

Protein folding: Much more intricate than we thought

 cen.acs.org/articles/95/i31/Protein-folding-Much-intricate-thought.html

Volume 95 Issue 31 | pp. 32-38

Issue Date: July 31, 2017

Cover Story

Scientists are still uncovering all the players that help keep proteins folded inside cells—and all the ways the process can go wrong

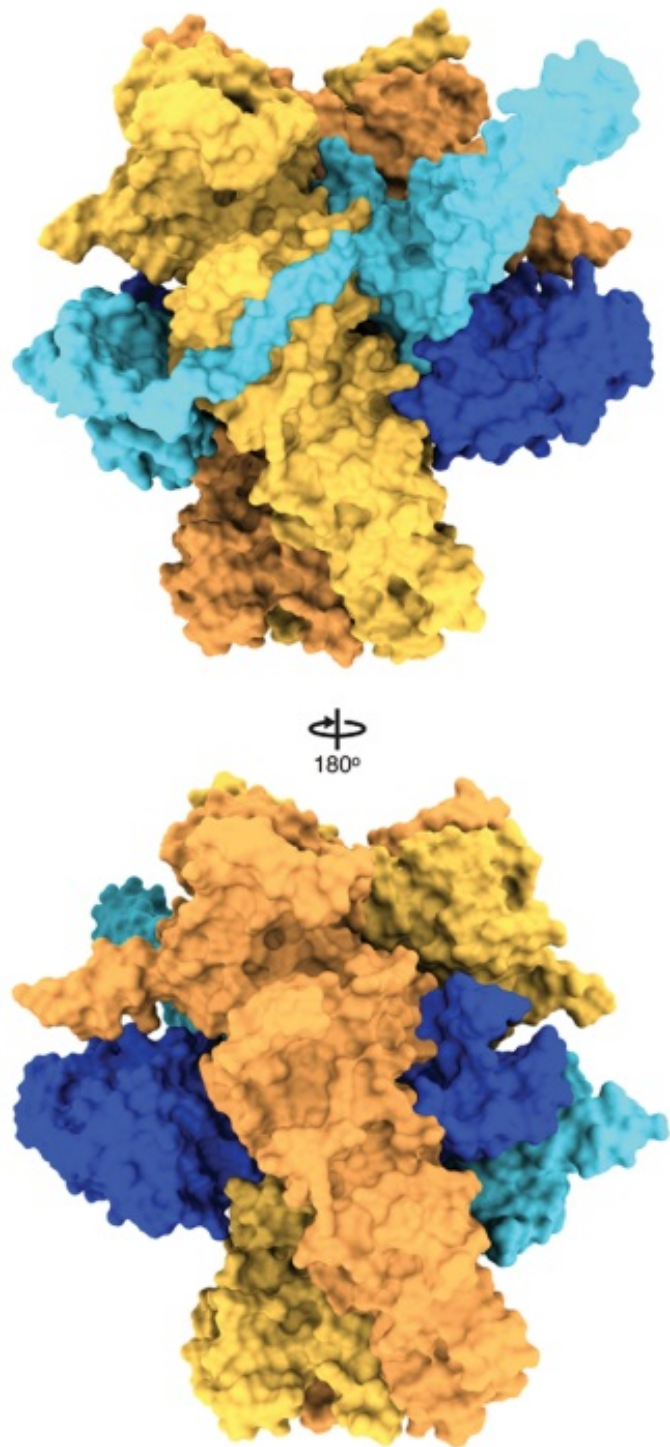
By Sarah Everts

With assistance from a helper protein (light blue), the chaperone protein HSP90 (yellow) captures a kinase protein client (dark blue) in a fleeting unfolded conformation to inactivate it.

Credit: *Science/AAAS*

To get down to business in a cell, proteins need to fold into a correct 3-D conformation. But getting into shape isn't always straightforward. In the packed, busy confines of a cell, hundreds of chaperone proteins are needed to micromanage the process and keep it on track. From the moment proteins are "born" in the ribosome to the moment they're targeted as trash, they're monitored by a cell's chaperones to keep them out of trouble, researchers are discovering. These, in addition to fundamental discoveries about the folding process, are providing insights about the basic workings of cells as well as informing those designing new proteins for synthetic biology.

When he wasn't playing the viola or sailing around the Chesapeake Bay during the early 1950s, National Institutes of Health scientist Christian Anfinsen was hard at work in the lab trying to understand proteins on a molecular level. It was during this period that he wrote down a law that every biochemist still has top of mind: The sequence of amino acids in a protein is sufficient to determine its three-dimensional structure.



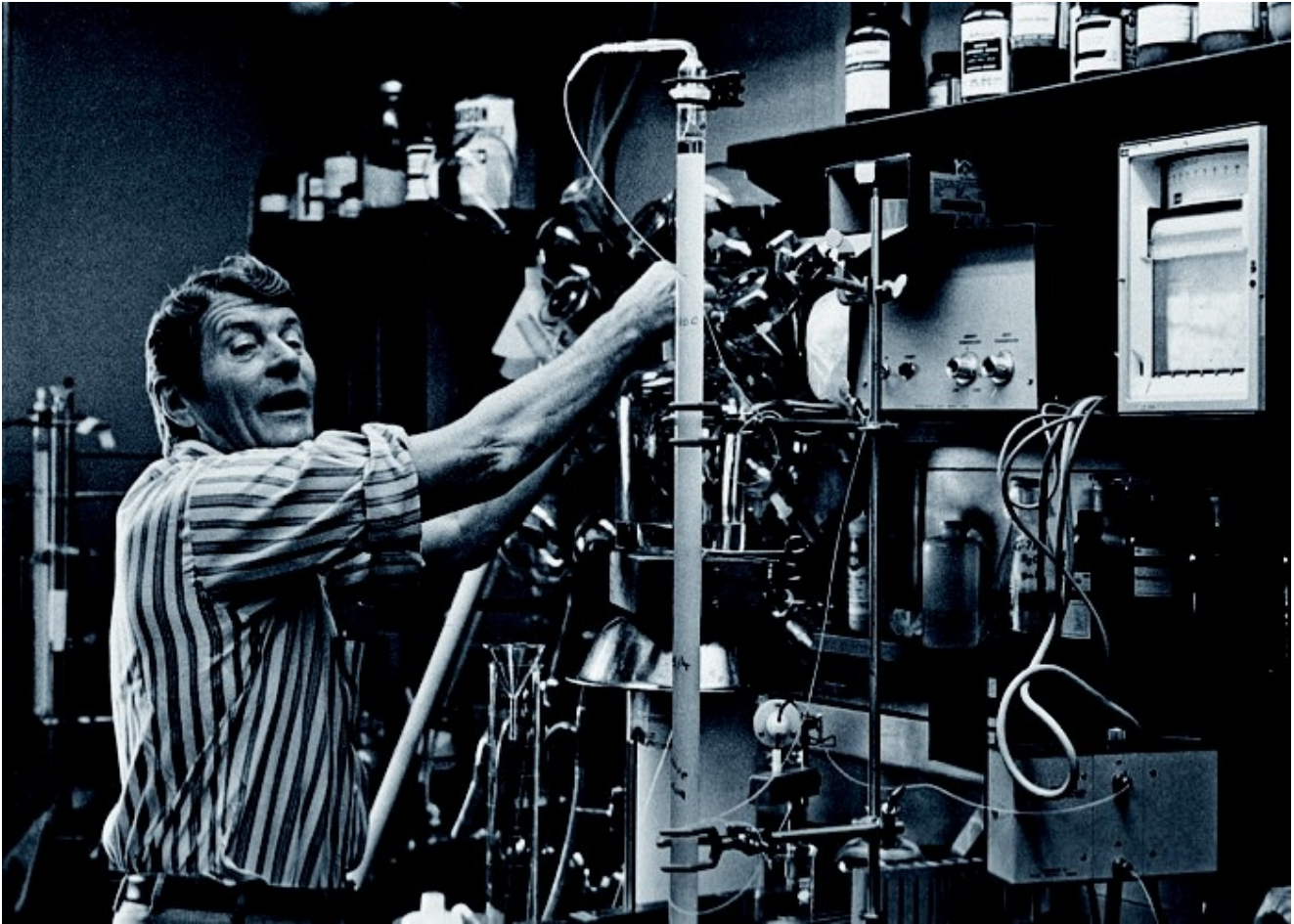
Life as we know it would be a bust if proteins didn't fold into functional 3-D structures, where α -helices, loops, and β -sheets conspire to catalyze most of the chemical reactions in a cell. Yet after Anfinsen's credo became widely accepted, many biologists gave little thought to the process proteins undergo to adopt a 3-D structure—protein folding seemed incidental compared with the exquisite biology performed by the protein thereafter. "Protein folding was thought to be somewhat esoteric, something that physical chemists but not biologists might spend their time looking at," says Christopher Dobson, a chemist at the University of Cambridge.

But over the past several decades, Dobson adds, researchers have discovered that protein folding is much more than just a perfunctory, fleeting performance before the real biology begins. Protein folding is a constantly ongoing, complicated biological opera itself, with a huge cast of performers, an intricate plot, and dramatic denouements when things go awry.

In the packed, busy confines of a living cell, hundreds of chaperone proteins vigilantly monitor and control protein folding. From the moment proteins are generated in and then exit the ribosome until their demise by degradation, chaperones act like helicopter parents, jumping in at the first signs of bad behavior to nip misfolding in the bud or to sequester problematically folded proteins before their aggregation causes disease. "People often mistakenly think that proteins are free to live out their lives in a cell," says Stanford University's Judith Frydman. "Instead, for many proteins, existence in a cell is more like life in a totalitarian state. They are never really released from the clutches of the chaperones to find their independent way" inside the cell.

As it becomes increasingly clear that folding is not a once-in-a-lifetime event for proteins but instead a part of day-to-day life in the cell, scientists are discovering that problems in this sophisticated system are implicated in diseases as diverse as cancer, diabetes, and Alzheimer's. In June, leaders in the protein-folding field gathered near Stockholm at a Nobel Foundation-sponsored meeting to discuss many of the recent advances in our understanding of how proteins fold, such as newly captured atomic-resolution snapshots of chaperones in action and strategies for tweaking protein folding as a basis for disease-fighting drugs. They also shared newly discovered rules of folding—insights into the physical chemistry of this process that could enable the design of entirely new proteins by synthetic biologists.

It's not that Anfinsen's 1950s credo was wrong. It's just that his initial physical and chemical analyses did not entirely account for the reality that biology tends to make life incredibly complicated.



Christian Anfinsen working at the bench.

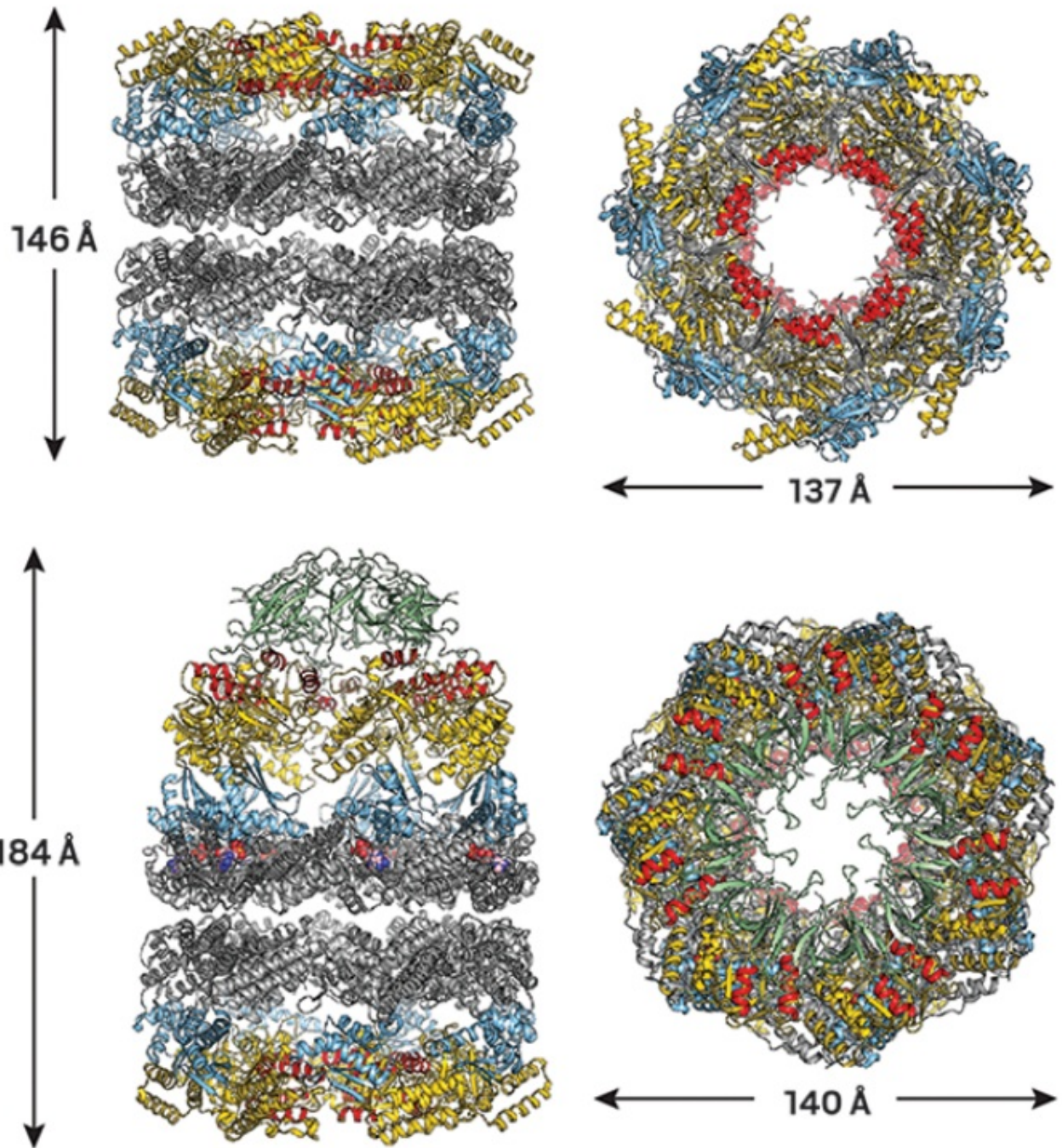
Credit: National Institutes of Health

Protein-folding trajectories

Anfinsen's work, in fact, was important enough to get him the Nobel nod: He was one of the 1972 winners of the Nobel Prize in Chemistry for his efforts to study the folding of a small, hardy 100-amino-acid-long protein called ribonuclease A (RNase A). In the wake of Anfinsen's work, RNase A, an enzyme that chops up RNA, became a go-to protein for many a folding experiment.

Researchers did everything they could to unfold RNase A—using temperature, chemicals, or both—so they could watch the protein refold rapidly and exquisitely into its functional 3-D, so-called “native,” state. Thanks to RNase A and other small, resilient proteins, researchers figured out why a polypeptide chain, which could adopt any of an astronomical number of conformations, instead chooses to collapse into a specific ordered structure. Proteins defy entropy and its drive toward disorder because of the greater energetic benefit of sequestering their hydrophobic side chains inside a compact package and away from the

polar environment of a biological cell. At the same time, researchers were developing techniques to delineate important 3-D intermediates adopted by protein chains as they ventured toward their final form.



The chaperone GroEL (side and top-down views shown) entices unfolded proteins inside its cylindrical core with the help of greasy hydrophobic lures (red, left). After a capping protein arrives to isolate the unfolded chain, the lures retract (right), making the chaperone core polar so that the protein can fold.

Credit: Modified from *Trends in Biochemical Sciences*

Yet even by the late 1970s, “it was generally assumed that protein folding occurred spontaneously in a cell,” says Franz-Ulrich Hartl, a director at the Max Planck Institute of Biochemistry. “Many biologists thought that a ribosome made a protein, it folded, and then things got interesting,” says Arthur Horwich, a biomedical researcher at Yale University. Horwich and Hartl were the first to prove that this perspective was far from the truth.

Hints that biology might be exerting some control over the folding process did exist, Horwich adds. “In the 1970s, people in biotech started to use *Escherichia coli* to make large batches of clinically important proteins. In many cases, the bacterial cells produced only aggregated, inactive masses of the relevant proteins,” he says.

Anfinsen and many others had studied protein folding in a test tube, where the unfolded protein could refold “in an ocean of solvent,” explains Gary Pielak, a chemist at the University of North Carolina, Chapel Hill, who says he has “loved protein folding since Richard Nixon was President.” In reality, however, “cells are as crowded with proteins as you can pack oranges in a crate,” Pielak says.

In a cell, “if you denature a protein and if a lot of proteins are nearby, the unfolded parts can glom on to their neighbors and suddenly you end up with a fried egg”—a tangled mess of protein that can’t be refolded properly, Pielak continues. Researchers began to realize that in these jam-packed environments, an unfolded protein might not have a lot of wiggle room to refold on its own without exposing its hydrophobic regions and risking the formation of tangled aggregates.

Scientists had further hints that biology might need to meddle in the folding process, particularly for the many proteins in a cell that are larger than RNase A, or those located in a membrane. For instance, cell membrane proteins have large stretches of hydrophobic residues exposed on the cell surface—even in a folded state—so that they can remain anchored into greasy membranes. When these hydrophobic sections emerged from the ribosome into the polar environment of the cell, why were they not aggregating into unfolded protein messes?

As biologists pondered these puzzles, Hartl and Horwich focused on yet another conundrum, this one involving the energy-generating organelles in cells: the mitochondria. Thousands of proteins used within the mitochondria are made by ribosomes outside the mitochondria. To be ushered across the organelles’ double membrane and inside, these proteins would need to be unfolded first, Hartl says. Clearly the cell had to be micromanaging the folding process to make this happen.

In 1989, Hartl, Horwich, and their colleagues reported the first folding micromanager: a chaperone protein called heat shock protein 60 (HSP60). Working in yeast, the team showed that HSP60 was responsible for refolding proteins once they arrived safely inside the mitochondria. A decade later, Horwich collaborated with the late structural biologist Paul

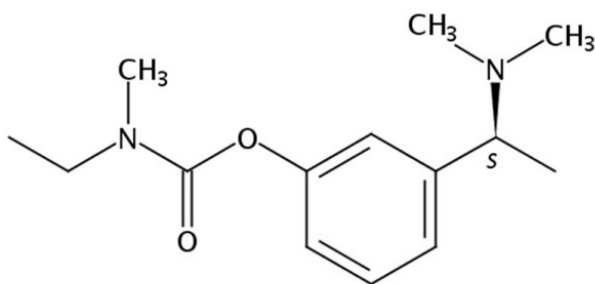
Sigler, also at Yale, to solve the crystal structure of HSP60's bacterial equivalent, GroEL (pronounced Grow-E-L because the protein was essential for *E. coli* to grow), establishing that this chaperone is a cylindrical megamachine.

It turned out that GroEL entices unfolded proteins into a core compartment lined with greasy, hydrophobic amino acids, Hartl explains. After an unfolded chain is inside, a protein called GroES caps GroEL's cylinder. During this step, GroEL undergoes a large-scale conformational rearrangement that retracts the hydrophobic residues, making the interior of the cage polar—much like the cell's inner environment. This provides a safe, isolated place for the protein to refold before being ejected.

When GroEL was first reported, "a lot of people didn't believe us," Horwich says. "Many thought it was heretical, that it disagreed with Anfinsen's principles." The thing is, Hartl explains, the existence of GroEL doesn't negate Anfinsen's rule that sequence determines structure. And it doesn't change the rules of physics that drive a protein to adopt its 3-D fold. Instead, Hartl says, chaperones that sequester unfolded proteins from the distracting, packed environment inside the cell act more like a catalyst. "They increase the rate of the folding reaction, at least for some proteins," Hartl says. He suspects that the structural intermediates adopted by proteins on the way to their final 3-D conformation that scientists have observed in test tubes are likely also present during the folding process inside GroEL. "Chaperones just give protein folding a little kinetic kick," Horwich says.

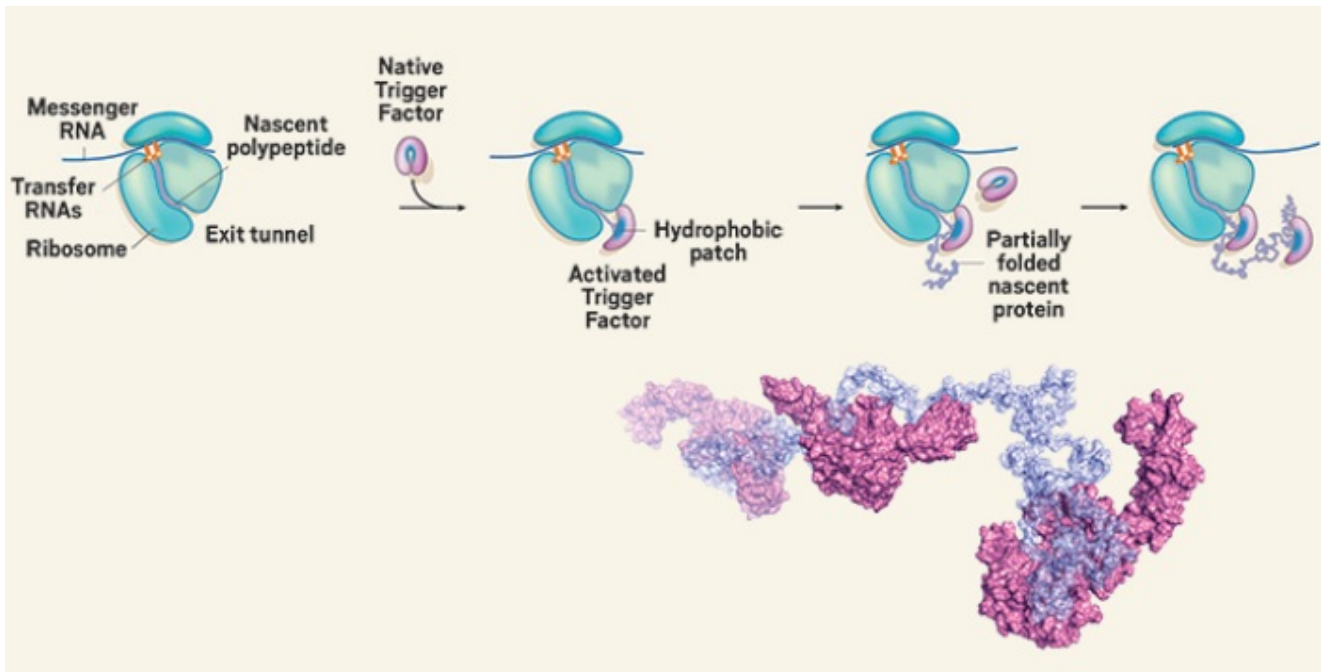
The secrets of chaperones

In the intervening decades since chaperones were first reported, scientists have discovered that protein-folding chaperones come in many shapes and sizes. They vary from a few kilodaltons to enormous megadalton machines, says Charalampos Kalodimos, who uses nuclear magnetic resonance spectroscopy to study protein folding at St. Jude Children's Research Hospital. Even the ribosome can be considered a chaperone because its exit channel provides the first secluded environment for a nascent polypeptide chain to fold, says Martin Gruebele, a biophysical chemist at the University of Illinois, Urbana-Champaign.



Some chaperones are physically associated with the ribosome, hovering near the exit channel, waiting expectantly for newly synthesized peptide chains to emerge. One such chaperone in bacteria, called Trigger Factor, has been likened, for

better or worse, to both a crouching dragon and a midwife. This chaperone pounces on hydrophobic sequences or delicately wraps them up, depending on your preferred analogy. Either way, it sequesters the unfolded chain from the cell's polar environment to prevent aggregation or misfolding until the entire peptide chain has emerged.



In this scheme, the chaperone Trigger Factor prevents nascent polypeptides from misfolding as they emerge from the ribosome by sequestering their hydrophobic residues from the cell's polar interior. The model at bottom right depicts a recent NMR structure of an unfolded polypeptide chain (light purple) stabilized by three Trigger Factor proteins (pink).

Credit: *Nature* (Scheme); *Science* (Protein structure)

In 2014, a team led by Kalodimos reported an NMR-based structure of Trigger Factor showing that it binds to an unfolded peptide in at least four spots where the peptide has stretches of six to 10 hydrophobic residues (*Science*, DOI: 10.1126/science.1250494). "The binding sites have a flexible local architecture that allows interaction with a large and diverse population of peptide stretches with unrelated primary sequences," the team notes in the paper.

Researchers have also discovered that some chaperones have ubiquitous roles in cells. A prime example is heat shock protein 70 (HSP70), one of the most important chaperones in a cell, says Lila Gierasch, a biochemist at the University of Massachusetts, Amherst. Although it was among the first chaperones to be discovered, HSP70 is still revealing its tricks to scientists.

Researchers have known that like Trigger Factor, HSP70 greets nascent chains coming off the ribosome, and it helps shuttle unfolded proteins from the ribosome to the mitochondria for transport inside. In times of heat or other cellular stress, HSP70 acts as a scaffold for partially unfolded proteins—it reduces or minimizes their aggregation before guiding nascent chains to another chaperone for refolding, Gierasch says. More recently, scientists learned that when unfolding in a cell gets out of control, HSP70 directs hopeless cases to a cell's janitorial systems, including the ubiquitin and autophagy pathways (*Curr. Opin. Cell Biol.* 2014, DOI: 10.1016/j.ceb.2013.12.006).

Perhaps one of the most intriguing recent discoveries about chaperones is that some of them are involved in the day-to-day functioning of a protein. About half of all kinases and receptors for the glucocorticoid and other steroid hormones function this way, says David Agard, a structural biologist at the University of California, San Francisco. Last year, using cryo-electron microscopy, Agard's team reported the first atomic-resolution structure of a chaperone called heat shock protein 90 (HSP90) in action on a protein folding client (*Science* 2016, DOI: 10.1126/science.aaf5023).

"HSP90 gets to work very late in the folding process, when a protein has already acquired substantial structure," Agard says. In fact, the chaperone regulates the function of many kinases, holding these clients in a fleeting unfolded conformation to inactivate them. That is, until a different chaperone—likely an HSP70 protein—can get the kinase back into a conformation where it can do its job again.

"It's clear that chaperone machinery is doing a lot more than enabling folding or rescuing proteins from stressful situations," Agard explains. "There's a layer of functional regulation enhanced by chaperone folding." Unlike many other chaperones, HSP90 does not recognize the large, exposed hydrophobic sections of massively unfolded proteins, he says. Instead, HSP90 likely recognizes relatively rare structural conformations.

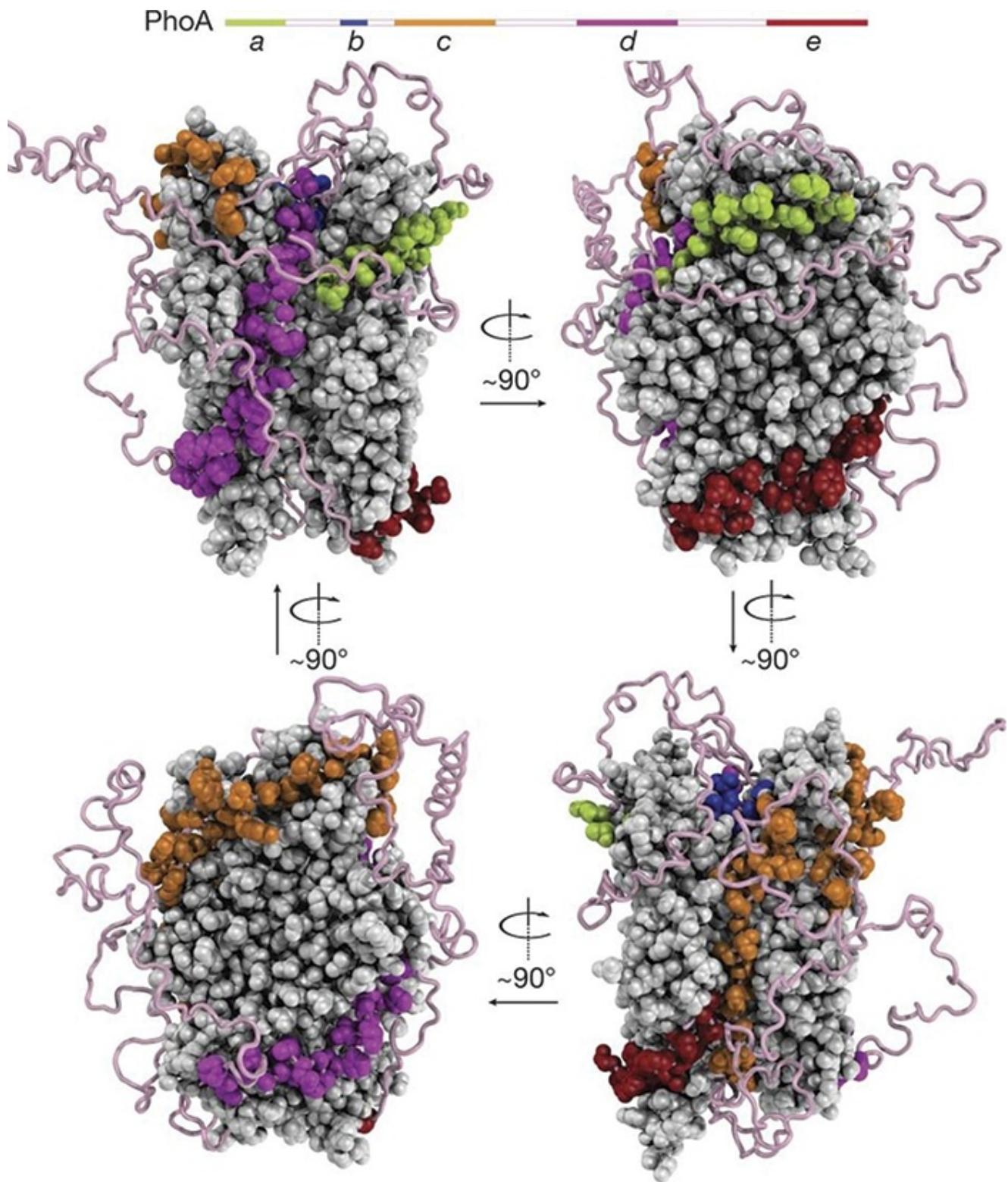
Scientists now know that in addition to regulating important cellular processes, chaperones such as HSP90 probably play a profound role in our evolution. The logic goes as follows: For evolution to happen, mutations in proteins are necessary. But mutations often destabilize a protein. One of the great pioneers of the chaperone field, the late Susan Lee Lindquist at Massachusetts Institute of Technology, proposed that chaperones enable evolution by stabilizing mutations while the cell has a chance to see if the mutation is beneficial. Recently, Stanford's Frydman and colleagues



used modeling to further support Lindquist's theory, showing that chaperones "promote protein evolvability by buffering the destabilizing effect of mutations" (*PLOS Comput. Biol.* 2014, DOI: 10.1371/journal.pcbi.1003674).

Proteostasis buzz

This deep involvement of protein chaperones in most aspects of a protein's life has elevated the family from being merely midwives at a protein's birth to being intricate players in a protein's entire life cycle—what many in the protein science field call proteostasis, a portmanteau of the words protein and homeostasis. Since proteostasis was first introduced in a 2008 review article in *Science*, more than 3,000 papers have used the term (DOI: 10.1126/science.1141448). Most people in the protein-folding field have their own pet definition for proteostasis, but Wikipedia's entry is perhaps most straightforward: "Proteostasis ... is the concept that there are competing and integrated biological pathways within cells that control the biogenesis, folding, trafficking and degradation of proteins present within and outside the cell."



The chaperone SecB1 wraps an unfolded protein around itself to prevent hydrophobic sequences from aggregating in crowded cells.

Credit: *Nature*

As you can imagine, the proteostasis network in a cell is complex, and chaperones do not work alone to manage it, Frydman says. This also means proteostasis can become dysfunctional in a number of ways, leading to a variety of diseases, including cancer and neurodegeneration, and aging.

Consider the controversy about whether misfolded protein aggregates called amyloid fibrils cause Alzheimer's disease or whether they are just a downstream consequence of another toxic event. Either way you look at it, the proteostasis network is out of whack. "Proteostasis is all about the quality of proteins" and keeping the proteome robust, says Northwestern University's Rick Morimoto, one of the scientists to coin the term.

Morimoto and others argue that many aging and age-related protein aggregation diseases could be the result of chaperones losing their tight control on the complex array of productive protein conformations in a cell. This view is grounded in the fact that "as people age, our cells start producing fewer chaperones," Morimoto says. In our twilight years, proteostasis systems are dysregulated and likely get overwhelmed.

Chaperones sequester misfolded proteins to localized sites in a cell, shuttling them either to other chaperones for refolding or to cellular janitors for degradation, so it's easy to imagine how a cell might be worse for wear if any of these steps are out of sync or overloaded.

Meanwhile, the mutations that produce problematically folded proteins might be overloading the proteostasis network—especially later in life when fewer chaperones are around to keep the cell's folding in check, explains Jeffery Kelly, a chemist at Scripps Research Institute California.

Molecules that kick-start or bolster the proteostasis network might be good therapeutics for age-related aggregation diseases, Kelly adds. For example, many groups are trying to find molecules that activate a janitorial process in cells called autophagy, where misfolded proteins are sent to the lysosome for degradation.

Of course, not everyone is convinced that tweaking proteostasis is a good idea. "If something is very tightly controlled, nature has done it for a reason," says Christine Queitsch, an evolutionary biologist who studies protein folding at the University of Washington. She points to studies where researchers have created transgenic plants with increased chaperone levels to make them more heat tolerant. But the plants were dwarves, she says, which is counterproductive for agriculture.

The trick might be to carefully nudge the proteostasis network instead of hitting its major players head-on, Agard says. For example, when researchers in search of new anticancer drugs have targeted HSP90, they have typically interfered with the site where adenosine triphosphate (ATP) binds to starve the protein of this chemical fuel. Agard suspects that the

reason this strategy has failed is because targeting the ATP binding pocket is too “blunt” of a strategy. “This drives all of the chaperone’s clients toward degradation—not just the cancer-causing kinases,” he says.

Agard hopes the recent atomic-resolution HSP90 structure his group solved “will give us clues on how to get selectivity for potential drugs,” he says. For example, because HSP90 relies on a protein cochaperone to help make it bind specific kinase clients, “I’m intrigued with the possibility of interfering with that interaction to get selectivity,” he says. “You want to hit cancer without doing nasty things to the rest of the chaperone’s clients.”

Although many putative drugs targeting the proteostasis network have been so far unsuccessful, Kelly and colleagues have developed a therapeutic called Tafamidis. This compound ameliorates a misfolding disease by stabilizing a target protein’s properly folded state—a strategy that many researchers developing drugs to avert protein-misfolding diseases have embraced.

The drug, currently approved by regulators in Europe and Japan, treats a rare genetic disease called familial amyloid neuropathy. In this disease, a single-point mutation in a protein called transthyretin results in misfolding and the consequent production of harmful aggregates that can cause life-threatening enlargement of the heart and irregular heartbeats. In healthy individuals, transthyretin is a team player—it travels through the blood as a tetramer, a group of four identical proteins. The disease mutation destabilizes the tetramer, causing it to break down to its four individual monomers, which in turn tend to misfold to form disease-causing aggregates. Tafamidis stabilizes the tetramer, preventing the formation of misfolded protein aggregates and slowing progression of the disease’s symptoms.

New avenues in protein folding

Even as some researchers are working to intervene in protein folding to fight disease, others are discovering new strategies that nature uses to control this most fundamental of processes. Case in point: Biochemists have long wondered why there is redundancy in the three-nucleotide sequences called codons in genetic blueprints that correspond to particular amino acids in a protein. Namely, several codons in messenger RNA all result in the same amino acid being added to a growing protein chain. Mounting evidence now suggests some redundant codons result in fast protein synthesis and others slow it down. These so-called “slow” codons often occur on mRNA in between regions that code for segments of a protein that need to fold independently. The pausing during protein synthesis that these codons enable “is like a stutter that allows individual regions to fold,” explains the University of California, Berkeley’s Susan Marqusee, a biochemist who studies protein folding.

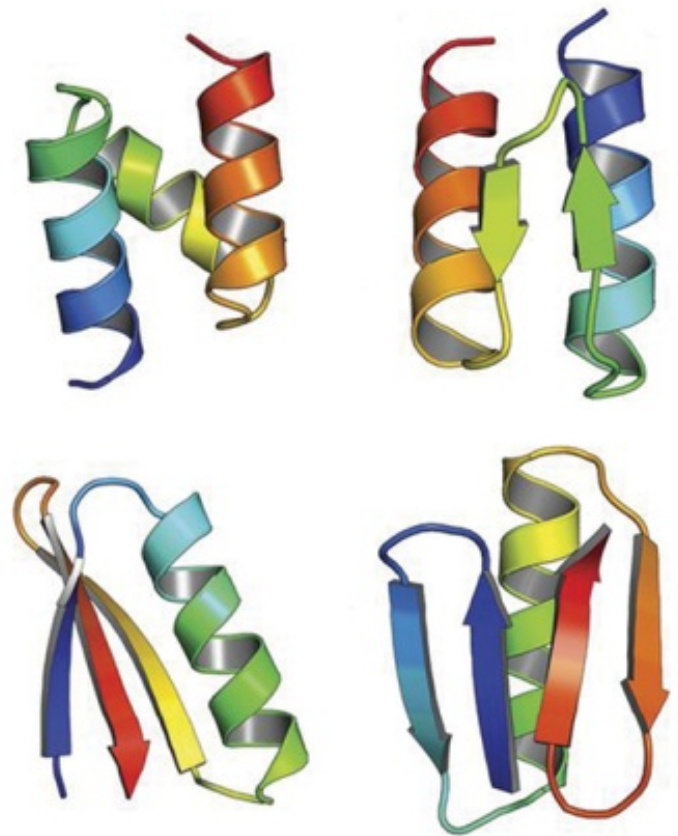
And then there's new, fundamental protein-folding work that recalls Anfinsen's "sequence determines structure" dogma and also strides toward the future of synthetic biology.

[+]Enlarge

Researchers have designed four new synthetic protein folds never-before-evolved by nature.

Credit: *Science*/AAAS

Although researchers have long understood why proteins are driven to fold—to bury hydrophobic side chains—they have not been particularly successful at predicting how a protein sequence will fold nor at designing entirely new sequences that stably collapse into 3-D structures, explains the University of Washington's David Baker. After decades of work in the area, his group reported success on both fronts this year.



In January, Baker and colleagues predicted the 3-D conformations of 12% of the protein families that still had unknown structure, thanks to a marriage of machine learning, big data from microbiome projects, Baker's Rosetta protein folding algorithm, and a distributed network of volunteer computers. The 600 newly determined protein family structures included 100 protein folds not found in the Protein Data Bank and 200 membrane proteins (*Science* 2017, DOI: 10.1126/science.aah4043). To prove their predictions were accurate, Baker and his team initially deposited some of their forecasted structures into public databases. In the months afterward, structural biologists serendipitously solved—and confirmed—the 3-D conformation of six of these unknown protein families.

Then in July, the team reported that it had designed protein sequences that adopted four protein topologies never before seen in nature. Although the proteins were small—shorter than 50 amino acids long—the feat was a milestone for synthetic biology. Because many of the designed proteins "are more stable than any comparably sized monomeric proteins in the [Protein Data Base]," the sturdy scaffolds might be used in a variety of biotechnology and medical applications, the team notes in its *Science* paper (2017, DOI: 10.1126/science.aan0693).

To achieve the milestone, the researchers used computer protein design and high-throughput synthesis to produce 15,000 miniproteins, of which 2,500 folded stably (without the help of chaperones). Then they analyzed the characteristics of the stably folded proteins to quantify what success looked like. "If you just make random sequences, it's rare that they would fold," Baker says. The team found that stable proteins need to bury a surface area of 30 Å² for each residue of buried hydrocarbon, a criterion that paves the way for tactical design of additional novel folds.

Anfinsen would have surely been fascinated by the new work. Before his death in 1995, the biochemist exchanged letters with chaperone expert R. John Ellis at Warwick University, expressing his interest and excitement about the then-nascent chaperone field and his desire to include this new research in classes he taught to students. If he were still alive today, Anfinsen would marvel at the way the field has developed—the protein-folding pioneer might even be inspired to unfold the sails of his boat and go for a reflective glide.

Sign up for C&EN's must-read weekly newsletter

Chemical & Engineering News

ISSN 0009-2347

Copyright © 2019 American Chemical

Society

Copyright © 2019 American Chemical Society. All Rights Reserved.