

NEWS

Problem Solved*

(*sort of)

Researchers have toiled for decades to understand how floppy chains of amino acids fold into functional proteins. Learning many of those rules has brought them to the verge of being able to make predictions about proteins they haven't even discovered

IN 1961, CHRISTIAN ANFINSEN, A BIOCHEMIST at the U.S. National Institutes of Health, saw something that continues to perplex and inspire researchers to this day. Anfinsen was studying an RNA-chewing protein called ribonuclease (RNase). Like all proteins, RNase is made from a long string of building blocks called amino acids that fold up into a particular three-dimensional (3D) shape to give RNase its chemical abilities.

Anfinsen raised the temperature of his protein, causing it to unravel into a spaghetti-like string. When he cooled it back down again, the protein automatically refolded itself into its normal 3D shape. The implication: Proteins aren't folded by some external cellular machine. Rather, the subtle chemical push and pull between amino acids tugs proteins into their 3D shapes. But how? Anfinsen's insights helped earn him a share of the 1972 Nobel Prize in chem-

istry—and laid the foundation for one of biology's grand challenges.

With an astronomical number of ways those chains of amino acids can potentially fold up, solving that challenge has long seemed beyond hope. But now many experts agree that key questions have been answered. Some even assert that the most daunting part of the problem—predicting the structure of unknown proteins—is now within reach, thanks to the inexorable improvements in computers and computer networks. “What was called the protein-folding problem 20 years ago is solved,” says Peter Wolynes, a chemist and protein-folding expert at the University of California, San Diego.

Most researchers won't go quite that far. David Baker of the University of Washington, Seattle, believes that such notions are “dangerous” and could undermine interest in the field. But all agree

Big picture. Experimental data helped computer modelers nail down the structure of the nuclear pore complex.

that long-standing obstacles are beginning to fall. “The field has made huge progress,” says Ken Dill, a biophysicist at the University of California, San Francisco (UCSF).

The work has huge implications for medicine. Misfolded proteins lie at the heart of numerous diseases, including Alzheimer's and Creutzfeldt-Jakob disease. Understanding how proteins fold could shed light on why they sometimes misfold and could suggest ways to intervene. Accurate protein models can also lead to the development of more-conventional drugs that block or enhance the work of key proteins in the body.

Twin challenges

Today, the protein-folding challenge boils down to two separate but related questions. First, what general rules govern how, and how quickly, proteins fold? Second, can researchers predict the 3D shape that an unknown protein will adopt?

These simple questions open the door to a world of mind-boggling complexity. Because two neighboring amino acids can bind to each other at any one of three different angles, a simple protein with 100 amino acids can fold in 3^{200} different ways. Somehow, a folding protein sorts through all those possibilities to find the correct, or “native,” conformation.

And it's not by trial and error. Even if a folding protein could try out one different conformation every quadrillionth of a second, it would still take 10^{80} seconds—60 orders of magnitude longer than the age of the universe—to find the right solution. Because most proteins fold in milliseconds to seconds, something else is clearly going on.

Molecular biologist Cyrus Levinthal pointed out this paradox in 1969 and concluded that proteins don't follow a random set of wiggles to find their native conformation. But figuring out what path they do take hasn't been easy. Early on, researchers largely assumed that a protein follows a set path as it folds, wending its way through certain intermediate states as it coils up into its particular arrangement of helices and sheets and so on. But in the mid-1980s, Wolynes and Dill suggested that proteins, rather than working like origami—in which one fold in a sheet of paper leads to the next until the final shape is reached—actually break the problem down into many pieces. Local clusters, each containing a handful of amino acids, initially pull

CREDIT: F. ALBER ET AL., NATURE (29 NOVEMBER 2007)

Downloaded from www.sciencemag.org on August 11, 2008

and repel one another. As these clusters begin to fold, neighboring clusters come together, and so on.

To explain this process, Wolynes and Joseph Bryngelson, both then at the University of Illinois, Urbana-Champaign (UIUC), suggested that as proteins fold they follow an energy landscape, akin to water flowing downhill. The result is a more energetically stable arrangement. Dill pushed the same notion and later came up with an image of an energy funnel, showing how proteins can follow many possible different pathways to their native conformation at the bottom of the funnel (see top figure, p. 786).

Dill's funnel explained how proteins could avoid Levinthal's paradox and fold quickly. It also led to a testable hypothesis. The time it takes for a protein to fold depends on the energetic obstacles in its path. With fewer amino acids, most small proteins fold more quickly than larger ones that can get caught up on energetic plateaus before finding another downhill run. If the simpler-is-faster rule were true, researchers realized, then it should be possible to make some proteins fold faster by mutating the amino acids that slowed things down.

In 2007, Martin Gruebele, a chemist at UIUC, and colleagues set a record for such streamlining when they tracked the folding of both native and mutated conformations of a protein called λ repressor. After making their proteins, Gruebele's team cooled them down to unravel them and then zapped them with a laser. The nanosecond burst of heat caused the proteins to begin refolding, which the Illinois

team could watch by tracking their fluorescence. Certain mutations enabled the protein to fold in just 2.5 microseconds, 200 times faster than the natural protein does.

Such mutations, however, often disrupt the protein's chemical function. The reason, Gruebele says, is that in most proteins, hydrophobic amino acids tend to shield themselves from interacting with water—an energetically favorable arrangement—by nestling in the center. Charged, or polar, amino acids by contrast tend to stick out into the water that surrounds proteins. These groups tend to be more chemically active and commonly play key roles in the protein's reactive center. And as proteins wiggle into shape, the polar interactions are often slower to make their adjustments. Changing some of those amino acids speeds things up but alters the chemistry. "They evolved to do a job, not to fold fast," Gruebele says.

The next step

Thanks to this and many other related studies, Andrej Sali, a biochemist and protein modeler at UCSF, says that most protein-folding experts now believe they understand the general rules for how proteins fold and how they fold so quickly. But the next step—predicting how a specific set of amino acids will fold—remains a much bigger challenge. "We have not been able to transfer our conceptual understanding into [a] prediction of how specific amino acid

sequences will fold," Sali says.

There have been some successes. Every 2 years since 1994, for example, computer modelers have vied to determine the 3D structure of an unknown amino acid sequence in a competition known as the Critical Assessment of Techniques for Protein Structure Prediction. At first, only about half of the modelers

came close to predicting the structures of moderately difficult target proteins (see figure, below). In 2006, however, 80% did. Most of the predictions still can't match

the resolution of an x-ray crystal structure, which can pinpoint the position of atoms down to a couple of tenths of a nanometer. Nevertheless, Dill says, "it's gotten to the point where if you have a reasonably small protein, you can get a good structure."

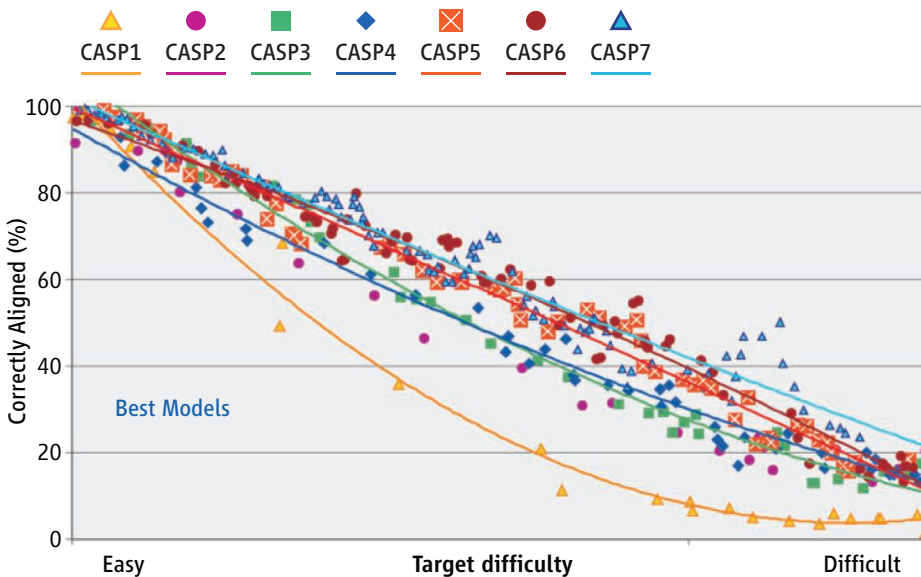
Computer models for predicting protein structures come in many varieties. But they generally fall in two camps: Ab initio models start by specifying the attractive and repulsive forces on each point and then calculate a structure by cranking through the calculations until they find the lowest energy state. Homology models, by contrast, make their predictions by comparing the target protein with proteins with closely related sequences whose structures are already known. More powerful computers and search algorithms have recently given ab initio models a major boost, but Dill says homology models still hold the upper hand.

The accuracy of those predictions depends on a model's resolution (whether it aims to map the position of individual atoms or just of individual amino acids) and how thoroughly it samples the energy landscape to find the lowest energy configuration. As a result, modelers face a tradeoff. Increase the resolution by mapping out all the atoms, and you limit the amount of sampling a computer, or network, is able to carry out. Increase the sampling rate, and you limit the resolution.

To simplify the computations, some researchers bolster their computer models with experimental data that narrow the search for the protein's lowest energy configuration. In the 25 March issue of the *Proceedings of the National Academy of Sciences*, for example, Baker and 15 colleagues in the United States and Canada described a new technique for using nuclear magnetic resonance (NMR) data to boost the speed and accuracy of protein simulations with atomic resolution (see bottom figure, p. 786).

"What was called the protein-folding problem 20 years ago is solved."

—PETER WOLYNES, UCSD



Steady rise. Computer modelers have slowly but steadily improved the accuracy of the protein-folding models.

Challenges in Theoretical Chemistry

Like x-ray crystallography, NMR has long been used to map proteins in atomic detail. But the technique typically works only with small proteins. It usually requires taking at least two separate types of NMR data, an easily acquired data set known as the chemical shifts and a much slower technique called the Nuclear Overhauser effect (NOE).

In their new work, Baker and his many colleagues dispensed with NOE data and fed chemical shift data for 16 proteins into a computer prediction program known as ROSETTA. The resulting atomic-scale models closely resembled structures previously solved by either NMR or x-ray crystallography. As a control experiment, the researchers also solved the structures for nine proteins for which the NMR and x-ray structures were still being worked out. Those results, too, ended up in tight agreement. "Our joint hope is by combining our methods with NMR data we can work to larger and larger proteins," Baker says. He and his colleagues are adding other types of experimental data, such as data from lower resolution electron cryo-microscopy and from mutation experiments that highlight amino acids sitting next to each other in the folded protein.

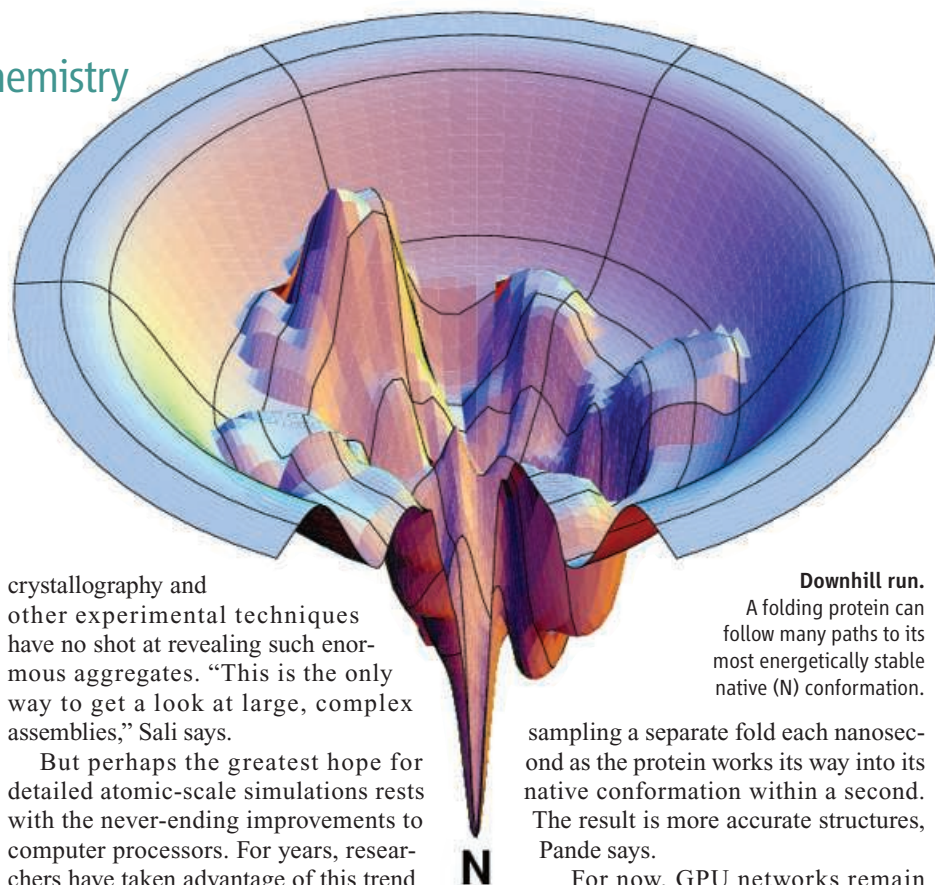
The multipronged approach is paying off for other researchers as well. In the 29 November 2007 issue of *Nature*, Sali and colleagues in the United States, the Netherlands, and Germany predicted the structure of the nuclear pore complex, an assembly of 456 separate proteins, by integrating 10 different biophysics and proteomic data sets into a model. The resolution was low by the standards of x-ray crystallography. However,

crystallography and other experimental techniques have no shot at revealing such enormous aggregates. "This is the only way to get a look at large, complex assemblies," Sali says.

But perhaps the greatest hope for detailed atomic-scale simulations rests with the never-ending improvements to computer processors. For years, researchers have taken advantage of this trend by joining thousands of processors together to build powerful supercomputers, such as IBM's Blue Gene machines, that have long excelled at protein-folding simulations.

More recently, Baker, Vijay Pande of Stanford University in Palo Alto, California, and other researchers have created distributed supercomputers. They rely on computer users from around the world to download software that lends the computer's central processing unit (CPU) to folding calculations when the computer is not in use. Today, Pande's Folding@home network counts more than 250,000 active participants, and Baker's Rosetta@home totals more than 300,000.

Pande says these networks have sped up protein predictions 100,000-fold. Other recent improvements to search algorithms have boosted speeds another 1000 times. And most recently, distributed networks have begun turning to ultrafast graphics processors known as GPUs to gain another 100- to 1000-fold advantage. Taken together, these improvements now allow distributed networks to follow a protein through a billion gyrations,



Downhill run.

A folding protein can follow many paths to its most energetically stable native (N) conformation.

sampling a separate fold each nanosecond as the protein works its way into its native conformation within a second. The result is more accurate structures, Pande says.

For now, GPU networks remain smaller than their CPU counterparts. Pande's GPU network, for example, counts only 10,000 participants. But these are likely to grow quickly as modelers gain experience with writing code to take advantage of their talents. "The sampling part of the problem will soon be an obsolete issue," Pande predicts.

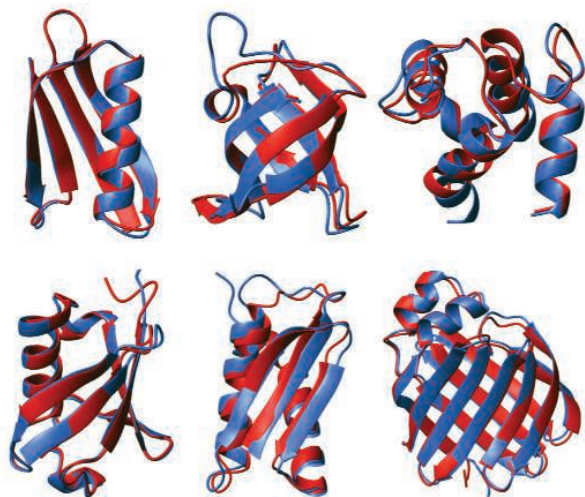
So is the protein-folding problem solved? Not quite. But Pande and others say researchers are getting tantalizingly close. "The practical issue in protein folding is now an engineering issue," says Sali. Pande agrees: "We're on the verge of being able to tackle the complete problem."

Crossing that line won't solve all the problems in protein folding. Drug designers in particular have a tough challenge, because they often need to know the position of atoms in an active site of a protein at an ultrahigh resolution in order to design drugs to block or enhance the protein's work.

Still, Dill calls the progress in his field a revolution. But he says most scientists haven't noticed because it has occurred so slowly. "Progress in science comes out as news when there are big steps," he says. "In protein folding, there have been a huge number of folks involved and lots of incremental steps. And that doesn't usually make news."

Maybe not. But now it's set to make a difference for scientists, physicians, and their patients.

—ROBERT F. SERVICE



Tight fit. Adding data from nuclear magnetic resonance experiments improves the accuracy of computer models of how proteins fold.