Protein folding and misfolding

Christopher M. Dobson

University of Cambridge, Department of Chemistry, Lensfield Road, Cambridge CB2 1EW, UK (e-mail: cmd44@cam.ac.uk)

The manner in which a newly synthesized chain of amino acids transforms itself into a perfectly folded protein depends both on the intrinsic properties of the amino-acid sequence and on multiple contributing influences from the crowded cellular milieu. Folding and unfolding are crucial ways of regulating biological activity and targeting proteins to different cellular locations. Aggregation of misfolded proteins that escape the cellular quality-control mechanisms is a common feature of a wide range of highly debilitating and increasingly prevalent diseases.

One of the defining characteristics of a living system is the ability of even the most intricate of its component molecular structures to self-assemble with precision and fidelity. Uncovering the mechanisms through which such processes take place is one of the grand challenges of modern science[1]. The folding of proteins into their compact three-dimensional structures is the most fundamental and universal example of biological self-assembly; understanding this complex process will therefore provide a unique insight into the way in which evolutionary selection has influenced the properties of a molecular system for functional advantage. The wide variety of highly specific structures that result from protein folding and that bring key functional groups into close proximity has enabled living systems to develop astonishing diversity and selectivity in their underlying chemical processes. In addition to generating biological activity, however, we now know that folding is coupled to many other biological processes, including the trafficking of molecules to specific cellular locations and the regulation of cellular growth and differentiation[2]. In addition, only correctly folded proteins have long-term stability in crowded biological environments and are able to interact selectively with their natural partners. It is therefore not surprising that the failure of proteins to fold correctly, or to remain correctly folded, is the origin of a wide variety of pathological conditions. In this article we explore the underlying mechanism of protein folding and of the nature and consequences of misfolding and its links with disease.

The fundamental mechanism of protein folding

The concept of an energy landscape

The mechanism by which a polypeptide chain folds to a specific three-dimensional protein structure has until recently been shrouded in mystery. Native states of proteins almost always correspond to the structures that are most thermodynamically stable under physiological conditions[3]. Nevertheless, the total number of possible conformations of any polypeptide chain is so large that a systematic search for this particular structure would take an astronomical length of time. However, it is now clear that the folding process does not involve a series of mandatory steps between specific partly folded states, but rather a stochastic search of the many conformations accessible to a polypeptide chain[3–5]. The inherent fluctuations in the conformation of an unfolded or incompletely folded polypeptide chain enable even residues that are highly separated in the amino-acid sequence to come into contact with one other. Because, on average, native-like interactions between residues are more stable than non-native ones, they are more persistent and the polypeptide chain is able to find its lowest-energy structure by a process of trial and error. Moreover, if the energy surface or 'landscape' has the right shape (see Fig. 1) only a small number of all possible conformations needs to be sampled by any given protein molecule during its transition from a random coil to a native structure[3–6]. Because the landscape is encoded by the amino-acid sequence, natural selection has enabled proteins to evolve so that they are able to fold rapidly and efficiently.

Such a description, based more on the ideas of statistical mechanics and polymer physics than on those of classic chemical dynamics, is often referred to as the 'new view' of protein folding[7]. As well as providing a firm conceptual basis for folding, it has shown that many of the earlier phenomenological descriptions of the folding process are important limiting cases of a general mechanism. These ideas are stimulating the investigation of the most elementary steps in the folding process by both experimental and theoretical procedures. For example, biophysical measurements and computer simulations have revealed that many of the local elements of protein structures can be generated very rapidly; for example, individual α-helices are able to form in less than 100 ns, and β-turns in as little as 1 μs (refs 8, 9). Indeed, the folding *in vitro* of some of the simplest proteins, such as small helical bundles, is completed in less than 50 μs (refs 10, 11). Intriguingly, some other small proteins, particularly those based on β-sheet structures, can take many orders of magnitude longer to fold, as we see below, but such rate changes can be understood to a significant extent in terms of the characteristics of the native structures[12].

A key question is how does the correct fold emerge from such fundamental steps; that is, how is the energy landscape unique to a specific protein defined by its amino-acid sequence. The structural transitions taking place during folding *in vitro* can be investigated in detail by a variety of techniques, ranging from optical methods to NMR spectroscopy[3], some of which can now even be used to follow the behaviour of single molecules[13]. The latter capability is of particular significance in the context of probing the stochastic nature of the folding process (see Fig. 1). Studies of a series of small proteins, typically with 60–100 residues, have been crucial for investigating the most basic steps in folding because these proteins convert from their unfolded states to their native states without the complication of highly populated intermediates. For these systems, monitoring the effects of specific mutations on the kinetics of folding and unfolding has proved to be a seminal technique, because of its ability to probe the role of individual residues in the folding process[14]. Particular insight has come from the use of this approach to analyse the transition states for folding, namely the critical regions of energy surfaces through which all molecules must pass to reach the native fold (see Fig. 1). The results of many studies of these species suggest that the

© 2003 Nature Publishing Group

Petitioner Merck, Ex. 1137, p. 884

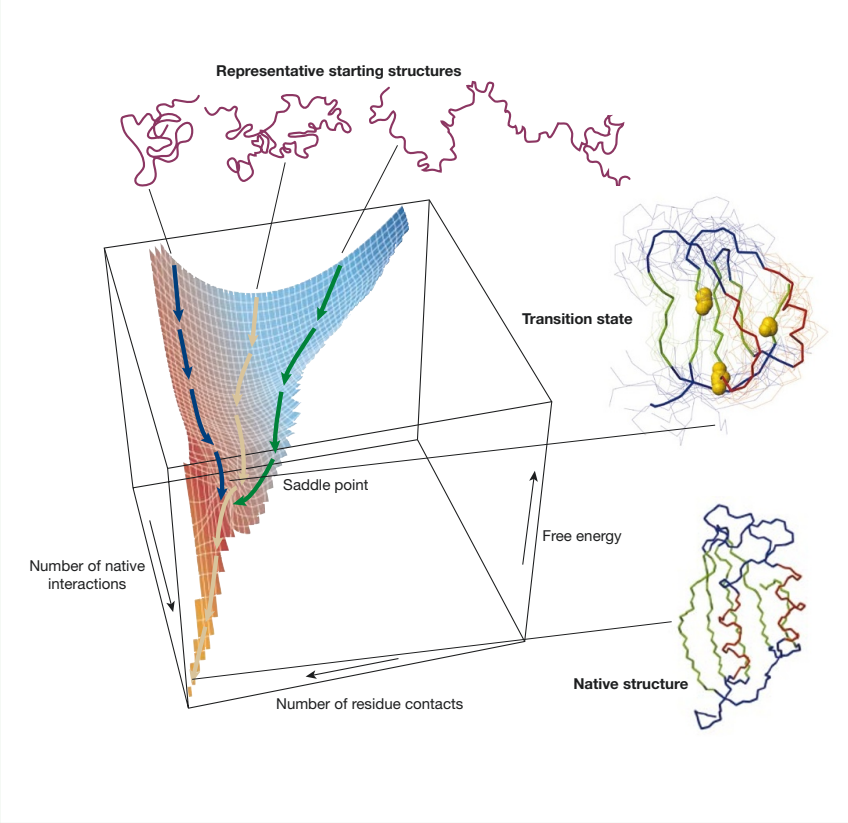

Figure 1 A schematic energy landscape for protein folding. The surface is derived from a computer simulation of the folding of a highly simplified model of a small protein. The surface 'funnels' the multitude of denatured conformations to the unique native structure. The critical region on a simple surface such as this one is the saddle point corresponding to the transition state, the barrier that all molecules must cross if they are to fold to the native state. Superimposed on this schematic surface are ensembles of structures corresponding to different stages of the folding process. The transition state ensemble was calculated by using computer simulations constrained by experimental data from mutational studies of acylphosphatase[18]. The yellow spheres in this ensemble represent the three 'key residues' in the structure; when these residues have formed their native-like contacts the overall topology of the native fold is established. The structure of the native state is shown at the bottom of the surface; at the top are indicated schematically some contributors to the distribution of unfolded species that represent the starting point for folding. Also indicated on the surface are highly simplified trajectories for the folding of individual molecules. Adapted from ref. 6.

fundamental mechanism of protein folding involves the interaction of a relatively small number of residues to form a folding nucleus, about which the remainder of the structure rapidly condenses[15].

More details of how such a mechanism is able to generate a unique fold have emerged from a range of theoretical studies, particularly involving computer simulation techniques[16]. Of particular significance are investigations that compare the simulation results with experimental observations[6,17]. One approach incorporates experimental measurements directly into the simulations as restraints limiting the regions of conformational space that are explored in each simulation; this strategy has enabled rather detailed structures to be generated for transition states[18] (see Fig. 1). The results suggest that, despite a high degree of disorder, these structures have the same overall topology as the native fold. In essence, interactions involving the key residues force the chain to adopt a rudimentary native-like architecture. Although it is not yet clear exactly how the sequence encodes such characteristics, the essential elements of the fold are likely to be determined primarily by the pattern of hydrophobic and polar residues that favours preferential interactions of specific residues as the structure becomes increasingly compact. Once the correct topology has been achieved, the native structure will then almost invariably be generated during the final stages of folding[18]. Conversely, if these key interactions are not formed, the protein cannot fold to a stable globular structure; this mechanism therefore acts also as a 'quality-control' process by which misfolding can generally be avoided.

The determinants of protein folds

Secondary structure, the helices and sheets that are found in nearly every native protein structure, is stabilized primarily by hydrogen bonding between the amide and carbonyl groups of the main chain. The formation of such structure is an important element in the overall folding process, although it might not have as fundamental a role as the establishment of the overall chain topology[19]. Perhaps the most dramatic evidence for such a conclusion is the observation of a remarkable correlation between the experimental folding rates of a wide range of small proteins and the complexity of their folds, measured by the contact order[12]. The latter is the average separation in the sequence between residues that are in contact with each other in the native structure. The existence of such a correlation can be rationalized by the argument that a stochastic search process will be more time consuming if the residues that form the nucleus are further away from each other in the sequence. This evidence strongly supports the conclusion that there are relatively simple underlying principles by which the sequence of a protein encodes its structure[20]. Not only will the establishment of such principles reveal in more depth how proteins are able to fold, but it should advance significantly our ability to predict protein folds directly from their sequences and to design sequences that encode novel folds.

For proteins with more than about 100 residues, experiments generally reveal that one (or more) intermediate is significantly populated during the folding process. There has, however, been considerable discussion about the significance of such species: whether they assist the protein to find its correct structure or whether they are traps that inhibit the folding process[21–23]. Regardless of the outcome of this debate, the structural properties of intermediates provide important evidence about the folding of these larger proteins. In particular, they suggest that these proteins generally fold in modules, in other words, folding can take place largely independently in different segments or domains of the protein[6,14]. In such cases, interactions involving key

© 2003 Nature Publishing Group

Petitioner Merck, Ex. 1137, p. 885

residues are likely to establish the native-like fold within local regions or domains and also to ensure that the latter then interact appropriately to form the correct overall structure[23,24]. The fully native structure is only acquired when all the native-like interactions have been 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[25]. This modular mechanism is appealing because it suggests that highly complex structures might be assembled in manageable pieces. Moreover, such a principle can readily be extended to describe the assembly of other macromolecules, particularly nucleic acids, and even large 'molecular machines' such as the ribosome.

Protein folding and misfolding in the cell

In a cell, proteins are synthesized on ribosomes from the genetic information encoded in the cellular DNA. Folding *in vivo* is in some cases co-translational; that is, it is initiated before the completion of protein synthesis, whereas the nascent chain is still attached to the ribosome[26]. Other proteins, however, undergo the major part of their folding in the cytoplasm after release from the ribosome, whereas yet others fold in specific compartments, such as mitochondria or the endoplasmic reticulum (ER), after trafficking and translocation through membranes[27,28]. Many details of the folding process depend on the particular environment in which folding takes place, although the fundamental principles of folding, discussed above, are undoubtedly universal. But because incompletely folded proteins must inevitably expose to the solvent at least some regions of structure that are buried in the native state, they are prone to inappropriate interaction with other molecules within the crowded environment of a cell[29]. Living systems have therefore evolved a range of strategies to prevent such behaviour[27–29].

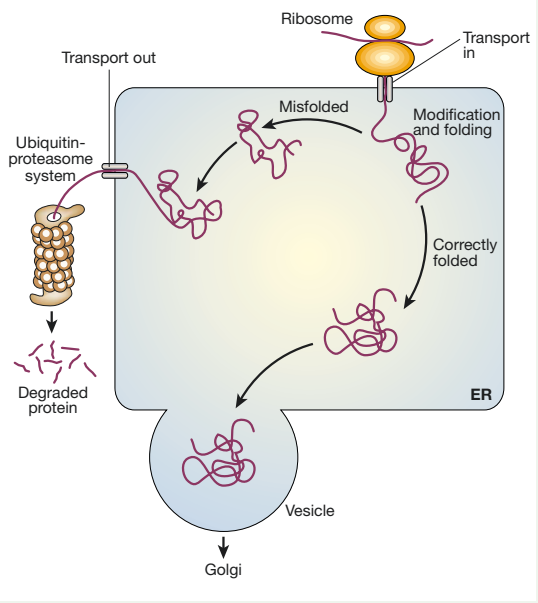

Figure 2 Regulation of protein folding in the ER. Many newly synthesized proteins are translocated into the ER, where they fold into their three-dimensional structures with the help of a series of molecular chaperones and folding catalysts (not shown). Correctly folded proteins are then transported to the Golgi complex and then delivered to the extracellular environment. However, incorrectly folded proteins are detected by a quality-control mechanism and sent along another pathway (the unfolded protein response) in which they are ubiquitinated and then degraded in the cytoplasm by proteasomes. Adapted from ref. 32.

Of particular importance in this context are the many molecular chaperones that are present in all types of cells and cellular compartments. Some chaperones interact with nascent chains as they emerge from the ribosome, whereas others are involved in guiding later stages of the folding process[27,28]. Molecular chaperones often work in tandem to ensure that the various stages in the folding of such systems are all completed efficiently. Many of the details of the functions of molecular chaperones have been determined from studies of their effects on folding *in vitro*. The best characterized of the chaperones studied in this manner is the bacterial complex involving GroEL, a member of the family of 'chaperonins', and its 'co-chaperone' GroES. Many aspects of the sophisticated mechanism through which this coupled system functions are now well understood[27,28]. Of particular interest is that GroEL, and other members of this class of molecular chaperone, contains a cavity in which incompletely folded polypeptide chains can enter and undergo the final steps in the formation of their native structures while sequestered and protected from the outside world.

Molecular chaperones do not themselves increase the rate of individual steps in protein folding; rather, they increase the efficiency of the overall process by reducing the probability of competing reactions, particularly aggregation. However, there are several classes of folding catalyst that accelerate potentially slow steps in the folding process. The most important are peptidylprolyl isomerases, which increase the rate of *cis–trans* isomerization of peptide bonds involving proline residues, and protein disulphide isomerases, which enhance the rate of formation and reorganization of disulphide bonds[30]. Despite these factors, given the enormous complexity and the stochastic nature of the folding process, it would be remarkable if misfolding never occurred. Clear evidence that molecular chaperones are needed to prevent misfolding and its consequences comes from the fact that the concentrations of many of these species are substantially increased during cellular stress; indeed, the designation of many as heat shock proteins (Hsps) reflects this fact. It is also clear that some molecular chaperones are able not only to protect proteins as they fold but also to rescue misfolded and even aggregated proteins and enable them to have a second chance to fold correctly[27,28]. Active intervention in the folding process requires energy, and ATP is required for most of the molecular chaperones to function with full efficiency.

In eukaryotic systems, many of the proteins that are synthesized in a cell are destined for secretion to the extracellular environment. These proteins are translocated into the ER, where folding takes place before secretion through the Golgi apparatus. The ER contains a wide range of molecular chaperones and folding catalysts, and in addition the proteins that fold here must satisfy a 'quality-control' check before being exported (Fig. 2)[31,32]. Such a process is particularly important because there seem to be few molecular chaperones outside the cell, although one (clusterin), at least, has recently been discovered[33]. This quality-control mechanism involves a remarkable series of glycosylation and deglycosylation reactions that enables correctly folded proteins to be distinguished from misfolded ones[31]. The importance of these regulatory systems is underlined by recent experiments that suggest that a large fraction of all polypeptide chains synthesized in a cell fail to pass this test and are targeted for degradation[34]. Like the 'heat shock response' in the cytoplasm, the 'unfolded protein response' in the ER is also stimulated (upregulated) during stress and, as we shall see below, is strongly linked to the avoidance of misfolding diseases[35].

Folding and unfolding are the ultimate ways of generating and abolishing specific types of cellular activity. In addition, processes as apparently diverse as translocation across membranes, trafficking, secretion, the immune response and regulation of the cell cycle are directly dependent on folding and unfolding events[2]. Failure to fold correctly, or to remain correctly folded, will therefore give rise to the malfunctioning of living systems and hence to disease[36–38]. Some of these diseases (such as cystic fibrosis[36] and some types of cancer[39]) result from proteins folding incorrectly and not being able to exercise

© 2003 Nature Publishing Group

Petitioner Merck, Ex. 1137, p. 886

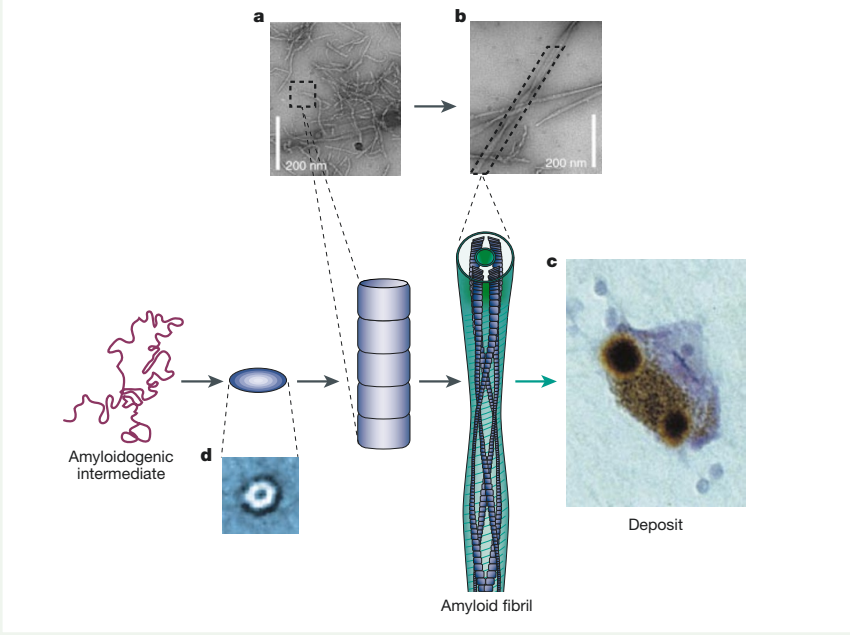

Figure 3 A schematic representation of the general mechanism of aggregation to form amyloid fibrils. Unfolded or partially unfolded proteins associate with each other to form small, soluble aggregates that undergo further assembly into protofibrils or protofilaments (**a**) and then mature fibrils (**b**, top electron microscope images from ref. 44). The fibrils often accumulate in plaques or other structures such as the Lewy bodies associated with Parkinson's disease (**c**, electron microscope image on right). Some of the early aggregates seem to be amorphous or micellar in nature, although others form ring-shaped species with diameters of approximately 10 nm (**d**, electron microscope image from ref. 53). Adapted from a figure provided by H. A. Lashuel and P. T. Lansbury Jr.

their proper function; many such disorders are familial because the probability of misfolding is often greater in mutational variants. In other cases, proteins with a high propensity to misfold escape all the protective mechanisms and form intractable aggregates within cells or (more commonly) in extracellular space. An increasing number of disorders, including Alzheimer's and Parkinson's diseases, the spongiform encephalopathies and type II diabetes, are directly associated with the deposition of such aggregates in tissues, including the brain, heart and spleen[37,38,40,41]. In the next section we look at the formation of these species.

Protein aggregation and amyloid formation

Each amyloid disease involves predominantly the aggregation of a specific protein, although a range of other components including additional proteins and carbohydrates are incorporated into the deposits when they form *in vivo*. In neurodegenerative diseases, the quantities of aggregates involved can sometimes be so small as to be almost undetectable, whereas in some systemic diseases literally kilograms of protein can be found in one or more organs[40]. The characteristics of the soluble forms of the 20 or so proteins involved in the well-defined amyloidoses are very varied — they range from intact globular proteins to largely unstructured peptide molecules — but the aggregated forms have many characteristics in common[42]. Amyloid deposits all show specific optical behaviour (such as birefringence) on binding certain dye molecules such as Congo red. The fibrillar structures typical of many of the aggregates have very similar morphologies (long, unbranched and often twisted structures a few nanometres in diameter) and a characteristic 'cross-β' X-ray fibre diffraction pattern. The latter reveals that the organized core structure is composed of β-sheets whose strands run perpendicular to the fibril axis[42]. The ability of polypeptide chains to form amyloid structures is not restricted to the relatively small number of proteins associated with recognized clinical disorders, and it now seems to be a generic feature of polypeptide chains[37,43]. The most compelling evidence for the latter statement is that fibrils can be formed *in vitro* by many other peptides and proteins, including such well-known molecules as myoglobin, and also by homopolymers such as polythreonine or polylysine[37,44].

Although no structure of an amyloid fibril has yet been determined in atomic detail, increasingly convincing models based on data from techniques such as X-ray fibre diffraction[42], cryoelectron microscopy[45] and solid-state NMR[46] are emerging. The core structure of the fibrils seems to be stabilized primarily by interactions, particularly hydrogen bonds, involving the polypeptide main chain. Because the main chain is common to all polypeptides, this observation explains why fibrils formed from polypeptides of very different sequence seem to be so similar[42,43]. In some cases only a handful of the residues of a given protein might be involved in this structure, with the remainder of the chain being associated in some other manner with the fibrillar assembly; in other cases almost the whole polypeptide chain seems to be involved. The generic amyloid structure contrasts strongly with the highly individualistic globular structures of most natural proteins. In these latter structures the interactions associated with the very specific packing of the side chains seem to override the main-chain preferences[43,44].

Even though the ability to form amyloid fibrils seems to be generic, the propensity to do so under given circumstances can vary markedly between different sequences. The relative aggregation rates for a wide range of peptides and proteins correlates with the physicochemical features of the molecules such as charge, secondary-structure propensities and hydrophobicity[47]. In a globular protein the polypeptide main chain and the hydrophobic side chains are largely buried within the folded structure. Only when they are exposed, for example when the protein is partly unfolded (for example, at low pH) or fragmented (for example, by proteolysis), will conversion into amyloid fibrils be possible. Experiments *in vitro* indicate that their formation is then generally characterized by a lag phase, followed by a period of rapid growth[48,49]. Such behaviour is typical of nucleated processes such as crystallization; the lag phase can be eliminated by the addition of preformed aggregates to fresh solutions, a process known as seeding. An interesting recent suggestion is that seeding by chemically modified forms of proteins, resulting for example from deamidation or oxidative stress, might in some cases be an important factor in triggering the aggregation process and the onset of disease[50].

There are striking similarities in the aggregation behaviour of different peptides and proteins (Fig. 3)[48,49]. The first phase in amyloid

© 2003 Nature Publishing Group

Petitioner Merck, Ex. 1137, p. 887

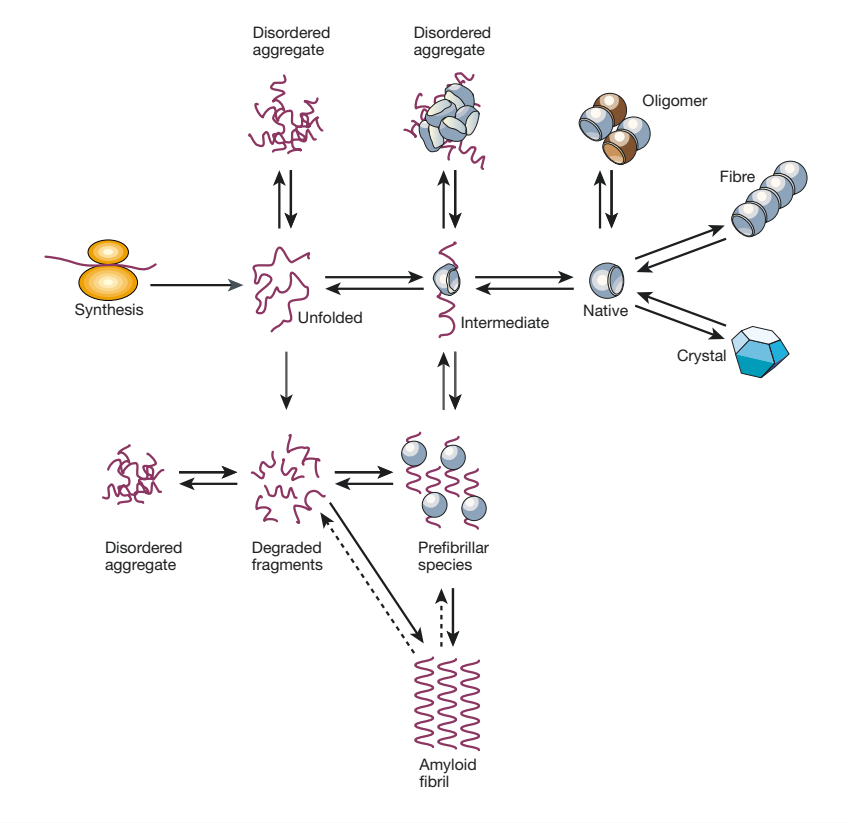

Figure 4 A unified view of some of the types of structure that can be formed by polypeptide chains. An unstructured chain, for example newly synthesized on a ribosome, can fold to a monomeric native structure, often through one or more partly folded intermediates. It can, however, experience other fates such as degradation or aggregation. An amyloid fibril is just one form of aggregate, but it is unique in having a highly organized 'misfolded' structure, as shown in Fig. 3. Other assemblies, including functional oligomers, macromolecular complexes and natural protein fibres, contain natively folded molecules, as do the protein crystals produced *in vitro* for X-ray diffraction studies of their structures. The populations and interconversions of the various states are determined by their relative thermodynamic and kinetic stabilities under any given conditions. In living systems, however, transitions between the different states are highly regulated by the environment and by the presence of molecular chaperones, proteolytic enzymes and other factors. Failure of such regulatory mechanisms is likely to be a major factor in the onset and development of misfolding diseases. Adapted from ref. 54.

formation seems to involve the formation of soluble oligomers as a result of relatively nonspecific interactions, although, in some cases, specific structural transitions, such as domain swapping, might be important[51]. The earliest species visible by electron or atomic-force microscopy generally resemble small bead-like structures, sometimes linked together, and often described as amorphous aggregates or as micelles. These early 'prefibrillar aggregates' then transform into species with more distinctive morphologies, often called 'protofilaments' or 'protofibrils'. 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 reorganization. The aggregates that form first are likely to be relatively disorganized structures that expose to the outside world a variety of segments of the protein that are normally buried in the globular state[52]. In some cases, however, these early aggregates appear to adopt quite distinctive structures, including well-defined annular species[53] (see Fig. 3).

Molecular evolution and the control of protein misfolding

The state of a protein that is adopted under specific conditions depends on the relative thermodynamic stabilities of the various accessible conformations and on the kinetics of their interconversion (Fig. 4)[37,54]. Amyloid fibrils are just one of the types of aggregate that can be formed by proteins, although a significant feature of this particular species is that its highly organized hydrogen-bonded structure is likely to give it unique kinetic stability. Thus, once formed, such aggregates can persist for long periods, allowing a progressive build-up of deposits in tissue, and indeed enabling seeding of the subsequent conversion of additional quantities of the same protein into amyloid fibrils. It is therefore not surprising that biological systems have almost universally avoided the deliberate formation of such material. Nevertheless, there is increasing evidence that the unique properties of amyloid structures have been exploited by some species, including bacteria, fungi and even mammals, for specific (and carefully regulated) purposes[55–57].

There is evidence that evolutionary selection has tended to avoid amino-acid sequences, such as alternating polar and hydrophobic residues, that favour a β-sheet structure of the type seen in amyloid fibrils[58]. Moreover, recent studies suggest that the aggregation process that results in amyloid fibrils is nucleated in a similar manner to that of folding, but that the residues involved might well be located in different regions of the sequence from those that nucleate folding[59]. Such 'kinetic partitioning' means that mutations that occur during evolution could be selected for their ability to enhance folding at the expense of aggregation. However, it is apparent that biological systems have become robust not just by careful manipulation of the sequences of proteins but also by controlling, by means of molecular chaperones and degradation mechanisms, the particular state adopted by a given polypeptide chain at a given time and under given conditions. This process can be thought of as being analogous to the way in which biology regulates and controls the various chemical transformations that take place in the cell by means of enzymes. 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[35,60].

The ideas encapsulated in Fig. 4 therefore serve as a framework for understanding the fundamental events that underlie misfolding diseases. For example, many of the mutations associated with the familial forms of deposition diseases increase the population of partially

© 2003 Nature Publishing Group

Petitioner Merck, Ex. 1137, p. 888

unfolded states, and hence the propensity to aggregate, by decreasing the stability or cooperativity of the native state[37,41,61,62]. Other familial diseases are associated with the accumulation of amyloid deposits whose primary components are fragments of native proteins; such fragments can be produced by aberrant processing or incomplete proteolysis, and are unable to fold into aggregation-resistant states. Other pathogenic mutations enhance the propensities of such species to aggregate, for example by increasing their hydrophobicity or decreasing their charge[47]. And, in the prion disorders such as Kuru or Creutzfeldt–Jakob disease, it seems that ingestion of pre-aggregated states of an identical protein, for example by voluntary or involuntary cannibalism or through the use of contaminated pharmaceuticals or surgical instruments, can markedly increase the inherent rate of aggregation through seeding and hence can generate a mechanism for transmission[48,63].

In some aggregation diseases, the large quantities of insoluble protein involved can physically disrupt specific organs and thereby cause pathological behaviour[40]. But for neurodegenerative disorders, such as Alzheimer's disease, the primary symptoms almost certainly result from a 'toxic gain of function' associated with aggregation[64]. The early prefibrillar aggregates of proteins associated with such diseases are highly damaging to cells; by contrast, the mature fibrils are usually relatively benign[48,65]. Moreover, experiments have recently suggested that similar aggregates of proteins that are not connected with any known diseases could be equally cytotoxic[52]. The generic nature of such aggregates and their effects on cells has recently been supported by the remarkable finding that antibodies can cross-react with early aggregates of different peptides and proteins, and moreover inhibit their toxicity[66]. It is possible that there are specific mechanisms for this toxicity, for example as a result of annular species (Fig. 3) that resemble the toxins produced by bacteria that form pores in membranes and disrupt the ion balance in cells[53]. However, it is likely that the relatively disorganized prefibrillar aggregates are also harmful to cells, probably through a less specific mechanism, for example as a result of the exposure of non-native hydrophobic surfaces stimulating aberrant interactions with cell membranes or other cellular components[67].

Future directions

In normal circumstances the molecular chaperones and other 'housekeeping' mechanisms are remarkably efficient in ensuring that such potentially toxic species as prefibrillar aggregates are neutralized before they can do any damage[28,68]. This neutralization could result simply from the efficient targeting of misfolded proteins for degradation, but it seems that molecular chaperones are also able to alter the partitioning between harmful and harmless forms of aggregates (Fig. 4)[69]. If the efficiency of these protective mechanisms is impaired, however, the probability of pathogenic behaviour increases[35,68]. Such a process would explain why most of the amyloid diseases are associated with old age, when there is likely to be an increased tendency for proteins to become misfolded or damaged, coupled with a decreased efficiency of the molecular chaperone and unfolded proteins responses[70]. It is ironic that through our success in increasing the life expectancy of the populations of the developed world, we are now seeing the limitations of our proteins and of the regulatory mechanisms that control their behaviour[71]. It is therefore essential that we use our developing understanding of misfolding and aggregation to find effective strategies for combating these increasingly common and highly debilitating diseases[54]. Fortunately, there is now real evidence to suggest that modern science will rise successfully to this tremendous challenge. □

doi:10.1038/nature02261

1. Vendruscolo, M., Zurdo, J., MacPhee, C. E. & Dobson, C. M. Protein folding and misfolding: a paradigm of self-assembly and regulation in complex biological systems. *Phil. Trans. R. Soc. Lond.* **361,** 1205–1222 (2003).
2. Radford, S. E. & Dobson, C. M. From computer simulations to human disease: emerging themes in protein folding. *Cell* **97,** 291–298 (1999).
3. Dobson, C. M., Sali, A. & Karplus, M. Protein folding: a perspective from theory and experiment. *Angew. Chem. Int. Ed. Eng.* **37,** 868–893 (1998).
4. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. Navigating the folding routes. *Science* **267,** 1619–1620 (1995).
5. Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nature Struct. Biol.* **4,** 10–19 (1997).
6. Dinner, A. R., Sali, A., Smith, L. J., Dobson, C. M. & Karplus, M. Understanding protein folding via free energy surfaces from theory and experiment. *Trends Biochem. Sci.* **25,** 331–339 (2000).
7. Baldwin, R. L. Protein folding: matching speed and stability. *Nature* **369,** 183–184 (1994).
8. Eaton, W. A., Munoz, V., Thompson, P. A., Henry, E. R. & Hofrichter, J. Kinetics and dynamics of loops, α-helices, β-hairpins, and fast-folding proteins. *Acc. Chem. Res.* **31,** 745–753 (1998).
9. Snow, C. D., Nguyen, H., Pande, V. S. & Gruebele, M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **420,** 102–106 (2002).
10. Yang, W. Y. & Gruebele, M. Folding at the speed limit. *Nature* **423,** 193–197 (2003).
11. Mayor, U. *et al.* The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **421,** 863–867 (2003).
12. Plaxco, K. W., Simons, K. T. & Baker, D. Contact order, transition state placement and the refolding rates of single domain proteins. *J. Mol. Biol.* **277,** 985–994 (1998).
13. Schuler, B., Lipman, E. A. & Eaton, W. A. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419,** 743–747 (2002).
14. Fersht, A. R. *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding* (W.H. Freeman, New York, 1999).
15. Fersht, A. R. 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* **97,** 1525–1529 (2000).
16. Shea, J. E. & Brooks, C. L. From folding surfaces to folding proteins: a review and assessment of simulation studies of protein folding and unfolding. *Annu. Rev. Phys. Chem.* **52,** 499–535 (2001).
17. Fersht, A. R. & Daggett, V. Protein folding and unfolding at atomic resolution. *Cell* **108,** 573–582 (2002).
18. Vendruscolo, M., Paci, E., Dobson, C. M. & Karplus, M. Three key residues form a critical contact network in a transition state for protein folding. *Nature* **409,** 641–646 (2001).
19. Makarov, D. E. & Plaxco, K. W. The topomer search model: a simple, quantitative theory of two-state protein folding kinetics. *Protein Sci.* **12,** 17–26 (2003).
20. Baker, D. A surprising simplicity to protein folding. *Nature* **405,** 39–42 (2000).
21. Roder, H. & Colon, W. Kinetic role of early intermediates in protein folding. *Curr. Opin. Struct. Biol.* **7,** 15–28 (1997).
22. Sanchez, I. E. & Kiefhaber, T. Evidence for sequential barriers and obligatory intermediates in apparent two-state protein folding. *J. Mol. Biol.* **325,** 367–376 (2003).
23. Khan, F., Chuang, J. I., Gianni, S. & Fersht, A. R. The kinetic pathway of folding of barnase. *J. Mol. Biol.* **333,** 169–186 (2003).
24. Vendruscolo, M., Paci, E., Karplus, M. & Dobson, C. M. Structures and relative free energies of partially folded states of proteins. *Proc. Natl Acad. Sci. USA* **100,** 14817–14821 (2003).
25. Cheung, M. S., Garcia, A. E. & Onuchic, J. N. Protein folding mediated by solvation: water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl Acad. Sci. USA* **99,** 685–690 (2002).
26. Hardesty, B. & Kramer, G. Folding of a nascent peptide on the ribosome. *Prog. Nucleic Acid Res. Mol. Biol.* **66,** 41–66 (2001).
27. Bukau, B. & Horwich, A. L. The Hsp70 and Hsp60 chaperone machines. *Cell* **92,** 351 –366 (1998).
28. Hartl, F. U. & Hayer-Hartl, M. Molecular chaperones in the cytosol: from nascent chain to folded protein. *Science* **295,** 1852–1858 (2002).
29. Ellis, R. J. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr. Opin. Struct. Biol.* **11,** 114–119 (2001).
30. Schiene, C. & Fischer, G. Enzymes that catalyse the restructuring of proteins. *Curr. Opin. Struct. Biol.* **10,** 40–45 (2000).
31. Hammon, C. & Helenius, A. Quality control in the secretory pathway. *Curr. Opin. Cell. Biol.* **7,** 523–529 (1995).
32. Kaufman, R. J. *et al.* The unfolded protein response in nutrient sensing and differentiation. *Nature Rev. Mol. Cell Biol.* **3,** 411–421 (2002).
33. Wilson, M. R. & Easterbrook Smith, S. B. Clusterin is a secreted mammalian chaperone. *Trends Biochem. Sci.* **25,** 95–98 (2000).
34. Schubert, U. *et al.* Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature* **404,** 770–774 (2000).
35. Bence, N. F., Sampat, R. M. & Kopito, R. R. Impairment of the ubiquitin–proteasome system by protein aggregation. *Science* **292,** 1552–1555 (2001).
36. Thomas, P. J., Qu, B. H. & Pedersen, P. L. Defective protein folding as a basis of human disease. *Trends Biochem. Sci.* **20,** 456–459 (1995).
37. Dobson, C. M. The structural basis of protein folding and its links with human disease. *Phil. Trans. R. Soc. Lond. B* **356,** 133–145 (2001).
38. Horwich, A. Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions. *J. Clin. Invest.* **110,** 1221–1232 (2002).
39. Bullock, A. N. & Fersht, A. R. Rescuing the functions of mutant p53. *Nature Rev. Cancer* **1,** 68–76 (2001).
40. Tan, S. Y. & Pepys, M. B. Amyloidosis. *Histopathology* **25,** 403–414 (1994).
41. Kelly, J. W. Alternative conformation of amyloidogenic proteins and their multi-step assembly pathways. *Curr. Opin. Struct. Biol.* **8,** 101–106 (1998).
42. Sunde, M. & Blake, C. C. F. The structure of amyloid fibrils by electron microscopy and X-ray diffraction. *Adv. Protein Chem.* **50,** 123–159 (1997).
43. Dobson, C. M. Protein misfolding, evolution and disease. *Trends Biochem. Sci.* **24,** 329–332 (1999).
44. Fändrich, M. & Dobson, C. M. The behaviour of polyamino acids reveals an inverse side-chain effect in amyloid structure formation. *EMBO J.* **21,** 5682–5690 (2002).
45. Jiménez, J. L. *et al.* Cryo-electron microscopy of an SH3 amyloid fibril and model of the molecular packing. *EMBO J.* **18,** 815–821 (1999).
46. Petkova, A. T. *et al.* A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl Acad. Sci. USA* **99,** 16742–16747 (2002).
47. Chiti, F., Stefani, M., Taddei, N., Ramponi, G. & Dobson, C. M. Rationalization of mutational effects on protein aggregation rates. *Nature* **424,** 805–808 (2003).
48. Caughey, B. & Lansbury, P. T. Jr. Protofibrils, pores, fibrils, and neurodegeneration: separating the

© 2003 Nature Publishing Group

Petitioner Merck, Ex. 1137, p. 889

responsible protein aggregates from the innocent bystanders. *Annu. Rev. Neurosci.* **26,** 267–298 (2003).
49. Bitan, G. *et al.* Amyloid β-protein (Aβ) assembly: Aβ40 and Aβ42 oligomerize through distinct pathways. *Proc. Natl Acad. Sci. USA* **100,** 330–335 (2003).
50. Nilsson, M. R., Driscoll, M. & Raleigh, D. P. Low levels of asparagine deamidation can have a dramatic effect on aggregation of amyloidogenic peptides: implications for the study of amyloid formation. *Protein Sci.* **11,** 342–349 (2002).
51. Schlunegger, M. P., Bennett, M. J. & Eisenberg, D. Oligomer formation by 3D domain swapping: a model for protein assembly and misassembly. *Adv. Protein Chem.* **50,** 61–122 (1997).
52. Bucciantini, M. *et al.* Inherent cytotoxicity of aggregates implies a common origin for protein misfolding diseases. *Nature* **416,** 507–511 (2002).
53. Lashuel, H. A., Hartley, D., Petre, B. M., Walz, T. & Lansbury, P. T. Jr. Neurodegenerative disease: amyloid pores from pathogenic mutations. *Nature* **418,** 291 (2002).
54. Dobson, C. M. Protein folding and disease: a view from the First Horizon Symposium. *Nature Rev. Drug Discov.* **2,** 154–160 (2003).
55. True, H. L. & Lindquist, S. L. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407,** 477–483 (2000).
56. Chapman, M. R. *et al.* Role of *Escherichia coli* curli operons in directing amyloid fiber formation. *Science* **295,** 851–855 (2002).
57. Kelly, J. W. & Balch, W. E. Amyloid as a natural product. *J. Cell Biol.* **161,** 461–462 (2003).
58. Broome, B. M. & Hecht, M. H. Nature disfavours sequences of alternating polar and non-polar amino acids: implications for amyloidogenesis. *J. Mol. Biol.* **296,** 961–968 (2000).
59. Chiti, F. *et al.* Kinetic partitioning of protein folding and aggregation. *Nature Struct. Biol.* **9,** 137–143 (2002).
60. Macario, A. J. L. & Macario, E. C. Sick chaperones and ageing: a perspective. *Ageing Res. Rev.* **1,** 295–311 (2002).
61. Ramirez-Alvarado, M., Merkel, J. S. & Regan, L. A systematic exploration of the influence of the protein stability on amyloid fibril formation in vitro. *Proc. Natl Acad. Sci. USA* **97,** 8979–8984 (2000).
62. Dumoulin, M. *et al.* A camelid antibody fragment inhibits amyloid fibril formation by human lysozyme. *Nature* **424,** 783–788 (2003).
63. Prusiner, S. B. Prion diseases and the BSE crisis. *Science* **278,** 245–251 (1997).
64. Taylor, J. P., Hardy, J. & Fischbeck, K. H. Toxic proteins in neurodegenerative disease. *Science* **296,** 1991–1995 (2002).
65. Walsh, D. M. *et al.* Naturally secreted oligomers of amyloid β protein potently inhibit hippocampal long-term potentiation *in vivo*. *Nature* **416,** 535–539 (2002).
66. Kayed, R. *et al.* Common structure of soluble amyloid oligomers implies common mechanisms of pathogenesis. *Science* **300,** 486–489 (2003).
67. Stefani, M. & Dobson, C. M. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *J. Mol. Med.* **81,** 678–699 (2003).
68. Sherman, M. Y. & Goldberg, A. L. Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases. *Neuron* **29,** 15–32 (2001).
69. Muchowski, P. J. *et al.* Hsp70 and Hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. *Proc. Natl Acad. Sci. USA* **97,** 7841–7846 (2000).
70. Csermely, P. Chaperone overload is a possible contributor to 'civilization diseases'. *Trends Genet.* **17,** 701–704 (2001).
71. Dobson, C. M. Getting out of shape—protein misfolding diseases. *Nature* **418,** 729–730 (2002).

Acknowledgements I should like to thank in particular the Wellcome Trust, the Leverhulme Trust and the UK Research Councils for generous support over many years, without whom my own research activities in this area of science could not have been carried out.

© 2003 Nature Publishing Group

Petitioner Merck, Ex. 1137, p. 890