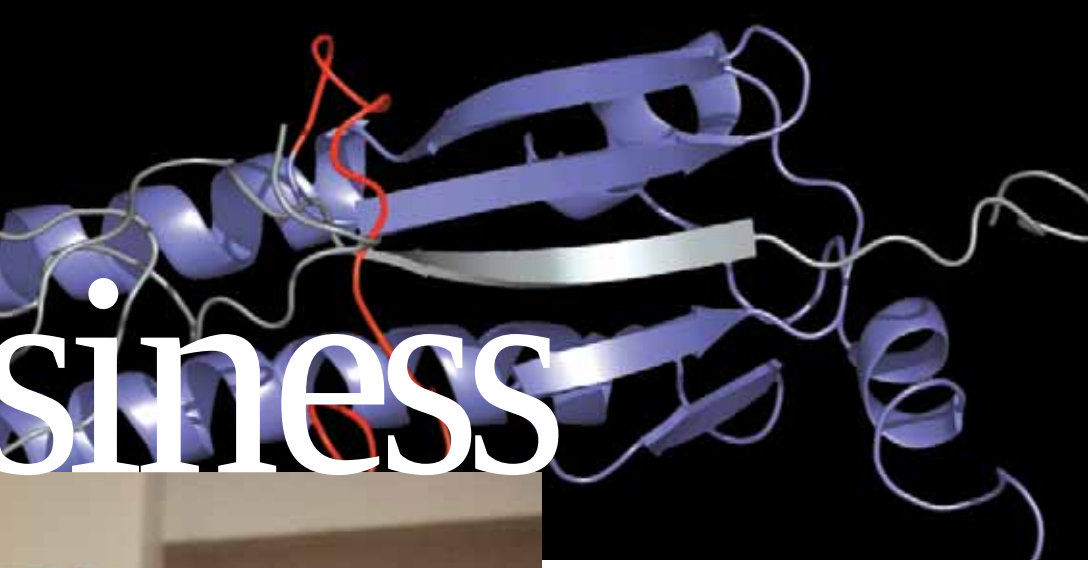
The Family Bu

National Institute of General Medical Sciences



RYAN HARRISON/PROTEIN DATA BANK

siness

DAN LAMONT

"I feel very lucky to be here."

David Baker is a computational biologist at the University of Washington in Seattle. Baker custom designs computer software to predict the three-dimensional shapes of proteins.

By Emily Carlson

Although he's only 42, David Baker is already a grandfather. Well…sort of.

Baker is raising a second generation of scientists in his lab at the University of Washington in Seattle. The Baker scientific family tree now includes scientists all across the country linked by a common goal: a driving curiosity to predict the shapes of proteins, the basic building blocks of our bodies.

Would you believe that all this happened in just 10 years?

Baker's remarkable enthusiasm for science and endless energy to solve hard problems keeps the family growing. This combination easily attracts new students to Baker's lab, and his captivating way with people keeps them there.

As with any good parent, Baker instills a sense of independence in his scientific children. After they leave the nest, most continue the journey in their own labs, where they raise their own research families.

Like glue, good communication holds everything together. Baker, a computational biologist, believes conversation gives birth to great ideas. Starting open discussions in the lab, he says, is one of his most important jobs.

"I remember a very lively energy in [David's] lab," says Jeffrey Gray, who started working with Baker 5 years ago. "David was the catalyst that increased the flow of ideas."

Gray, now a biomolecular engineer at Johns Hopkins University in Baltimore, Maryland, carries on the family tradition. Modeling his own career after Baker, Gray mentors many young scientists. Among them is a student who, as a high school senior under Gray's mentorship, placed fifth in the 2005 national Intel Science Talent Search competition (see sidebar, page 13).

Birth of an Idea

Baker has spent much of his life tucked between two mountain ranges. A Seattle native, he hikes the local trails in the summertime and skis the slopes when the snow starts to fall.

"This is the greatest place on Earth!" Baker says. "The mountains are one of the great advantages of living here."

But for Baker, the mountains offer more than just pretty scenery and recreational opportunities. They symbolize an inner passion to achieve.

"David approaches science like he does a mountain," says Gray. "He finds the highest peak and heads toward it."

But what Baker now heads toward isn't what he originally set out to find. At first, he thought modeling the shapes of proteins was, well, boring.

"I remember writing a report for a college biochemistry course and thinking, 'Protein folding seems like a neat problem, but not much is happening in the field,'" Baker says.

He admits that his opinion changed during graduate school when he began studying how cells organize their many parts, which of course include proteins.

Our bodies consist of billions of proteins, large molecules made of smaller components called amino acids. Anywhere from a few to tens of thousands of amino acids link up in a particular sequence, and then each amino acid sequence folds into a unique three-dimensional structure.

It's this shape that really determines a protein's job. When a protein attaches to other molecules, it triggers a host of chemical reactions that run all of our biological machinery.

"Proteins are incredibly organized and do amazing things," says Baker.

But sometimes the things go wrong. Altering just one amino acid in the chain can change the entire shape of the protein. This switch can lead to life-threatening disorders like sickle cell disease or cystic fibrosis.

If we want to treat and prevent diseases, Baker says, we need to know what proteins look like. Having this information will help scientists custom-design medicines to target proteins and fix health problems.

The Shape of Things

At first glance, determining the structure of a protein from its amino acid sequence seems like it would be easy. But things have not turned out to be so simple. Score one for Mother Nature.

If a protein is really big, scientists can spend months or even years trying to determine its structure. Sometimes, knowing the shape of a similar protein and using that as a guide can speed up the process. But part of the problem is that researchers only know the structures of a small fraction of the proteins in the human body.

So where does that leave scientists who want to find out a protein's shape? They either do physics-based experiments with X rays or huge magnets, or they use computer models to make good guesses.

LUKASZ JOACHIMIAK, BRIAN KUHLMAN

One of Baker's passions is exploring the mountain ranges near Seattle, Washington.

Both approaches have their drawbacks. X-ray crystallography and NMR spectroscopy, the methods from physics, are labor-intensive and can be expensive. The computer modeling approach can be inaccurate and unreliable.

For Baker, the Holy Grail is developing software programs that generate high-resolution models of proteins. Ideally, these models would reveal every feature of a protein's landscape, including its atoms, hydrogen and other bonds, and all the places where important chemical reactions occur.

With refined pictures, researchers can examine single proteins and track their interactions with other molecules. Accurate models could ultimately let researchers make entirely new proteins with custom functions, motions, and chemical reactions.

Scientists have been trying to develop accurate computer models for years. But the models rarely capture all the details, instead creating mostly "rough sketches" of how protein parts fold together into complex structures.

"Simplifying the model of a protein is like smoothing out a mountain until you have rolling hills instead of sharp peaks and deep valleys," explains Baker, adding that a lot of extra work goes into finessing the computer's output.

Quality Time

With an intense interest in trying to solve what others find unsolvable, Baker splits his time between his two families—his wife and two children and his lab personnel. During the week, he spends regular work hours with about a dozen postdoctoral fellows, 10 graduate students, and a handful of other researchers. Many come from different

countries and different scientific fields, like chemistry, engineering, and medicine.

"It's a privilege to walk out of my office and talk to really smart people interested in the same problem I am," he says. "I feel very lucky to be here."

Together, the group focuses on a computer software program that Baker developed called Rosetta. Just like the famous stone of the same name once helped linguists decipher ancient languages, Baker and his group hope their Rosetta will decode the mysterious shapes of proteins and even help them build new and better versions.

Basically, Rosetta uses information about a protein's amino acid sequence to predict its possible shape. It breaks the protein into small chunks of amino acid sequences, searches for all the different shapes each chunk could assume, and then mixes and matches them until it finds a perfect fit. Rosetta may create up to 10,000 simulations and run for 100 days before honing in on the structure of even the simplest protein.

Baker and his team have created many Rosetta flavors, each of which can answer a different question, such as how a protein interacts with another protein or with a DNA sequence. Some varieties incorporate experimental data or the structural information of other, similar proteins.

One version of Rosetta being developed could run on the computers in University of Washington dorm rooms when students aren't using them. This could add up to 10,000 processors to the team's protein structure prediction effort and make the work go faster.

Because Baker wants as many minds as possible working on the problem, he gives Rosetta to other scientists for free.

Baker and his students take the Rosetta models and go to work refining them. Sometimes, they run into problems, but that doesn't stop progress. When this happens, Baker says he knows it's time to talk, and he brings the lab together to troubleshoot.

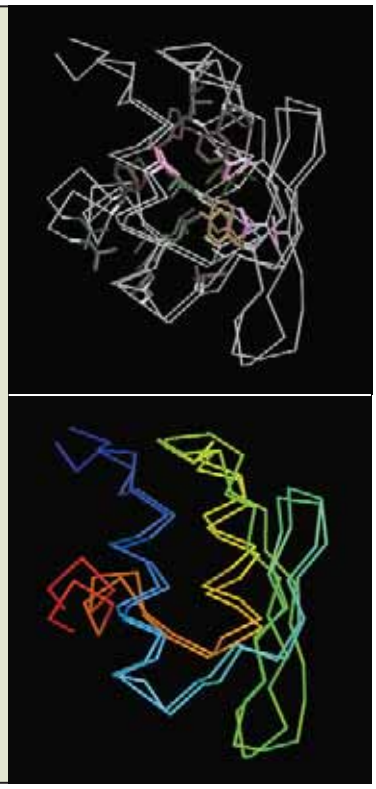
"I think the human factor is one of the most important elements [of science]," Baker says.

Baker is developing a version of Rosetta that can run on University of Washington dorm room computers, adding processing power to his protein prediction experiments.

Outward Bound

Whether the lab takes the day off to go hiking or sits around the lunch table trying to solve a problem, Gray says, "David has the energy to push people beyond their boundaries to explore new ideas."

Baker's ultimate goal is predicting a detailed structure at a level of resolution, or clarity, of 2 angstroms, or 200 billionths of a millimeter.



During the 2004 Critical Assessment of Techniques for Protein Structure Prediction (CASP) community-wide experiment, Baker's team successfully modeled the structure of a protein they had never seen before. Their computer model (top) was strikingly similar to this protein's actual X-ray crystallographic structure (bottom). The model highlighted even more protein detail (pink, brown, and dark gray, top) than did the X-ray experimental data.

PHILIP BRADLEY, DAVID BAKER

"If we can successfully model protein structures with a level of accuracy so that biologists are confident the models are right, we could compute all the protein structures that already exist without [doing] experiments," Baker says.

This, he notes, would save researchers a lot of time and money.

Some people may find it ironic that Baker, who never took a computer course in his life, not only developed a tremendously useful software program but also spends his days (and sometimes nights) thinking about computers.

But he doesn't see it that way.

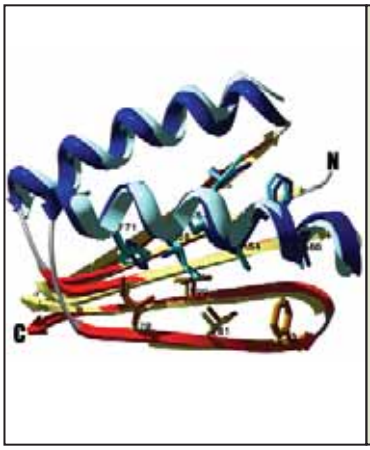
Baker says he finds a problem he wants to solve, meets up with a great group of people, and then learns whatever he needs to know along the way.

Every year, Baker recruits more adventurers to join him on his climb to model proteins. Only together, he says, can they reach the top—better models of protein structures that may lead to new drugs and vaccines for keeping us healthy.

New Direction

One of the latest Baker family franchises involves designing new proteins not found in nature. Unlike the usual approach of starting with an amino acid sequence and then building a structure, Baker and his team are working backwards. They're designing proteins from scratch.

Like architects who design a house before drawing up blueprints, Baker and his trainees began with a sketch of a made-up protein structure. Using Rosetta, they pieced together a string of amino acids that most likely would link up to create the new protein, and then made the actual protein in the lab. In an early experiment, the researchers found that the real protein was virtually identical to the one Baker had imagined.



BRIAN KUHLMAN, GAUTAM DANTAS, DAVID BAKER

Baker used his computer program, Rosetta, to design a small protein not found in nature. His computer model of the protein's structure (dark blue, red) is virtually identical to its lab-determined structure (light blue, yellow).

The scientific community recognized this work as monumental and Baker and his research group received a prestigious prize for the best paper published in a 2003 issue of the journal *Science*.

Next, Baker wants to design proteins that cause particular chemical reactions on demand.

"This would open up a whole new world of functional proteins," says Baker.

The ability to create proteins made to order offers a promising route for developing custom proteins that could interrupt or enhance a particular reaction inside a cell.

Community Center

When it comes to modeling protein structures, Baker and his group have proven that they can climb with the best. Every other year, the group enters a friendly competition

BRIAN KUHLMAN

Baker's annual scientific family reunions always include a mountain trek, like this one to Dragontail Peak in Washington's Wenatchee National Forest.

called CASP (Critical Assessment of Techniques for Protein Structure Prediction). They go head to head with hundreds of labs worldwide to see who can make the best predictions.

In December 2004, scientists from more than 200 labs gathered in Italy, submitting a total of 15,000 predictions for selected protein structures. The only people who knew what the proteins actually looked like were the judges.

Baker's group used Rosetta to develop their models, and as in previous years, Baker's team did very well. One of the post-doctoral researchers in Baker's lab modeled a protein structure with a very small average error of 1.59 angstroms.

"I like working on the methods and seeing them pay off," says Baker. "CASP is very collegial and a great experience for the people making the predictions."

Although there are some competitive aspects, everyone walks away from CASP with a prize—the opportunity to work together, learn about current challenges, set future goals, and assess the methods and technology used to predict protein structures. For this reason, scientists prefer to call CASP a "community-wide experiment" instead of a contest.

Despite this progress, and even with Baker's many successes, a lot still needs to be done. When it comes to accuracy, many current, low-resolution models are in the ballpark, Baker says, but they have a way to go.

As Baker and his team continue to work on the problem, one thing stands in their way: insufficient computing power.

"For a long time, the problem was not having accurate descriptions of proteins and their interactions," explains Baker. "But now the problem is that we don't have enough computer power to run the simulations."

For example, Rosetta can run for months before it finally spits out a model that closely resembles the real thing. Not only does this take computer time, it also takes a lot of computing power, Baker says.

Making really accurate predictions, and lots of them, means having a herd of computers that can quickly process data. Currently, Baker is talking to large computer companies to try

to get his hands on more machines, especially ones with faster processors.

Family Reunion

Every summer, Baker invites his extended scientific family to join him in Seattle for what he calls "Rosetta Commons." For 2 days, they talk about prediction projects, challenges they're encountering, and potential ways to improve the software behind it all.

"We've all started labs that are working on different problems," says Gray, who attended the reunion last summer. "But we're still related by the Rosetta code."

On the third day, the group usually heads for the hills, something the former students fondly remember from their days in Baker's lab.

"If you're walking next to David, you're going to be talking about science," jokes Gray. "David focuses so much on science. It's what he does naturally."

Last summer, the group hiked up Dragontail Peak, which looms about 9,000 feet above sea level. The trail, recalls Gray, was quite ambitious.

Minus the time for picnicking and swimming in a crystal-blue lake, the group spent nearly the entire day climbing to the top. They were only halfway down when the sun started to set.

"It was 8:00 p.m., and we still had several hours of hiking," says Gray.

Without enough flashlights to guide their way down, Baker and former postdoctoral researcher Brian Kuhlman—by far the most experienced hikers in the group—volunteered to run back to the cars, drive into town, and bring back extra supplies. The two met up with the other hikers, still creeping their way down, with flashlights and chocolate.

Everyone finished, still in good spirits, remembers Gray.

"It was definitely a bonding experience!" he says. ■

The Next Generation

STEPHEN SPARTANA

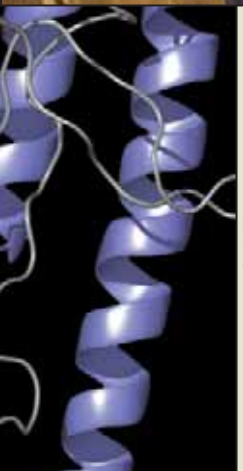
Every year, more than 1,500 U.S. high school seniors enter the ultimate science fair: the Intel Science Talent Search, dubbed the "Junior Nobel Prize."

"If you're doing high-level research in high school, it's expected that you'll apply to Intel," says Ryan Harrison, a recent graduate of Baltimore Polytechnic Institute in Maryland who ranked fifth in the 2005 national competition.

Harrison got his prize, a $25,000 scholarship, for work that follows in the footsteps of two generations of scientists who predict the shapes of proteins, molecules vital to our everyday health.

With the help of his mentor, Johns Hopkins University professor Jeffrey Gray, Harrison spent more than 2 years developing a version of Rosetta software that models protein structure in a particular pH environment.

Just 17 years old, Harrison already has won the respect and admiration of many scientists. Among them is David Baker, a computational biologist at the University of Washington in Seattle, who mentored Gray (see main story).



RYAN HARRISON/PROTEIN DATA BANK

"[I heard] Ryan give this great talk, and I thought he must be a graduate student," says Baker. "It turned out he was a high school senior!"

To that end, when Harrison talks about his science, he purposely omits details about his schooling. "I've learned that you have to fool people, otherwise they won't take you seriously," he explains. "You need to prove yourself first."

With a laid-back attitude and energy that keeps him bouncing in his chair, this scientific prodigy doesn't sacrifice fun for success. A self-described "goofball," Harrison says his main objective is to have a good time.

When he headed to Washington, DC, for the final round of the Intel competition, he says, "I just wanted to hang out with cool people. I never expected to take home a prize."—*E.C.*