Up Front

Science Behind the Screens

Tapping the power of distributed computing, biophysicists push the rate of their protein-folding simulations toward warp speed.

Although its name may evoke visions of washcloths and towels, the Folding@Home project is in fact serious science. Project volunteers offer the power of their home computers to help researchers study protein folding. In the process, they help speed the collection of data that could lead to answers about disorders such as Alzheimer's disease.

The really great thing about the project is the scale, says HHMI predoctoral fellow Christopher Snow. "If I have an interesting idea, I can come in over the weekend and get 10,000 computers working on it!"

Snow, a Ph.D. candidate in biophysics at Stanford University, can tap that kind of computing capacity courtesy of individuals around the world who loan him unused processing power online. Their common goal is to help scientists figure out how protein molecules fold into their final shapes—a challenge that ranks as one of the toughest and most important in biology.

Properly folded proteins are nature's "nanomachines," serving as molecular gatekeepers and cellular scaffolding, among other functions. Cells couldn't work without them. Indeed, life would not be possible without them.

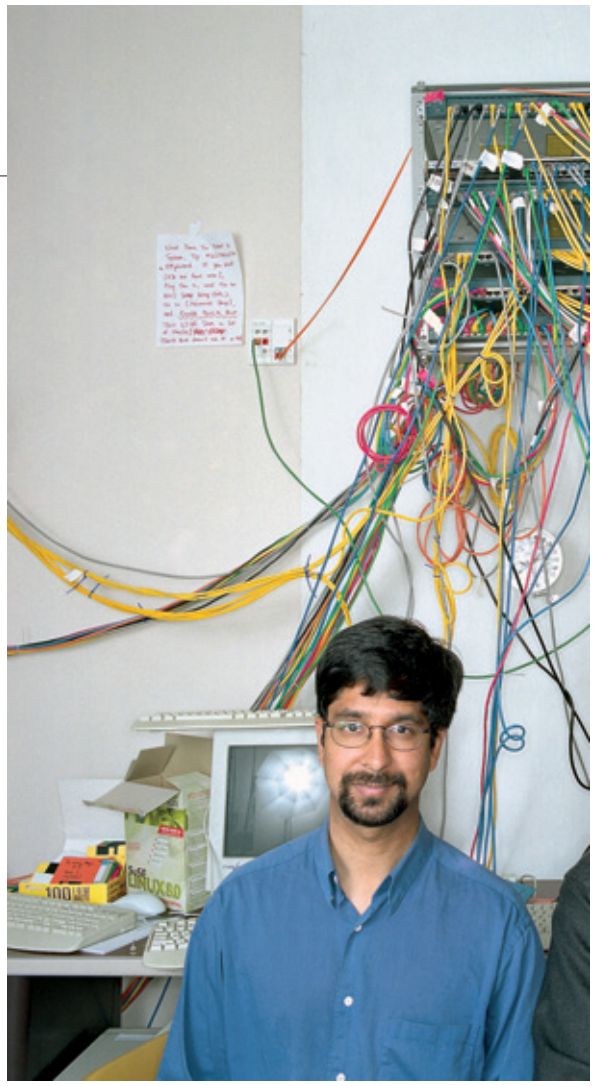
By contrast, improperly folded proteins are like tiny time bombs. The mutated molecules not only fail to perform their jobs in the cell, but also can form sticky, insoluble clots that arrest or even kill the cell. A wide variety of disorders—including Huntington's, Parkinson's, and Alzheimer's diseases—are thought to derive from improper protein folding. Folding@Home research could therefore help scientists discover the origins of, and possible treatments for, this rogues' gallery of maladies.

Ever-Morphing Molecule

Folding@Home software, which is free for the downloading, currently runs on an average of almost 100,000 personal computers around the world every day, and Snow could potentially draw on any or all of them. Not that he'd be bothering their owners in the slightest: From the user's perspective, Folding@Home is just a stark-looking screen saver in which a molecule of some sort keeps reappearing from different points of view. Like any good screen saver, Folding@Home vanishes at the first keystroke; it pops up again only after the machine has been sitting idle for a while.

The magic is in what's happening behind the scenes. This screen saver is actually a small but sophisticated molecular-simulation package that takes advantage of the computer's unused processing power. The image on the screen is actually a simulated protein molecule, temporarily frozen at one instant in time while the computer calculates what it will look like an instant later. The results are automatically transmitted back to the researchers at Stanford each time the PC connects to the Internet. And in the meantime, the user will get quite a show, gain some insight and enjoy the good feeling of contributing to scientific knowledge.

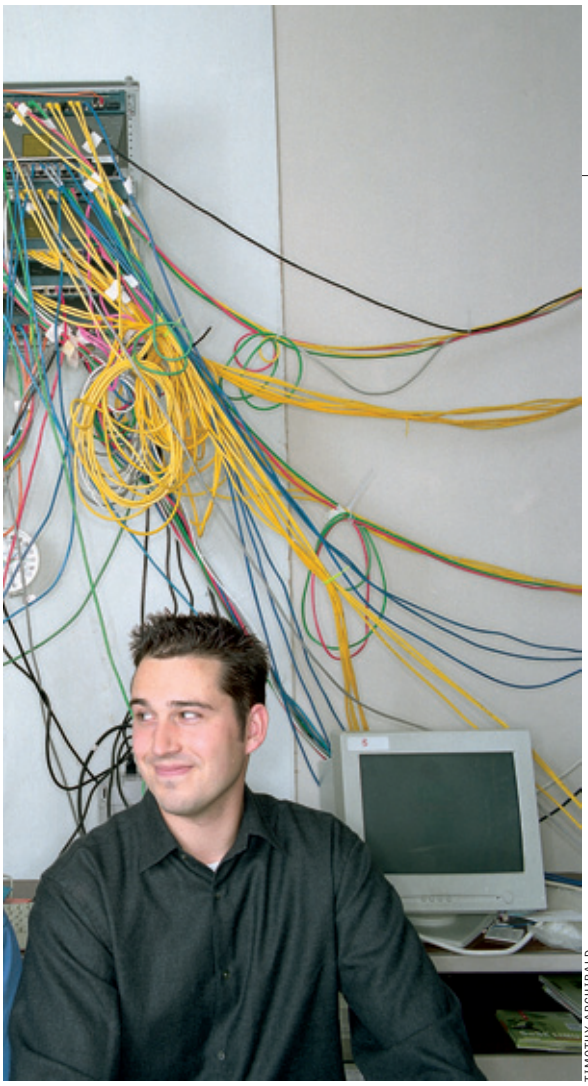
This is what draws people into the project, says Vijay Pande, the Stanford biophysicist who launched the Folding@Home project three years ago. When you look at who's downloading the software, he says, "it varies from people with an interest in computers, to people interested in biology, to people interested in helping fight disease"—not to mention a fair number of high school students, whose teachers find that Folding@Home is a unique way to get their classes excited about science. The users have also formed dozens of teams, with names such as "Dutch Power Cows" (706 members) or "Overclockers Australia" (3,184 members), and keep tabs on who has contributed the most processing time.



To help analyze protein folding, Vijay Pande (left) and Christopher Snow can borrow the power of tens of thousands of computers worldwide.

"This is like having a whole new kind of 'funding agency' for research—namely, the general public donating its computers," says Pande. "When you factor in the maintenance they're doing, the operating system upgrades, and so on, that's a gigantic resource!" It's also a godsend, he adds, because the scientists who work on this problem need all the help they can get.

When a protein molecule is constructed inside the cell, Pande explains, it starts out as a simple chain of amino acids: smaller compounds that are linked together like so many beads on a string. The sequence of amino acids—there are 20 different kinds—determines the type of protein. But even as the



TIMOTHY ARCHIBALD

protein is forming, subtle interatomic forces cause the chain to start tangling like a nanoscale telephone cord. Every chemical bond in the molecule is involved—potentially thousands of them—each one stretching, twisting and bending until at last the chain achieves stability. Indeed, the complexity of the folding process goes far beyond anything that can be measured in the laboratory, says Pande; the only way to understand it in detail is by computer simulation.

But simulation has its own problems. Proteins typically take about a millisecond (one thousandth of a second) to fold, yet this process is so complex that the fastest PC in existence can simulate only a nanosecond (one billionth of a second) in a day. "That's a million times slower," says Pande. "So it would take you a million days, or roughly 3,000 years, to finish—and then you'd only see one result." By being artful about the programming and the analysis, however, one could get away with "only" 10,000 days, he notes.

That's still 30 years. But if researchers can't wait that long, Pande says, they can find the capacity to do 10,000 simulations at once by breaking up the calculation into that number of pieces. Each of 10,000 computers may then work independently and simultaneously on one piece. So the only problem remaining, he recalls, was "Where were we going to get 10,000 computers?" Forget about buying them; a computer farm that large would cost $10 million or more and be a nightmare to administer. Pande and his colleagues would have to borrow them.

Previous distributed-computing projects, such as distributed.net and SETI@home, showed them how. Distributed.net was the first major Internet distributed-computing project, and its goal was factoring large numbers, important in cryptography and code breaking. SETI@home had already enticed some half-million people around the world to run screen savers that processed radio signals for the search for extraterrestrial intelligence. "So in October 2000," says Pande, "my group started Folding@Home." With some strategic help and advice from Adam Beberg (founder of distributed.net and now with the Cosm Project), Pande's team developed the software for Folding@Home. They found their first volunteers from Cosm's mailing list. From there, Folding@Home grew rapidly as a result of several reports in the press as well as by word of mouth. In the relatively brief period since then, more than 400,000 processors have contributed at least some time to the project, Pande reports.

The Speed of Simulation

Snow joined the group as a graduate student just a few months after the project began, and Pande gives him considerable credit for nursing the software through its infancy. "Chris did a lot to shape things up and make it more solid," he says. Snow then went on to help validate the approach by running multiple simulations of a small artificial protein, BBA5, that could easily be compared with experimental results.

The results, which agreed quite closely with those obtained in laboratory experiments by Martin Gruebele and his students at the University of Illinois at Urbana-Champaign, were published in the November 2002 issue of *Nature*. "What they did was almost unheard of," says a deeply impressed Gruebele. Until then, the longest protein simulation on record covered only about one microsecond (one millionth of a second), he explains. "But Vijay and his group just blew away that timescale completely. They got 700 microseconds"—long enough to simulate BBA5's folding all the way.

Of course, as Pande and everyone else on the project takes care to emphasize, they're still far away from simulating most real proteins—and even farther from finding cures for diseases. The BBA5 chain is only 23 amino acids long; naturally occurring proteins often have hundreds or even thousands of amino acids. This computational challenge dwarfs Folding@Home's resources.

Still, Folding@Home needs to grow in processing power only by another factor of 50 to 100 to be robust enough to take on the big molecules. Given computer science's famous Moore's Law, which states that machine processing power doubles every 18 months or so, "that's not a factor that frightens you," Gruebele points out. Nor is the Folding@Home group waiting around.

Snow, for example, is already planning to take on the amyloid precursor protein, a protein that seems to be critical in Alzheimer's disease. This protein misfolds and then forms aggregates; these aggregates are believed to lead to Alzheimer's disease. "Understanding the structure of these aggregates is fundamental to understanding Alzheimer's disease," says Pande, "and that's what we'd like to simulate."

In the meantime, graduate student Bojan Zagrovic, another HHMI fellow working in Pande's group, has been looking at the structure of proteins before they are fully folded. "For a long time this was ignored," says Zagrovic. "But thanks to the huge sampling from Folding@Home, our research seems to indicate that the unfolded state is not random." As the simulated molecules begin to bend and twist, he says, "you find that they very quickly crumple into all these weird-looking shapes"— like spaghetti as it softens in boiling water. "And if you average all that crumpled spaghetti, you find that the average shape is already very close to the final folded state!" That's a totally unexpected result, he says, and no one quite knows what to make of it. "But that" he says, "will be my thesis project."

—M. MITCHELL WALDROP

» For more information, see *folding.stanford.edu.*