A protein is a string of amino acids, which you can think of, roughly, as a necklace of multi-colored beads. A protein, however, twists and folds into a shape that’s less like a necklace than a pretzel, and each protein has its own unique pretzel shape, determined by which beads are on the necklace and in what sequence.

Scientists know that sometimes if you switch one bead for another, the pretzel doesn’t change much, but other times—depending on which bead you switch or which new one you substitute—the pretzel may radically alter its shape. When a gene mutation causes this to happen in people, it may—depending on the protein—lead to a debilitating disease. No wonder, then, that scientists want to decipher the underlying code of relationships between amino acids and protein shape, called the protein-folding problem. “It’s like saying you have a sentence and it has a meaning,” says computational chemist Carlos Simmerling. “How many words can you change before it doesn’t have that meaning anymore? How many of the amino acids in the protein can you change before it doesn’t fold properly and you have a disease?”

Simmerling, of the State University of New York at Stony Brook, and graduate student Melinda Layten used LeMieux, PSC’s terascale system, to see what happens when you switch beads on the necklace. With LeMieux, he’s been asking these questions with a special protein called Trp-cage, a short necklace of only 20 amino acids—compared to hundreds for most proteins. Because of its small size, Trp-cage’s folding is relatively simple and presents a model case for analysis.

Working in close collaboration with a laboratory team led by Niels Andersen at the University of Washington, Simmerling made efficient use of up to 1,000 of LeMieux’s processors to simulate Trp-cage, all of its atoms—first in its native state, then with several carefully chosen variations. He’s found, dramatically, that one switched amino acid transforms the pretzel to a floppy noodle, but another subtle switch changes it back to a pretzel. His simulations add detail and complexity to an emerging picture of sensitive relations between amino acids and structure in this special minimalist protein.
LESS IS MORE

Trp-cage, significantly, is the smallest protein known that has a stable, folded shape. In 2002, starting with a longer protein from Gila monster saliva, Andersen and his team created Trp-cage in their laboratory. Their work was lauded as among the biochemistry highlights of the year.

“It provides a model system,” says Simmerling, “for close collaboration between experiment and simulation on the same sequence, and until Trp-cage there was nothing like this. The proteins that we could simulate were too small to be stable, and the systems that were stable experimentally were too big to simulate. It’s an important meeting place for experiment and theory.”

Simmerling is one among a team of computational biochemists who have developed a widely used software package, called AMBER, that employs a method called molecular dynamics (MD) to simulate proteins and DNA. MD calculates the forces that act among all the atoms in the molecule and tracks their movement over time.

Before Andersen’s group released their experimental findings of Trp-cage’s structure, Simmerling used AMBER to accurately predict it. Starting with only the amino-acid sequence, his simulations arrived at a structure in excellent agreement with the configuration, as determined by NMR methods, that Andersen’s group subsequently published. “We demonstrated,” says Simmerling, “that MD simulations have come a long way, and are at a point where accurate structure prediction by simulation may soon be routine enough to contribute significantly to our understanding of folding.”

This simulation also suggests structural detail that goes beyond the experimentally determined shape. Based on these details, Andersen’s group is working to further analyze and refine the structural picture.

Carlos Simmerling, The State University of New York at Stony Brook.
RIDGES AND VALLEYS OF PROTEIN ENERGY

This graphic represents the free-energy landscape of Trp-cage. Color (increasing from dark blue to red) corresponds to altitude in the landscape. Stable proteins tend to form in structures that correspond to the low-energy valleys.

FOLDED, UNFOLDED, REFOLDED

These graphics from the simulation show a simplified representation of Trp-cage with the protein backbone (yellow ribbon) and selected amino acids. In the native, folded structure (Gly), spheres of cyan (carbon) and white (hydrogen) represent side-chains of tryptophan (blue, nitrogen) within a "cage" of two prolines. This characteristic fold gives Trp-cage its name. The small glycine (green) is at a bend in the fold.

In one of many structures (Lala) adopted by Trp-cage after alanine (purple side chain) replaces the glycine, the tryptophan no longer packs into the cage and the protein remains floppy. A slight change to the alanine side-chain permits the tryptophan to pack in the cage, forming a structure (Dala) similar to native Trp-cage.
To predict Trp-cage’s shape, Simmerling relied on a well-tested axiom of molecular structure. A molecule tends to move toward being in a shape in which the atoms expend the least possible energy to maintain structure. Although Simmerling’s structure-prediction simulation was remarkably successful, proteins in the real world aren’t represented solely by their low-energy state.

“There’s a native lowest energy structure,” explains Simmerling, “and at the same time, there may also be non-native, unfolded structures. We need to know not just the native structure, but we want to know its relative probability. Is it native 30 percent of the time? Or 90 percent? All we can say from the first simulation is this structure is best, but we don’t know what that means in terms of real stability.”

To produce this information on relative probabilities, Simmerling employed a method called “replica exchange,” a relatively recent innovation in MD simulation that he implemented as part of AMBER. In this approach, designed to exploit a massively parallel system such as LeMieux, many separate simulations of a single protein run at the same time on different processors. The processors exchange information to arrive, eventually, at a picture of the protein’s “energy landscape”—a map of the relation between possible shapes and their likelihood.

Using LeMieux, Simmerling carried out replica-exchange simulations of Trp-cage, showing that it’s in its native, folded state 90 percent of the time, a very stable structure. In further studies, using up to 1,000 LeMieux processors (at 90 percent parallel efficiency), he simulated three altered versions of the protein—each with a single change of amino acid—which the Andersen group also looked at experimentally. One of these changes produced a dramatic result in the stability of the folded structure.

Among amino acids, glycine is the smallest, and is distinctive in having no attached chemical group, called a side chain. For this reason, glycine provides structural flexibility and often appears at a tight turn in a protein’s fold. In experiments, changing a Trp-cage glycine to alanine—an amino acid only slightly larger—prevented folding. “We can take something that’s over 90 percent folded,” says Simmerling, “make one small change, and it won’t fold anymore. Alanine has only a small side chain—a single methyl group (CH3)—and yet it has a huge effect.”

Somewhat differently from the experiments, the replica-exchange simulations showed that with alanine the folded structure still exists, but it’s highly unstable. Spurred by these simulations, further experiments found that this altered Trp-cage retains some areas of structure.

Based on analysis of these results, the researchers tried another switch. They replaced the alanine with its mirror image, called d-alanine—which flips alanine’s side-chain from one side to the other of the protein. With this slight change, simulations showed that Trp-cage regained nearly all of its folded stability. Experiments confirm that this switch restores nearly full stability.

“There’s wide interest in glycine,” says Simmerling, “and how it’s involved in folding. And there hasn’t been data, especially with respect to how it compares with d-alanine. We think this study, relying on both experiment and simulation, will be among the first to show that not only do we have this model system, Trp-cage, that’s sensitive to change, but also that simulations on computers like LeMieux are helping us to understand and begin to predict the effects that gene mutations will have on these key molecules of life.”

MORE INFORMATION:
http://www.psc.edu/science/simmerling.html