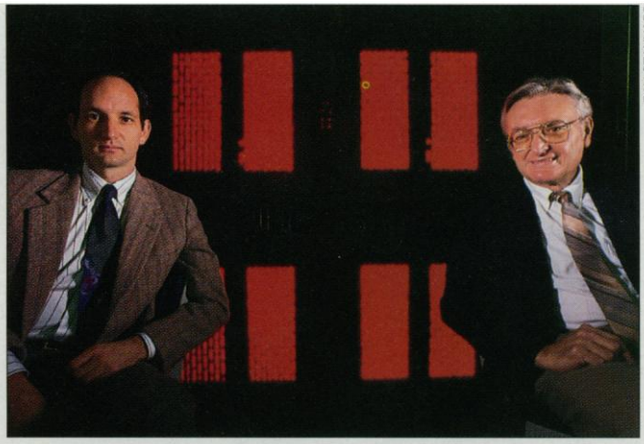
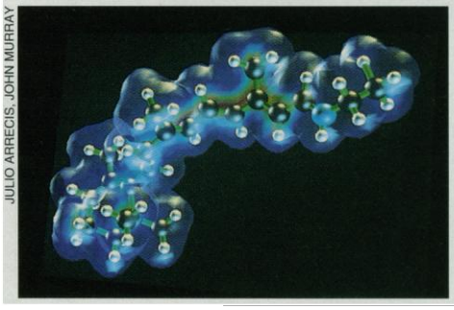
Biology Approaches the Teraflop Era

Should computational biology put all its eggs in a single basket—even one that could perform a trillion operations per second? Some say yes; some say no; some say "later"

Each time the obsidian-like monolith appears in *2001: A Space Odyssey*, casting a long rectangular shadow and a supernatural spell, it dramatically alters life on Earth. These days, computational biologists, who use powerful computers to simulate the structure and activity of proteins and other macromolecules, are preparing for the potential arrival of their monolith—a "teraflop" supercomputer that would be capable of performing a trillion floating point operations per second. The mere thought of a machine that would run about 1000 times faster than today's supercomputers has begun to transform the field. "Biology is undergoing a revolution that is at least as profound as the one that physics underwent with quantum mechanics," says crystallographer Lynn Ten Eyck, a senior scientist at the San Diego Supercomputing Center.



RAY STANYARD



JULIO ARRECIS, JOHN MURRAY

Teraflop apostles. Paul Bash *(left)* and Joe Lannutti would like computational biology to have the world's fastest computer, which could go far past current supercomputer images such as this one of the chemical retinal from bacteriorhodopsin.

That revolution will involve the application of enormous amounts of new computing power to problems like rational drug design and protein folding. As with all revolutions, though, there is considerable disagreement among the revolutionaries on which direction the movement should take. Torn by a love for small science and a lust for more computing power, computational biologists are drifting into three camps, says Ten Eyck: Those who oppose investing heavily in a teraflop supercomputer because it might rob them of "small-science" funding; those who want the power promised by a teraflop machine but are willing to wait until the mid-1990s, when advances in technology will make such a computer more affordable; and those who are so "obviously committed" to a teraflop machine that they're ready to lunge at the first opportunity to buy one.

One small wrench is jamming everyone's gears: Teraflop computers don't exist yet. However, computer firms have said they could begin building one this year at a cost of between $50 million and $100 million. Even if computational biologists wait a few years for the price tag to shrink, says theoretical chemist Paul Bash of Florida State University, it's likely that the field would be capable of purchasing only one machine devoted to computational biology.

The prospect of one machine for hundreds of researchers is forcing the three camps of computational biologists to unite under one flag. "As a field, we like small science," says theoretical biophysicist Klaus Schulten of the University of Illinois' Beckman Institute. And small science means many voices, but "now," says Schulten, "we're suddenly confronted with a new age in which we have to speak with one voice." One plan that has stirred the field—not producing consensus but certainly leading to lots of discussion—is a move by Bash to snag the first teraflop machine for his school. In January, Bash organized a gathering of computational biologists (about as rare an event as the coming of the Monolith) at Florida State, during which the scientists informally endorsed an effort to seek funding for a teraflop machine. Since then, Bash has put together a "steering committee" of elite computational biologists who plan to start writing a grant proposal for a teraflop in late summer. The teraflop machine would be the centerpiece of a national center for structural biology at Florida State. Says Bash: "This is a watershed period for us in the computational field."

With a teraflop machine looming over computational biology like the Superconducting Super Collider (SSC) did over high-energy physics a few years ago, it might be no surprise that a physicist spurred the 39-year-old Bash to Think Big. Shortly after Bash came to Florida State from Harvard in November 1990, Joe Lannutti, director of Florida State's Supercomputer Computations Research Institute, persuaded him to organize scientists to collaborate on a teraflop proposal. According to Lannutti, who served on a Department of Energy (DOE) high-energy physics advisory panel during the push to get the SSC built in Texas, injecting a "big science" way of thinking into the computational biology community seemed "the natural thing to do." To bolster their case for Florida State, Bash and Lannutti have begun feeling out state support—about $2 million to $3 million a year—for a staff to operate a teraflop.

Meanwhile, the Florida team has begun warming up for its major-league pitch to funding agencies, from which they hope to snag much of the estimated $50 million to $100 million needed to build a teraflop computer. Bash's plan is to cast rational drug design, a process stymied by a lack of computing power (see box on next page), as the best bet for a quick payoff with a teraflop machine. And he is making his arguments in terms designed to appeal to the politicians: A teraflop computer, he points out, would cost a fraction of what Americans spent on health care last year—about $800 billion—yet it might yield significant inroads into treatments for cancer, AIDS, tuberculosis, and other diseases for which pharmaceutical firms are busily screening chemicals. According to William Wilson, manager of the Center for Computational Engineering at Sandia National Laboratories and one of the scientists working with Bash on a teraflop proposal, "These enormous computers could help us understand some of the basic mechanisms that underlie the efficacy of drugs against cancer."

Bash and his colleagues are well aware that rational drug design is already a favored topic in Washington. "Drug design is a very strong plus," says Thomas Kitchens, an official in the DOE's applied mathematics program who helps apportion computational biology funding. The attractive thing about

it, says Kitchens, is that advances in drug design can be transferred readily to the pharmaceutical industry, thereby boosting the U.S. economy. This will make a teraflop proposal "extremely saleable" on Capitol Hill, says Bash.

If reactions like those are any indication, Bash may have hit on a winner—picking the right moment to steer the field toward a megamachine, and, in rational drug design, having found a good banner to crusade under. Some computational biologists, however, aren't buying. They worry that a teraflop could bring to their field some of the same problems that high-energy physics has experienced with the SSC. Overselling, for example: "If in a few years we haven't designed a drug, or folded a protein, would the whole venture be deemed a failure?" asks Barry Honig, a molecular biophysicist at Columbia University.

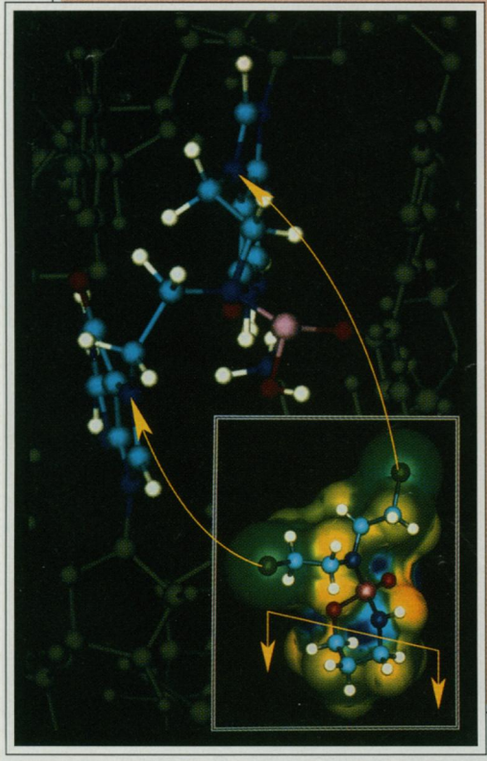
Indeed, Honig argues that overselling is virtually inevitable because the "protein-folding problem"—using computers to predict the tertiary structure of a string of amino acids as it folds into a protein—is too mathematically complex to be solved even with a teraflop computer. Bash concedes the point: "I doubt if [a teraflop computer] alone will give us a real general method that would give us a reasonably accurate [tertiary] atomic structure of proteins." Nevertheless, he maintains, the field is "perched" on the edge of deciphering other computational information more relevant to predicting the activity of potential drugs: For example, the kinetics of enzyme catalysis, and receptor/ligand interactions.

Echoes of the SSC also turn up in the fears on the part of some computational biologists that a national center for structural biology would devour grant money that would otherwise go to individual researchers. According to Honig, the thought of tens of millions of dollars being spent on a teraflop machine begs the question, "Where does the money come from?" Few scientists, he contends, would want to sacrifice money that could be invested in personal workstations for a teraflop computer. Like their physicist colleagues who pushed the SSC, Bash and Lannutti acknowledge that the teraflop funding would have to be "new money." The coffer they hope to raid is an estimated $803 million in the 1993 federal budget for the High Performance Computing and Communications (HPCC) initiative, a program run by several federal agencies that is intended to fund high-power computer projects deemed crucial to "U.S. technological leadership." And, as it happens, drug design is one of several "grand-challenge" projects identified in the HPCC initiative.

Finally, like the SSC, a teraflop computer, if it were to be used efficiently, probably would not be able to accommodate more than a handful of research projects at one time, with the result that, in the words of Herbert Hauptman (a theoretical biophysicist at the Medical Foundation of Buffalo who shared the 1985 Nobel Prize in chemistry for his work on the phase problem in crystallography), "a few people will be happy, and many will be sad." Then there's the related problem pointed out by Caltech chemist William Goddard III at the January meeting: "It may be

Computerized Drug Design: Still Promising, Not Yet Here

When computer programs that allow researchers to simulate molecular structures in living color made their debut in the 1970s, pharmaceutical scientists thought they had a tool that would allow them to formulate new drugs on these computer screens without going through the laborious hit-or-miss process of sifting through thousands of compounds. But although recently it was reported that San Diego's Agouron Pharmaceuticals claims to have designed an anticancer drug from scratch on the computer, for the most part technicians must still screen many, many compounds to find their magic bullets. "There's been a lot of hubris in the last couple of dec-ades about sitting at a computer and designing a drug from scratch," says Michael Colvin, a physical chemist at Sandia National Laboratories. "We're not able to do that yet."



CENTER FOR COMPUTATIONAL ENGINEERING/SANDIA

Like a glove. Supercomputer-generated image of anticancer drug phosphoramide mustard binding to guanine nucleotides.

But Colvin and other computational scientists who work in the field of drug design haven't given up. In fact, they're hoping that a teraflop computer, which would be about 1000 times faster than current computers (see main story), might transform molecular modeling from a theoretical exercise into the wonder-working tool that researchers had hoped for from the beginning. Now, however, a computer is seen as something to be used in collaboration with experiment. "The way experimental and computational researchers seem to be working most efficiently now is together," says Susan Ludeman, a physical-organic chemist at Johns Hopkins Oncology Center.

Ludeman knows that kind of working together at first hand. She is part of a group headed by oncologist O. Michael Colvin (father of Sandia's Colvin) that is collaborating with Sandia to study the molecular dynamics of cyclophosphamide, an anticancer drug. The Johns Hopkins researchers alter the drug's biological potency by changing its chemical structure. Meanwhile, Sandia computational biologists use supercomputers to see how the structural changes affect the interaction between cyclophosphamide's active metabolite, phosphoramide mustard, and DNA.

According to William Wilson, manager of the Center for Computational Engineering at Sandia laboratories, recent calculations on phosphoramide mustard cross-linked to DNA are among the largest electronic-structure calculations ever performed. Coupled with molecular dynamics simulations done at the University of Texas at San Antonio, they have helped identify the structural features of phosphoramide mustard that account for its activity, says Ludeman.

These results haven't convinced some pharmaceutical scientists that faster computers could ever turn drug design into a field based primarily on equations. Drug design "will never become an engineering problem," says Peter Gund, senior director of chemical and biological systems at Merck Inc. of Rahway, New Jersey. But Gund agrees that simulations have their place. Indeed, Merck employs 24 scientists to perform simulations, a pursuit that Gund calls the "third paradigm of science that illuminates theory and drives experiment."

Even modest gains in the ability to predict drug activity from structural data will be enough to delight some computational biologists. "Developing drugs is a vague science in which you synthesize a large number of compounds," says Klaus Schulten, a theoretical biophysicist at the University of Illinois. "Even guiding [drug designers'] minds a little would help greatly."

–R.S.

that whoever selects the projects that go on the teraflop computer picks the wrong ones."

In view of these problems, many computational biologists think it's best to wait and let computing power get a lot cheaper before the field leaps in and commits itself to one big machine. "It would be possible now to put together something that could compute that fast, but it wouldn't be a balanced machine," says Ten Eyck. This year industry might be capable of building a teraflop computer, he says, but it's likely that some of such a machine's components wouldn't be technically advanced enough to keep it running at peak speed for more than a fraction of the time. By 1995, however, several firms are expected to offer well-balanced teraflop computers and the price should be much less than current estimates, says Rick Stevens, director of math and computer science at Argonne National Laboratory. According to Ten Eyck, the San Diego Supercomputing Center is looking at several supercomputers that might be expandable to teraflop speed by 1995.

Concerns over the hardware's cost and quality apply to software, too, says Peter Wolynes, a theoretical biophysicist at the University of Illinois. "A lot of thinking will have to go into how to utilize all the masses of data that will come out of [a teraflop computer]," he says. Scientists in computational fields already have begun developing such software, adds Stevens. For example, he says, atmospheric scientists "are spending enormous energy getting retooled to take advantage of teraflop machines" in order to do computationally intense climate modeling, he says. Computational chemistry, automotive design, and high-energy physics are among the fields that "aren't standing still either," Stevens says.

With computational biologists at odds over nearly every issue engendered by a teraflop computer, it should come as no surprise that they don't all think Florida State is the logical place to put one. "It's not obvious to me that one should start a separate center from the National Science Foundation centers," says Wolynes. Another concern raised by some computational biologists is Bash himself. "A number of individuals have raised a concern about Paul's experience," one computational biologist told *Science*.

But some prominent scientists think that location and track record may not be the key things. "What I think is important is to have [a teraflop machine] in a place where people are hungry, interested in the problems, and ready to get on with the job," says Frederic Richards, a structural biologist at Yale University. Zerner agrees: "Location isn't crucial—I flip a switch and I can log onto Florida State." And some scientists outside computational biology think Bash could be just the person for the job. "As much as there's the technological problems to solve, there's the sociological problems—people used to workstations are constrained by that mindset," says Stevens. "Paul's trying to expand that mindset, get people thinking about what sort of problems can be solved on a teraflop."

Whether or not Bash persuades funding officials—not to mention other computational biologists—that his proposal will lead the field down the right path, computational biology already is beating a trail toward a teraflop computer. It's just a question of when—and how—the teraflop revolution takes place.

–Richard Stone

Model of Computing's Future?

Behind the veneer of yellow and green lights that flash when data is transferred between its 528 microprocessors, the 16-foot-long Touchstone Delta supercomputer looks like *2001's* monolith tipped on its side. But poised behind the computer's physical setup is another structure that, although it's not visible, is just as crucial to the machine's success: the Concurrent Supercomputing Consortium (CSC), which owns the Delta. The CSC is a group of 10 universities and national laboratories, as well as Intel Corp., which made the machine. And that organization could provide a model for computational biologists who might want to merge their funds and buy a teraflop computer (see main story).

"We all hoped that by forming this consortium we'd be getting a computational resource we couldn't individually get," says Paul Messina, a Caltech computer scientist and executive director of CSC. The consortium formed in November 1990; by last May, the Delta, capable of a record-breaking speed of 32 gigaflops, was up and running at Caltech.

Not only have computational scientists got the machine they wanted, but "big chunks of time are getting allocated to individual projects," says Messina, who sits on the machine's time-allocation committee. The reason is that there are relatively few users, and the result is that consortium members (who get time in proportion to their financial contribution to the purchase of the $15 million supercomputer) are pleased. "If you open it up to everyone, it's no longer a supercomputer," says Hans Kaper, a senior mathematician at Argonne National Laboratory, one of the consortium members. And that same tune—purposeful organization in the service of efficiency—is a tune computational biologists hope to be singing if they get their teraflop machine.

–R.S.

MEETING BRIEFS

Chemists Storm San Francisco

This year's Spring meeting of the American Chemical Society (ACS), held 5 through 10 April in San Francisco, was so huge that it took three sizable volumes–a total of 5 pounds, 1.5 ounces of paper–to hold the 6200 abstracts. Like a city within a city, 15,943 attendees scurried to and from social gatherings and sessions that offered something for every taste, from nuclear waste disposal to the birth of the solar system.

Superconductors That See Red, Green, and Blue

What do you get when you mix a high-temperature ceramic superconductor with biochemicals that capture light of specific wavelengths? Most chemical novices will get an ugly pile of grit, but in the hands of chemist John T. McDevitt and his colleagues at the University of Texas at Austin, the combo becomes a color-sensitive optical detector. The researchers' aim is to come up with new optoelectronic devices for, say, detecting faint light signals in astronomy or defense or even sophisticated data-storage devices.

McDevitt and his colleagues are following a trail laid in the late-1980s by workers who used thin films of high-temperature superconducting ceramic as sensitive, rapid-response light detectors. The detectors work like this: Light warms up the superconductor and degrades its ability to conduct electricity without resistance. Light thus causes a change in conductivity, which is easy to monitor. But these devices respond indiscriminately to any wavelength from ultraviolet through visible into the infrared. "We wanted to make devices that respond more selectively," says McDevitt.

That's where the light-sensitive biochemicals come in. McDevitt and his colleagues liked the idea of using porphyrins—a family of molecules whose most famous derivative is chlorophyll, the light harvesting pigment of photosynthesis. Because of their ability to capture photons of specific wavelengths, McDevitt thought porphyrins would be just the thing for making the superconducting light detectors more choosy.

Atop a thin film of ceramic superconductor such as yttrium barium copper oxide, the Austin researchers fashioned small superconducting junctions and then coated them with porphyrin-based dyes and other organic pig-