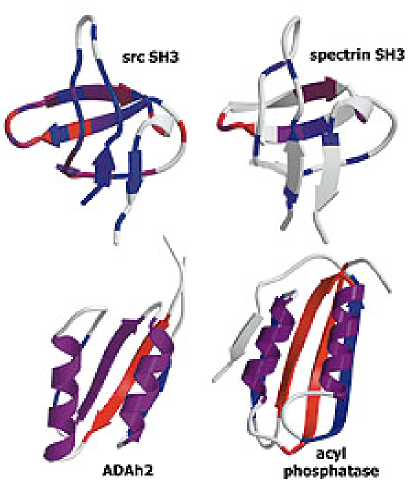
Computing Structural Biology

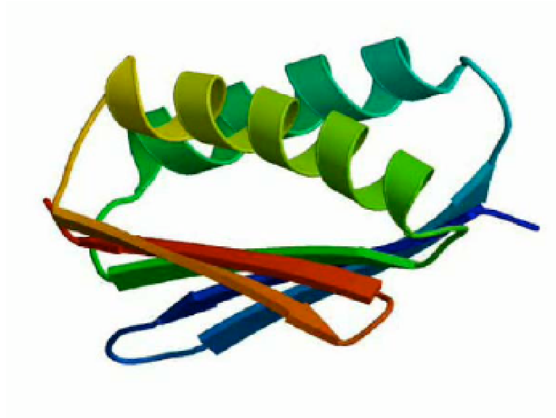
With enough computing power, Dr. David Baker believes he and other scientists could unlock the mysteries of structural biology by being able to better predict and design macromolecular structures and interactions.

"This is the era of molecular biology," said Baker, UW associate professor of biochemistry. "All of biology and medicine are being understood in molecular terms—how the body works, what goes on with diseases, how pathogens work."

Understanding this would allow scientists to design better drugs and treatments for major diseases such as AIDS and malaria and invent enzymes that could catalyze novel chemical reactions, allowing, for example, development of new methods for detoxifying poisonous substances in the environment.

To get there Baker needs more computing power and better algorithms. "In principle, we should be able to compute all of structural biology," Baker said. "If we could, it would be a fundamental test of our understanding and have huge practical relevance."





Solving three-dimensional jigsaw puzzles

Baker is a world leader in efforts to predict the three-dimensional structures of proteins from their sequences of amino acids. This problem is considered one of the great challenges of modern biology. In blind international challenges called CASP (Critical Assessment of Techniques for Protein Structure Prediction), Baker and his colleagues have consistently proven to be the best in the world.

Baker compares this challenge to solving a three-dimensional jigsaw puzzle. Proteins begin as linear chains of amino acids, but quickly fold into different three-dimensional shapes. These shapes relate to their function and determine what other proteins they can interact with and connect to.

While there are only 20 standard amino acids, a protein can be made up of hundreds of them folding into vast combinations of shapes. This presents a major computational challenge. To tackle it takes a combination of sophisticated algorithms and intensive computing power. Baker and his colleagues have produced a software program called *Rosetta*, which currently is one of the most effective tools scientists have for sorting through the possibilities to come up with a right answer.

Successful prediction and design

Baker is also working on predicting protein-protein interactions, which are essential to most biological processes. In another international blind challenge called CAPRI (Critical Assessment of Prediction of Interactions) held in December 2004, Baker proved uniquely successful at taking the structures of two proteins comprising a complex and then predicting the structure of that complex.

Designing new proteins is another key effort of Baker and his colleagues. Their article on this work recently won the 2003-2004 AAAS Newcomb Cleveland Prize for an outstanding article in the journal *Science*. Baker is now trying to design proteins that will be functional and useful—such as improved vaccines for HIV—and creating novel enzymes that will catalyze new chemical reactions. This last effort could lead to better ways to detoxify poisonous environmental substances and make the syntheses of complicated molecules easier and cleaner.

Additionally, Baker is trying to develop proteins that will interact in novel and specific ways with DNA code, to do such things as recognize and destroy DNA from pathogenic organisms or regulate how an organism develops.

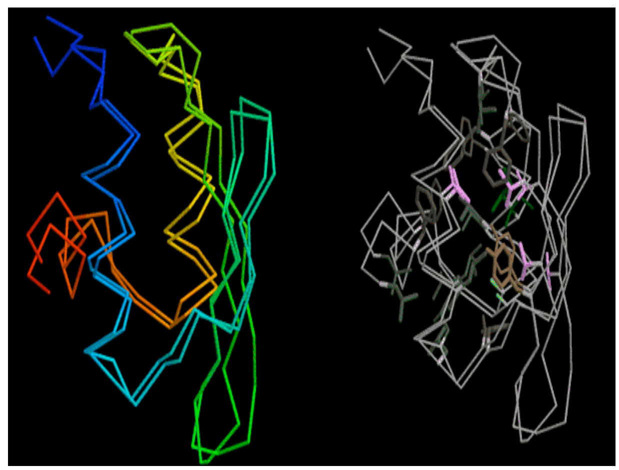
“We aren’t there yet,” Baker said. “But I can start to taste the answer. We’re getting closer.”

Access to computing power

What Baker needs now is computing power. He believes that more computing power could revolutionize the way structural biology is done. Currently, structural biology is a completely experimental science requiring millions of dollars in instrumentation.

“The revolution would be that you won’t have to do experiments to determine structural biology, you would be able to compute it.” said Baker, “It would be more elegant and cheaper.”

The catch is that in order to prove this theory, Baker needs the computing power. And he needs it on a scale found only at super-computing centers like those in San Diego and Pittsburgh. Baker needs help to bring those resources to the UW or, more likely, to get access to computing power housed elsewhere.



Future data center space

In the meantime, Baker is working on securing a larger computer cluster. He’s optimistic he will be able to get more machines, but if this happens, he will need a place to house them.

Baker’s current machines are in the UW’s central data center, operated by Computing & Communications at the 4545 Building, which soon will be out of capacity. In the near term, Baker is looking to add 500 to 1,000 new machines. The building might have enough space to house them, but Baker wonders about the future.

Data centers are highly specialized environments with unique requirements including physical security, electrical and HVAC (heating, ventilation and air conditioning) capacity, fire detection and suppression, central monitoring, and more. Providing this capacity is expensive. Baker said offering this type of specialized computing space is one important way the university can help researchers. Demand for this space is growing rapidly. "Biology will require a lot of computing," Baker said. "Housing could become a real issue, and soon."

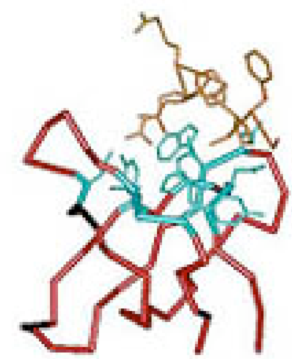
Better algorithms

Expertise in global optimization is needed to improve Baker's algorithmic approaches. Currently, he uses the Monte Carlo minimization method to search for the most promising solutions. In this method, each processor carries out an independent search. Baker thinks coordination between the searches could be helpful. "How do you make this search process the most efficient and how do you coordinate between searches," Baker said. "That's the challenging computational problem." Baker believes that someone with expertise in global optimization could help.

Sharing UW expertise

Baker suspects that there are people with global optimization expertise at the UW, if only he could find and engage them. The university has some forums to help scientists reach across disciplines, but Baker said it would be helpful to have more ways to connect people.

Additionally, the university could do more to promote interdisciplinary study, for example, expanding joint advising of students by faculty in different disciplines and then encouraging students to pursue that direction.



The future requires computing

The future of structural biology will involve computing at a level not yet available, according to Baker. "With more computing power we could solve these problems," Baker said. "I lie awake at night thinking about how to get more computers."

For more information see:
http://depts.washington.edu/bakerpg/

Cyberscience Needs

- *More computing power (on the scale found at a supercomputing center)*
- *Expertise in global optimization*
- *Increased data center space*
- *Forums for sharing expertise*
- *Ways to encourage interdisciplinary study*