

The shaded/solid rendering of monensin, a naturally occurring antibiotic, is based on structural information determined through x-ray crystallography.

Rich Friesner

## COmputational CHEmistry

**C**OMPUTATIONAL CHEMISTS ARE ON THE VERGE OF being able to create, on demand, materials that have specific properties, whether of flexibility, durability, or the ability to turn an alluring shade of lavender when the late-afternoon light strikes them at a certain angle.

BY PAULINA BORSOOK

A possibly apocryphal insight (variously attributed to Nobel Prize-winning physicists Paul Dirac, Edward Teller, and others) states, "The work of physicists is finished; what is left is engineering." Dirac, one of the fathers of quantum mechanics, said in 1929 that "The underlying physical laws necessary . . . for a large part of physics and the whole of chemistry are completely known, and the difficulty is only that the exact application of these laws to equations much too complicated to be soluble." No chemist has had the time or the means to perform the calculations, so no chemical engineer has been able to use the solutions to build new materials. It is turning out that advances in computer hardware and software may prove Dirac wrong: supercomputers and new ways to use them may finally be able to solve the equations that explain and describe the material universe.

Computational chemists are beginning to make runs at these quantum mechanical equations, now that supercomputers can help relieve the burden of complex calculations. In theory, it is possible to apply the equations to understand precisely how it is that a piece of wood is brown and hard. Once that mathematical solution is understood, it should be a matter of engineering to create other substances that are equally brown and hard.

Until now, in spite of advances in basic science, chemists and biotechnologists have still largely made their discoveries on the basis of empirical methods of trial and error. Computational science could change all that.

Molecular design (as one branch of computational science is called) is already considered to have strategic importance. The Office of Naval Research is bringing together chemists, physicists, engineers, and computer scientists to consider techniques for mastering the material universe at levels never before possible. Computational science will have an immediate impact on quality of life in areas as diverse as animal rights (no more animal testing for drugs or cosmetics), use of natural resources, and workplace safety. The secondary impacts — the way new materials will change our minds and our societies — are harder to predict.

### It happened in Tallahassee

A conference (one attendee called it "molecular Woodstock") brought together computational scientists from diverse disciplines at Florida State University in Tallahassee in January 1992. This "Workshop on High-Performance Computing and Grand Challenges in

Structural Biology" felt like being present at the creation of something. The feeling in the air was that of approaching a major threshold in capabilities that scientists have long sought. Computers are finally fast enough for Something Important to happen in several different fields. (Although it pays to remember that Artificial Intelligence [AI] has been threatening to be at a similar threshold for the last 20 years.)

Structural biology, now largely a branch of computational science, is concerned with the physical/chemical structures of biological compounds. Crick and Watson's uncovering of the double-helix structure of DNA is perhaps the best-known example of structural biology. The people who cracked the structure of hemoglobin, the first protein whose structural configuration ("conformation," in structural-biology vernacular) was understood, won a Nobel Prize for their work, too. Yet, thirty years later, scientists are still not exactly sure how oxygen really binds within the hemoglobin molecule — and the binding of oxygen within hemo-

globin is among the most basic physiological functions. As things stand now, it's like not really understanding what it means to breathe, though we still take it on faith that breathe we must. Structural biologists try to figure out similar life processes at the level of physics.

A central problem facing the field of structural biology is the issue of protein folding — understanding (and therefore being able to predict) why proteins, as dictated by the ordering principles of genetic instructions, take the shapes they do in order to perform the functions they have been assigned by nature.


Protein folding fascinates structural biologists and computational scientists from other disciplines because it is a highly visible problem, with huge payoffs for those who solve it. Understanding protein folding is necessary to understand how biological processes work. Once the rather routine mapping from the highly funded, highly publicized Human Genome project is completed, the hard part will begin: knowing how the proteins associ-

ated with a particular gene are biologically active. The same principle, of structure determining function, also applies to research into how drugs take action in the body. In spite of all the leaps in pharmacological research, no one truly understands the specific mechanics of drug action, the particularities of why an effective medicine binds to a particular molecule, and why a similar but ineffective medicine doesn't.

Structural biologists and other computational scientists are

**The goal of computational science is to make theoretical breakthroughs about fundamental physical principles by using computers to extend scientists' abilities to deal with complex phenomena. Computational biology is scary enough, as Kevin Kelly's and Steven Levy's articles in this issue demonstrate. Computational chemistry might be equally powerful, promising, and scary, in a different way. What new powers, tools, weapons, and side effects might result when the capability of creating molecules to perform certain tasks on command is translated into technology?**

**Paulina Borsook is finishing her MFA thesis at Columbia, "a book-length series of interconnected short stories that delineate how the new information technologies (focusing on but not limited to email) deform relationships." The title story of the collection, "Virtual Romance," features the WELL, and was nominated for a Pushcart Prize. She describes herself as "a Luddite who has made a living for the past ten years writing about computers and communications." Her email address is [loris@well.sf.ca.us](mailto:loris@well.sf.ca.us). —Howard Rheingold**



drawn to the mysteries behind protein folding. Computational scientists (as do all good scientists) have the sense that if this problem can be posed in an aesthetically pleasing way, an equally aesthetic solu-

tion will be on hand. Particle physicists are migrating to this problem in part because of their traditional attraction to aesthetically pleasing problems, but also because breaking the code of pro-

tein folding at the atomic level, of breaking down into its physics, is a problem that can — and most likely will — be solved. Not so the origins of the universe, the mysteries of time, and where we all came from and are going to. Particle physicists are leaping on the problem of protein folding because so many obvious, elegant, and soluble problems have already been solved in their native field.

## OTHER LINES OF INQUIRY IN COMPUTATIONAL SCIENCE

**Protein folding isn't the only area of interest for computational scientists. Computational chemistry, as manifested in molecular engineering, hopes to create mathematical models able to predict the behavior of experimental substances. This could change the shake-and-bake process of chemical invention.**

**Molecular engineering — controlling the microworld — could transform every area of industrial production. With supercomputers and supercomputer software that did not exist as recently as five years ago, theoretical chemists are poised to invent solutions to a number of problems:**

- **New products could be designed from the molecular level to be biodegradable. Substitutes for environmentally destructive compounds could be created that work as well as, or even better than, the fluorocarbons they replace.**
- **Solar energy may finally replace fossil fuels. Efficient photovoltaic cells could replace those available today, manufactured from materials designed from scratch at the molecular level to absorb energy from the sun.**
- **Metal alloys can be created that could, for example, make jet engines simultaneously far more powerful, lighter, and safer.**
- **Drug design could allow pharmaceutical companies to assess the usefulness of candidate medicines long before they are tested on human subjects.**
- **New methods of catalysis (speeded-up chemical reactions used in ordinary manufacturing processes such as the cracking of crude oil to make gasoline) can be achieved through mathematical modeling. Each new process so discovered is potentially worth billions of dollars.**
- **Polymers, the basis of plastics and other synthetics, can be custom-designed to achieve specific ends. At last, pantyhose that really won't run!**

### Dimensions of the problem

There are from anywhere from 60 to 500 constituent amino acids in a particular protein, with the sequences of the amino acids determining its structure. With sickle-cell anemia, for example, the difference between one amino acid being present as opposed to another makes the crucial difference in oxygen uptake — and serious illness in the life of the sickle-cell-anemia sufferer. Scientists as yet can't predict the three-dimensional structure (conformation) of a protein from the sequence of the amino acids in it.

Each amino acid is at an angle relative to the next one in a sequence. Initially, conformations were sought experimentally through x-ray crystallography, but scientists prefer the intellectual satisfaction that would come from finding an overarching principle that would generally predict conformations. Further, making crystals is a black art, not a science, and there are whole classes of proteins, such as those embedded in the fatty ripply media of membranes, where it is almost impossible to create crystals. Hence the current turn to computers to seek a general principle of protein folding.

Theoretically, there are about five different possibilities for the difference in position from one amino acid to the next; extrapolate this, and you have a fine computational problem to be solved,

for there could be 5-to-the-100th-power different variations of amino-acid angle variations for a particular protein. No computer on the market today (or projected to be around in the not-so-near future) could run through this many iterations to find what the conformation would be for a protein. Which is where things get interesting: researchers are attempting to narrow the search through the creation of clever software algorithms.

Computational scientists have ideas, of course, on how to prune the search for likely conformations. Some use familiar computational methods, such as decision trees, or Monte Carlo, a computer program that takes all the known information about a physical system and systematically and repetitively introduces an element of chance

(hence the name: like rolling the dice at Monte Carlo) into the model created by the data and probabilities fed into it. AI has also been thrown around as a way to solve the problem.

More interesting, Peter Wolnyes, a scientist at the University of Illinois, has a more interdisciplinary approach. He borrows from neural-networking models of how the brain works, using the 500 or so protein configurations already known, to reason by analogy. His novel method brings prior information into the search, creating a learning algorithm.

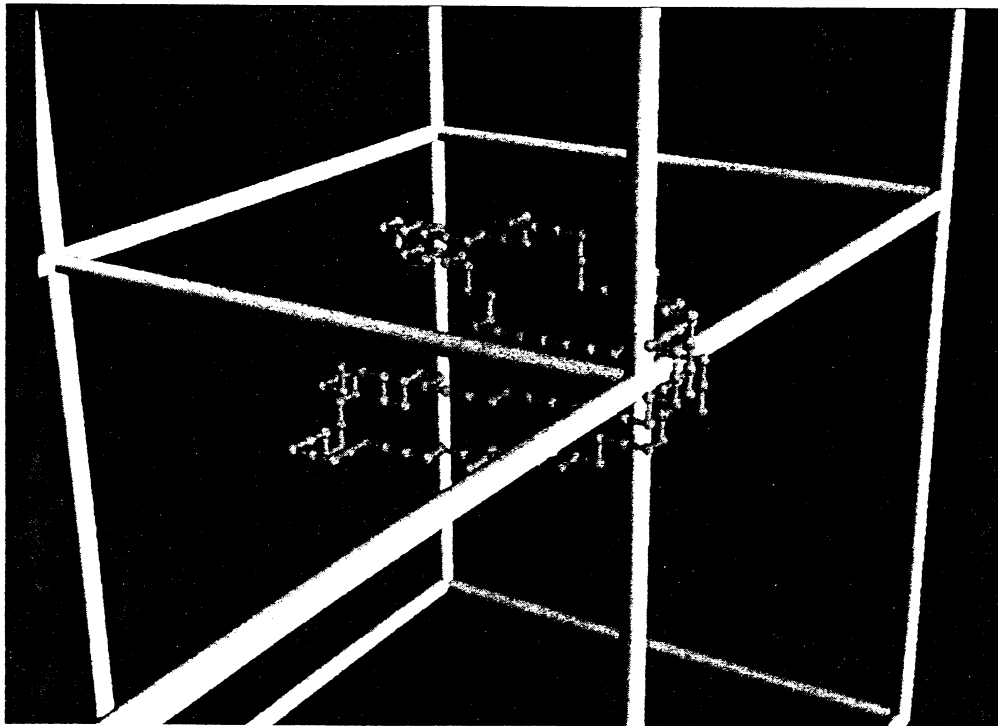
Everyone devoted to the problem seems to have a different intuitive sense of how the search ought to be pruned. The feeling at Tallahassee was that computer hardware had

improved enough in the last ten years and that massively parallel machines have finally matured enough, that somebody might happen on a truly interesting solution in the next year or so.

### The birth of a new science

The conference in Tallahassee was expected to attract 50 people; more than 300 came. The birth of a new science was evident in several regards: structural biologists, with their concerns about algorithms, and discussions of energy fields, had more in common with physicists than their more traditional colleagues in evolutionary biology or in field ecology. Biologists, chemists, physicists, and computer scientists were speaking a common

E. J. van Kesterburg, Hwa Lin, M. T. Carroll, FSU



In gel electrophoresis, electrical potential applied to DNA fragments placed in a gel causes the fragments to move through the gel and separate by size.

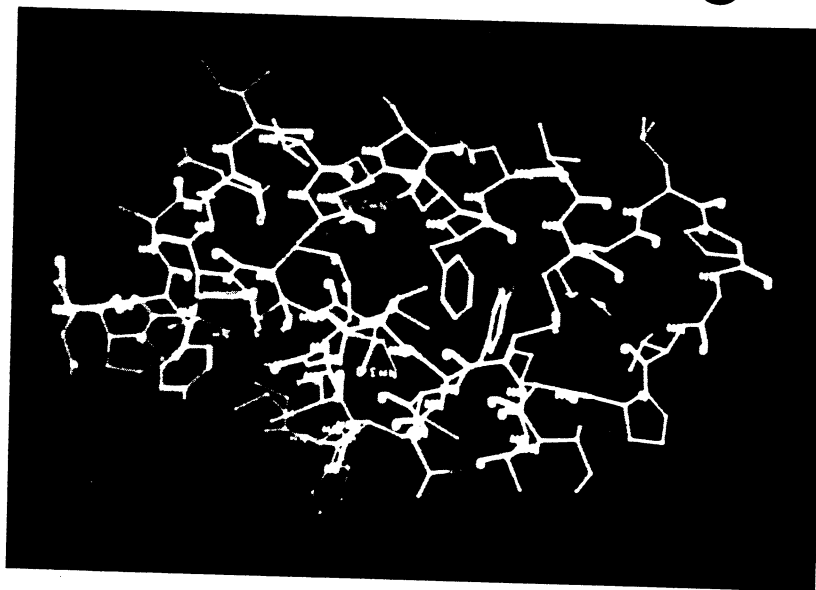
This schematic shows a knotted DNA fragment within a gel, with the gel represented as a regular lattice. Simulating gel electrophoresis helps researchers better attack the cataloging of the functions of the thousands of genes found in the human cell.

language of problem-solving through modeling on supercomputers.

The conference also raised the hot issue of big science and little science, centralized versus distributed computing. Some of the most passionate argument centered on whether, in a time of decreased national funding for research and development, an effort should be made to lasso \$20 million for a nifty new supercomputer, or have that money put into having the equivalent of a Sun or Silicon Graphics workstation on every grad student's and every post-doc's desk.

The people who want the supercomputer say that there are kinds of problems that can only be solved by big iron; that having access to the unique ways massively parallel machines solve problems will create new understandings and new ways to try to understand reality; that those who have never played on one of these magnificent instruments cannot appreciate the difference a Cray or Kendall Square or Connection Machine would make. ["Traditional" supercomputers use one very powerful processor that processes many instructions "serially" — one at a time. "Massively parallel" supercomputers use thousands of relatively less powerful processors that operate simultaneously — "in parallel." The use of large numbers of less powerful processors makes it possible to produce supercomputers at much lower cost — "desktop supercomputers."] The best and brightest scientists say that the best science gets done when scientists know how to make use of new technologies, leaving behind endless recapitulations of what they learned in grad school. They say, bring on the supercomputers.

Yet the genius of the microcomputer revolution (and the way some of the most intellectually interesting science has historically gotten done) has been by someone noodling around in a modest, relatively low-tech fashion. Personal computers



Rich Friesner

The main value to computational chemists of crambin, a very simple protein, is its simplicity: crambin was one of the first proteins whose structure has achieved high resolution at the atomic level. This visualization was done on a Silicon Graphics workstation.

gave individuals routine access to computing power. Funding for access to a pricy new supercomputing center necessarily means that only certain people can use that computer time, in rationed amounts — and the question remains, with fiscal technology policy being what it is in the US: might that money be better spent letting graduate students try out their ideas for the minimal price of a piece of small-scale local hardware? No supercomputer center overhead costs, no telecommunications charges, no competition for scarce computing time with folks from other universities. The power to solve the problem would remain on the desktop.

Still, revolutions in technology are as much about changes in toolmaking as they are about the creation of Wonder Widgets. Think of the zillions of toolmaking innovations that had to take place in order for the automobile to hit the road with any reliability and force. Since software is the tool of this technological revolution, it might well be true that it is only on big, massively parallel machines that the new tools, and new ways of doing business,

can be created to attack the problems of computational science (see sidebar, p. 50).

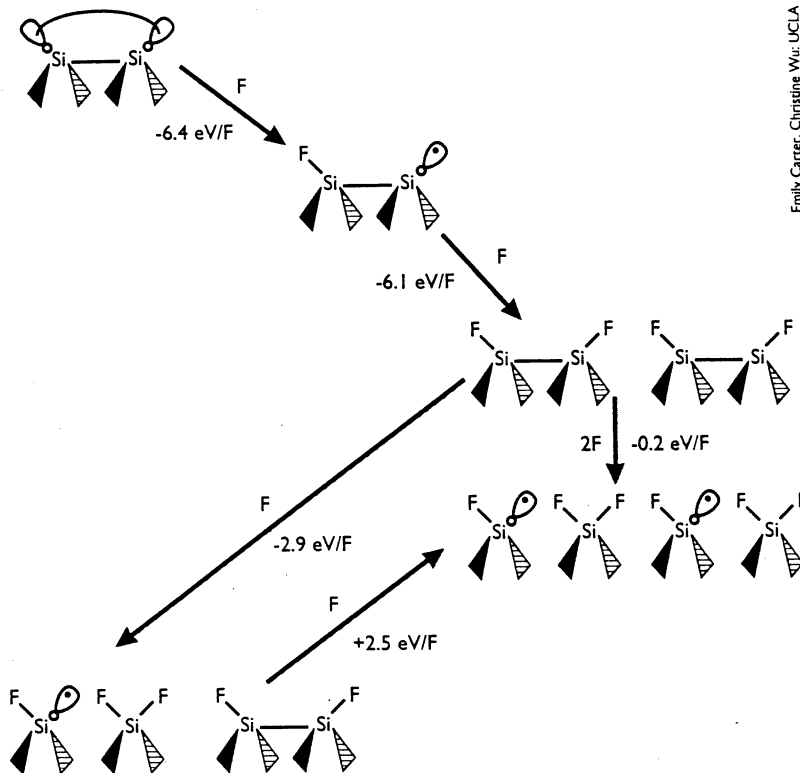
This problem of toolmaking is related to a curious blending of research and commerce unique to the field. Supercomputer companies, most notably Thinking Machines, have on staff Ph.D.s whose jobs involve cranking out papers and doing research on computational problems in physics, chemistry, and biology, as well as servicing accounts and tending to customers' needs. The synergy appeals to academic scientists: at last, they think, computer-company staffers who speak our language and can understand the problems we are working on. But the covert marketing message this sends is that the presence (and publication histories) of scientists on staff at the computer companies are existence proofs that the machines can be used to solve problems similar to those academic researchers are struggling with. This matters supremely in this new field of computational science and with this new technology of massively parallel machines: regardless of what vendors say, programming

and software development for these new machines is, as the scientific jargon goes, not trivial. And scientists take comfort in the indirect assurance that the machines can be made to work and that functional software-tooling is at hand.

Predictably, venture capital is flowing into the field — but interestingly enough, it is into the companies creating the software tools. The market for these tools is already grossing around \$50 million per year, and there are already mergers and acquisitions of these toolmaking companies.

As usual, there is a potential downside to the fine invisible hand of capitalism. Will the dictates of private enterprise determine what kinds of problems get worked on, and which ones get slighted? Further, this is a field in which petrochemical and pharmaceutical companies have invested large amounts of money from the very beginnings of computational science — so the question remains whether the free flow of scientific information will be affected by the hybrid mixture of basic research and commercial investment. Will unfavorable results not be reported? If an in-house scientist doing molecular modeling for a drug company solves the protein-folding problem, for example, will patent concerns prevent word from getting out?

Regardless, computational science is poised to take off the way biotechnology did ten years ago. Even *Business Week* did a recent cover story on the topic (titled "The New Alchemists"). Professionals in the field, the academics researching these frontiers, are beginning to form companies capitalizing on the means to predict and design the behavior of any new compound desired by science, industry, or 100 well-trafficked shopping malls across the country. And the Tallahassee conference has spawned a task force charged with trying to wrest money from Congress to support this Grand Challenge in Struc-



Emily Carter, Christine Wu, UCLA

Fluorine atoms reacting with a silicon surface — an important general chemical reaction because it is the basis for all computer-chip technology. Computer chips are etched with fluorine to create integrated circuits. The manufacture of chips has become sufficiently advanced that patterns can be etched at the level of submicrons; it has thus become critical to understand what goes on at the atomic level. Once that understanding is achieved, integrated-circuit patterns could be created at the atomic level, accelerating the miniaturization of computer technology.

Here, silicon bonds are being broken and silicon-fluorine bonds are being formed. Computational chemistry on an FPS Computing model 521-EA superminicomputer solved the equations that created this model.

tural Biology that, they argue, can only be met by funding a super-computer center.

In the short term, physical chemists wonder if, to get the support to develop the software that will underpin the twenty-first-century industry of alchemy, they'll have to beg for a Big Science project like a Structural Biology Supercomputer Center, instead of simply being able to ask for more funding for more support for more graduate students who will use more workstations. Meanwhile, computational chemists feel they may be able to deliver

on the promise that nanotechnology has been making: engineering in the microworld. And if they succeed, they may overturn our entire post-industrial way of life: if diamonds or gold or what appears to be clear-heart redwood can be manufactured as easily as plastic wrap is now, then the dream of enough of everything for everyone might be achieved. Money and material wealth might become something very different, but chemists, or whoever employs them, would become the new Lords of Creation. And that could be pretty scary. ☛