Alzheimer’s disease. Cystic fibrosis. Mad Cow disease. An inherited form of emphysema. Even many cancers. Recent discoveries show that all these apparently unrelated diseases result from protein folding gone wrong. As though that weren’t enough, many of the unexpected difficulties biotechnology companies encounter when trying to produce human proteins in bacteria also result from something amiss when proteins fold.

What exactly is this phenomenon? We all learned that proteins are fundamental components of all living cells: our own, the bacteria that infect us, the plants and animals we eat. The hemoglobin that carries oxygen to our tissues, the insulin that signals our bodies to store excess sugar, the antibodies that fight infection, the actin and myosin that allow our muscles to contract, and the collagen that makes up our tendons and ligaments (and even much of our bones)—all are proteins.

To make proteins, ‘machines’ known as ribosomes string together amino acids into long, linear chains. Like shoelaces, these chains loop about each other in a variety of ways (i.e., they fold). But, as with a shoelace, only one of these many ways allows the protein to function properly. Yet lack of function is not always the worst scenario. For just as a hopelessly knotted shoelace could be worse than one that won’t stay tied, too much of a misfolded protein could be worse than too little of a normally folded one. This is because a misfolded protein can actually poison the cells around it.

**Early Studies**

The importance of protein folding has been recognized for many years. Almost a half-century ago, Linus Pauling discovered two quite

W. A. (Bill) Thomasson, Ph.D., is a science and medical writer based in Oak Park, IL. Jonathan A. King served as science writer. This series is available on FASEB’s Public Policy Home Page at http://www.faseb.org/opa/ or as reprints from FASEB’s Office of Public Affairs, 9650 Rockville Pike, Bethesda, MD 20814.
simple, regular arrangements of amino acids—the α-helix and the β-sheet (see the box, Fundamental Patterns of Protein Structure)—that are found in almost every protein. And in the early 1960s, Christian Anfinsen showed that the proteins actually tie themselves: If proteins become unfolded, they fold back into proper shape of their own accord; no shaper or folder is needed.

Of course, neither Pauling nor Anfinsen nor the committees that awarded them their respective Nobel prizes knew at the time that these discoveries would be so important for understanding Alzheimer’s disease or cystic fibrosis. And when Pauling, at least, was doing his breakthrough studies, he could hardly have imagined the enormity of today’s biotechnology industry. What scientists did know is that any process that was so fundamental to life as protein folding would have to be of the utmost practical importance.

But research did not stop with Pauling and Anfinsen. Indeed, we now know that Anfinsen’s conclusions needed expansion: Sometimes a protein will fold into a wrong shape. And some proteins, aptly named chaperones, keep their target proteins from getting off the right folding path (see the box, Molecular Chaperones). These two small but important additions to Anfinsen’s theory hold the keys to protein folding diseases.

We’ve known since antiquity (but didn’t know we knew) that protein folding can go wrong. When we boil an egg, the proteins in the white unfold. But when the egg cools, the proteins don’t return to their original shapes. Instead, they form a solid, insoluble (but tasty) mass. This is misfolding. Similarly, biochemists have always cursed the tendency of some proteins to form the insoluble lumps in the bottom of their test tubes. We now know that these, too, were proteins folded into the wrong shapes.

Until recently, biochemists lacked the tools to study these insoluble lumps. Nor did they expect such masses would be particularly interesting. The prevailing view at the time was that the lumps were just hopelessly tangled and completely amorphous masses of protein fibers (aggregation). Researchers eventually discovered that these aggregates of incorrect folding could be highly structured, but before this crucial insight and before proper investigative tools were developed, biochemists simply threw their fouled test tubes away.

‘Gunking Up’ Tissues

As far back as the start of this century, physicians have been noticing that certain diseases are characterized by extensive protein deposits in certain tissues. Most of these diseases are rare, but Alzheimer’s is not. It was Alois Alzheimer himself who noted the presence of “neurofibrillary tangles and neuritic plaque” in certain regions of his patient’s brain. Tangles are more or less common in diseases that feature extensive nerve cell death; plaque, however, is specific to Alzheimer’s. The major question, which has only recently been answered, is
Fundamental Patterns of Protein Structure

More than a half century ago, evidence began to accumulate that a major part of most proteins’ folded structure consists of two regular, highly periodic arrangements, designated $\alpha$ and $\beta$. In 1951 researchers worked out the precise nature of these arrangements.

The key to both structures is the hydrogen bond. A hydrogen atom is nothing more than a proton with a surrounding electron cloud. When one of these atoms is chemically bonded to an electron-withdrawing atom such as nitrogen or oxygen, much of the electron cloud moves toward the nitrogen or oxygen. The proton is thus left almost bare, with its positive charge largely unshielded. If it comes close to another atom with a bit of extra negative charge (typically, an oxygen or nitrogen atom), the partial positive and negative charges will attract each other. It is this attraction that produces the hydrogen bond and stabilizes the $\alpha$ and $\beta$ structures.

The $\alpha$ structure, now called $\alpha$-helix, is a right-hand spiral stabilized by hydrogen bonds between each amino acid's nitrogen atom and the oxygen atom of the fourth one up the chain. This means that there are 3.6 amino acids for each turn of the helix. The main part of the amino acid (the side chain, designated “R” in the figure) sticks out from this spiral backbone like the bristles on a bottle brush.

The $\beta$ structure is now called $\beta$-sheet. It is essentially flat, with the side chains sticking out on alternate sides. $\beta$-sheet is also stabilized by hydrogen bonds between nitrogen and oxygen atoms. In this case, however, the hydrogen-bonded atoms belong to different amino acid chains running alongside each other. The sheets are 'parallel' if all the chains run in the same direction and are ‘antiparallel’ if alternate chains run in opposite directions. Antiparallel sheets are often, but not always, formed by a single chain looping back upon itself.

When a single chain loops back on itself to form an antiparallel $\beta$-sheet, the one to three amino acids linking the two strands are known as a $\beta$-turn. Today, scientists recognize the $\beta$-turn as one of the fundamental elements of protein structure. All other local arrangements of amino acids are described as ‘random coil,’ although they are random only in the sense of not being periodic.
whether plaque causes Alzheimer’s or, like tangles, is a consequence of it.

Further investigation showed that neuritic plaque (unrelated to the plaque that clogs atherosclerotic blood vessels and causes heart attacks) is composed almost entirely of a single protein. Deposits of large amounts of a single, insoluble protein around the degenerating nerve cells of Alzheimer’s disease eventually provided a key to understanding the disorder.

It was development of the biotechnology industry that unexpectedly spurred interest in insoluble protein gunk. This industry can produce proteins (often otherwise difficult-to-obtain human proteins) quickly and economically in bacteria. To their surprise, however, scientists who worked for biotech companies often found two things: protein that was supposed to be soluble instead precipitated as insoluble inclusion bodies within the bacteria and proteins that were supposed to be secreted into the surrounding medium instead got stuck at the bacterial cell wall.

This puzzling activity led scientists, almost for the first time, to seriously study just what goes wrong during protein folding.

Further Studies

In the decades after Anfinsen’s work, the National Institutes of Health and the National Science Foundation continued to finance research in several laboratories. Working in relative obscurity, these protein biochemists tried to discover how a completely unfolded protein, with hundreds of millions of potential folded states to choose from, consistently found the correct one—and did so within seconds to minutes.

Could there be specific, critical intermediates (partially folded chains) in the folding process? This turned out to be a difficult question to answer. Partially folded chains don’t stay that way very long; they become fully folded chains in a fraction of a second. Nevertheless, by the early 1980s researchers had not only found clear evidence for the existence of partially folded proteins, but also realized the key role these played in the folding process.

One study involved the difficulty in getting bovine growth hormone to fold properly. Although the unfolded proteins were not sticky, and the fully folded proteins were not sticky, the partially folded molecules stuck to each other—a first clue as to the origins of misfolded lumps (at least for purified proteins in test tubes). It still remained unclear why misfolding occurred in cells under certain circumstances but not under others.

Temperature Sensitivity

The early 1980s also saw one of the first serious investigations of protein misfolding. These studies focused on temperature-sensitive mutations (mutations allowing growth at 75-F but not at 100-F) in the tailspike protein of bacteriophage P22. Neither bacteriophage P22, a virus that infects certain bacteria, nor its tailspike protein has any practical importance in themselves. Faced with thorny problems,
Molecular Chaperones

The beauty and the frustration of science are that they are constantly producing surprises. Almost three decades after Christian Anfinsen had won the Nobel Prize for demonstrating that protein folding is governed solely by the protein itself, other scientists discovered that some proteins have helped in the process. This help consists of proteins called chaperones (or chaperonins) that are associated with the target protein during part of its folding process. However, once folding is complete (or even before) the chaperone will leave its current protein molecule and go on to support the folding of another.

Proper folding of some proteins appears to call for not just one chaperone, but several. Especially clear evidence for such multi-step chaperoning is provided by test-tube experiments on a protein known as rhodanese. Proper folding of this protein, the experiments show, requires five different chaperone-type proteins acting at two distinct steps in the operation. Early in the folding process, rhodanese binds to a chaperone known as DnaK; the complex that binds a further chaperone: DnaJ. Somewhat later, a protein known as GrpE catalyzes transfer of the partially folded rhodanese to another chaperone, GroEL, and its partner, GroES. These latter two proteins then see rhodanese all the way through to its properly folded state.

Several lines of evidence suggest that chaperones’ primary function may be to prevent aggregation. For example, a chaperone found in the ‘power plant’ organelles of mammalian cells (but otherwise similar to GroEL) has been shown to consist of 14 protein chains arranged as two doughnut holes stacked on top of each other (see figure). The chaperoned protein sits inside the two doughnut holes, safely sequestered from other molecules with which it might aggregate.

A role for chaperones in preventing aggregation is also suggested by what happens to mammalian proteins produced in bacteria. Although bacteria have chaperones, they are not the same as those in mammals. It is thus easy to imagine that they may be relatively ineffective toward mammalian proteins, and that this results in the aggregation so often seen. Indeed, there has been one case in which bacteria engineered to overproduce their own chaperones successfully produced a mammalian protein that otherwise irretrievably aggregated. Unfortunately, this approach has failed in other cases. And no one has yet reported introduction of mammalian chaperones into bacteria to help produce soluble mammalian proteins. Yet this, along with the introduction of mutations that block the aggregation pathway and the discovery of small molecules that prevent aggregation, is one of the most promising ways to overcome the roadblocks that biotechnology companies have so often encountered.
however, scientists often look for experimental systems that will allow them to get a foothold or find a way around them. In this case, they thought that a large protein, whose folding passes through multiple stages, would be a good system for looking at folding pathways within cells. Many temperature-sensitive mutations had already been isolated in bacteriophages, but never examined for their effect on folding. Their hopes were realized: The majority of the temperature-sensitive mutations they found, despite having only one amino acid altered, caused the tailspike protein to end up as insoluble gunk at high temperatures. Since these folding failures were occurring in bacterial cells that were growing in the laboratory, it was now possible to analyze what went wrong in a protein’s folding process.

The obvious guess at the time was that the mutant proteins were less stable. After all, the temperature scale is fundamentally defined by how much atomic-scale shaking or motion is going on; in other words, the higher the temperature, the more shaking there is. This implies that a less stable protein is more likely to fall apart at elevated temperatures and might therefore be more likely to end up (like cooked eggs) as insoluble gunk.

But this turned out not to be the case. If the mutant chains were allowed to fold up at low temperature, and were then heated, they were as stable as wild-type. It turned out to be a partially folded intermediate, on the route from the random shoelace to the correctly folded protein, that was sensitive to temperature. At higher temperatures these intermediates would stick to themselves and be unable to reach the properly folded state.
This turned out to be a general problem in the folding of many proteins: They have to pass through partially folded states in which they are delicately poised between folding all the way to the correct state or becoming seriously stuck as a result of premature entanglement with other molecules. Recognizing that it was the intermediates and not the fully folded protein that were in trouble opened the way to understanding some aspect of a range of diseases.

**Familial Amyloidotic Polyneuropathy**

Over the past several years, collaborators have conducted similar studies in connection with a human disease. The minor differences between their results and others’ are very revealing.

In the hereditary disease familial amyloidotic polyneuropathy (FAP), peripheral nerves and other organs are damaged by deposits of amyloid-type protein. Although the disorder is quite rare, extensive genetic studies have shown that the disease results from mutations in the protein transthyretin. As with the P22 tailspike protein, transthyretin contains large amounts of β-sheet structure and normally consists of several identical amino acid chains (four in this case) associated into a single, three-dimensional structure.

FAP results from any of more than 50 distinct mutations within the transthyretin protein, each altering a single amino acid. After studying several of these, scientists found that their four-chain structure is less stable under mildly acid conditions than is the wild-type structure. This contrasts with the P22 tailspike mutations, which fold slowly but are stable once folded. It also appears that transthyretin aggregation takes place from a monomeric unfolding intermediate, rather than the folding intermediate involved in P22 tailspike aggregation (the pathway may or may not be the same in both directions).

In both cases, however, the single-chain intermediates have structures that nature has designed for association with other chains of the same type. It apparently takes only a very small change in the shape of these
intermediates to alter their normal linkage with two or three other chains into an endless series of linkages that creates insoluble gunk.

There is yet another contrast between the P22 tailspike mutations and those in transthyretin: From the P22 virus’s view, the problem with the tailspike mutations is that not enough normal protein is made. People with transthyretin mutations, on the other hand, have all the normal transthyretin they need to carry out its usual function (transporting the thyroid hormone). The problem is that, as the protein is being broken down, it forms insoluble gunk, and the insoluble gunk poisons the tissues where it is deposited.

**Alzheimer’s Disease**

FAP is a rare disease; not so Alzheimer’s, which afflicts 10 percent of those over 65 years old and perhaps half of those over 85. Every year Alzheimer’s not only kills 100,000 Americans, but also costs society $82.7 billion to care for its victims.

In 1991, several different research groups found that individuals with specific mutations in their amyloid precursor protein developed Alzheimer’s disease as early as age 40. The body processes amyloid precursor protein into a soluble peptide (small protein) known as Aβ; under certain circumstances, Aβ then aggregates into long filaments that cannot be cleared by the body’s usual scavenger mechanisms. These aggregates then form the β-amyloid, which make up the neuritic plaque in Alzheimer patients. So the consistent association of amyloid precursor protein mutations with early-onset Alzheimer’s has finally answered a long-debated question: the deposition of neuritic plaque is part of the pathway leading to the disease, not a late consequence of it.

To help understand the Aβ aggregation process, researchers chemically synthesized fragments of the 40-amino-acid-long peptide. By using these fragments, they showed that the key step is getting started. Specifically, the precursor fragments have to form a specific nucleus, which then grows into the amyloid process. Possibly the slowness of this first step is why Alzheimer’s disease is almost entirely limited to older people, and it could be that the mutations in amyloid precursor protein that lead to early-onset Alzheimer’s are the ones that make it progress more quickly and easily.

Even so, Aβ remains soluble in most people. Most individuals who develop Alzheimer’s disease have the normal form of amyloid precursor protein, indistinguishable from that in people who never acquire the disorder. Why the same form of Aβ aggregates in some people’s brain but not in others’ remains a mystery, although a recent discovery has suggested an intriguing possibility.

We know that people with different genetic variants of the protein apolipoprotein E (apoE) have quite different risks of developing Alzheimer’s disease. Compared to those with the most common variant, known as apoE3, those with the apoE4 variant are
significantly more likely to develop the disease. Some studies suggest that those with the apoE2 variant may be at lower risk, although other studies disagree.

These findings are particularly surprising because apoE is best known as part of the complex that transports cholesterol and other fatty materials in the bloodstream. What could a fat-transporting protein have to do with Alzheimer’s disease? It may be significant that small amounts of this protein are associated with neuritic plaque and that apoE binds to Aβ in the test tube. The results of this binding are in dispute, however.

Researchers report that adding apoE to a test-tube solution of soluble Aβ causes rapid formation of plaque-type β-amyloid fibers—and that apoE4 does so more rapidly than apoE3. Others, however, have obtained opposite results: apoE prevents fibril formation. Thus, whereas some suggest that apoE acts as a pathological chaperone, one that actually promotes misfolding, other researchers believe that it exerts a normal chaperone’s protective effect. In either case, apoE’s influence on the folding of Ab may play a major role in development of Alzheimer’s disease.

Mad Cow and Other Species

Perhaps the most interesting example of a protein folding disorder is Mad Cow disease and its human equivalent, Creutzfeldt-Jacob disease. These diseases, along with the sheep version known as scrapie, have had the scientific community in an uproar for years. They are infectious diseases transmitted by prions, or protein particles. Prions seem to be pure protein; they contain neither DNA nor RNA. Yet an infectious agent is necessarily self-replicating. How, scientists asked themselves, could a pure protein replicate itself?

The answer now starting to emerge may be viewed as a variation on the concept of the pathological chaperone, only in this case the protein serves as its own chaperone.

The protein whose aggregation damages nerve cells in Mad Cow disease is constantly being produced by the body. Normally, though, it folds properly, remains soluble, and is disposed of without problem. But suppose that somehow a small amount misfolds in a particular way so as to become a scrapie prion. If this scrapie prion bumps into a normal-folding intermediate, it shifts the folding process in the scrapie direction and the protein, despite its perfectly normal amino acid sequence, ends up as more scrapie prion. And the process continues:
So long as the body keeps producing the normal protein, a little bit of scrapie prion can keep on creating more and then more. In effect, the prion is “replicating” itself without needing any nucleic acid of its own.

What old-school scientists find even more strange is that the process resembles something akin to genetics. Different strains of these diseases, with somewhat different clinical symptoms, ‘breed true’ as they are transmitted from one animal or human to another. Moreover, these strain differences are associated with slight differences in the protein deposits that apparently cause the disease. (Scientists have recently used these strain differences to show that a few Britons truly have Mad Cow disease, the form seen in cattle, rather than the usual human form of Creutzfeldt-Jacob disease.)

Just as replication can occur without DNA or RNA, other experiments have shown how ‘genetics’ is possible without nucleic acids. Thus, when researchers mix seed quantities of two different scrapie prion strains in separate test tubes with large amounts of normal protein, each test tube produces more of the specific scrapie prion strain that was added. That is, each strain induces the normal protein to fold in exactly the same way as the original seed. The strain breeds true in the test tube, just as it does in the body. Odd as it may seem, genetics without nucleic acid is truly possible in the world of protein folding.

**Too Little, Too Late**

Despite the examples of FAP, Alzheimer’s disease, and Mad Cow disease, in which the problem derives from accumulation of toxic, insoluble gunk, many human diseases arise from protein misfolding leaving too little of the normal protein to do its job properly. The most common hereditary disease of this type is cystic fibrosis.

Recent research has clearly shown that the many, previously mysterious symptoms of this disorder all derive from lack of a protein that regulates the transport of the chloride ion across the cell membrane. More recently scientists have shown that by far the most common mutation underlying cystic fibrosis hinders the dissociation of the transport-regulator protein from one of its chaperones. Thus, the final steps in normal folding cannot occur, and normal amounts of active protein are not produced.

A hereditary form of emphysema shows an even greater analogy to the mutations studies in P22 tailspike protein. Investigators have found that one of the most common mutations producing this disorder greatly slows the normal folding process, just as the P22 temperature-sensitive mutations do. As with the tailspike mutations, the resulting buildup of a crucial folding intermediate leads to aggregation, which deprives affected individuals of enough circulating α₁-antitrypsin to protect their lungs. Emphysema is the result.

As intriguing as these examples may be, there is a far more com-
mon instance of misfolding, which leaves too little normal protein to do its job. In this case, the protein’s job is to block cancer development.

Over the past couple of decades, scientists have learned that most cancers result from mutation in the genes that regulate cell growth and cell division. The most common of these genes, involved in roughly 40% of all human cancers, is p53. The sole function of the p53 protein appears to prevent cells with damaged DNA from dividing before the damage is repaired (or to induce them to destroy themselves, if the damage cannot be fixed). In other words, p53 exists to prevent cells from becoming cancerous.

p53 mutations associated with cancer fall into two classes. The first keeps the protein from binding to DNA; the other makes the folded form of the protein less stable. In the second group, there is simply never enough properly folded protein around to block the division of DNA-damaged cells. It will be interesting to see how many of the p53 mutants fall into this second class and whether some way can be found to stabilize them.

**Treating Protein Misfolding**

The purpose of studying any human disease is to find ways to treat it. The story of protein folding has not yet led to treatments for the diseases involved, but this could happen within the next decade.

The key is to find a small molecule, a drug that can either stabilize the normally folded structure or disrupt the pathway that leads to a misfolded protein. Although many molecular biologists and protein chemists believe this will be quite difficult, others are more optimistic.

It is difficult to pinpoint where the search for treatment currently stands, however. One scientist notes that the bulk of that work is tied up in the patent stage: companies are pursuing it but have published little on the subject. Nevertheless, one research group has shown that both thyroid hormone and the related compound TIP (2, 4, 6-triiodophenol) can stabilize transthyretin. Since TIP neither blocks the action of thyroid hormone nor exerts any hormone-like effects of its own, it

---

*Folding and aggregation during protein renaturation. Correct folding reactions, leading to the native state [(1), (2)]. Irreversible aggregation reactions, starting from different conformations during the renaturation process [(3), (4)]. [From FASEB J. 10, 52 (1996)]*
appears to be a promising treatment for FAP.

Developing small-molecule therapies is quite straightforward for proteins like transthyretin that naturally bind small molecules, but these therapies are more difficult to apply to proteins that do not have a small-molecule binding site.

One of the few other groups currently publishing their research on small-molecule structure stabilizers is working to stabilize p53, an acknowledged ‘difficult target’. In fact, one laboratory has obtained encouraging results by using two different approaches.

Treatments based on our growing knowledge and continued research of protein folding are on the way. When they arrive, the saga that began with Pauling’s fundamental studies of protein structure and Anfinsen’s investigation of what some call ‘the second genetic code’ will reach its practical fruition.

Suggested Readings


