The physical basis of

protein misfolding disorders

Tuomas P. J. Knowles,
Michele Vendruscolo, and
Christopher M. Dobson

Investigations of the fundamental physical principles underlying Alzheimer's and Parkinson's diseases are providing new insights into their origins and are suggesting rational strategies through which to combat their onset and progression.

Over the past 50 years, major advances in the diagnosis and treatment of disease have dramatically increased human longevity. In many countries, average life expectancies have gone from less than 60 years to over 80 years.[1] One statistic that puts such changes into perspective is that of all the people who have ever reached the age of 65, more than half are estimated to be alive today. But despite the increasing success in reducing the burden of human disease, there are exceptions to that general trend, as we discuss in this article.

In recent years the number of cases of dementia and other neurodegenerative disorders has risen significantly; it is projected to double every 20 years and reach 115 million worldwide by 2050.[2] The most prevalent and rapidly increasing of the disorders is

Tuomas Knowles, **Michele Vendruscolo**, and **Christopher Dobson** work in the department of chemistry at the University of Cambridge in Cambridge, UK.

Alzheimer's disease, which accounts for some 80% of dementia cases. Related conditions include Parkinson's and Creutzfeldt–Jakob diseases; one form of Creutzfeldt–Jakob, the human analogue of mad cow disease, has been associated with the consumption of contaminated beef. Alzheimer's disease in particular is only now being widely recognized as a growing epidemic. And dementias in general are becoming the most costly and widely feared of all diseases in many parts of the world.[2] No effective therapies exist to combat those disorders, and in many cases reliable diagnostics are not available except after a disease has already reached advanced stages.

Current studies of the nature of neurodegenerative disorders are bringing together the most re-
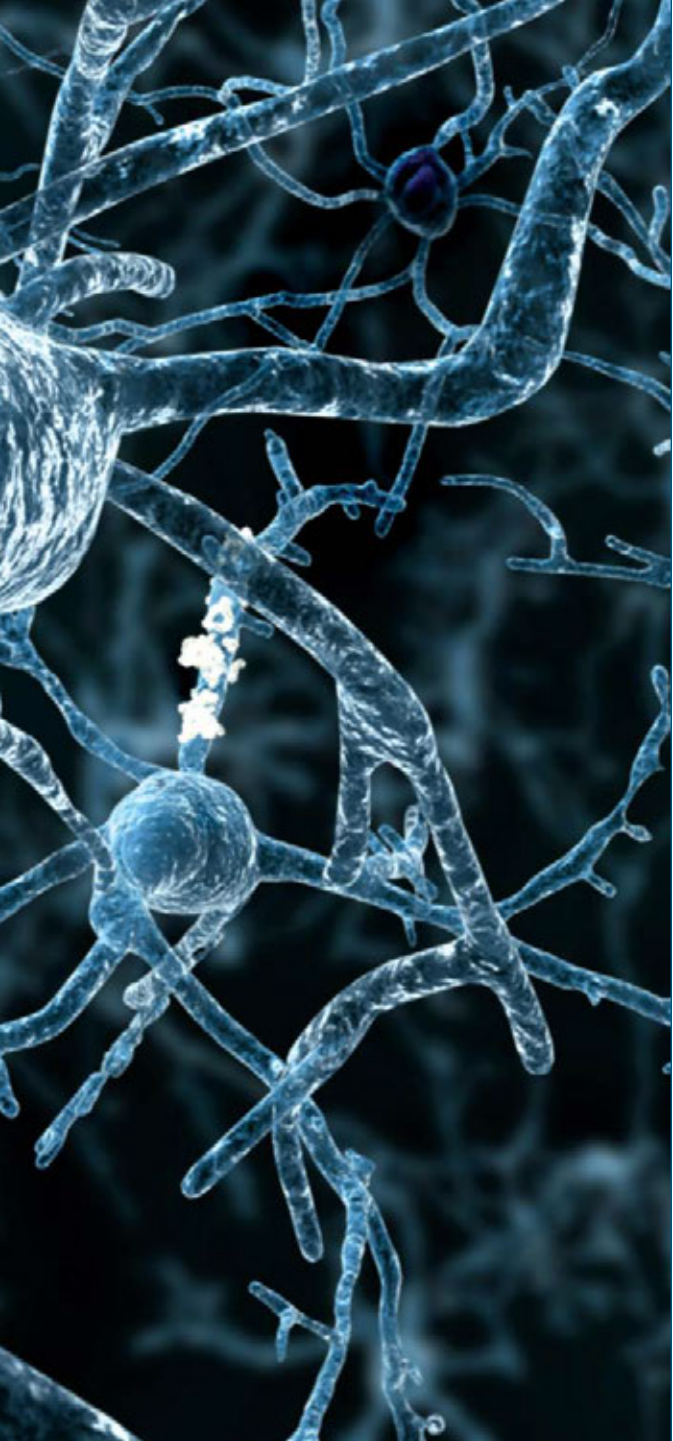
SELVANEGRA
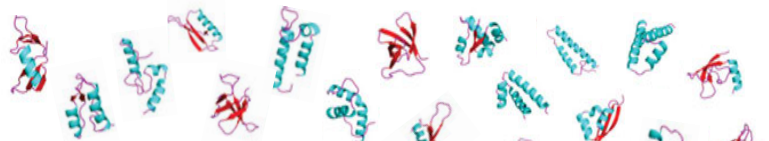
Box 1. Protein folding

After its biosynthesis in a ribosome, a newly assembled protein starts out as a long linear chain of simpler molecules called amino acids. But to become functional, the protein must find its way into a specific structure, its native state.[6] As the many images here illustrate, native-state structures are characterized by an intricate assembly of substructures known as α helices and β sheets. The α helix is a coil that spirals around a central axis; each amino acid is hydrogen bonded to another one three positions away along the chain. The β sheet contains two or more extended segments of the protein, called β strands, that are laterally linked to each other, also through hydrogen bonds. Ribbon representations, like the ones here, show α helices as spring-like ribbons (cyan) and β strands as flat arrows (red); nonrepeating parts of the protein are shown as thin tubes.

Remarkably, each type of protein—our bodies produce nearly 10 000 of them—adopts a unique structure through the process of protein folding. The protein sequence, the number and order of the amino acids that characterize the protein, determines the structure of its native state and its folding process—that is, the order of events leading to the acquisition of the native state itself. A vast amount of research has been carried out to define the process of protein folding, and to a great extent, the input of ideas from physics, particularly statistical mechanics, has allowed us to understand the basic principles reasonably well.[6]

The different amino acids have different shapes and chemical properties, and for any given protein, evolutionary selection has found one way of packing the amino acids together that is more favorable than all others in terms of free energy. The process of protein folding involves random thermal fluctuations to efficiently explore numerous packing configurations until the lowest free-energy arrangement is found.
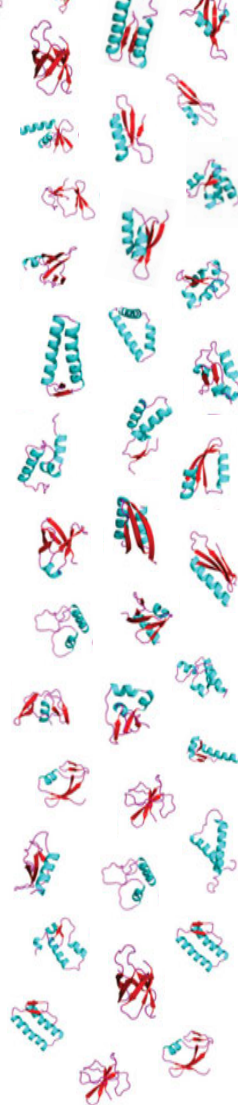
cent advances in genetic profiling with the latest methods in chemistry and physics.[3] One of the most important findings has been that the underlying molecular origins of the disorders share common aspects.[4] Perhaps even more remarkable, the aberrant biochemical processes that cause them are essentially the same ones responsible for a range of other debilitating conditions. One such disorder is type 2 diabetes, which is associated with obesity and is emerging as another modern epidemic that could halt or even reverse the rapidly increasing human life expectancy.[4]

When proteins misfold

The common feature behind a generic group of more than 50 disorders is a phenomenon known as protein misfolding. Our genetic information is encoded in the DNA in cells that make up every part of our bodies, from our livers, to our muscles, and to our brains. However, DNA just stores information, which has to be translated into the different proteins that carry out the vast number of biochemical processes that enable our bodies to function correctly.[5] (See the article by Alexander Dunn and Andrew Price, PHYSICS TODAY, February 2015, page 27.)

Protein molecules are formed by connecting amino acids into long chains.[6] Twenty amino acids, often differing by just a few atoms from each other, make up the ingredients list for creating proteins. Each type of protein is distinguished by the order in which the amino acids are joined together to form a sequence. That is the information that is encoded in
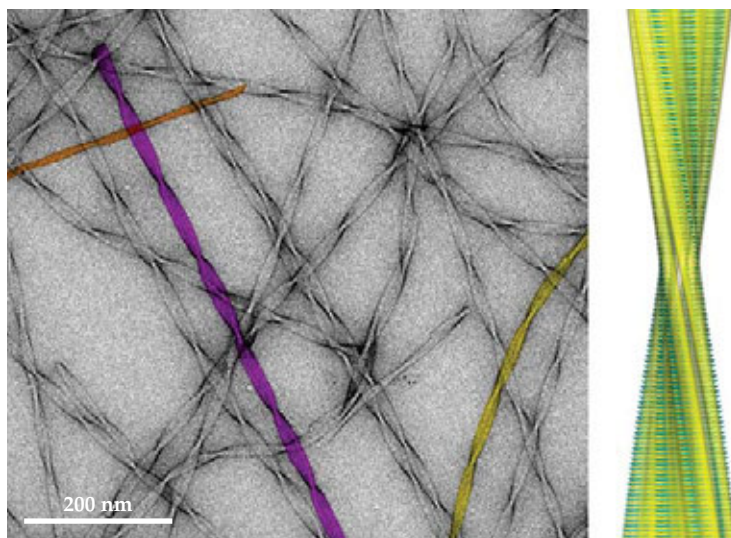
Figure 1. Amyloid fibrils formed from peptide fragments of transthyretin, a protein associated with several misfolding diseases. In the transmission electron microscopy image (left), three of the fibrils are shown in color to highlight differences in fibril widths corresponding to two (orange), three (yellow), and four (purple) peptide protofilaments, the subfilaments that form an amyloid fibril. Atomic-resolution structures of the amyloid fibrils were recently determined by combining data from multiple biophysical techniques. (See PHYSICS TODAY, June 2013, page 16.) As with correctly folded native states of proteins, amyloid structures are closely packed and highly ordered. However, they fundamentally differ from native states, as they possess a generic architecture (right) formed exclusively by paired, very long, β-sheet structures (seen here as the stacks of teal ridges) within the protofilaments that twist around a central axis and contain thousands of protein molecules. By contrast, the folds of native states are highly diverse with various and often highly intricate topologies as shown in box 1. (Adapted from ref. 17.)

our DNA. Proteins typically have 100–1000 individual amino acids. Mixing and matching the different types of amino acids is how so many different molecules are made. Indeed, human biology requires rather few of the vast number—on the order of 10^{1000}—of possible protein sequences.

Cellular protein factories, called ribosomes, assemble each of our proteins by joining the amino acids in the right order. Newly minted proteins emerge from the ribosomes as long, thin, and rather flexible chains of the component building blocks. The extended protein chains must then usually fold up into a unique and compact structure to function properly.[6] Box 1 gives an overview of protein folding. In any living system, whether a simple bacterium or a human being, a sufficient number of each different type of protein must reach the most favorable folds to allow the intricate biochemical processes needed for life to take place.

The human disorders discussed in this article, from Alzheimer's and Parkinson's diseases to type 2 diabetes, result when one or more types of proteins fail either to fold properly or to remain correctly folded in their optimally packed and functional states. Hence those disorders are called protein-misfolding diseases.[4] In other words, they stem from an occasional failure in one of the most fundamental and well-defined types of molecular processes that occur in our bodies. If we could understand what causes such failures to occur, we might be able to devise rational approaches to the treatment of many debilitating and at present largely incurable diseases.

In addition to being nonfunctional, misfolded proteins often cling together to form a range of aggregates.[4,7–10] The most familiar examples of such aggregates are the plaques that accumulate in the brain and are a hallmark of Alzheimer's disease. But similar amyloid deposits are present in all other misfolding diseases—sometimes in trace amounts, sometimes in kilogram quantities. A further remarkable feature of the disease-associated aggregates is that they are not just shapeless masses. As shown in figure 1, they are made up of large numbers of long, thin, thread-like structures, called amyloid fibrils, that weave through the organs and tissues in which they form. Interestingly, the structures look almost the same regardless of the type of protein that forms aggregates, in contrast to the structures of functional states of proteins, which are all different.

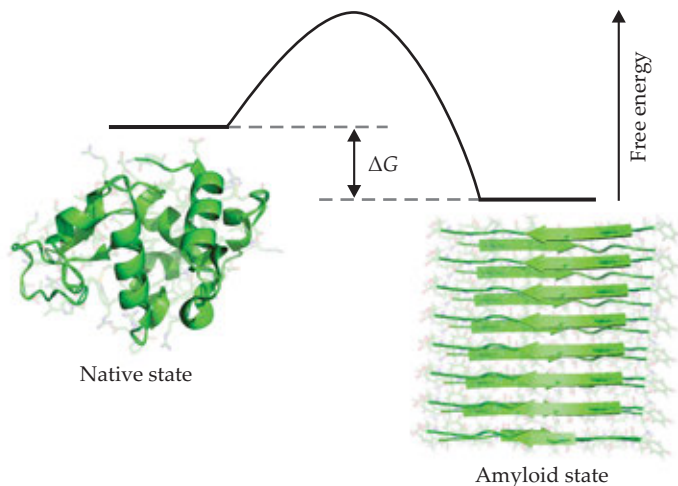To give a more personal perspective, our group's research in protein misfolding began by

Figure 2. The relative thermodynamic stability of a protein's native and amyloid states is determined by the free energy difference, ΔG, of the two forms. As free energy of the amyloid state depends on the protein concentration, whereas that of the native state is generally independent of it, at some critical concentration, the stability of the amyloid state is the same as that of the native state. A protein is more stable in the amyloid state than in its native state at concentrations exceeding that critical concentration, but the native state can still persist temporarily if, as sketched here, there are high free-energy barriers that hinder the transition into the amyloid state. Under such conditions the native state is kinetically metastable. (Adapted from ref. 12.)

chance some 20 years ago. We had been studying the folding of one particular protein, called lysozyme, as a means of investigating the principles of how proteins fold. A medical colleague told us of some patients he had with enormous quantities of amyloid fibrils composed of lysozyme. To our amazement, the protein turned out to be associated with an important misfolding disease, known as lysozyme amyloidosis.[11] In addition, all those patients had mutations in the gene encoding lysozyme, which showed that the condition is a rare genetic disease. We then realized that the mutations, each resulting in a change in one of the amino acids in the protein sequence, caused the protein to have a high propensity to misfold and assemble into the thread-like amyloid fibrils. Indeed, we were able to work out many details of the exact way the misfolding occurred and how it led to disease.[11]

Relatively few proteins have been directly associated with misfolding diseases. An obvious explanation is that those proteins could have unusual features in their sequences that somehow allow them to fold to amyloid forms as well as to their functional states. But another accidental discovery gave a clue to a different possibility. We were carrying out laboratory experiments to define the detailed folding process of a protein not connected with disease. We discovered to our great surprise that under some circumstances, the protein converted from its normal soluble state into the thread-like amyloid fibrils we had previously only seen in disease-associated proteins. A series of subsequent experiments, by us and others, showed that one could take many—perhaps nearly all—ordinary proteins and turn them into fibrils. We realized that amyloid fibrils must represent a generic, alternative structural state for proteins. And unlike with the correctly folded functional state, the formation of the amyloid state and its overall architecture do not depend on the nature of the amino acids that make up a particular protein or the protein sequence.

The physics of misfolding

Why then don't amyloid fibrils appear all the time in living systems? The answer is that the cellular environment, through a series of protective measures collectively known as protein homeostasis,[12,13] inhibits the conversion of correctly folded proteins into the misfolded structures. In addition, proteins known as molecular chaperones, large multiprotein complexes, called proteasomes, and enzyme-containing organelles, called lysosomes, act to degrade and destroy any misfolded molecules that slip through the regulatory processes.[3,14]

As illustrated in figure 2, native protein molecules in living systems are not always thermodynamically stable.[12] When aggregation reactions become thermodynamically favorable, they must be slowed down dramatically to enable proteins to exist in long-lived, metastable, functional states. Box 2 describes the fundamental elements of the kinetics of amyloid formation. It is clear why the conversion of correctly folded proteins to amyloid fibrils must be avoided. The amyloid state does not have the individual and uniquely folded structure needed for a given protein to carry out its specific biological function. And amyloid structures are usually highly insoluble, so they sometimes accumulate in large quantities and cause organ failure, as in the case of patients with systemic amyloid disease.

Further experiments have revealed another crucial factor that is particularly important for neurodegenerative diseases. Misfolded states of proteins are often toxic because they interact inappropriately with some of the numerous other cell components and cause them to malfunction. The effect is much

---

Box 2. Kinetics of amyloid formation

Thermodynamics describes whether or not a system can spontaneously transition from one state to another, whereas kinetics addresses the question of how fast such a transformation will take place (see PHYSICS TODAY, August 2011, page 18). The kinetics of the formation of amyloid fibrils is normally monitored in experiments by fluorescence-based assays that provide a signal proportional to the amount of fibril present. Typically in those assays fibril formation exhibits a nearly sigmoidal time course (left panel, blue line), in which a lag time, τ_{lag}, during which almost no signal is observed, is followed by rapid growth with rate, r_{max} (red line); the growth finally stops when all the soluble proteins have been used. Kinetic measurements have historically been one of the main tools for elucidating reaction mechanisms, both for small-molecule reactions and for enzymology.[6]
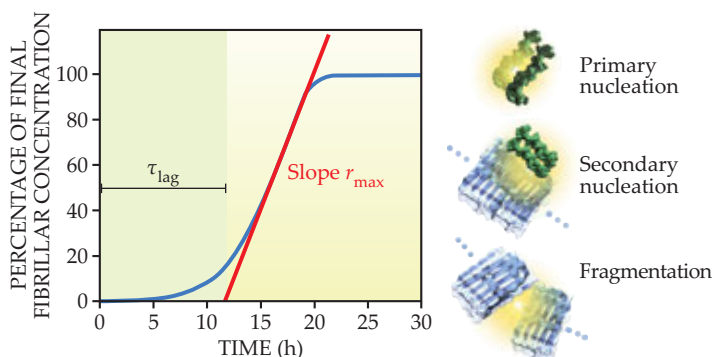
A powerful way to connect the elementary microscopic aggregation processes with their macroscopic manifestations is to consider the different ways in which new aggregates can be formed. They can be built from scratch out of individual protein molecules through primary nucleation, or they can come from fragmentation of existing fibrils. They can also be made from a combination of individual and aggregated species through secondary nucleation (right panel). Areas colored yellow in the images indicate where the reactions take place. A convenient bookkeeping strategy to track the cumulative effect of all those processes is to write a master equation

$$\frac{\partial f}{\partial t} = F_{growth} + F_{source},$$

where the time derivative of the fibrillar concentration, f, is expressed as a combination of growth processes, F_{growth}, which include the addition of monomeric proteins to existing fibrils, and source processes, F_{source}, which include primary nucleation, secondary nucleation, and fragmentation. The analytical integration of the master equation offers the possibility to connect in a fully quantitative manner the microscopic processes to macroscopic observations.[12] One can then investigate ways to modulate the aggregation process by using a range of biological macromolecules such as antibodies, molecular chaperones, and therapeutic small molecules. (Figure adapted from ref. 12.)
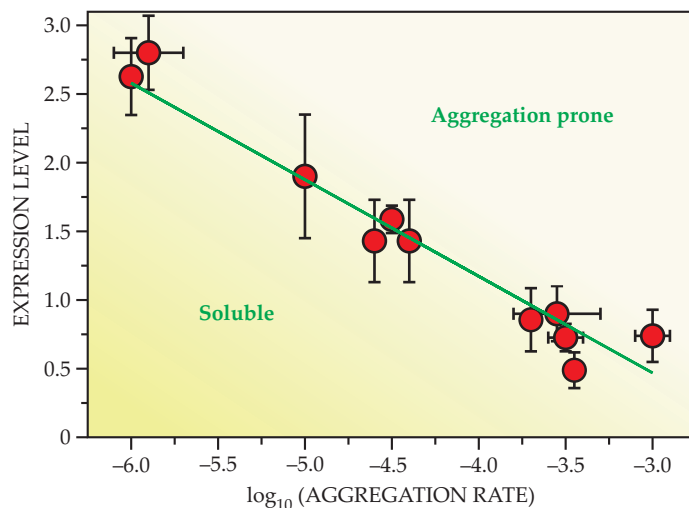
Figure 3. Living on the edge of protein solubility. Expression levels, which indicate the amounts of proteins present in a cell, are correlated with the rate at which those proteins aggregate *in vitro* under near-physiological conditions. As the aggregation rate is related to the solubility of a protein, the decreasing expression level with increasing aggregation rate indicates that our proteins have evolved to be soluble at the concentrations at which they are present in the cell, but only just—therefore a small change in the cellular environment can tip the balance in favor of aggregation and amyloid formation. (Adapted from ref. 18.)

the same as with synthetic drugs, which almost always cause side effects when they are introduced to the highly complex and well-organized environment of a living organism. In addition, the initial aggregates containing just a few protein molecules are more toxic than the well-defined fibrils into which they convert, whether or not the proteins concerned are associated with disease.[15] That observation can be rationalized on the basis that in small aggregates a far greater fraction of their misfolded molecular components are exposed and can interact with other molecules than in the well-defined mature fibrils.[12]

Taken together, the body of recent experimental findings suggests a new conceptual framework for understanding the origin of protein-misfolding diseases. As a result of evolutionary pressure over vast lengths of time, our proteins have been selected to fold into, and remain in, specific structures that enable them to efficiently carry out specific functions. However, in the past 50–100 years, we have rapidly prolonged our lifespans and changed our lifestyles, and the systems that regulate and protect our functional proteins have not kept up with those changes. In particular, many of the neurodegenerative diseases occur late in life, when such systems are at greater risk of becoming impaired or overwhelmed. Other disorders, such as diabetes, are strongly associated with changes in diet, lack of exercise, and increasing obesity—again, situations that challenge the mechanisms that protect our molecular components. So as we are better able to avoid other dis-

eases, the inability of our molecular components and their protective environments to always function correctly becomes more apparent.

Although current ideas about the nature of protein-misfolding diseases emerged from results of test-tube experiments, they have been strongly supported by experiments in various living systems and by clinical data.[3] In essence, our proteins and their environments have evolved together to ensure that misfolding disorders do not occur during the biologically important years of our lives, so that we have enough time to pass on our genes and protect our offspring until they can look after themselves.[12] But evolution cannot make anything better than it needs to be or anticipate future changes. Over the past century or so, we have made enormous advances that take us out of the evolutionary phase of our development into a postevolutionary era in which humankind is in control of its destiny to an unprecedented degree.

Novel therapeutic strategies

Mounting evidence indicates that protein-misfolding disorders are a result of the inability of cellular defenses to deal with conditions in which too many proteins fail to fold correctly or to remain correctly folded.[12–14,16] Proteins, at the concentrations at which they are present in our cells, have an inherent tendency to convert from their functional soluble states to insoluble aggregates, but various protein homeostasis mechanisms combat that tendency and prevent such a conversion or degrade the aggregates once the conversion is initiated. Normal and aberrant behaviors are, however, finely balanced—people are, in molecular terms, living on the edge (see figure 3)—and slight changes in conditions can cause the transition from one to the other;[12] a small mutation in a single gene, a change in diet, or a few extra years of life can trigger aberrant behavior. And the nature of the aggregation process turns out to be such that once it begins in earnest, it is extremely difficult to stop.

The corollary is that a therapeutic strategy might only need to generate small changes in the protein homeostasis system to induce a transition from pathogenic to normal behavior. Several emerging strategies based on that principle might be extremely powerful in preventing or limiting the advances of protein-misfolding disorders. Moreover, approaches that prevent the initial events that drive the misfolding and the aggregation process appear likely to be more effective than those that attempt to control the aggregation process once it has got underway.[12]

Thus the ideal approach should be to enhance the natural protective systems that prevent protein misfolding from occurring in the first place. One strategy that we are actively pursuing is based on our fundamental understanding of the fundamental physicochemical processes involved (see box 2). Because the early stages of misfolding appear to be the most crucial, drug-discovery strategies could involve the search for therapeutic molecules that interact with the soluble forms of the proteins whose aggregation is involved in disease and that help prevent them

from becoming misfolded. We are using that type of approach initially to explore a possible treatment for Parkinson's disease, and we are also engaged in a similar approach for Alzheimer's disease. We have already screened numerous compounds and identified potential therapeutic ones. In cellular models, some of the compounds are able to reduce the amount of aggregates and enable the cells' natural defense systems to cope with the resulting lower populations of toxic aggregated species. The next stage is to use the most promising compounds in more detailed trials and explore their potential for clinical use.

A hopeful future

The approach of tackling protein-misfolding disorders by defining their detailed molecular origins and then designing ways of perturbing the underlying microscopic processes appears to be making progress. By developing a more complete understanding of the nature of those phenomena, the medical research community is well on its way to devising novel strategies that might prevent, or at least slow down, the onset of misfolding diseases. A particularly exciting discovery is that the origins of protein misfolding appear to be based on generic physical and chemical properties, which can be perturbed slightly to prevent the transition of proteins to pathogenic states. In that context, physics is becoming increasingly important in clarifying the thermodynamics and kinetics of protein misfolding and aggregation through the development of a range of theoretical and experimental methods that provide much needed quantitative tools to analyze the phenomena.

References

1. M. Dobson, *Disease: The Extraordinary Stories Behind History's Deadliest Killers*, Metro Books (2013).
2. M. Prince et al., *The Global Impact of Dementia 2013–2050*, Alzheimer's Disease International (December 2013).
3. R. I. Morimoto, D. J. Selkoe, J. W. Kelly, eds., *Protein Homeostasis*, Cold Spring Harbor Laboratory Press (2011).
4. C. M. Dobson, *Trends Biochem. Sci.* **24**, 329 (1999).
5. To access a publicly available protein image database, see http://www.proteinatlas.org.
6. A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding*, Macmillan (1999).
7. J. Hardy, D. J. Selkoe, *Science* **297**, 353 (2002).
8. C. A. Ross, M. A. Poirier, *Nat. Med.* **10**, S10 (2004).
9. H. W. Querfurth, F. M. LaFerla, *N. Engl. J. Med.* **362**, 329 (2010).
10. D. Eisenberg, M. Jucker, *Cell* **148**, 1188 (2012).
11. D. R. Booth et al., *Nature* **385**, 787 (1997).
12. T. P. J. Knowles, M. Vendruscolo, C. M. Dobson, *Nat. Rev. Mol. Cell Biol.* **15**, 384 (2014).
13. W. E. Balch et al., *Science* **319**, 916 (2008).
14. F. U. Hartl, A. Bracher, M. Hayer-Hartl, *Nature* **475**, 324 (2011).
15. C. Haass, D. J. Selkoe, *Nat. Rev. Mol. Cell Biol.* **8**, 101 (2007).
16. E. T. Powers et al., *Annu. Rev. Biochem.* **78**, 959 (2009).
17. A. W. P. Fitzpatrick et al., *Proc. Natl. Acad. Sci. USA* **110**, 5468 (2013).
18. G. G. Tartaglia et al., *Trends Biochem. Sci.* **32**, 204 (2007). ∎