised proteins achieve their biologically functional states. Protein molecules, however, all have a finite tendency either to misfold, or to fail to maintain their correctly folded states, under some circumstances. This article describes some of the consequences of such behaviour, particularly in the context of the aggregation events that are frequently associated with aberrant folding. It focuses in particular on the emerging links between protein aggregation and the increasingly prevalent forms of debilitating disease with which it is now known to be associated.
© 2003 Elsevier Ltd. All rights reserved.

Keywords: Amyloid formation; Biological self-assembly; Energy landscape; Molecular evolution; Neurodegenerative disease

1. Introduction

Proteins are the most abundant molecules in biology other than water. We contain perhaps 100,000 different types of protein and they stimulate or control virtually every chemical process on which our lives depend [1]. Different proteins are distinguished by a different order of amino acids in the polymeric sequence of typically 300 such building blocks. Following their biosynthesis, the majority of proteins must be converted into tightly folded compact structures in order to function. As many of these structures are astonishingly intricate, the fact that folding is usually extremely efficient is a remarkable testament to the power of evolutionary biology. It is sobering to recognise that, because there are 20 different naturally occurring amino acids found in proteins, the total possible number of different proteins with the average size of those in our bodies is much greater than the number of atoms in the universe. Natural proteins are therefore a very special group of molecules indeed. Their properties are not typical of random sequences, but have been selected through evolutionary pressure to have specific characteristics—of which the ability to fold to a unique structure and hence to generate enormous selectivity and diversity in their functions—is a particularly important one. As we shall see later, however, under some conditions even natural proteins can revert to behaviour that is typical of polymers that have not been subject to such careful evolutionary selection.

The interior of a cell is an extraordinarily complex environment in which proteins and other macromolecules are present at a concentration of 300–400 mg/ml [2]. It is now known that within the cells of living organisms there are large numbers of auxiliary factors that assist in the folding process, including folding catalysts and molecular chaperones [3]. These factors serve to enable polypeptide chains to fold efficiently in the complex and crowded milieu of the cell but they do not determine their native structures; the latter are fully encoded by the amino acid sequences. The question of how proteins find their unique native states simply from the information contained within this code is at the heart of molecular biology. As well as its intrinsic significance, our ability to understand the process of protein folding will give us insight into the more general problem of the way that biological systems—ultimately whole organisms—have evolved the ability to self-assemble. The robustness and predictability of biological self-assembly in comparison to related processes in non-biological systems is arguably the most remarkable feature of living systems. Understanding protein folding, perhaps the most fundamental example of biological self-assembly, is therefore a first step on the path to resolving one of the most important questions that can be addressed by modern science [4].

2. The underlying mechanism of protein folding

The mechanism by which even a simple protein could, even in principle, fold to a specific structure was until very recently shrouded in mystery [5]. There is considerable ev-

* Tel.: +44-1223-763070; fax: +44-1223-763418.
E-mail address: cmd44@cam.ac.uk (C.M. Dobson).

1084-9521/$ – see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.semcdb.2003.12.008

Petitioner Merck, Ex. 1136, p. 3
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00006

idence that the native state of a protein corresponds to the structure that is most stable under physiological conditions. Nevertheless, the total number of possible conformations of a polypeptide chain is so large that it would take an astronomical length of time to find this particular structure by means of a systematic search of all conformational space. Recent experimental and theoretical studies have, however, provided a resolution of this apparent paradox. It is now evident that the folding process does not involve a series of mandatory steps between specific partially folded states, but rather a stochastic search of the many conformations accessible to a polypeptide chain [5–8]. The conceptual basis of such a mechanism is shown in Fig. 1. In essence, the inherent fluctuations in the conformation of an incompletely folded polypeptide chain enable even residues at very different positions in the amino acid sequence to come into contact with one other. Because the correct (native-like) interactions between different residues are on average more stable than the incorrect (nonnative) ones, such a search mechanism is in principle able to find the lowest energy structure [9]. It is evident that this process is extremely efficient for those special sequences that have been selected during evolution to fold to globular structures, and indeed only a very small number of all possible conformations needs be sampled during the search process.

This stochastic description of protein folding is often referred to as the "new view" of this complex process [10]. It involves the concept of an "energy landscape" for each protein, describing the free energy of the polypeptide chain as a function of its conformational properties. To enable a protein to fold efficiently, the landscape required has been likened to a funnel because the conformational space

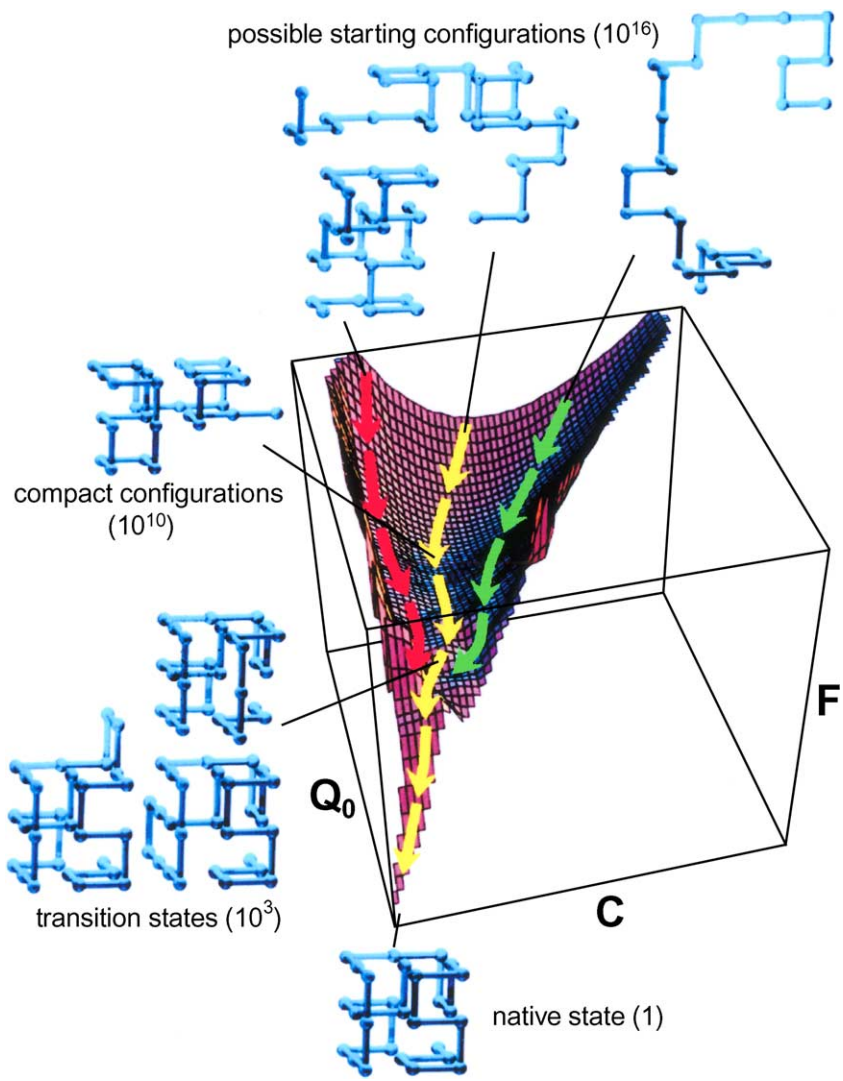

Fig. 1. Schematic energy landscape for protein folding. This surface is derived from a computer simulation of the folding of a highly simplified 27-residue protein. The free energy (F) of the system is shown as a function of the total number of contacts between residues (C) and the number of contacts that correspond to those of the most stable structure, denoted the native state (Q_0). The simulation shows that the surface serves to "funnel" the multitude of denatured conformations to the unique native state. From [9].

Petitioner Merck, Ex. 1136, p. 4
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00006

accessible to the polypeptide chain is reduced as the native state is approached [6]. In essence the high degree of disorder of the polypeptide chain is reduced as folding progresses, as the more favourable enthalpy associated with stable native-like interactions can offset the decreasing entropy as the structure becomes more ordered. The exact manner in which the correct overall fold can be achieved through such a process is emerging primarily from studies of a group of small proteins—most having less than 100 residues—that fold to their native states without populating significantly any intermediate states [11]. The most important experimental strategy has been to use site-directed mutagenesis to probe the roles of individual residues in the folding process [12,13]. The results of a wide range of studies suggest that the fundamental mechanism of folding can be described as "nucleation–condensation" in which a folding nucleus of a small number of key residues forms, about which the remainder of the structure can then condense [13,14].

Further insight into how such a mechanism can in principle generate a unique fold has emerged from experimental studies of the process of protein folding. Typically, a protein is placed in a solution of a chemical denaturant, such as urea, so that it unfolds. By returning the protein to conditions under which the native state is again stable (this can sometimes be achieved simply by diluting the solution) it is often possible to monitor the complex process of the recovery of the native structure over time by using biophysical techniques (e.g. circular dichroism or NMR spectroscopy) coupled with stopped or quenched flow methodologies [15]. Of equal importance in understanding the way that folding occurs have been theoretical approaches, particularly based on computer simulations of the events occurring during folding (or more often unfolding, as this process is easier to simulate but can be related to the complementary folding reaction) [16,17]. The ultimate objective of such studies is to define the complete energy landscape for the folding reaction, and to understand in detail how this is defined by the sequence. It has proved to be particularly fruitful in this regard to combine the results of experimental measurements with the conclusions from theoretical simulations [16,17]. Recently, experimental data have been incorporated directly in computer simulations of folding, and this approach has allowed structural ensembles representing the transition states of folding reactions to be defined in considerable detail [18]. Such studies suggest that a crucial aspect of the transition states of at least some proteins is that they have the same overall topology as the native fold. It appears that this topology can result from the acquisition of a native-like environment for the very small number of key residues that constitute the core of the folding nucleus of natural proteins; in essence, these interactions force the chain to adopt a rudimentary native-like architecture [18,19]. Once this topology has been achieved, the native structure is almost invariably generated when the remainder of the protein coalesces around this nucleus. Conversely, if these key interactions are not formed, the protein cannot usually fold directly to a stable globular structure. As all the protein molecules have to pass through the transition state region of the energy landscape prior to achieving their folded state, this mechanism therefore acts also as a "quality control" process by which misfolding can generally be avoided [20].

That chain topology plays a particularly critical role in folding is supported by an increasing number of experimental and theoretical studies [19]. Perhaps the most dramatic example of this point is the observation of a remarkable correlation between the folding rates of small proteins and the "contact order" of their structures [21]. The contact order describes the average separation in the sequence between residues that are in contact with each other in the native structure. Such a correlation appears to be largely independent of other details of the protein folds, such as their size and secondary structure content (i.e. the helices and sheets that are seen in almost all native protein structures [1]). The existence of a correlation between the behaviour of a wide range of proteins with very different architectures supports the concept of a common fundamental mechanism of folding [8]. Moreover, the correlation itself can be qualitatively rationalised by the argument that the search process will be more time consuming if the residues that form the nucleus are more distant from each other in the sequence. Similar conclusions are emerging from computer simulations of the energy landscapes of proteins based on rather simple descriptions of the sequence-dependent properties of polypeptide chains [22]. The success of such studies suggests that many of the underlying principles through which the sequence of a protein encodes its structure could soon emerge. Not only will such progress reveal in more depth how proteins are able to fold, but should advance significantly our presently rather limited ability to predict protein folds directly from their sequences.

Experiments show that the in vitro folding of proteins with more than about 100 residues involves a larger number of species than the fully unfolded and the fully folded states found to be populated in the simplest systems [8,13]. On the "new view" of folding, however, the underlying mechanism has not changed in any fundamental manner. In the case of small proteins one can consider that the collapse of the polypeptide chain to a stable compact structure can occur only if at least the large majority of the interactions involving the key residues have been formed. Once this has happened the conversion to the fully folded state is fast. For larger proteins, however, a variety of factors, including the higher proportion of hydrophobic residues, provides a greater driving force for chain collapse, and the latter can occur prior to the stage at which folding can progress rapidly to the fully native state [8]. Experiments show that the folding intermediates that result from such a scenario sometimes correspond to species in which segments of the protein have become highly native-like, whilst others have yet to achieve a folded state, as indicated in Fig. 2 [8]. In other cases the protein may have formed a significant proportion of non-native interactions, and hence becomes trapped at least transiently in

Petitioner Merck, Ex. 1136, p. 5
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00006

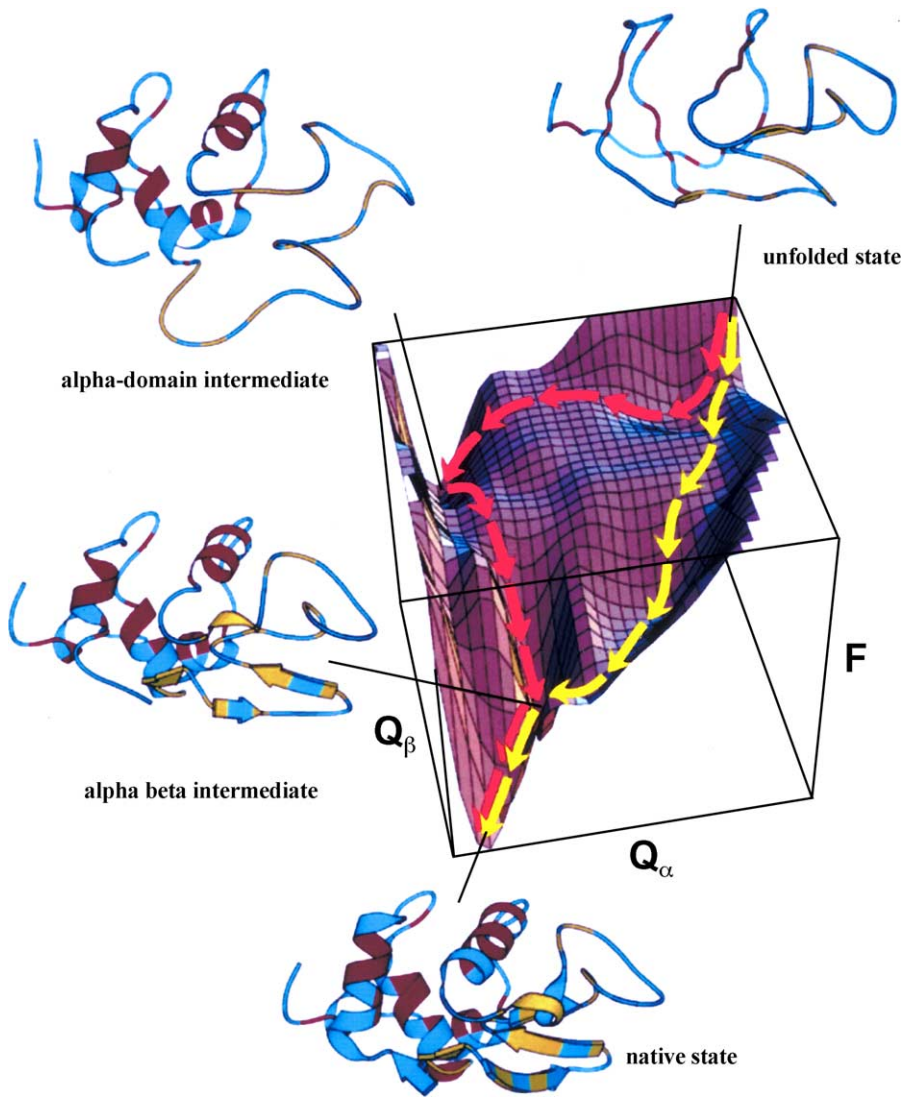

Fig. 2. Schematic energy landscape for the folding of lysozyme. This protein has 129 residues and its structure consists of two domains, denoted α and β. The free energy (F) is shown as a function of the number of native contacts (Q_α and Q_β) in each domain. The yellow trajectory represents a "fast track" in which the two domains fold concurrently and populate an intermediate (labelled alpha/beta) only briefly. The red trajectory represents a "slow track" in which the chain becomes trapped in a long-lived intermediate state with persistent structure only in the alpha domain. Further folding from this intermediate involves either a transition over a higher barrier, or partial unfolding to enable the remainder of the folding to take place along a fast track. The structures of the various species were generated by constraining molecular dynamics simulations with experimental data characterising each species, derived from NMR experiments. From [9].

a misfolded state [23]. One of the very attractive features of the new view of folding is that it brings together many of the earlier conceptual models of folding. These models can now often be seen as special cases of a unified general mechanism of folding that describes the behaviour of all protein molecules [24]. For example, some protein sequences have a very high propensity to form a given type of secondary structure (e.g. helices) with the result that these elements of the protein fold develop early in the folding process, and may be observed as essentially fully formed even before the overall architecture is defined by non-local interactions. The folding of such proteins is often very well described by the "diffusion–collision" model of folding, in which it is envisaged that the search for the native architecture involves the diffusional motion of pre-formed helices [25].

The topic of this review is one in which nomenclature can cause a great deal of confusion. The term "protein folding" is universally accepted to refer to the process that results in the acquisition of the native structure from a completely or partially unfolded state. I use the term "misfolding" to describe processes that result in a protein acquiring a sufficient number of persistent non-native interactions to affect its overall architecture and/or its properties in a biologically significant manner. Thus a monomeric intermediate populated during the normal folding process of a protein will not usually be described as misfolded, even though it may

have some non-native contacts between its residues. If a monomeric protein were to find its way into a compact stable structure with a non-native topology, however, such a species would be described as misfolded, as its architecture (and almost invariably its biological) would be demonstrably distinct from that of the native state. The aggregation process that leads to a peptide or protein to form an amyloid fibril will universally be described on this definition as "misfolding", as the interactions that determine the structure and properties of such fibrils are distinct from those that determine the structure and properties of the biologically active forms of the molecules concerned.

Both experimental and theoretical studies suggest that larger proteins generally fold in modules, i.e. that folding takes place largely independently in different segments or domains of the protein [8,26]. In such cases, key interactions are likely to define the fold within local regions or domains, and other specific interactions ensure that these initially folded regions subsequently interact appropriately to form the correct overall structure. The fully native structure is only acquired, however, when all the native-like interactions are formed both within and between the domains; this happens in a final cooperative folding step when all the side-chains become locked in their unique close-packed arrangement and water is excluded from the protein core [27]. Such a mechanism is appealing because it begins to explain how highly complex structures may be assembled in manageable pieces. Moreover, such a principle can readily be extended to describe the assembly of complexes containing a variety of different proteins, and in some cases other macromolecules, notably nucleic acids. Thus, even large molecular machines such as the ribosome or the proteasome can be assembled efficiently and with high fidelity. The detailed manner in which such species are able to achieve their functional states is a fascinating extension of the studies of the fundamental process of protein folding, as are other facets of the larger subject of self-assembly, such as the folding of proteins within membranes [28] and of other macromolecules such as RNA [29].

3. Protein folding and misfolding in the cell

Proteins are synthesised in cells on organelles known as ribosomes from the information contained within the cellular DNA. In vivo, protein folding can in principle begin whilst a nascent chain is still attached to the ribosome, and there is evidence that some proteins do fold at least partially in such a co-translational manner [30]. Recently, a combination of electron microscopy and X-ray crystallography has revealed detailed structures of these complex "molecular machines" [31], no mean feat given that they contain more than 50 different proteins and three large RNA molecules. One can be optimistic, therefore, that at least some information will soon be available about those folding events that are coupled to the process of translation [32,33]. Other proteins are known to undergo the major part of their folding in the cytoplasm after release from the ribosome, whilst yet others fold in specific compartments such as the endoplasmic reticulum (ER) following translocation through membranes [34]. Many details of the folding process will therefore depend on the environment in which folding takes place, although the fundamental principles of how the process occurs are unlikely to be changed in any significant manner. But as incompletely folded chains expose regions of the polypeptide molecule that are buried in the native state, such species are prone to inappropriate contacts with other molecules within the cell. There is evidence that in some cases non-native interactions form transiently to bury highly aggregation-prone regions such as exposed hydrophobic surfaces [23,35]. But to cope with this problem more generally, living systems have evolved a range of elaborate strategies to prevent interactions with other molecules prior to the completion of the folding process [3,34,36].

Of particular importance is the large number of molecular chaperones that are present in all types of cells and cellular compartments (see article by Barral et al. in this issue). Despite their similar general role in enabling efficient folding and assembly, their specific functions can differ substantially and it is evident that many types of chaperone work in tandem with one other [34]. Some molecular chaperones have been found to interact with nascent chains as they emerge from the ribosome, and bind rather non-specifically to protect aggregation-prone regions rich in hydrophobic residues. Others are involved in guiding later stages of the folding process, particularly for complex proteins including oligomeric species and multimolecular assemblies. The best characterised molecular chaperone is the bacterial "chaperonin", GroEL, and many of the details of the mechanism through which this molecule functions are now well understood [34,37]. Of particular significance for GroEL is the fact that it contains a cavity in which incompletely folded polypeptide chains can be sequestered and protected from the outside world [38]. In addition to molecular chaperones, there are several classes of folding catalyst that act to accelerate steps in the folding process that can otherwise be extremely slow. The most important are peptidylprolyl isomerases, that increase the rate of *cis*/*trans* isomerisation of peptide bonds involving proline residues, and protein disulphide isomerases that enhance the rate of formation and reorganisation of disulphide bonds within proteins [3,34].

Given the enormous complexity of the folding process, it would be remarkable if misfolding were never to occur. Clear evidence for the importance of molecular chaperones in preventing misfolding and its consequences comes from the fact that the levels of many of these species are very substantially increased during cellular stress. Indeed, many chaperones were first recognised in such circumstances, and their nomenclature as Hsps (heat shock proteins) reflects this fact [39]. It appears that some molecular chaperones are able to rescue misfolded proteins to enable them to have a second chance to fold correctly, and there are examples of

Petitioner Merck, Ex. 1136, p. 7
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00006

molecular chaperones that are able to solubilise some forms of aggregates under at least some circumstances [40]. Such active intervention requires energy, and not surprisingly ATP is required for many of the molecular chaperones to function correctly [34]. Despite the fact that many molecular chaperones occur at high levels only in stressed systems, it is clear that they have a critical role to play in all organisms even when present at lower levels under normal physiological conditions.

In eukaryotic systems, many of the proteins that are synthesised in a cell are destined for secretion to an extracellular environment. These proteins are translocated into the ER where folding takes place prior to secretion through the Golgi apparatus. The ER contains a wide range of molecular chaperones and folding catalysts to promote efficient folding, and in addition the proteins involved must satisfy a stringent "quality control" mechanism in the ER (see article by Weilh in this issue and [41]). The quality control mechanism involves a complex series of glycosylation and deglycosylation processes and prevents misfolded proteins from being secreted from the cell. In addition, unfolded and misfolded proteins are recognised and targeted for degradation through the ubiquitin–proteasome pathway [42]. The details of how these remarkable regulatory systems operate represent astonishing examples of the stringent mechanisms that biology has established to ensure that misfolding, and its consequences, are minimised. The importance of the quality control process is underlined by the fact that recent experiments indicate that up to half of all polypeptide chains fail to satisfy the quality control mechanism in the ER, and for some proteins the success rate is even lower [43]. Like the "heat shock response" in the cytoplasm, the "unfolded protein response" in the ER is also upregulated during stress and, as we shall see below, is strongly linked to the avoidance of misfolding diseases.

Folding and unfolding are the ultimate ways of generating and abolishing cellular activities, and unfolding is also the precursor to the degradation of proteins [44]. Moreover, it is increasingly apparent that some events in the cell, such as translocation across membranes, can require proteins to be in unfolded or partially folded states. Processes as apparently diverse as trafficking, secretion, the immune response and the regulation of the cell cycle, are in fact now recognised to be directly dependent on folding and unfolding [45]. It is not surprising therefore that failure to fold correctly, or to remain correctly folded, will give rise to the malfunctioning of living systems and therefore to disease. Indeed, it is becoming increasingly evident that a wide range of human diseases is associated with aberrations in the folding process (Table 1) [46,47]. Some of these diseases (e.g. cystic fibrosis) result from the simple fact that if proteins do not fold correctly they will not be able to exercise their proper functions. In other cases, misfolded proteins escape all the protective mechanisms discussed above and form intractable aggregates within cells or in the extracellular space. An increasing number of pathologies, including Alzheimer's and Parkinson's diseases, the spongiform encephalopathies and late-onset diabetes, are known to be directly associated with the deposition of such aggregates in tissue [46–50]. Diseases of this type are amongst the most debilitating, socially disruptive and costly diseases in the modern world, and they are becoming increasingly prevalent as our societies age and become more dependent on new agricultural, dietary and medical practices [51].

4. Protein aggregation and amyloid formation

One of the most characteristic features of many of the aggregation diseases is that they give rise to the deposition of proteins in the form of amyloid fibrils and plaques [46–50]. Such deposits can form in the brain, in vital organs such as the liver and spleen, or in skeletal tissue, depending on the disease involved. In the case of neurodegenerative

Table 1
Representative protein folding diseases (from [47])

Disease	Protein	Site of folding
Hypercholesterolaemia	Low-density lipoprotein receptor	ER
Cystic fibrosis	Cystic fibrosis *trans*-membrane regulator	ER
Phenylketonuria	Phenylalanine hydroxylase	Cytosol
Huntington's disease	Huntingtin	Cytosol
Marfan syndrome	Fibrillin	ER
Osteogenesis imperfecta	Procollagen	ER
Sickle cell anaemia	Haemoglobin	Cytosol
αl-Antitrypsin deficiency	αl-Antitrypsin	ER
Tay–Sachs disease	β-Hexosaminidase	ER
Scurvy	Collagen	ER
Alzheimer's disease	Amyloid β-peptide/tau	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

Petitioner Merck, Ex. 1136, p. 8
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00006

disorders, the quantity of such aggregates can be almost undectectable in some cases, whilst in systemic diseases kilograms of protein can be found in such deposits (see articles by Hirschfield and Goedert in this issue). Each amyloid disease involves the aggregation of a specific protein although a range of other components, including other proteins and carbohydrates, is also incorporated into the deposits when they form in vivo. The characteristics of the soluble forms of the 20 or so proteins involved in the well-defined amyloidoses are varied—they range from intact globular proteins to largely unstructured peptide molecules—but the aggregated forms have many common characteristics [52]. Amyloid deposits all show specific optical properties (such as birefringence) on binding certain dye molecules, notably Congo red; these properties have been used in diagnosis for over a century. The fibrillar structures that are characteristic of many of the aggregates have very similar morphologies (long, unbranched and often twisted structures a few nm in diameter) and a characteristic "cross-beta" X-ray fibre diffraction pattern [52]. The latter reveals that the organised core structure is composed of β-sheets having strands running perpendicular to the fibril axis, as indicated in Fig. 3 [53]. Fibrils having the essential characteristics of ex vivo deposits can be reproduced in vitro from the component proteins under appropriate conditions, showing that they can self-assemble without the need for other components.

For many years it was generally assumed that the ability to form amyloid fibrils with the characteristics described above was limited to a relatively small number of proteins, largely those seen in disease states, and that these proteins possess specific sequence motifs encoding the amyloid core structure. Recent studies have suggested, however, that the ability of polypeptide chains to form such structures is common, and indeed can be considered a generic feature of polypeptide chains [54,55]. The most direct evidence for the latter statement is that fibrils can be formed by many different proteins that are not associated with disease, including such well-known proteins as myoglobin [56], and also by homopolymers such as polythreonine or polylysine [57]. Remarkably, fibrils of similar appearance to those containing large proteins can be formed by peptides with just a handful of residues [58]. One can consider that amyloid fibrils are highly organised structures (effectively one-dimensional crystals) adopted by an unfolded polypeptide chain when it behaves as a typical polymer; similar types of structure can be formed by many types of synthetic polymer. The essential features of such structures are determined by the physicochemical properties of the polymer chain. As with other highly organised materials (including crystals) whose structures are based on repetitive long-range interactions, the most stable structures are those consisting of a single type of peptide or protein where such interactions can be optimised [59].

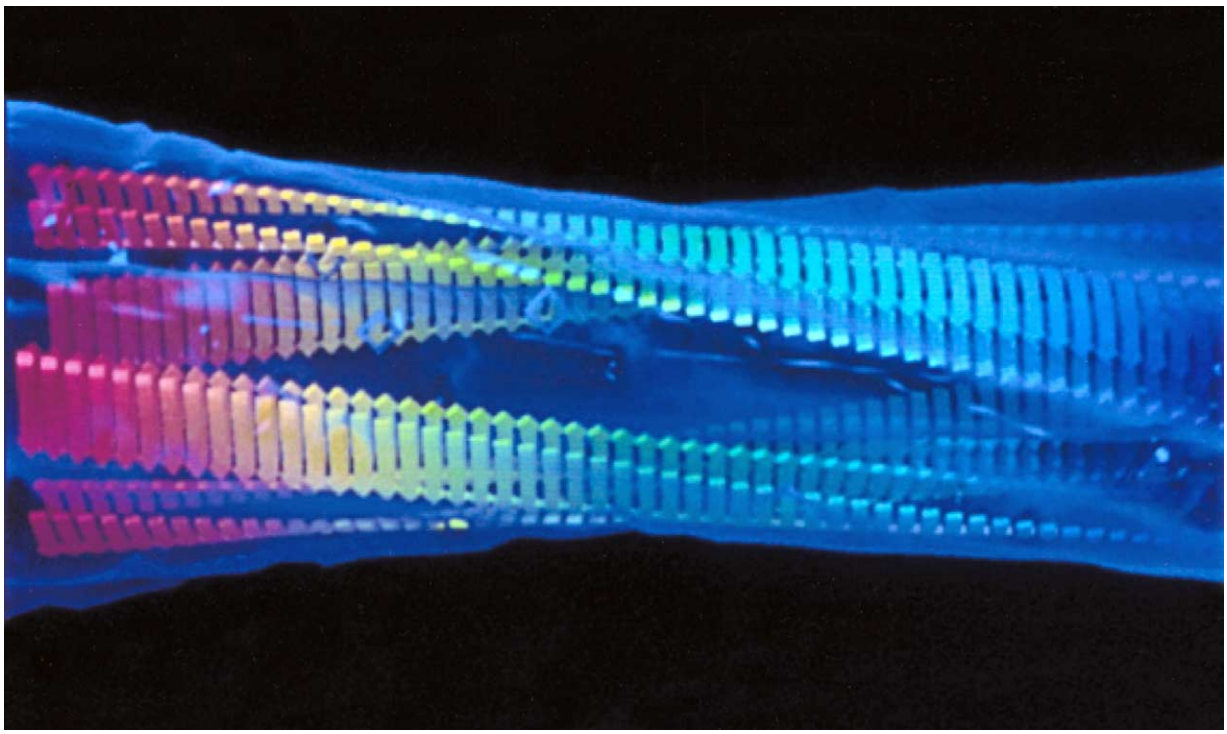

Fig. 3. A molecular model of an amyloid fibril. This model is derived from cryo-EM analysis of fibrils grown from an SH3 domain [53]. The fibril consists of four "protofilaments" that twist around one another to form a hollow tube with a diameter of approximately 6 nm. The model illustrates one way in which regions of the polypeptide chain involved in β-sheet formation could be assembled within the fibrils. From [55], kindly provided by Helen Saibil, Birkbeck College, London.

Petitioner Merck, Ex. 1136, p. 9
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00006

Studies of both disease-associated fibrils and of those formed from other proteins has enabled many of the features of these structures to be defined [52,53,60–63], although no complete structure has yet been determined in atomic detail. It is clear that the core structure of the fibrils is stabilised primarily by interactions, particularly hydrogen bonds, involving the polypeptide main chain. As the main chain is common to all polypeptides, this observation explains why fibrils formed from polypeptides of very different amino acid sequence are similar in appearance. The side chains are likely to be incorporated in whatever manner is most favourable for a given sequence within the amyloid structures; they affect the details of the fibrillar assembly but not their general structure [64]. In addition, the proportion of a polypeptide chain that is incorporated in the core structure can vary substantially; in some cases only a handful of residues may be involved in the core structure, with the remainder of the chain associated in some other manner with the fibrillar assembly. This generic type of structure contrasts strongly with the globular structures of most natural proteins [55]. In these globular structures the interactions associated with the highly specific packing of the side chains can sometimes override the main chain preferences [55–57]. The strands and helices so familiar in the structures of native proteins [1] are then the most stable structures that the main chain can adopt in the folds that are primarily defined by the side chain interactions. If the solution environment (pH, temperature etc.) in which the molecules are found is, however, such that these side chain interactions are insufficiently stable, the structures can unfold and may then, at least under some circumstances, reassemble in the form of amyloid fibrils.

Although the ability to form amyloid fibrils appears generic, the propensity to do so can vary dramatically between different sequences [65]. At the most fundamental level, some types of amino acid are much more soluble than others, such that the concentration that is required to be reached before aggregation occurs will be much greater for some polypeptides than for others. In addition the aggregation process, like crystallisation, needs to be nucleated and the rate at which this process takes place can be highly dependent on many different factors. It is clear that even single changes of amino acid in protein sequences can change the rates at which the unfolded polypeptide chains aggregate by an order of magnitude or more. It has proved possible to correlate the changes in aggregation rates cause by such mutations with changes in simple properties that result from such substitutions, such as charge, secondary structure propensities and hydrophobicity [66]. As this correlation has been found to hold for a wide range of different sequences (Fig. 4), it strongly endorses the concept of the generic nature of amyloid formation. In accord with such ideas, those proteins that are completely or partially unfolded under normal conditions in the cell have sequences that are predicted to have very low propensities to aggregate [65]. An interesting and potentially important additional observation is that the residues that nucleate the folding of a globular protein appear to be distinct from those that nucleate its aggregation into amyloid fibrils [67]. Such a characteristic, that may reflect the different nature of the partially folded species that initiate the two types of assembly processes, offers the opportunity for evolutionary pressure to select sequences that favour folding over aggregation.

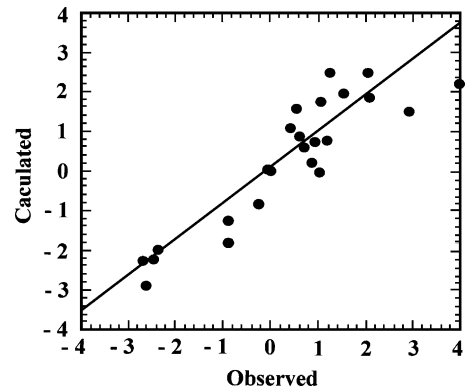

Fig. 4. Rationalisation of the effects of mutations on the aggregation rates of peptides and proteins. The observed experimental aggregation rates of a variety of short peptides or natively unfolded proteins, including amylin, amyloid β-peptide, tau and α-synuclein (see Table 1 for a summary of the diseases with which these are associated) are shown plotted against rates calculated from an algorithm derived from extensive mutational studies of the protein acylphosphatase. The numbers are the natural logarithims of the changes in aggregation rate. Such a correlation argues strongly for a common mechanism for amyloid formation and provides a platform both to predict the effects of natural mutations and to design polypeptides with altered aggregation properties. From [66].

One of the crucial aspects of the formation of amyloid fibrils is the mechanism by which they are assembled from the precursor species. In globular proteins the polypeptide main chain is largely buried within the folded structure, and it is necessary for it to be exposed prior to the formation of fibrillar species. Thus, conditions that favour formation of amyloid fibrils are ones in which proteins involved are at least partially unfolded, for example, low pH [68]. Because of the importance of the globular fold in preventing aggregation, the fragmentation of proteins, through proteolysis or other means, is a ready mechanism to stimulate amyloid formation. Indeed many amyloid disorders, including Alzheimer's disease, involve aggregation of fragments of larger precursor proteins that are unable to fold in the absence of the remainder of the protein structure. Experiments in vitro indicate that the formation of fibrils, by appropriately destabilised or fragmented proteins, is then generally characterised by a lag phase, followed by a period of rapid growth [69]. Such behaviour is typical of nucleated processes such as crystallisation; as with crystallisation, the lag phase can be eliminated by addition of pre-formed fibrils to fresh solutions, a process known as seeding [69]. Although the details of the events taken place during fibril growth are not yet elucidated in any detail, it is becoming possible to simulate the overall kinetic profiles using relatively simple

Petitioner Merck, Ex. 1136, p. 10
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00006

models that incorporate well-established principles of nucleated processes [70].

One of the key findings of such studies is that there are many common features in the behaviour of those systems that have so far been examined [50,69,71,72]. The first phase of the aggregation process involves the formation of oligomeric species as a result of relatively non-specific interactions, although in some cases specific structural transitions, such as domain swapping [73], may be involved if such processes increase the rate of aggregation. The earliest species visible by electron or atomic force microscopy often resemble small bead-like structures, sometimes described as amorphous aggregates or as micelles. These early "prefibrillar aggregates" then appear to transform into species with more distinctive morphologies, sometimes described as "protofibrils" or "protofilaments" [61,72,74]. These structures are commonly short, thin, sometimes curly, fibrillar species that are thought to assemble into mature fibrils, perhaps by lateral association, accompanied by some degree of structural reorganisation [75]. The extent to which dissolution and reassembly of monomeric species is involved at the different stages of assembly is not clear, but it could well be important in the slow growth conditions under which the most highly structured fibrils are formed. The earliest aggregates are likely to be relatively disorganised structures that expose to the outside world a variety of segments of the protein that are normally buried in the globular state. In other cases these early aggregates appear to be quite distinctive structures, including well-defined "doughnut" shaped species seen in a number of systems [72,76].

5. A generic description of protein self-assembly

In the discussion of the various conformational states that can be adopted by proteins under different conditions, we have emphasised the similarities as well as the differences in the behaviour of different proteins. This approach can be summarised in a schematic representation of at least some of the different states that a given polypeptide chain can in principle adopt, illustrated in Fig. 5 [16,36,77]. From such a description, the state of a given protein that is populated under specific conditions will depend on the relative thermodynamic stabilities of the different states (in the case of oligomers and aggregates the concentration will be a critical parameter) and on the kinetics of the various interconversion processes. In this diagram, amyloid fibrils are included as just one of the types of aggregate that can be formed by proteins, although it has particular significance in that its highly organised hydrogen-bonded structure gives it unique kinetic stability. This type of diagram emphasises that biological systems have become robust by controlling and regulating the various states accessible to a given polypeptide chain at given times and under given conditions, just as they regulate and control the various chemical transformations that take place in the cell [36]. The latter is achieved primarily through enzymes, and the former by means of the molecular chaperones and degradatory mechanisms mentioned above. And just as the aberrant behaviour of enzymes can cause metabolic disease, the aberrant behaviour of the chaperone and other machinery regulating polypeptide conformations can contribute to misfolding and aggregation diseases [78].

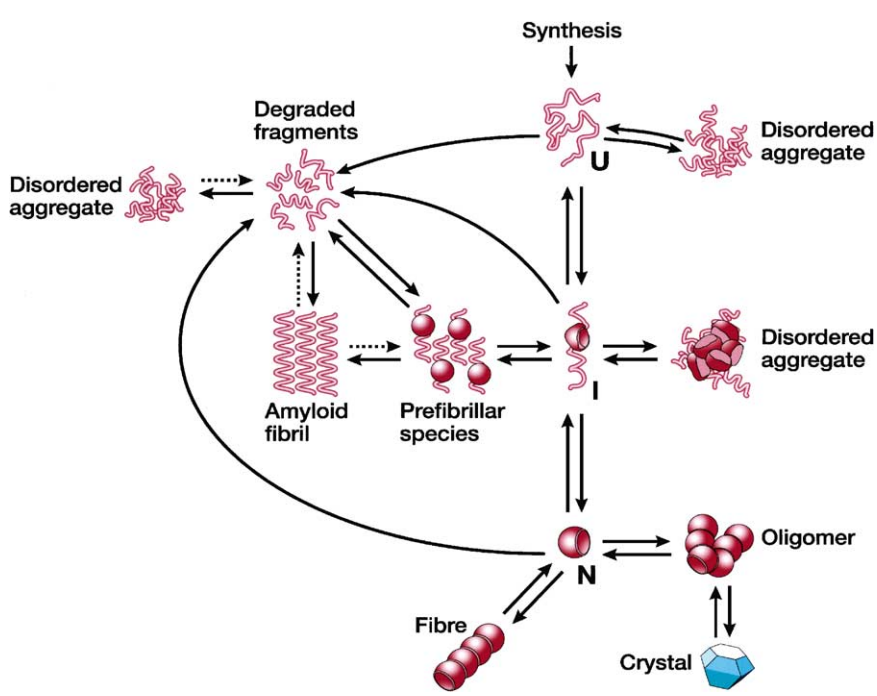

Fig. 5. Schematic representation of some of the states accessible to a polypeptide chain following its synthesis on a ribosome. The relative populations of the different states depend on the kinetics and thermodynamics of the various equilibria shown in the diagram. In living systems the fate of a given protein molecule is closely regulated by molecular chaperones and other quality control mechanisms, rather as metabolic pathways in cells are controlled by enzymes and associated molecules such as cofactors. From [36].

Petitioner Merck, Ex. 1136, p. 11
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00006

The details of the regulatory, chaperone and quality control mechanisms, the nature of specific diseases and the various possible approaches to prevention and therapy, will be discussed in detail in later articles in this volume. But the type of diagram shown in Fig. 5 serves as a framework for understanding the fundamental molecular events that underlie all of these topics, and indeed the principles that can be used to intervene for therapeutic purposes [36]. As we have discussed above, it is partially or completely unfolded polypeptides that are highly aggregation-prone. Such species are inherent in the folding process, and a variety of molecular chaperones is present in abundance in the cellular compartments wherever such processes occur. At the present time, only one extracellular species (clusterin) has been reported to act in a similar manner [79]. It is therefore essential that proteins are correctly folded prior to their secretion from the cell; hence the need for a highly effective system of quality control in the ER. In this context it is interesting that the majority of the diseases associated with amyloid formation involve deposits that are extracellular in nature [48]. As biology is a dynamic process, there is a continuous need to degrade as well as synthesise proteins, and the degradatory mechanisms target misfolded as well as redundant proteins. It is during such processes, which require unfolding and proteolysis of polypeptide chains, that aggregation may be particularly likely. Degradation pathways, such as those of the ubiquitin–proteasome system, are therefore highly regulated in order to avoid such events [80,81].

In recent years our understanding of the detailed mechanism of all the processes associated with the complete lifespans of proteins—from their synthesis to their degradation—has advanced dramatically through progress in both cellular and structural biology. High-resolution structures of ribosomes, proteasomes, molecular chaperones and other complexes are revealing the details of how these molecular machines operate and are regulated. Structural techniques are also providing an increasing amount of information about the various states of the proteins with which they are associated. This task is particularly challenging as the majority of these species (such as unfolded and partially folded states and the less organised types of aggregates) are ensembles of more or less highly disordered structures, and in addition may have only a transient existence. Nevertheless, it is beginning to become possible to identify the specific species that regulate the different types of transition in different organisms, cell types and cellular compartments, and to understand at least in general terms the molecular basis of their functions [34].

In order to understand misfolding and aggregation diseases we need to know not just how such systems function efficiently, but also why they fail to do so under some circumstances [81–83]. The effects of many pathogenic mutations can be particularly well understood from the schematic representation given in Fig. 5. Many of the mutations associated with the familial deposition diseases increase the population of partially unfolded states, either within or outside the cell, by decreasing the stability or cooperativity of the native state [84–86]. Cooperativity is in fact a crucial factor in enabling proteins to remain soluble, as it ensures that even for a protein that is marginally stable, the equilibrium population of unfolded molecules, or of unfolded regions of the polypeptide chain, is minimal. Other familial diseases are associated with the accumulation of fragments of native proteins, often produced by aberrant processing or incomplete degradation; such species are unable to fold into aggregation-resistant states. Other pathogenic mutations act by enhancing the propensities of such species to aggregate, for example by increasing their hydrophobicity or decreasing their charge [66]. In the case of the transmissible spongiform encephalopathies, it is likely that ingestion of pre-aggregated states of an identical protein (e.g. by cannibalistic means or by contamination of surgical instruments) increases dramatically the inherent rate of aggregation, and hence underlies the mechanism of transmission [69,87,88]. Such seeding events are also likely to be the reason why some deposition conditions such as Alzheimer's disease progress so rapidly once the initial symptoms are evident [69,89].

6. Evolution and the origins of misfolding diseases

Despite this outline knowledge, the manner by which protein aggregation results in pathological behaviour is not yet understood in detail. In the case of systemic disease, the sheer mass of insoluble protein may physically disrupt the functioning of specific organs [49]. In other cases it may be that the loss of functional protein results in the failure of some crucial cellular process [46]. But for neurodegenerative disorders, such as Alzheimer's disease, it appears that the primary symptoms results from the destruction of cells such as neurons by a "toxic gain-of-function" that results from the aggregation process [50,72]. It has recently become apparent that the early pre-fibrillar aggregates of proteins associated with neurological diseases can be highly damaging to cells; by contrast the mature fibrils are relatively benign [72,90]. It has become clear, however, that the toxic nature of protein aggregates is not restricted to species formed from the peptides and proteins associated with pathological conditions. Experiments have recently indicated that prefibrillar aggregates of several proteins that are not connected with any known diseases are as cytotoxic as those of Aβ [91]. The concept of a generic nature of such aggregates, and their effects on cells, has recently been reinforced through experiments with antibodies that cross-react with early aggregates of different peptides and proteins—and inhibit their toxicity [92]. It is possible that there are specific mechanisms for such toxicity, for example through the "doughnut" shaped aggregates that resemble the toxins produces by bacteria that form pores in membranes and disrupt the ion balance in cells [76]. It is also possible that disorganised early aggregates are toxic through a less specific mechanism; for example, the exposure of non-native hydrophobic surfaces may stim-

Petitioner Merck, Ex. 1136, p. 12
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00006

ulate aberrant interactions with membranes or other cellular components [83,93].

Such findings raise the question as to how cellular systems are able to tolerate the intrinsic tendency of incompletely folded proteins to aggregate. The answer is almost certainly that under normal circumstances the molecular chaperones and other "housekeeping" mechanism are remarkably efficient in ensuring that such potentially toxic species are neutralised [34,83]. Molecular chaperones of various types are able to shield hydrophobic regions, to unfold some forms of aggregates, or to alter the partitioning between different forms of aggregates. The latter mechanism, for example, could convert the precursors of amyloid fibrils into less intractable species, allowing them to be refolded or disposed of by the cellular degradation systems. Indeed, evidence has been obtained that such a situation occurs with polyglutamine sequences associated with disorders such as Huntington's disease [94]. In this case the precursors of amyloid fibrils appear to be diverted into amorphous species by the action of molecular chaperones. If such protective processes fail, it may be possible for potentially harmful species to be sequestered in relatively harmless forms, such as inclusion bodies in bacteria or aggresomes in eukaryotic systems. Indeed, it has even been suggested that the formation of mature amyloid fibrils, whose toxicity appears to be much lower than that of their precursors as we have discussed above, may itself represent a protective mechanism in some cases [50,72].

Most of the aggregation diseases are, however, not associated with either genetic mutations or infectious agents but with sporadic events and particularly old age. The ideas summarised in this review offer a qualitative explanation of why this is the case. We have seen that all proteins have an inherent tendency to aggregate unless they are maintained in a highly regulated environment. Selective pressure during evolution has resulted in protein molecules that are able to resist aggregation during our normal lifespan—enabling us to develop, pass on genes and to give appropriate protection to our offspring—but evolutionary pressure can do no better than is needed to achieve such an end [51]. There is no reason to suppose that random mutations will in general reduce further the propensity to aggregate; indeed it will generally increase it—just as random mutations generally reduce the stability of native proteins. We can see, therefore, that our recent ability to prolong life is leading to the proliferation of these diseases. It is intriguing, however, to speculate, however, that favourable mutations in aggregation-prone proteins might be the reason that some individuals in the population do not readily succumb, even in extreme old age, to diseases such as Alzheimer's [66].

The link with ageing is likely to involve more than just a greater probability of aggregation taking place as we get older. It is likely to be linked more fundamentally to the failure of the "housekeeping" mechanisms in our bodies [51,78,95]. This failure may be, in part, a result of the need for greater protective capacity in old age as aggregation becomes more prevalent, perhaps as a result of the increasing accumulation of misfolded and damaged proteins, leading to chaperone overload [95]. But as we age it is likely that the activity of our chaperone response and degradatory mechanisms declines, and that this decline results in the increasing probability that the protective mechanisms are overwhelmed. Similarly, the rapidity with which we have introduced practices that are not experienced previously in history—including new agricultural practices (BSE [87]), a changing diet (associated with the prevalence of Type II diabetes [96]) and new medical procedures (iatrogenic diseases such as CJD [87], and amyloid deposition in hemodialysis during which the concentration of β-2 microglobulin in serum increases [97])—means that we have not had time to evolve effective protective mechanisms [51]. This phenomenon can be compared to the lack of ability of our bodies to deal with heavy metals such as mercury and cadmium, whose levels in the biosphere have recently increased dramatically as a result of modern mining and industrial procedures; all the abundant elements in the earth with soluble compounds are essential to life but most of the rare elements or those with insoluble compounds are toxic [98]. This lack of adaptation can be viewed in the same way as the effects of introducing exotic species into previously isolated regions of the world—such as rabbits in Australia—where the consequences of the lack of effective evolved regulatory mechanisms can be seen on the organismal rather than the molecular scale.

7. Concluding remarks

Proteins have evolved to fold efficiently and to remain correctly folded and soluble, despite their inherent tendency to aggregate, as a result of the natural selection of sequences and the co-evolution of the environments in which they function. But evolution can achieve no more than is necessary for successful competition with other individuals or life-forms, and once we step outside the boundaries within which we have developed—for example by curbing infectious diseases or introducing new medical procedures or radically changing our lifestyles—we begin to see the limitations of our present levels of molecular evolution. As we are now in effective control of our own evolution, and increasingly that of other life-forms on earth, it is up to us to find means to respond to these limitations. Indeed, a host of exciting approaches is being developed, representing intervention at many of the steps in aggregation represented in Fig. 5, and many show considerable promise [86,99–103]. Fortunately, therefore, we can be optimistic that the better understanding of protein misfolding and aggregation that is developing from recent research of the type discussed in this article, will enable us to rise to this tremendous challenge [36]. As we develop further our knowledge of the mechanism of protein folding, and of the way that it is enhanced and regulated within the cellular environment, we shall be

Petitioner Merck, Ex. 1136, p. 13
Merck Sharp & Dohme LLC v. Halozyme Inc.
PGR2025-00006

able to answer with increasing conviction the more general question of how evolution has enabled even the most complex biological systems to self-assemble with astonishing fidelity. Such knowledge will represent a very significant step towards understanding at a molecular level one of the most fascinating and fundamental characteristics of life itself.

Acknowledgements

The ideas in this article have emerged from extensive discussions with outstanding students, research fellows and colleagues over many years. I am most grateful to all of them; they are too numerous to mention here but the names of many appear in the list of references. The structure of this article also owes much to the Horizon Symposium on Protein Folding Diseases that took place in Italy in October 2002, organised by Nature in conjunction with Aventis. I am particularly grateful to Adam Smith, editor of Nature Reviews Drug Discovery, who was the driving force behind this meeting, and with whom I have had many stimulating discussions. This article is indeed a revised and extended version of a Nature Insight Review that was published in association with the Horizon Symposium (*Nature*, 426, 844–890 (2003)). The research of CMD is primarily supported by Programme Grants from the Wellcome Trust and the Leverhulme Trust, as well as by the BBSRC, EPSRC and MRC.

References

[1] Branden C, Tooze J. Introduction to protein structure, 2nd ed. New York: Garland Publishing; 1999.
[2] Ellis RJ, Minton AP. Join the crowd. Nature 2003;425:27–8.
[3] Gething M-J, Sambrook J. Protein folding in the cell. Nature 1992;355:33–45.
[4] Vendruscolo M, Zurdo J, MacPhee CE, Dobson CM. Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems. Phil Trans R Soc Lond A 2003;361: 1205–22.
[5] Karplus M. The Levinthal paradox, yesterday and today. Fold Des 1997;2:569–76.
[6] Wolynes PG, Onuchic JN, Thirumalai D. Navigating the folding routes. Science 1995;267:1619–23.
[7] Dill KA, Chan HS. From Levinthal to pathways to funnels. Nature Struct Biol 1997;4:10–9.
[8] Dobson CM, Sali A, Karplus M. Protein folding: a perspective from theory and experiment. Angew Chem Int Ed Eng 1998;37:868–93.
[9] Dinner AR, Sali A, Smith LJ, Dobson CM, Karplus M. Understanding protein folding via free energy surfaces from theory and experiment. Trends Biochem Sci 2000;25:331–9.
[10] Baldwin RL. Protein folding: matching speed and stability. Nature 1994;369:183–4.
[11] Jackson SE. How do small single-domain proteins fold? Fold Des 1998;3:R81–91.
[12] Matouschek A, Kellis JT, Serrano L, Fersht AR. Mapping the transition state and pathway of protein folding by protein engineering. Nature 1989;342:122–6.
[13] Fersht AR. Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. New York: WH Freeman; 1999.
[14] Fersht AR. Transition-state structure as a unifying basis in protein-folding mechanisms: contact order, chain topology, stability, and the extended nucleus mechanism. Proc Natl Acad Sci USA 2000;97:1525–9.
[15] Plaxco K, Dobson CM. Time-resolved biophysical methods in the study of protein folding. Curr Opin Struct Biol 1996;6:630–6.
[16] Dobson CM, Karplus M. The fundamentals of protein folding: bringing together theory and experiment. Curr Opin Struct Biol 1999;9:92–101.
[17] Fersht AR, Daggett V. Protein folding and unfolding at atomic resolution. Cell 2002;108:573–82.
[18] Vendruscolo M, Paci E, Dobson CM, Karplus M. Three key residues form a critical contact network in a transition state for protein folding. Nature 2001;409:641–6.
[19] Makarov DE, Plaxco KW. The topomer search model: a simple, quantitative theory of two-state protein folding kinetics. Prot Sci 2003;12:17–26.
[20] Davis R, Dobson CM, Vendruscolo M. Determination of the structures of distinct transition state ensembles for a β-sheet peptide with parallel folding pathways. J Chem Phys 2002;117:9510–7.
[21] Plaxco KW, Simons KT, Baker D. Contact order, transition state placement and the refolding rates of single domain proteins. J Mol Biol 1998;277:985–94.
[22] Baker D. A surprising simplicity to protein folding. Nature 2000;405:39–42.
[23] Capaldi AP, Kleanthous C, Radford SE. Im7 folding mechanism: misfolding on a path to the native state. Nature Struct Biol 2002;9: 209–16.
[24] Dobson CM. The nature and significance of protein folding. In: Pain RH, editor. Mechanisms of protein folding, 2nd ed. Oxford: Oxford University Press; 2000. p. 1–33.
[25] Karplus M, Weaver DL. Folding dynamics: the diffusion–collision model and experimental data. Prot Sci 1994;3:650–6.
[26] Panchenko AR, Luthey-Schulten Z, Wolynes PG. Foldons, protein structural modules, and exons. Proc Natl Acad Sci USA 1996;93:2008–13.
[27] Cheung MS, Garcia AE, Onuchic JN. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. Proc Natl Acad Sci USA 2002;99:685–90.
[28] Booth PJ, Curran AR. Membrane protein folding. Curr Opin Struct Biol 1999;9:115–21.
[29] Trieber DK, Williamson JR. Exploring the kinetic traps in RNA folding. Curr Opin Struct Biol 1999;9:339–45.
[30] Hardesty B, Kramer G. Folding of a nascent peptide on the ribosome. Prog Nucleic Acid Res Mol Biol 2001;66:41–66.
[31] Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, et al. Crystal structure of the ribosome at 5.5 Å resolution. Science 2001;292:883–96.
[32] Jenni S, Ban N. The Chemistry of Protein Synthesis and Voyage Through the Ribosomal Tunnel. Curr Op Struct Biol 2003;13:212–9.
[33] Gilbert RJC, Fucini P, Connell S, Fuller SD, Nierhaus KH, Robinson CV, Dobson CM, Stuart DL. Three-dimensional structures of translating ribosomes by cryo-EM, submitted for publication.
[34] Hartl FU, Hayer-Hartl M. Molecular chaperones in the cytosol: from nascent chain to folded protein. Science 2002;295:1852–8.
[35] Hore PJ, Winder SL, Roberts CH, Dobson CM. Stopped-flow photo-CIDNP observation of protein folding. J Am Chem Soc 1997;119:5049–50.
[36] Dobson CM. Protein folding and disease: a view from the first Horizon symposium. Nature Rev Drug Disc 2003;2:154–60.
[37] Ellis RJ, editor. The chaperonins. San Diego, CA: Academic Press; 1996.
[38] Ellis RJ. Importance of the Anfinsen cage. Curr Biol 2003;13: R881–3.
[39] Pelham HR. Speculations on the functions of the major heat shock and glucose-regulated proteins. Cell 1986;46:959–61.

[40] Parsall DA, Kowal AS, Singer MA, Lindquist S. Protein disaggregation mediated by heat-shock protein Hsp104. Nature 1994;372:475.
[41] Hammon C, Helenius A. Quality control in the secretory pathway. Curr Opin Cell Biol 1995;7:523.
[42] Kaufman RJ, Scheuner D, Schroder M, Shen X, Lee K, Liu CY, Arnold SM. The unfolded protein response in nutrient sensing and differentiation. Nature Rev Mol Cell Biol 2002;3:411–21.
[43] Schubert U, Anton LC, Gibbs J, Orbyry CC, Yewdell JW, Bennink JR. Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. Nature 2000;404:770–4.
[44] Matouschek A. Protein unfolding—an important process in vivo? Curr Opin Struct Biol 2001;13:98–104.
[45] Radford SE, Dobson CM. From computer simulations to human disease: emerging themes in protein folding. Cell 1999;97:291–8.
[46] Thomas PJ, Qu BH, Pedersen PL. Defective protein folding as a basis of human disease. Trends Biochem Sci 1995;20:456–9.
[47] Dobson CM. The structural basis of protein folding and its links with human disease. Phil Trans R Soc Lond B 2001;356:133–45.
[48] Tan SY, Pepys MB. Amyloidosis. Histopathology 1994;25:403–14.
[49] Pepys MB. In: Weatherall DJ, Ledingham JG, Warrel DA, editors. The Oxford textbook of medicine, 3rd ed. Oxford: Oxford University Press; 1995. p. 1512–24.
[50] Koo EH, Lansbury Jr PT, Kelly JW. Amyloid diseases: abnormal protein aggregation in neurodegeneration. Proc Natl Acad Sci USA 1999;96:9989–90.
[51] Dobson CM. Getting out of shape—protein misfolding diseases. Nature 2002;418:729–30.
[52] Sunde M, Blake CCF. The structure of amyloid fibrils by electron microscopy and X-ray diffraction. Adv Prot Chem 1997;50:123–59.
[53] Jiménez JL, Guijarro JI, Orlova E, Zurdo J, Dobson CM, Sunde M, et al. Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. EMBO J 1999;18:815–21.
[54] Chiti F, Webster P, Taddei N, Clark A, Stefani M, Ramponi G, et al. Designing conditions for in vitro formation of amyloid protofilaments and fibrils. Proc Natl Acad Sci USA 1999;96:3590–4.
[55] Dobson CM. Protein misfolding, evolution and disease. Trends Biochem Sci 1999;24:329–32.
[56] Fändrich M, Fletcher MA, Dobson CM. Amyloid fibrils from muscle myoglobin. Nature 2001;410:165–6.
[57] Fändrich M, Dobson CM. The behaviour of polyamino acids reveals an inverse side-chain effect in amyloid structure formation. EMBO J 2002;21:5682–90.
[58] Lopez de la Paz M, Goldie K, Zurdo J, Lacrois E, Dobson CM, Hoenger A, et al. De novo designed peptide-based amyloid fibrils. Proc Natl Acad Sci USA 2002;99:16052–7.
[59] Krebs MRH, Morozova-Roche L, Daniel K, Robinson CV, Dobson CM. Observation of sequence specificity in the seeding of amyloid fibrils. Submitted for publication.
[60] Serpell LC, Blake CCF, Fraser PE. Molecular structure of a fibrillar Alzheimer's Aβ fragment. Biochemistry 2000;39:13269–75.
[61] Jiménez JL, Nettleton EJ, Bouchard M, Robinson CV, Dobson CM, Saibil HR. The protofilament structure of insulin amyloid fibrils. Proc Natl Acad Sci USA 2002;99:9196–201.
[62] Petkova AT, Ishii Y, Balbach JJ, Antzutkin ON, Leapman RD, Delaglio F, et al. A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. Proc Natl Acad Sci USA 2002;99:16742–7.
[63] Wille H, Michelitsch MD, Guenebaut V, Supattapone S, Serban A, Cohen FE, et al. Structural studies of the scrapie prion protein by electron crystallography. Proc Natl Acad Sci USA 1999;99:3563–8.
[64] Chamberlain A, MacPhee CE, Zurdo J, Morozova-Roche LA, Hill HAO, Dobson CM, et al. Ultrastructural organisation of amyloid fibrils by atomic force microscopy. Biophys J 2000;79:3282–93.
[65] DuBay KF, Chiti F, Zurdo J, Dobson CM, Vendruscolo M. Predicting absolute amyloid aggregation rates of polypeptide chains. Submitted for publication.
[66] Chiti F, Stefani M, Taddei N, Ramponi G, Dobson CM. Rationalisation of mutational effects on peptide and protein aggregation rates. Nature 2003;424:805–8.
[67] Chiti F, Taddei N, Baroni F, Capanni C, Stefani M, Ramponi G, et al. Kinetic partitioning of protein folding and aggregation. Nature Struct Biol 2002;9:137–43.
[68] Kelly J. Alternative conformation of amyloidogenic proteins and their multi-step assembly pathways. Curr Opin Struct Biol 1998;8:101–6.
[69] Harper JD, Lansbury Jr PT. Models of amyloid seeding in Alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. Annu Rev Biochem 1997;66:385–407.
[70] Padrick SB, Miranker AD. "Islet amyloid" phase partitioning and secondary nucleation are central to the mechanism of fibrillogenesis. Biochemistry 2002;41:4694–703.
[71] Nettleton EJ, Tito P, Sunde M, Bouchard M, Dobson CM, Robinson CV. Characterization of the oligomeric states of insulin in self-assembly and amyloid fibril formation by mass spectrometry. Biophys J 2000;79:1053–65.
[72] Caughey B, Lansbury Jr PT. Protofibrils, pores, fibrils, and neurodegeneration: separating the responsible protein aggregates from the innocent bystanders. Annu Rev Neurosci 2003;26:267–98.
[73] Schlunegger MP, Bennett MJ, Eisenberg D. Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. Adv Protein Chem 1997;50:61–122.
[74] Bitan G, Kirkitadze MD, Lomakin A, Vollers SS, Benedek GB, Teplow DB. Amyloid β-protein (Aβ) assembly: Aβ40 and Aβ42 oligomerize through distinct pathways. Proc Natl Acad Sci USA 2003;100:330–5.
[75] Bouchard M, Zurdo J, Nettleton EJ, Dobson CM, Robinson CV. Formation of insulin amyloid fibrils followed by FTIR simultaneously with CD and electron microscopy. Protein Sci 2000;9:1960–7.
[76] Lashuel HA, Hartley D, Petre BM, Walz T, Lansbury Jr PT. Neurodegenerative disease: amyloid pores from pathogenic mutations. Nature 2002;418:291.
[77] Dobson CM. How do we explore the energy landscape for folding? In: Fraunfelder H, Deisenhofer J, Wolynes PG, editors. Simplicity and complexity in proteins and nucleic acids. Berlin: Dahlem University Press; 1999. p. 15–37.
[78] Macario AJL, de Macario EC. Sick chaperones and ageing: a perspective. Ageing Res Rev 2002;1:295–311.
[79] Wilson MR, Easterbrook-Smith SB. Clusterin is a secreted mammalian chaperone. Trends Biochem Sci 2000;25:95–8.
[80] Sherman MY, Goldberg AL. Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases. Neuron 2001;29:15–32.
[81] Bence NF, Sampat RM, Kopito RR. Impairment of the ubiquitin–proteosome system by protein aggregation. Science 2001;292:1552–5.
[82] Horwich A. Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions. J Clin Invest 2002;110:1221–32.
[83] Stefani M, Dobson CM. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. J Mol Med 2003;81:678–99.
[84] Ramirez-Alvarado M, Merkel JS, Regan L. A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro. Proc Natl Acad Sci USA 2000;97:8979–84.
[85] Booth DR, Sunde M, Bellotti V, Robinson CV, Hutchinson WL, Fraser PE, et al. Instability, unfolding and aggregation of human lysozyme variants underlying amyloid fibrillogenesis. Nature 1997;385:787–93.
[86] Larsson G, Last AM, Desmyter A, Decanniere K, Canet D, Spencer A, et al. A camelid antibody fragment inhibits amyloid fibril formation by human lysozyme. Nature 2003;424:783–8.

[87] Prusiner S. Prion diseases and the BSE crisis. Science 1997;278: 245–51.
[88] Chien P, DePace AH, Collins SR, Weissman JS. Generation of prion transmission barriers by mutational control of amyloid conformations. Nature 2003;424:948–51.
[89] Perutz MF, Windle AH. Cause of neuronal death in neurodegenerative diseases attributable to expansion of glutamine repeats. Nature 2001;412:143–4.
[90] Walsh DM, Klyubin I, Fadeeva JV, Cullen WK, Anwyl R, Wolfe MS, et al. Naturally secreted oligomers of amyloid beta protein potently inhibit hippocampal long-term potentiation in vivo. Nature 2002;416:535–9.
[91] Bucciantini M, Giannoni E, Chiti F, Baroni F, Formigli L, Zurdo J, et al. Inherent cytotoxicity of aggregates implies a common origin for protein misfolding diseases. Nature 2002;416:507–11.
[92] Kayed R, Head E, Thompson JL, McIntire TM, Milton SC, Cotman CW, et al. Common structure of soluble amyloid oligomers implies common mechanisms of pathogenesis. Science 2003;300: 486–9.
[93] Polverino de Laureto P, Taddei N, Frare E, Capanni C, Constantini S, Zurdo J. Protein aggregation and amyloid fibril formation by an SH3 domain probed by limited proteolysis. J Mol Biol 2003;334:14817–21.
[94] Muchowski PJ, Schaffar G, Sittler A, Wanker EE, Hayer-Hartl MK, Hartl FU. Hsp70 and Hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. Proc Natl Acad Sci USA 2000;97:7841–6.
[95] Csermely P. Chaperone overload is a possible contributor to 'civilization diseases'. Trends Genet 2001;17:701–4.
[96] Höppener JWM, Nieuwenhuis MG, Vroom TM, Ahrén B, Lips CJM. Role of Islet amyloid in Type 2 diabetes mellitus: consequence or cause? Mol Cell Endocrin 2002;197:205–12.
[97] Gejyo F, Homma N, Suzuki Y, Arakawa M. Serum levels of β2-microglobulin as a new form of amyloid protein in patients undergoing long-term hemodialysis. N Engl J Med 1986;314:585–6.
[98] Frausto da Silva JJR, Williams RJP. The biological chemistry of the elements. Oxford: Oxford University Press; 1991.
[99] Schenck D. Amyloid-β immunotherapy for Alzheimer's disease: the end of the beginning. Nature Rev Neurosci 2002;3:824–8.
[100] Soto C. Unfolding the role of protein misfolding in neurodegenerative diseases. Nature Rev Neurosci 2003;4:49–60.
[101] Pepys MB, et al. Targeted pharmocological depletion of serum amyloid P component for treatment of human amyloidosis. Nature 2002;417:254–9.
[102] Hammarstrom P, Wiseman RL, Powers ET, Kelly JW. Prevention of transthyretin amyloid disease by changing protein misfolding energetics. Science 2003;299:713–6.
[103] May BC, et al. Potent inhibition of scrapie prion replication in cultured cells by bisacridines. Proc Natl Acad Sci USA 2003;100: 3416–21.