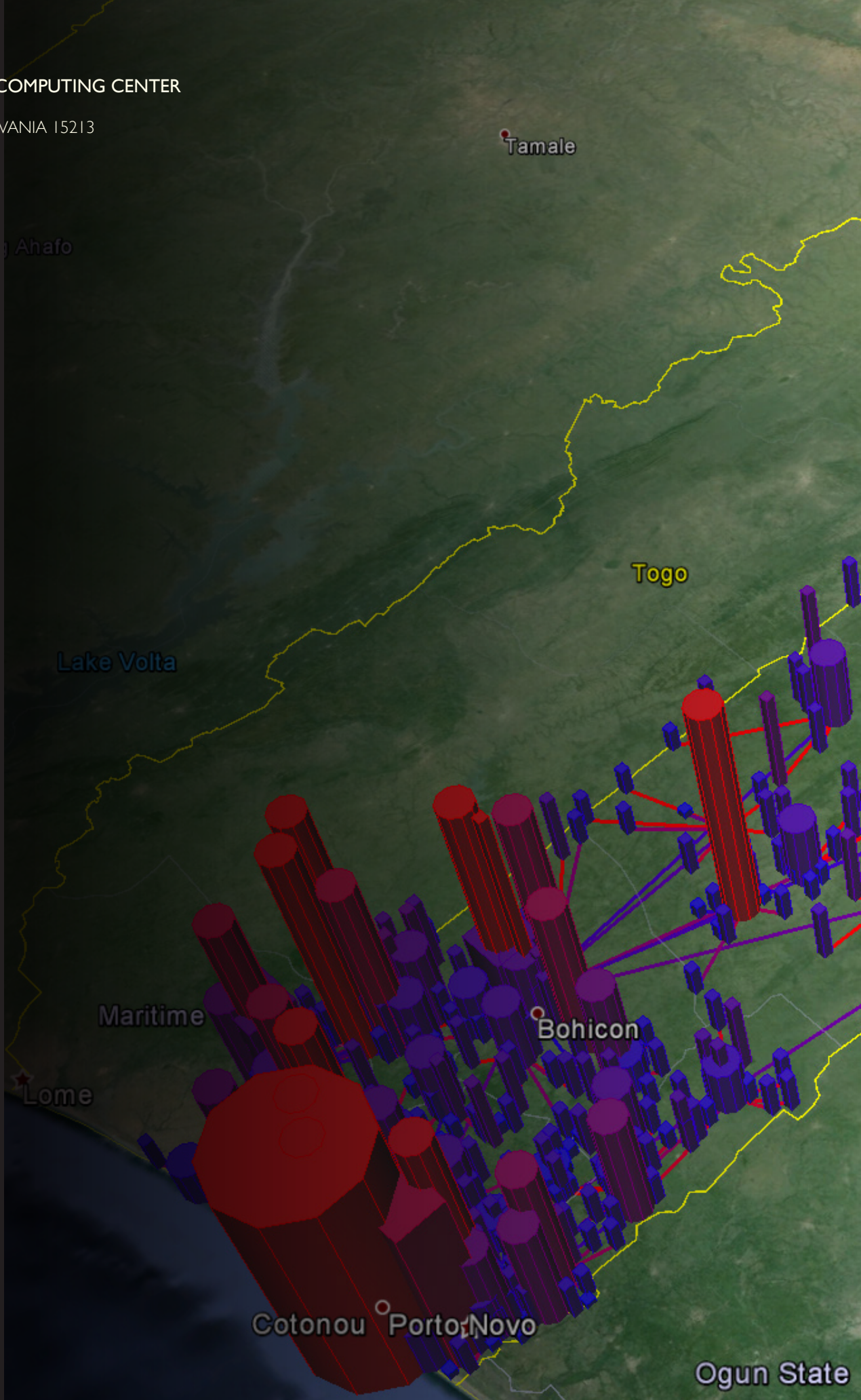
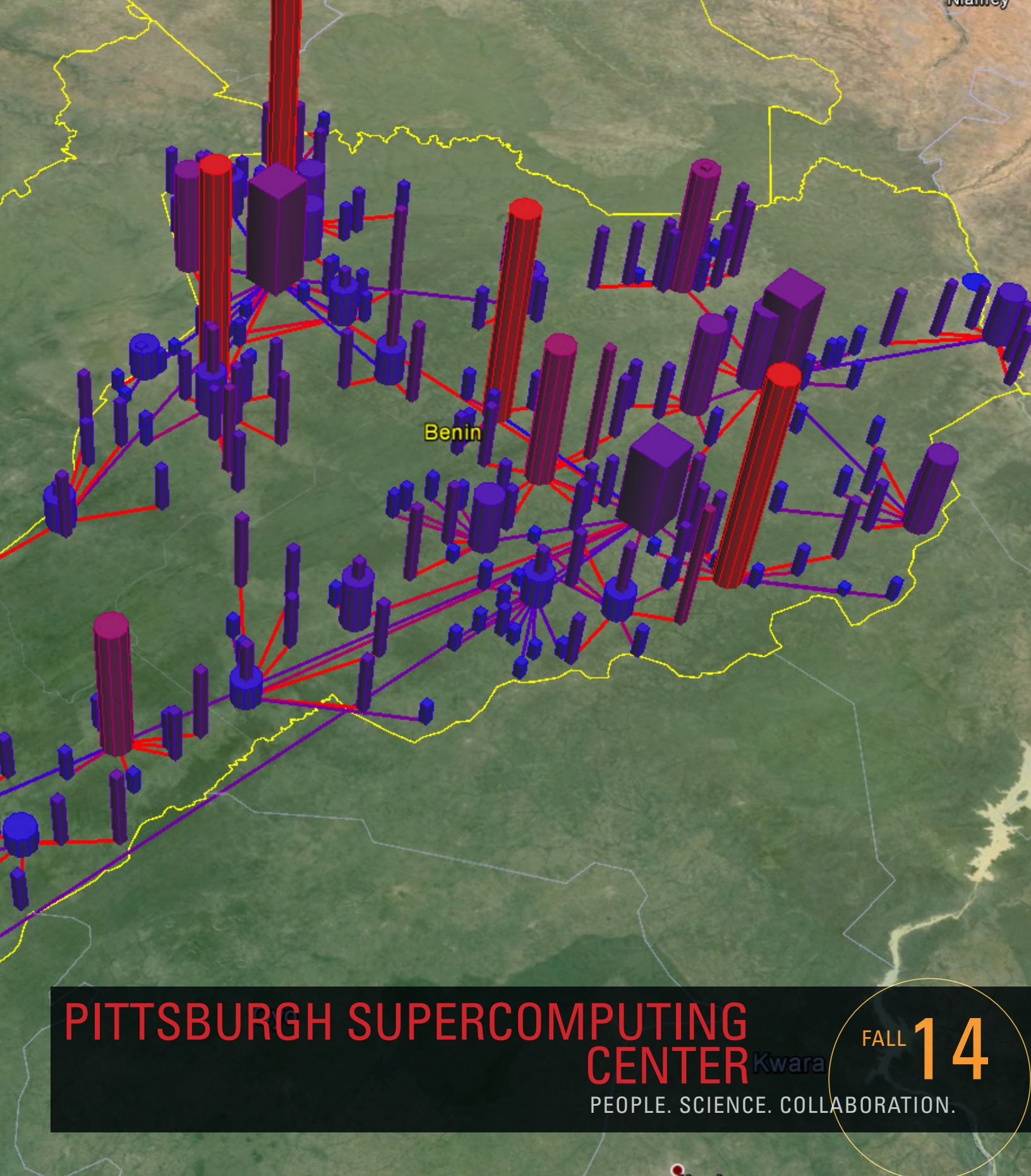
PITTSBURGH SUPERCOMPUTING CENTER
300 S. CRAIG STREET
PITTSBURGH, PENNSYLVANIA 15213

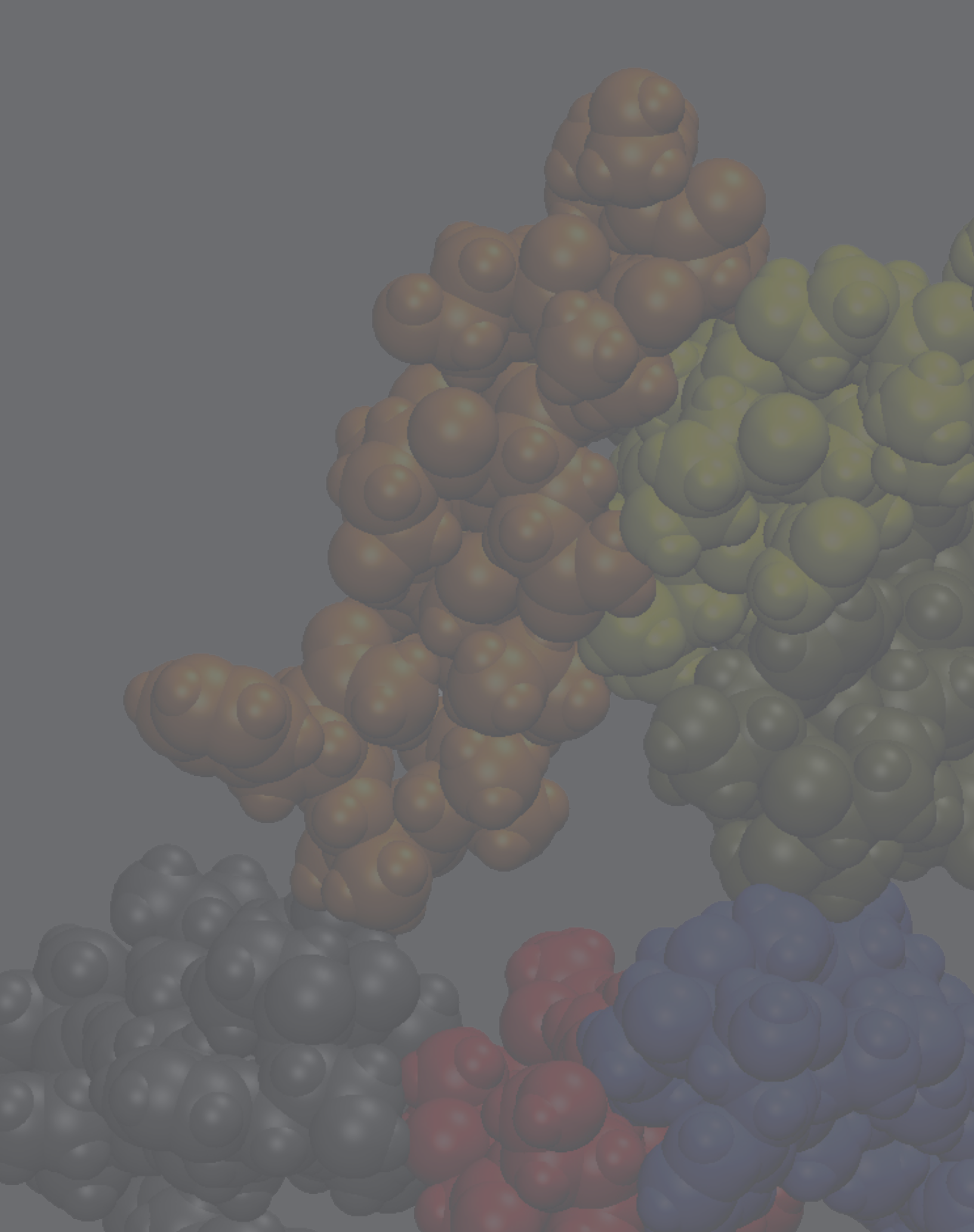


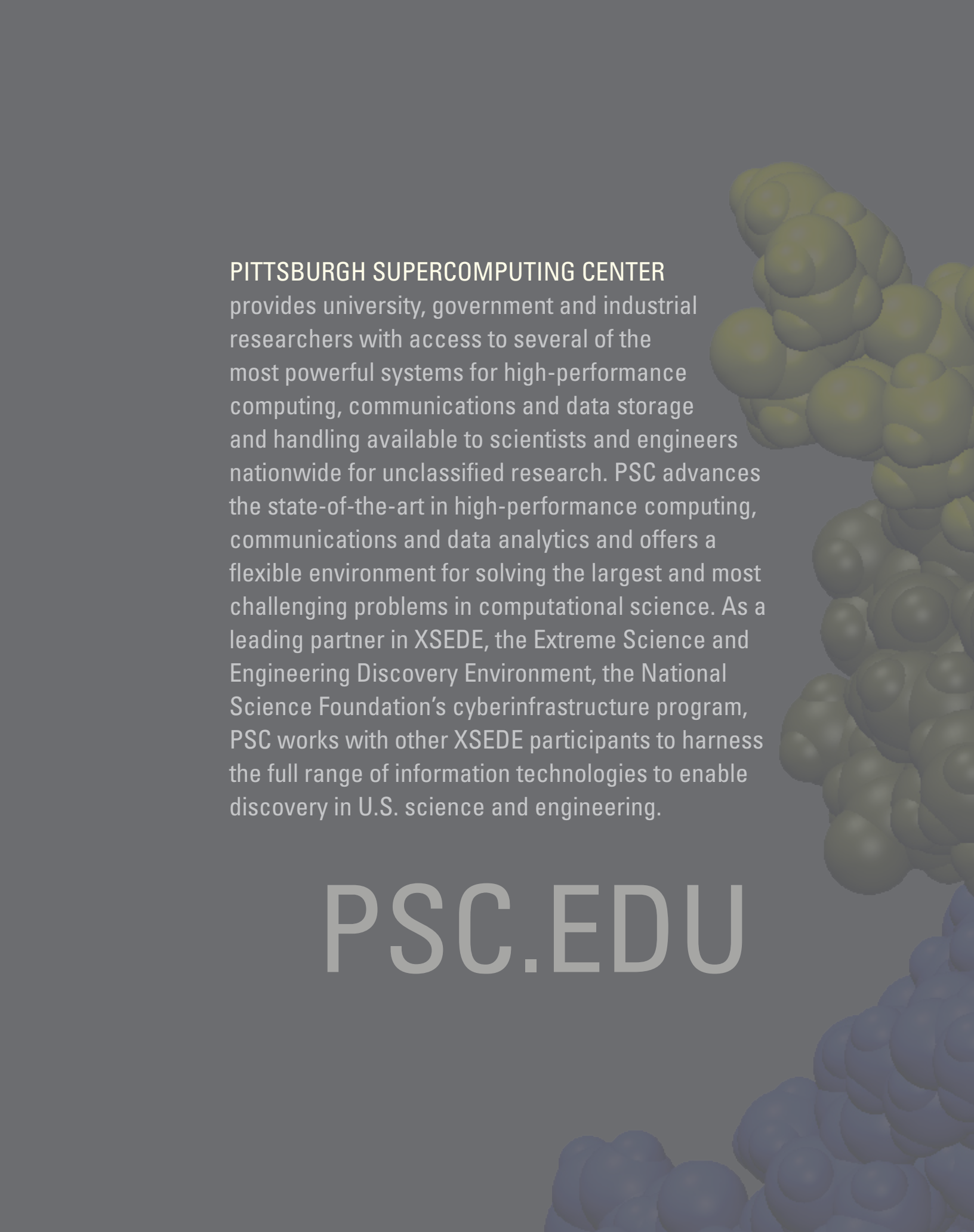


PITTSBURGH SUPERCOMPUTING CENTER

PEOPLE. SCIENCE. COLLABORATION.

FALL 14





PITTSBURGH SUPERCOMPUTING CENTER provides university, government and industrial researchers with access to several of the most powerful systems for high-performance computing, communications and data storage and handling available to scientists and engineers nationwide for unclassified research. PSC advances the state-of-the-art in high-performance computing, communications and data analytics and offers a flexible environment for solving the largest and most challenging problems in computational science. As a leading partner in XSEDE, the Extreme Science and Engineering Discovery Environment, the National Science Foundation's cyberinfrastructure program, PSC works with other XSEDE participants to harness the full range of information technologies to enable discovery in U.S. science and engineering.

PSC.EDU

FALL 2014

CONTENTS

FROM THE DIRECTORS

Ralph Roskies (left) and Michael Levine, PSC co-scientific directors

WELCOME TO THE FALL 2014 ISSUE OF PSC'S BIANNUAL REPORT!

The stories it contains reflect our expanding focus on a burgeoning need in modern research: data-driven analysis.

"Traditional" supercomputing focuses on scientific problems that simulate conceptually simple phenomena: How does air flow over an aircraft? While such simulations can create large amounts of output data, the focus is on the computational speed needed to work through complex equations.

Increasingly, a new generation of supercomputing researchers is tackling problems that are far more constricted by a limited ability to handle huge quantities of data than by the limits of processing speed:

- Biologists are taking the vast amount of information encoded in a biologically and medically important crustacean's DNA, comparing related species to find out what underlies their different responses to—and ability to transmit—the cholera bacterium (*p. 18*).

- Public health researchers are employing huge data input *and* output volumes to simulate the individual components—down to the last motorcycle—in a nation's vaccine supply chain, discovering means for introducing new vaccines and reducing costs without disrupting delivery of the currently provided vaccines (*p. 4*).

- We're helping researchers through a highly successful series of workshops on the major programming languages of supercomputers (*p. 10*). We're also fostering the next generation of researchers by helping high school teachers add bioinformatics to their curricula (*p. 20*).

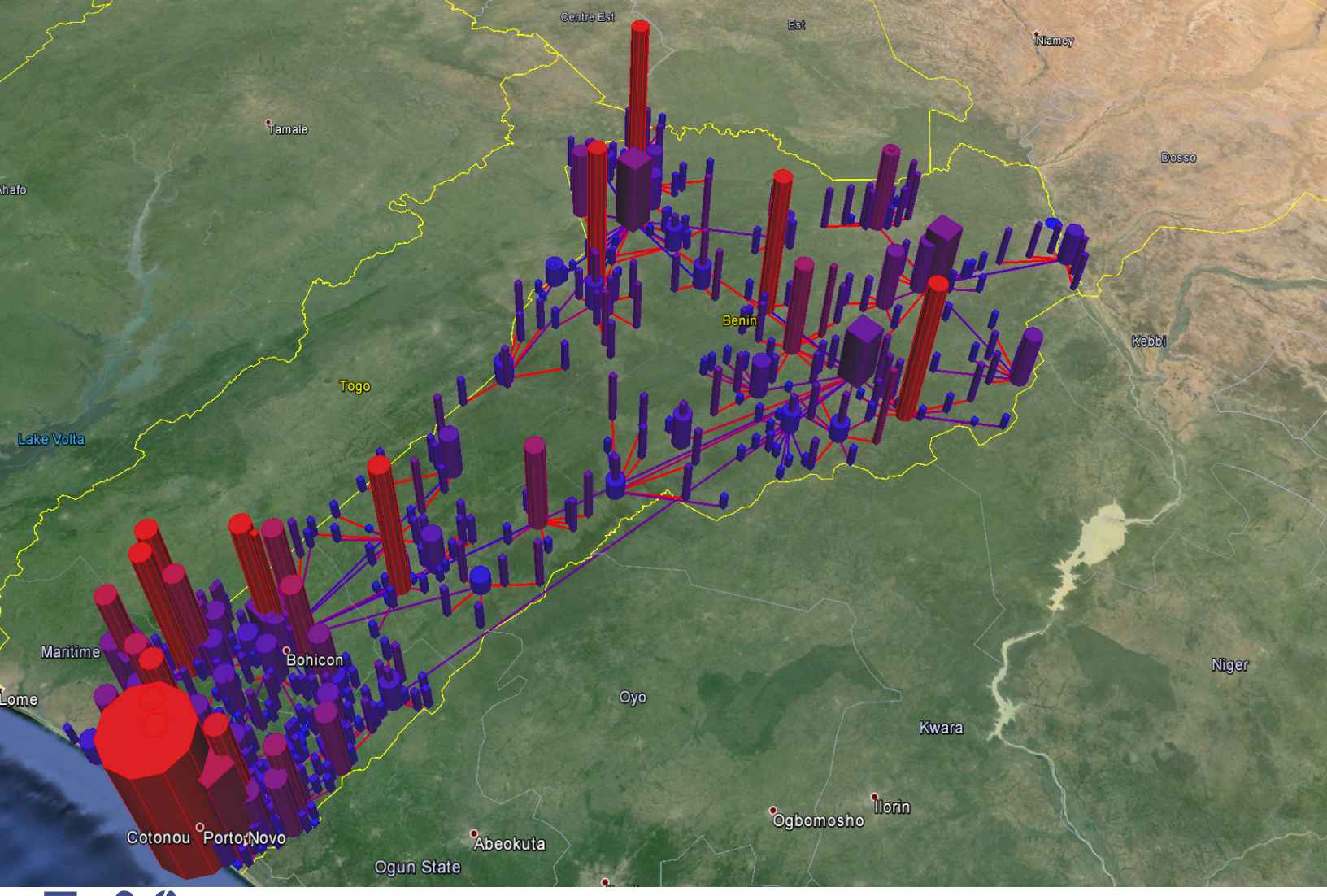
- In addition, researchers are using PSC resources to pursue problems such as understanding turbulent airflow over power-generating wind turbines (*p. 24*) or the functional consequences of how complex biomolecules move (*pp. 6 & 14*) — "traditional" supercomputing problems requiring large memory.

As analyzing, retrieving, moving and storing data have become central to a growing cadre of new supercomputing scientists, we have developed new hardware and software to overcome those challenges (see our list of resources on *Inside Back cover*).

Our staff members, as always, have been at the forefront of these efforts, their ingenuity and expertise making it all work.

These projects have also been made possible by a number of funding organizations and agencies. The Commonwealth of Pennsylvania has once again supported our educational, industrial outreach and operational programs. Our latest NSF grant will fund an enhancement of our Data Exacell (DXC) pilot project, itself announced in our Spring 2014 issue, to allow researchers to structure and manage their data for rapid analysis by PSC's Big Data analytical resources (*p. 26*). A new National Institutes of Health BD2K (Big Data to Knowledge) grant to the University of Pittsburgh will employ our computational expertise in using causal analysis to identify clinically important relationships without the attendant risk of false correlations inherent in working with large volumes of data. New grants will also support exciting collaborations between our Public Health Group and groups at Pitt studying influenza vaccination and surveillance to detect disease outbreaks.

We hope you find the following accounts of our efforts over the past 6 months to be enjoyable and edifying. We would like to hear any feedback you had, on our work or this publication. You can send any comments or suggestions via our feedback page at *psc.edu/index.php/feedback*. You can also contribute to PSC's nonprofit, academic mission at *psc.edu/donate*.



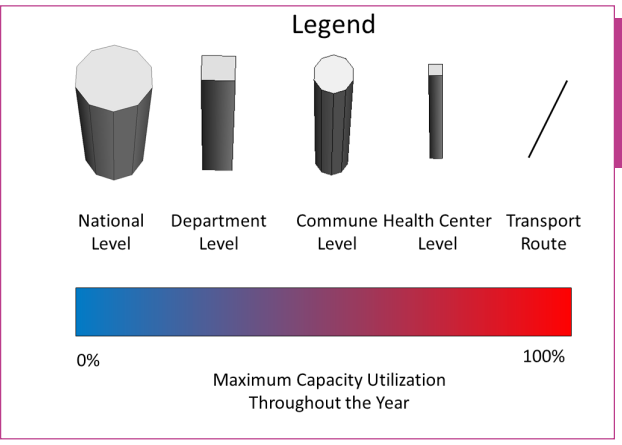
Life, through the Details

PSC, Johns Hopkins Computer Model Saving Lives through Details of Vaccine Supply

Rotavirus is a killer of children in much of the world. About half a million kids die of rotavirus-caused diarrhea every year, particularly in Africa and South Asia. It's a big reason diarrhea is the third leading cause of child deaths worldwide and why introducing rotavirus vaccine to many countries is so important.

Introducing lifesaving vaccines such as the one for rotavirus immunization could save millions of children's lives. It's an obvious international public health goal. But these vaccines have to get to children to help them. In countries that don't have an extra penny to spare, a new vaccine could easily overburden the system, crashing distribution of other vaccines and *increasing* sickness, suffering and death. To save lives with a new vaccine, you have to work out all the *details*.

PSC's Public Health Applications Group, working as part of the HERMES Logistics Modeling Team with researchers at the University of Pittsburgh School of Engineering and the Johns Hopkins Bloomberg School of Public Health, have helped the Republic of Benin in West Africa understand and control those details.



The HERMES model of the Republic of Benin's vaccine supply chain revealed depots and supply routes that were already operating at capacity as well as those that were under-utilized.

In a May paper in the journal *Vaccine*, they used their HERMES computer model to show how the Benin Ministry of Health could improve their ability to supply vaccines to their population, possibly saving $500,000 dollars through the year 2017 and adding a rotavirus vaccine in the bargain.

It's a "win-win" that the ministry has already successfully piloted in small scale and is now considering enacting nation-wide.

HERMES: TESTING THE OPTIONS

The HERMES Logistics Modeling Team has developed HERMES (Highly Extensible Resource for Modeling Supply-chains), a software platform that allows users to generate a detailed discrete event simulation model of any supply chain. This simulation model can serve as a "virtual laboratory" for policy makers, health officials, funders, investors, vaccine and other technology developers, manufacturers, distributors, logisticians and researchers. HERMES allows them all to address a variety of questions in supplying vaccines as well as other health supplies.

"The new *Vaccine* paper outlines our engagement with the Benin Ministry of Health in which we worked to choose among some key redesign options of their vaccine supply chain," says Shawn T. Brown, PhD, Director of Public Health Applications at PSC, Technical Lead of the HERMES team and first author. "It's a clear use of computational modeling and simulation to help a government figure out how to get the vaccines that are so desperately needed to every child they can."

The team used HERMES to evaluate different options of redesigning the Benin vaccine distribution system being considered by the country's Ministry of Health. These included maintaining the current national system; consolidating the nation's system of 80 "commune-level" supply depots to a system of 34 "health-zone" depots; eliminating the commune level entirely; and splitting the seven current high-level "department" and "regional" stores into 12 new departments. The investigators also evaluated the effects of changing transportation routes. In addition to optimizing delivery and storage of current vaccines being offered, the HERMES model helped show the impact of adding the rotavirus vaccine.

FINDING THE WIN-WIN

The computational model favored the health zone approach along with changing transportation routes. Together, these could save between $50,000 and $70,000 in initial expenses and $50,000 to $90,000 in annual costs compared with the other scenarios, while still reaching 99 to 100 percent of children. Through 2017, the improved plan would save Benin over $500,000 in total costs while improving vaccination rates.

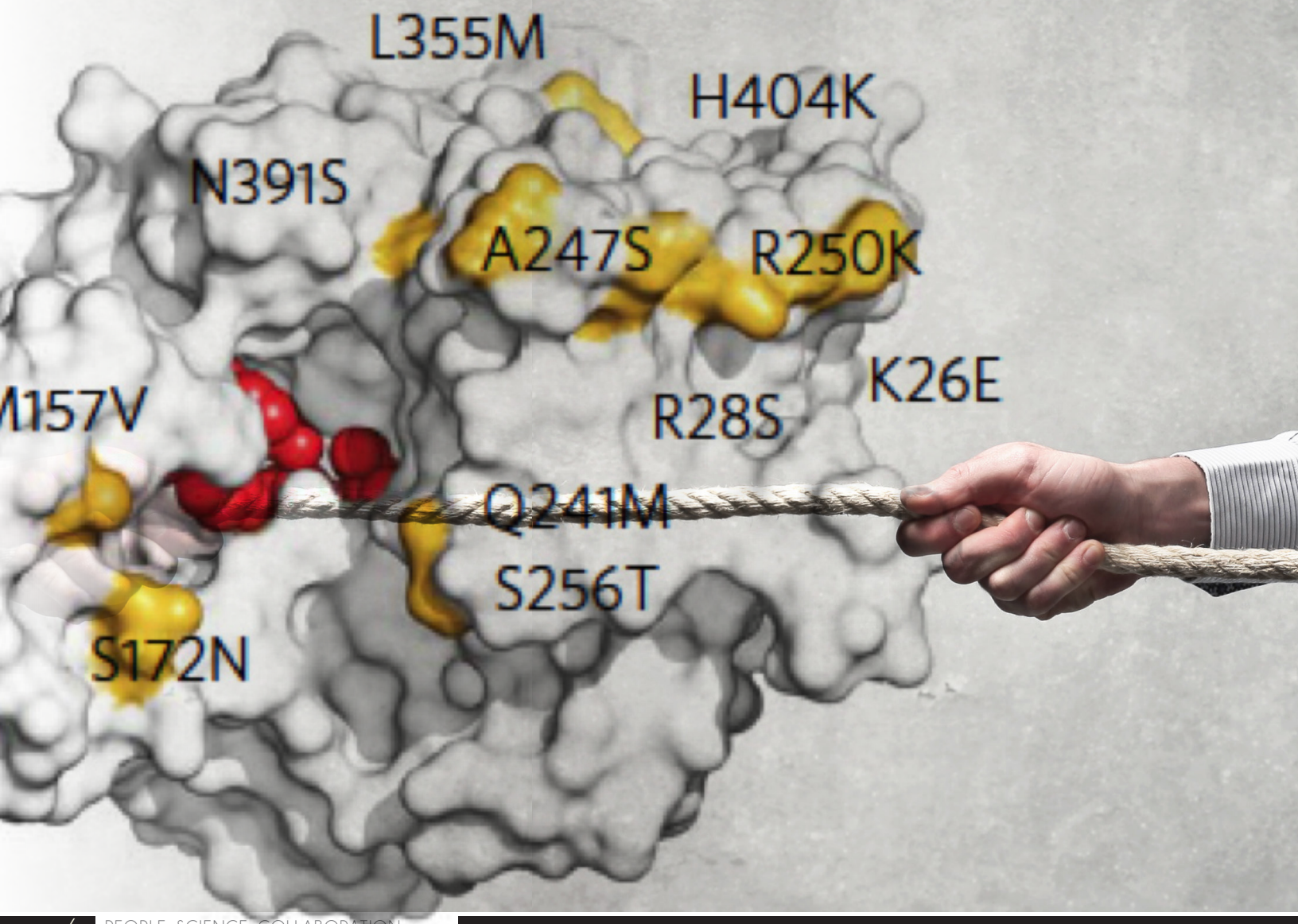
With the support of the Agence de Médecine Préventive, which participated in all steps of the project, UNICEF, the World Health Foundation and PATH, the Benin Ministry of Health has launched a successful demonstration following the group's pre-publication recommendations. The nation is now considering scaling the new system to the entire country.

"These are major policy decisions that could affect millions of lives," says Bruce Y. Lee, MD, MBA; Associate Professor of International Health and Director, Operations Research, International Vaccine Access Center, Johns Hopkins Bloomberg School of Public Health; Scientific Lead of the HERMES team; and co-author. "Our team's goal has been to develop state-of-the art computational technology to help such decision-making around the world."

"Life is not a plot; it's in the details." –Jodi Picoult

A TUG Felt Elsewhere

Anton Simulations Show How Drug-Producing Enzyme Is Enhanced by Changes Far from Reactive Site



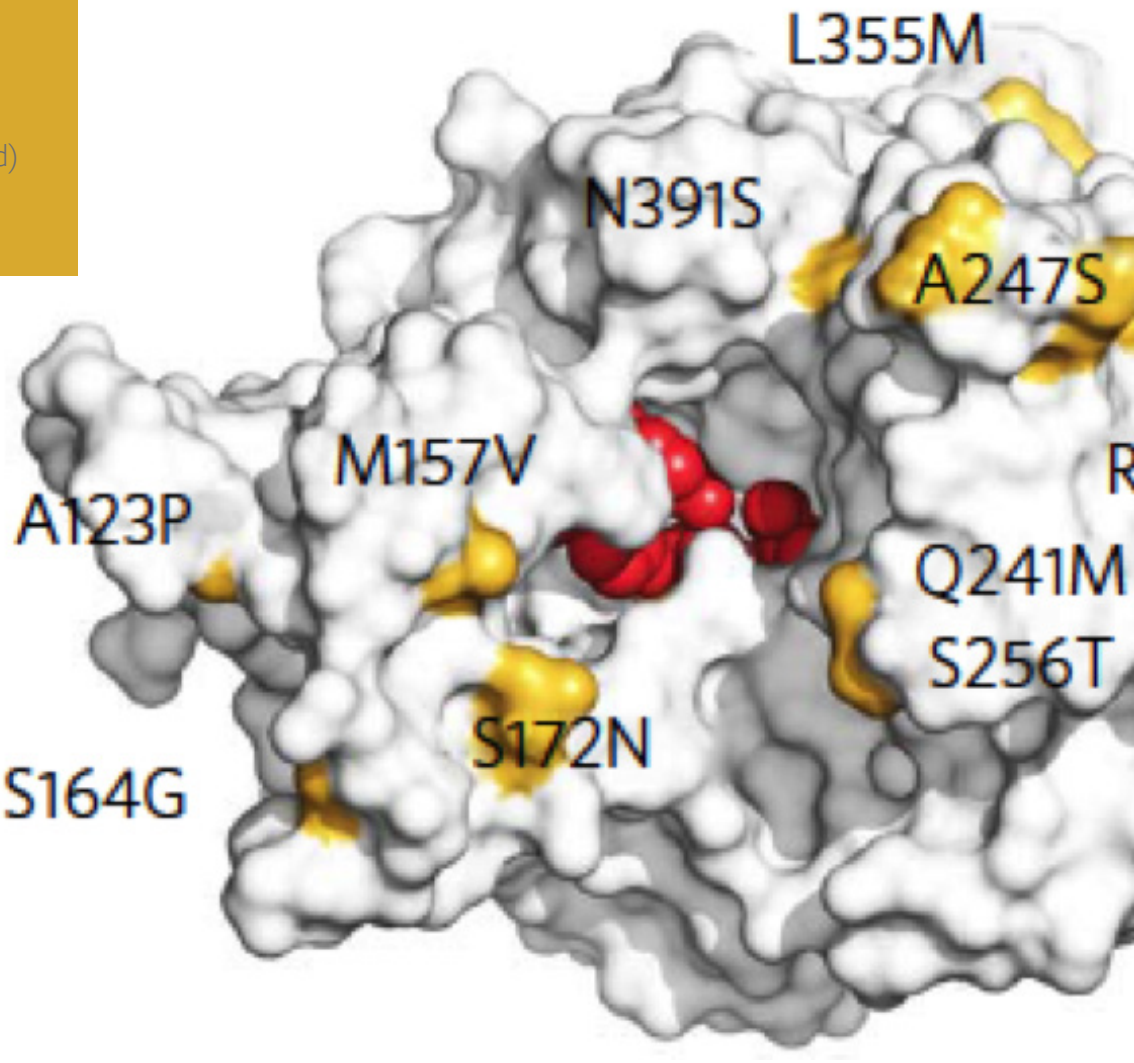
"No, no, no, don't tug on that.
You never know what it might be attached to."
—Buckaroo Banzai

Sometimes, getting what you want means giving a little tug, far from where your first instinct tells you.

Such is the case with the enzyme LovD, which performs a critical step in producing a family of cholesterol-lowering drugs.

Gonzalo Jiménez-Osés, working with Silvia Osuna and Xue Gao in the laboratories of Kendall N. Houk and Yi Tang at the University of California Los Angeles, investigated whether it was possible to engineer a better LovD to produce the medication simvastatin. Using a special-purpose supercomputer at PSC called Anton, the researchers found that the secret lay in modifying the enzyme in a location far distant from the enzyme's "active site." The active site is the part of an enzyme that directly interacts with and transforms the substrate—the enzyme's "raw material" molecule.

The Anton supercomputer was developed and made available without cost by D. E. Shaw Research. Support for operational funding was provided by the National Institute of General Medical Sciences via the National Center for Multiscale Modeling of Biological Systems.



Mutations (gold) far from the active site of LovD (red) affect the enzyme's activity.

A GREEN PATH, BUT DIFFICULT TO FOLLOW

For many reasons, enzymes—proteins that catalyze, or speed up, chemical reactions—are preferred over non-biologically derived catalysts for industrial processes. Enzymes are far more efficient. They are exquisitely specific in speeding a desired reaction but not side-product reactions. And they are inherently "green," not requiring toxic solvents or greatly toxic themselves. Unfortunately, designing tailor-made enzymes to carry out industrially important reactions has been harder than biochemists had initially hoped.

The natural version of LovD can only work on a substrate that is attached to another protein, LovF. This second protein is necessary in living cells. But it greatly increases the complexity of the reaction in a test tube. In turn this makes the reaction difficult to scale up for industrial use.

Jiménez-Osés used Anton at PSC to predict how mutations in the protein's amino acid code would change LovD's structure, particularly for explaining how mutations that removed the requirement for LovF worked. He performed the simulations in a back-and-forth investigation with his colleagues' laboratory experiments.

EXPLAINING THE MUTANTS

The researchers began by creating a set of mutations of the LovD enzyme in the laboratory using "directed evolution." In this method, they randomly mutated the protein. Then they selected mutants that were more active in the absence of LovF. In a series of these steps, they created versions of LovD that were more and more effective alone. Then they set out to understand why those versions worked without LovF's help using simulations on Anton.

"Previously researchers had tried to understand the LovD/LovF interaction using static X-ray structures," Jiménez-Osés says. "But they couldn't explain what was going on between some mutant enzymes that were active and some that were not. The static state was not telling the whole story."
In the Anton simulations, Jiménez-Osés was able to monitor how the structure of the protein moved with time. Because Anton was designed to dramatically increase the speed of molecular dynamics calculations, it offered simulations of far longer time periods than possible with conventional supercomputers.

"Transitions between the active and inactive conformations of the protein take hundreds of nanoseconds to occur," Jiménez-Osés says. "A non-specialized supercomputer would have required one or two months for each such simulation." Using Anton, though, the team was able to produce simulations as long as a thousand nanoseconds in one or two days, he adds.

AN AMAZING TRANSFORMATION

The long-timescale simulations revealed an amazing transformation. Without the LovF "helper protein" to stabilize it, an arm of the un-mutated LovD that is present in the enzyme's active site tended to move away at about 800 nanoseconds. This motion disturbed the active site and made it unable to act on the substrate.

What's more, LovF accomplished this stabilization task by interacting with a part of LovD that is far from the active site. One reason for the 800-nanosecond delay was that it took time for the disruption of losing LovF stabilization to work its way through the protein's structure to the active site.

The group reported their results in *Nature Chemical Biology* in April 2014.

Jiménez-Osés is quick to point out that the team has not yet produced an industrial-grade product, though their work is on its way toward that goal. "Ultimately, you want to keep the active site as untouched as you can and modify positions quite far away to create the changes you need," he says. "It's a neat finding, but really complicated to enact. It represents a new paradigm in enzyme design."

It Isn't Rocket Science

PSC Group Brings Supercomputer Programming within Reach of Researchers

We'll start out with what's only a slight exaggeration: many brain surgeons and rocket scientists think programming a supercomputer is really, really hard.

Which is a shame, according to John Urbanic, PSC parallel computing specialist and a lead instructor in the National Science Foundation's XSEDE supercomputing network's programming workshop series.

"Some of the technology—in particular MPI—has a reputation for being very intimidating and difficult to use," Urbanic says. MPI, or message passing interface, is the language necessary to split computational instructions for use by supercomputers with tens of thousands of processors. "That can turn off researchers who aren't computer scientists and just want to do their projects. But at every workshop, we take 150 people and convince them otherwise."

Obviously, the students at the XSEDE workshops are highly motivated, intelligent researchers and engineers. Moreover, in many fields, using supercomputers is no longer a "nice to have:" it's required to remain at the cutting edge of research.

Helping scientists realize that despite sometimes fearsome reputations, supercomputing languages and technologies are well within their reach has been a central mission for Urbanic and his PSC team. Tom Maiden, PSC user support and outreach coordinator, handles class logistics; Steve Cunningham, PSC network administrator, ensures the smooth operation of the workshops' connections to remote classrooms at other XSEDE supercomputing sites across the country.

The group teaches several different workshops. A two-day class tackles MPI, whose classic applications include weather and climate prediction and modeling, aerodynamic testing and visualizing brain circuits.

Another, one-day class focuses on OpenMP, a language optimized for supercomputers that have "shared memory," like PSC's Blacklight. Such supercomputers are quickest for solving scientific problems that require many processors to work off the same, large pool of data, such as genome sequencing, cosmology simulations and new applications like extracting data from handwriting in cursive script.

"It was clear that everyone at PSC had thought through all these details and designed a great workshop with the goal of demystifying MPI and encouraging users to write parallel code."

A one-day OpenACC workshop focuses on a third major supercomputing language used to program "heterogeneous" supercomputers. These supercomputers trade calculations between traditional central processing units and GPUs, or graphics processing units. The latter were developed to process the sophisticated images in modern computer games. But they have found an expanded use in heterogeneous computing, tackling complex scientific problems in which separate parts of the problem run faster in each of the two types of processor.

A new workshop began this year in association with YarcData, manufacturer of Sherlock, PSC's novel supercomputer. Sherlock focuses on Big Data applications such as searching unstructured networks for critical information.

The idea is to offer practical workshops to get researchers up to speed in whatever flavor of supercomputing their projects require, in a realistic computational setting. PSC's membership in XSEDE and the workshop series' XSEDE setting allow the classes to leverage some of the premier supercomputers in the nation for each purpose.

"We do things with whichever machine is the best platform to teach students," Urbanic says. The workshops give OpenACC students access to Blue Waters at the National Center for Supercomputing Applications, a premier heterogeneous supercomputer. MPI classes employ Stampede at the Texas Advanced Computing Center, with over 100,000 processors. "And we teach OpenMP on the world's best shared memory machine, which happens to be Blacklight here at PSC."

"These workshops are well known for being of the best quality," agrees Galen Collier, a computational scientist at Clemson University in South Carolina. "Everybody looks forward to John's classes."

While Collier has his own research projects, his chief role is to help other researchers on the Clemson campus get up to speed in using supercomputing resources—whether Clemson's own Palmetto supercomputing cluster or XSEDE resources such as Blacklight. He says that PSC's workshops help him use his own time more effectively.

"Students and faculty pick up the basic concepts" at the workshops, he says. "John makes sure their code works; I can take them a step forward and make sure they're using a specific instrument efficiently."

"The course was really good because it had a lot of practical applications," says Mary Krauland, a research programmer at the University of Pittsburgh's Public Health Dynamics Laboratory. "In the practical component we actually got to do the programming ourselves on several problems. It was a really nice intro." Krauland models disease transmission as part of the Models of Infectious Disease Agent Study—MIDAS—at Pitt, PSC and other institutions.

"The workshop made MPI programming a lot less intimidating by focusing on the most essential functionalities," adds Berhane Temelso, a postdoctoral associate at Bucknell University. He models how clusters of molecules in the atmosphere grow into aerosols, with applications ranging from cloud formation to pollution effects to climate change. "It was clear that John and everyone at PSC had thought through all these details and designed a great workshop with the goal of demystifying MPI and encouraging users to write parallel code."

After several years of offering the workshops in person—traveling across the country to hold them at different XSEDE sites—this year the PSC group began offering them as well to distance learners. Sites within and outside XSEDE are now providing local equipment for video conferencing as well as on-site teaching assistants. The PSC group hosts about 12 of these sites monthly on a rotating basis. This year the classes also culminated in the first summer "boot camp" featuring all four classes.

PSC's John Urbanic

"We've always focused on the classroom environment and have been reluctant to scale up to something that might not reproduce that quality," Urbanic says. "It's very important that it be smooth, because we have zero tolerance for breakdowns. That's one of the reasons we waited so long to make this jump. With the advent now of robust video conferencing, we can reproduce that classroom experience that we're looking over your shoulder."

JUMBO SHRIMP

Tiny Copepod Offers Outsized Insight on Cholera Spread

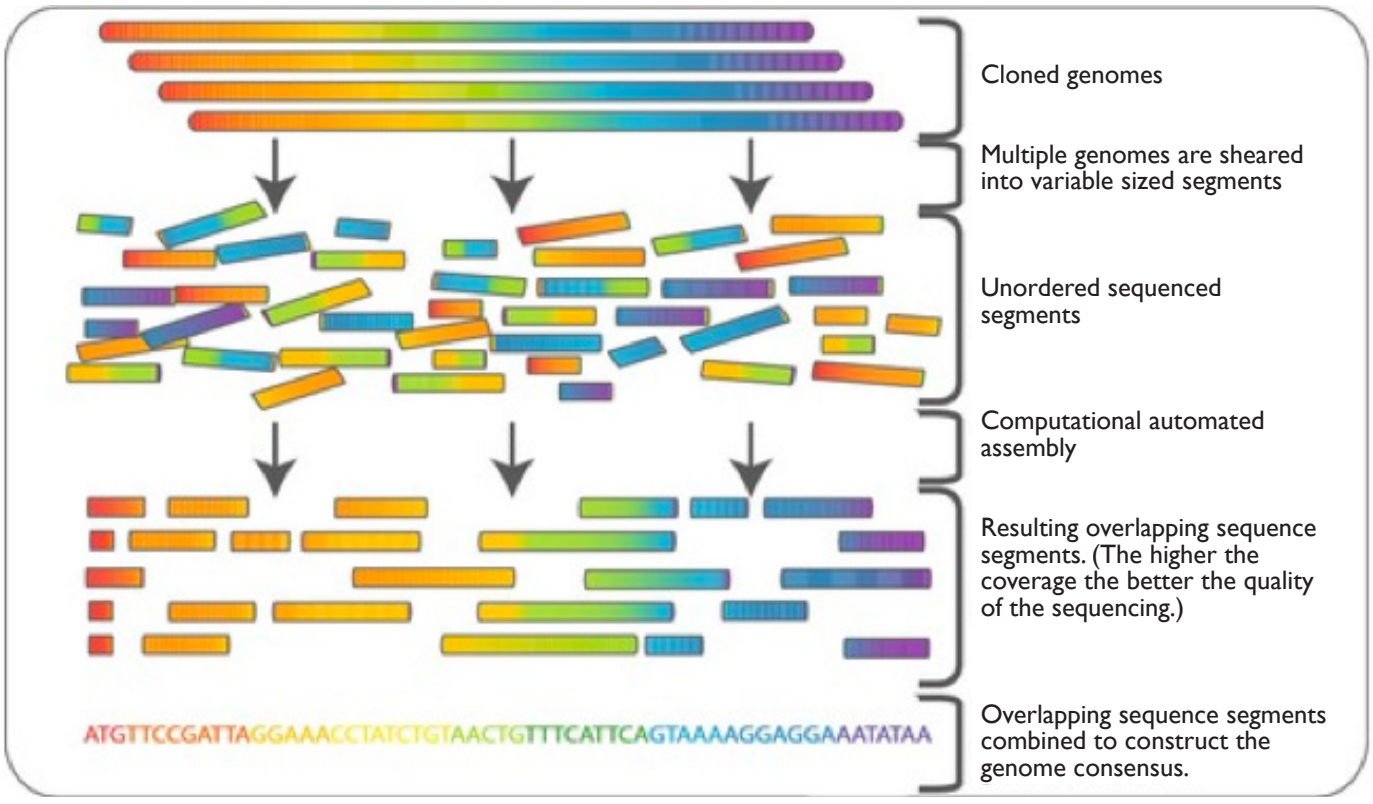
WHY IT'S IMPORTANT

The tiny copepod—a shrimp about the size of a grain of rice—carries outsized importance in the Earth's food web. These crustaceans are so abundant that they are the largest animal biomass on the planet. They are a fundamental food source for aquatic organisms ranging from tiny fish to baleen whales. But they are also important to human health, as they are host and conveyance for *Vibrio cholerae*, the deadly cholera bacterium.

Amalia Aruda Almada, working in the laboratory of Ann Tarrant at the Woods Hole Oceanographic Institution, set out to understand how a copepod species that lives in brackish waters reacts to *Vibrio* bacteria and whether and how it fights being infected by *Vibrio*.

"We're still in the process of analyzing our data, but we're seeing that the copepods are actually able to recognize and distinguish between the type of bacteria colonizing them … They really seem to have an impact on microbial activity."

—Amalia Aruda Almada, Woods Hole Oceanographic Institution



From Wikimedia Commons: Commins, J., Toft, C., Fares, M. A. "Computational Biology Methods and Their Application to the Comparative Genomics of Endocellular Symbiotic Bacteria of Insects." Biol. Procedures Online (2009).

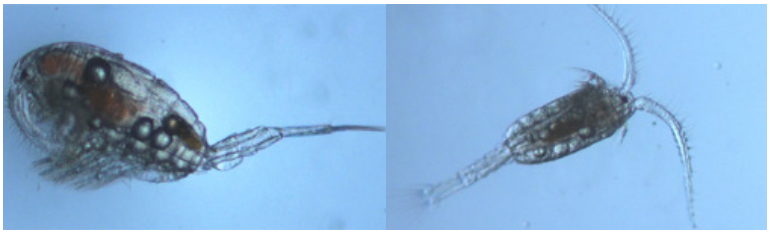


Figure. (above)

Aruda employed Blacklight to perform whole genome sequencing on the copepod genetic code.

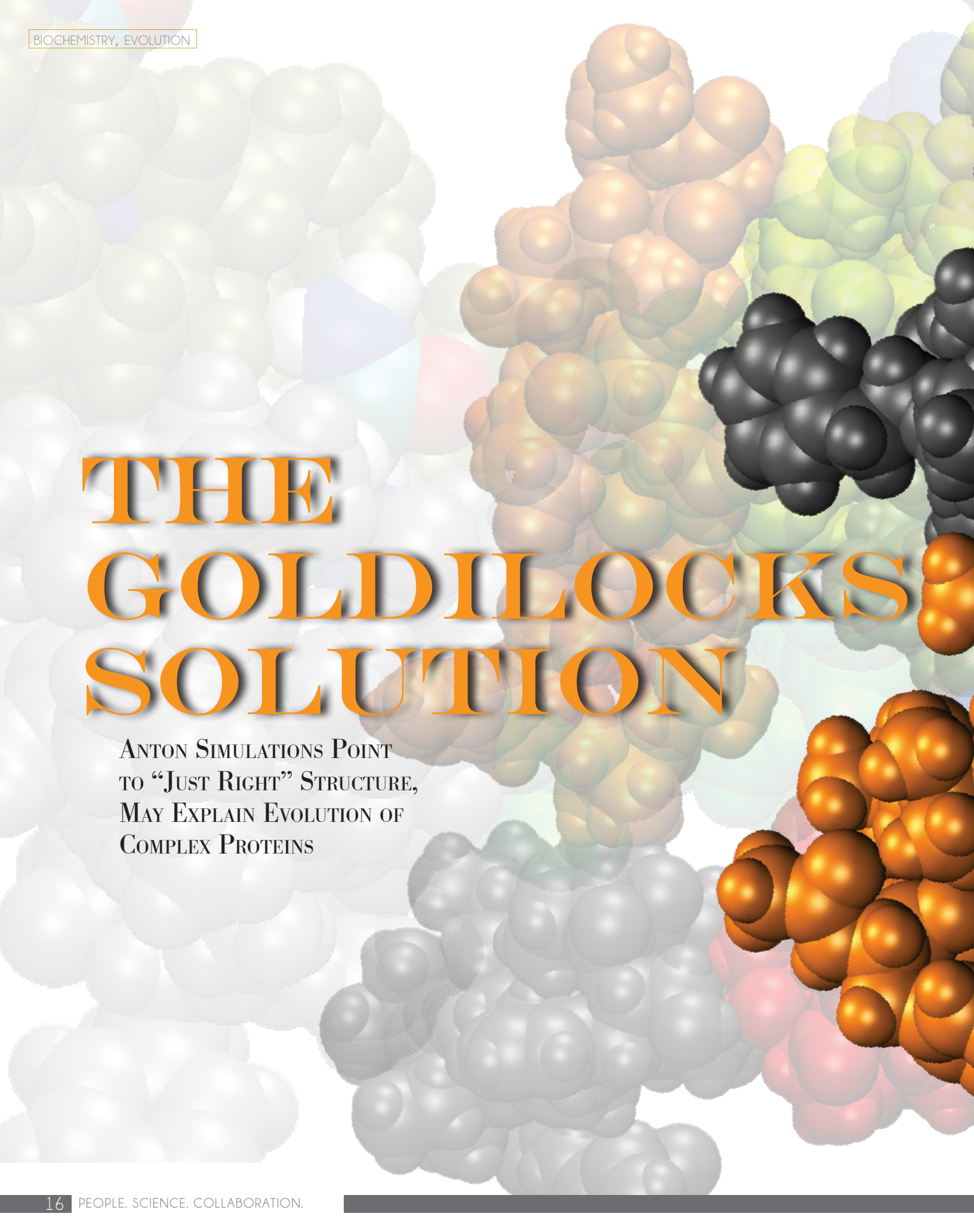
As shown above, she sequenced random fragments of DNA, using Blacklight's large memory to match up the many partly overlapping and often redundant fragments. Then she used Blacklight to compare the genes activated in different copepod species when colonized by *Vibrio* bacteria (not shown).

HOW BLACKLIGHT HELPED

To test the copepod's ability to stave off colonization by *Vibrio*, Aruda tested which genes the organism activated when grown alone or with either of two species of *Vibrio*. One bacterium was known to colonize and eat the copepod's chitin shell but not otherwise harm the shrimp. The other kills some small crustaceans outright.

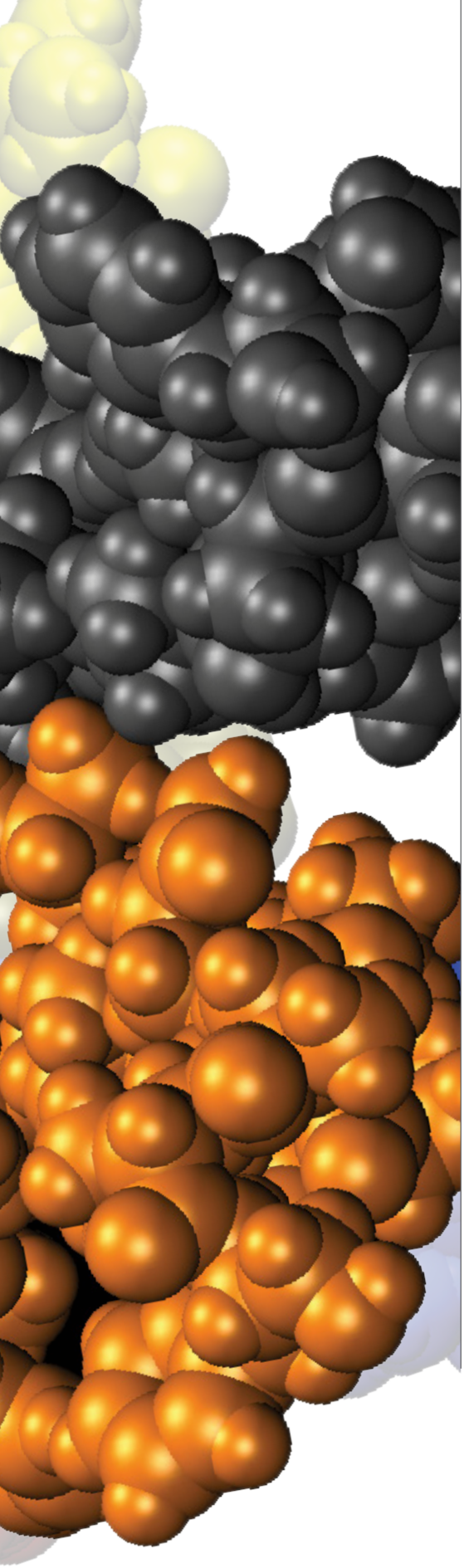
The project posed a computational problem. Little genetic data exists for copepods, so Aruda would have to determine the DNA code of the active genes without a known code as a guide. Testing billions of possible DNA combinations from samples derived from the copepods was a brute-force task requiring vast computer memory—a best-in-class strength of PSC's Blacklight supercomputer.

Using Blacklight, Aruda found to her surprise that the copepods activated their immunity genes far more strongly in response to the "harmless" *Vibrio* than to the "killer" one. The reasons for this aren't yet clear, but it showed that the copepods react differently to different microbial threats. The ongoing work may eventually offer insights on limiting the spread of cholera.



THE GOLDILOCKS SOLUTION

Anton Simulations Point to "Just Right" Structure, May Explain Evolution of Complex Proteins



Life is, well, *complicated.*

Every simple activity of a living creature is a tightly organized ballet of millions of molecules interacting in exquisite complexity that's difficult to fully grasp. In particular, we don't yet fully understand how the *there* of simple molecules on an early Earth got to the *here* of a planet teeming with sophisticated life forms.

Supercomputer simulations at PSC have revealed how six copies of a small, primitive fungus protein can assemble to punch holes in the membranes of attacking bacteria. By showing how very simple molecules can self-assemble in a way that rivals the function of much larger, more complex proteins, Michael Wilson, Andrew Pohorille and colleagues at the NASA Ames Research Center; the University of California, San Francisco; and the University of London have shown how life may have taken some of its earliest steps toward complexity.

The researchers carried out their simulations on an Anton supercomputer developed and provided by D. E. Shaw Research and hosted at PSC, under the auspices of the National Center for Multiscale Modeling of Biological Systems funded by the National Institute of General Medical Sciences.

"BIOWARFARE" IN PRIMITIVE ORGANISMS

The object of Wilson's and Pohorille's study was antiamoebin, a toxic molecule produced by a fungus as a kind of bioweapon to defend itself from bacteria. Researchers have known for some time that the small antiamoebin protein punches holes in a bacterium's cell membrane by assembling into a channel that allows vital potassium ions to leak out. Early studies on timescales of 150 nanoseconds—about one 10-millionth of a second—indicated that multiple copies of antiamoebin assembled to create a larger structure that looked like a "normal" channel: a lasting hole in the target cell's membrane. But researchers weren't sure how many copies of antiamoebin it took to form a working channel, so they couldn't be sure the simulations were accurate.

"For a long time, it was a mystery how ion channels work, because people were not able to get crystal structures of *any* ion channels," says Pohorille, a professor at UCSF and a research scientist at Ames. "Therefore we did not even have a starting point for structural mechanical studies."

But the Ames researchers realized they did have a good way to tell if their simulations were in the right neighborhood. Were the simulated channels able to conduct simulated ions the same way the real channels did?

"That's something that we know very well about ion channels," he says. "If our computer models produce the correct ion conductance, then we're probably doing a good job."

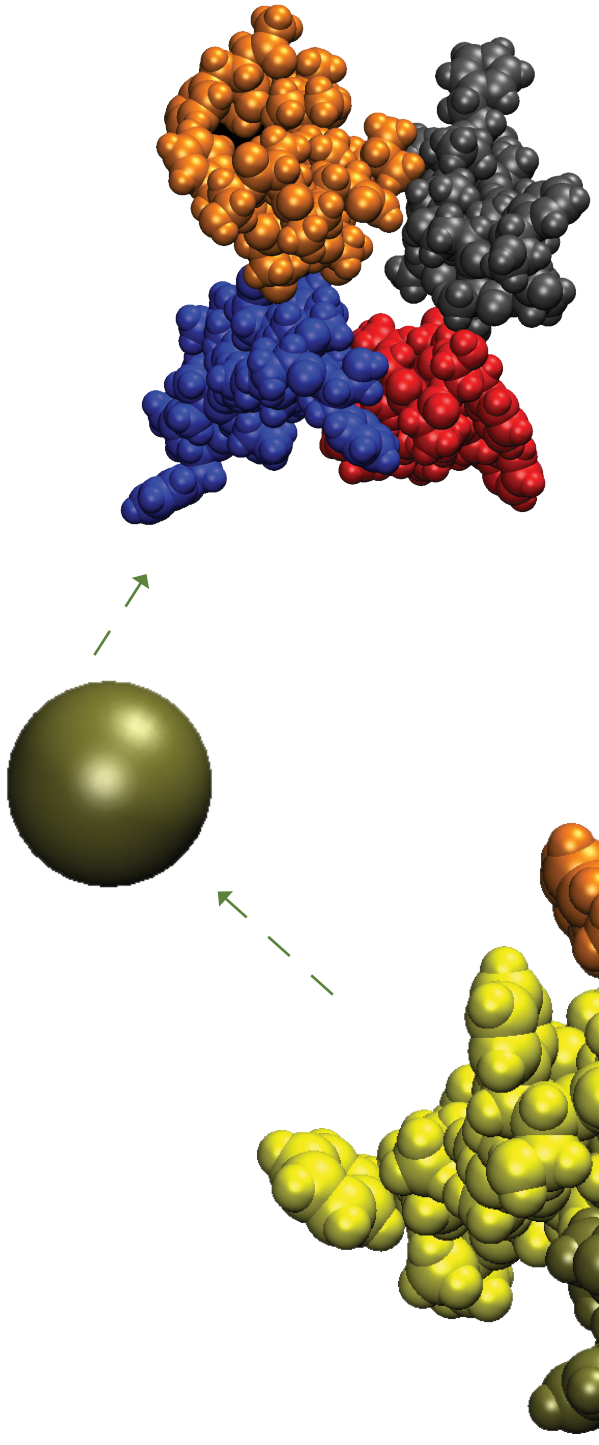
THE GOLDILOCKS SOLUTION

Before the current work, most researchers thought it took four copies of antiamoebin to create a channel. To cover the most likely possibilities, the Ames team carried out far longer simulations of antiamoebin, assuming the protein assembled in groups of four, six, or eight.

Anton proved instrumental because the microsecond simulations it provided—a millionth of a second or more—showed unexpected chaos in antiamoebin's structure. Instead of a stable channel, antiamoebin formed a moving, changing structure that *sometimes* created a hole in the bacterial membrane. Critically, though, the averaged-out conductance of potassium ions across the membrane predicted for this "sometimes-channel" gave important clues about how antiamoebin assembles to form the channel.

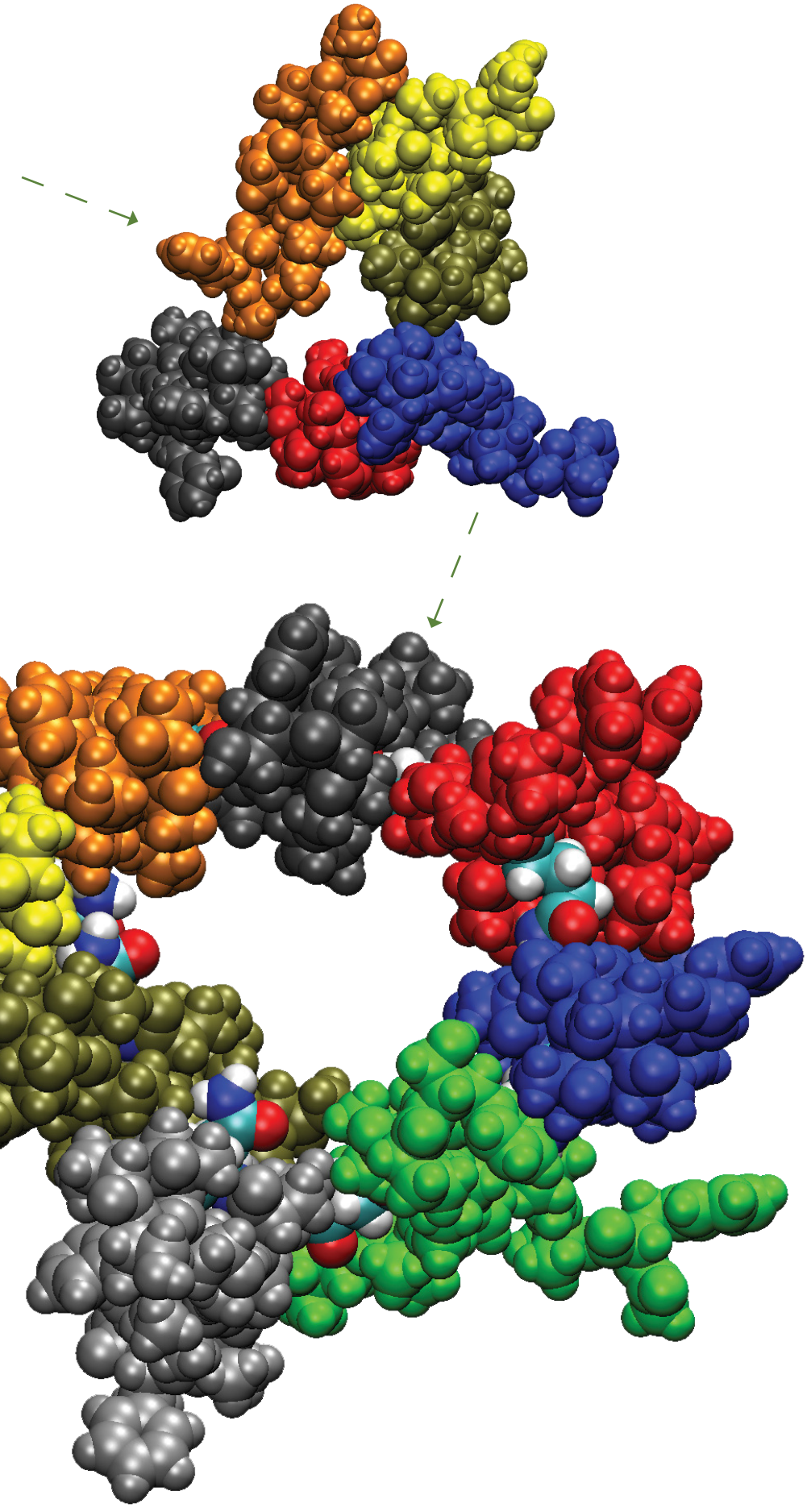
The conductance of the four-member antiamoebin assembly was virtually zero. The eight-member assembly showed a much higher conductance than that of the real-life antiamoebin: It was a channel, but it was too big. But the simulated six-member assembly, like Goldilocks' porridge, allowed ions to cross at a conductance close to the real channel's.

"I think none of the simulated structures is completely right," says Pohorille. "But one is much closer to the truth than the others." The six-member assembly will be the starting point of a new series of studies to understand antiamoebin and similar proteins better.



The simulated antiamoebin structures made up of four (*upper left*), six (*upper right*), and eight (*lower right*) copies of the protein produced internal channels that dictated how fast a potassium ion (*left center*) could pass through. Only the channel produced by the six-membered ring produced a simulated potassium flow that matched laboratory measurements of the protein.

Predicted antiamoebin structures



A WOBBLY WINDOW TO THE DAWN OF LIFE?

One thing about the "correct" antiamoebin structure that surprised the researchers was how chaotic the channel's structure proved to be. The copies of the protein fit and function together, but not particularly well—not as tightly and stably as more complex proteins, optimized by evolution, tend to do.

"This channel is not rigid; it's amazing how flexible it is," Pohorille says. "It doesn't look like any ion channel anybody had seen previously. But at the same time, it doesn't fall apart, it carries out its function."

Another strange thing about antiamoebin is that it isn't constructed solely with the standard amino acids that are contained in most proteins. Two of its ingredients, alpha amino isobutyric acid and isovaline, are "non-standard" amino acids that scientists believe would have been very common on the early Earth. Researchers think the fungi may have held onto these primitive building blocks because they make antiamoebin harder for target bacteria to destroy. Its other amino acids are standard but "boring"—very simple ones that also should have been present early on. Overall, antiamoebin looks a lot like early proteins may have, before they specialized and adopted more rigid and complicated structures.

"We're not saying that this is an ancient protein," Pohorille explains, as no one knows what the actual ancient proteins were. "It's more a model for an ancient protein."

THE BEST TEACHERS

PSC-Trained Teachers Bring Their Students 21st-Century Bioinformatics Skills

Rachel Evrard had a problem: Her current high school science curriculum emphasizes standardized tests that stress information, not skills.

Trained as a microbiologist, Evrard knew that academic advancement and employment in the life sciences depend more and more on analytic skills, not on recalling facts.

Then she saw an email about the BEST–Better Educators of Science for Tomorrow– program at PSC. BEST trains high school teachers in bioinformatics, the technology for analyzing and understanding vast amounts of biological data.

Evrard had to find out more.

"I'm always looking for ways to improve my own knowledge and for real-world applications that I can bring into my classroom," says the Lincoln High School, Ellwood City, Pa., teacher. "Then every student can at least get a glimpse of what bioinformatics is, even if they don't take a full course on it."

She became one of four Pittsburgh-region science teachers in PSC's fifth annual BEST workshop last summer.

MAKING THE CONNECTION

PSC's BEST program grew from the center's other major bioinformatics training initiative, the National Institutes of Health-funded MARC (Minority Access to Research Careers) program. "I joined PSC in 2006 and began talking to Alex and Hugh about the need to extend to high schools what MARC offered to university science courses," says Pallavi Ishwad, PSC's education program director. Alex Ropelewski, PSC biomedical code coordinator, and Hugh Nicholas, senior computational scientist, had been heavily involved with the creation of MARC, with the late David Deerfield, who had been director of PSC's Biomedical Program. They began with a series of introductory lectures on bioinformatics for high school science, technology, engineering and math (STEM) teachers.

"We pulled them aside and asked, 'Do you think we can introduce some of these concepts to your high school students?'" Ishwad says. "They said, 'Yes, this would be an extension to the molecular biology concepts we currently teach.'"

Funding from the Buhl Foundation provided stipends for the teachers to train in bioinformatics as well as covering other costs. In 2009, with PSC's help and ongoing Buhl Foundation funding, three schools conducted a 22-week bioinformatics course pilot project, offering it as an independent elective.

PSC's Pallavi Ishwad (center), pictured (l - r) with Lorraine Horner (Central Catholic), Rachel Evrard (Ellwood City), Jeremy Dawson (Fort Cherry) and Shubahada Bhamre (Plum).

OPENING HORIZONS: BEST SUCCESS STORIES

Rachel Evrard is working with her school district to try to launch a bioinformatics class, but BEST teachers can start to integrate bioinformatics into their biology classes immediately after they finish the program. Gaining buy-in by administrators to launch a freestanding bioinformatics class has taken most schools between a year and 18 months.

Schools that have taken the plunge report some impressive results. "We limited it in the beginning to just juniors and seniors, but we experimented with a few sophomores taking it," says Karen Runtich of Our Lady of the Sacred Heart. "They excelled, and so we opened the class to them." The PSC team also visits students in their classrooms, adding practical topics such as gene sequence analysis to expose them to real-life research arenas. "Two particular students came to me and said they wanted to do something exciting in the summer with research." Runtich explains. "The PSC folks developed summer internship/research projects for them."

This hands-on research experience was a resounding success. Annie Kayser and Danielle Auth presented their findings at the TeraGrid 2011 conference at Salt Lake City, Utah. Auth has since gone on to major in bioinformatics at Gannon University and is considering graduate school in the field when she graduates in 2015. Two other Frazier bioinformatics students, Creg Milko and Shilo Crook, also conducted summer research at PSC. Their teacher, Rebecca Day, presented their work at the 2010 International Supercomputing Conference in New Orleans.

Once the flame catches among a few students, Runtich says, the program's popularity can really grow. "I can go around and praise bioinformatics, but it's just another adult telling them," she explains. "But when they hear it through the students' network and talk with each other, there it is—there's the real sales pitch." One of her students who presented at a TeraGrid conference, for example, had recruited the other. The important thing, she says, is to get the students started: They'll take it from there.

BEST PLANS FOR A NATIONAL FUTURE

Following the BEST program's success in the Pittsburgh region, PSC has initiated outreach at four MARC program partners to bring its bioinformatics curriculum to the national stage. Tennessee State University, North Carolina Agricultural & Technical University, Jackson State University and the University of Puerto Rico at Mayaguez are establishing committees to launch the BEST curriculum in their affiliated high schools.

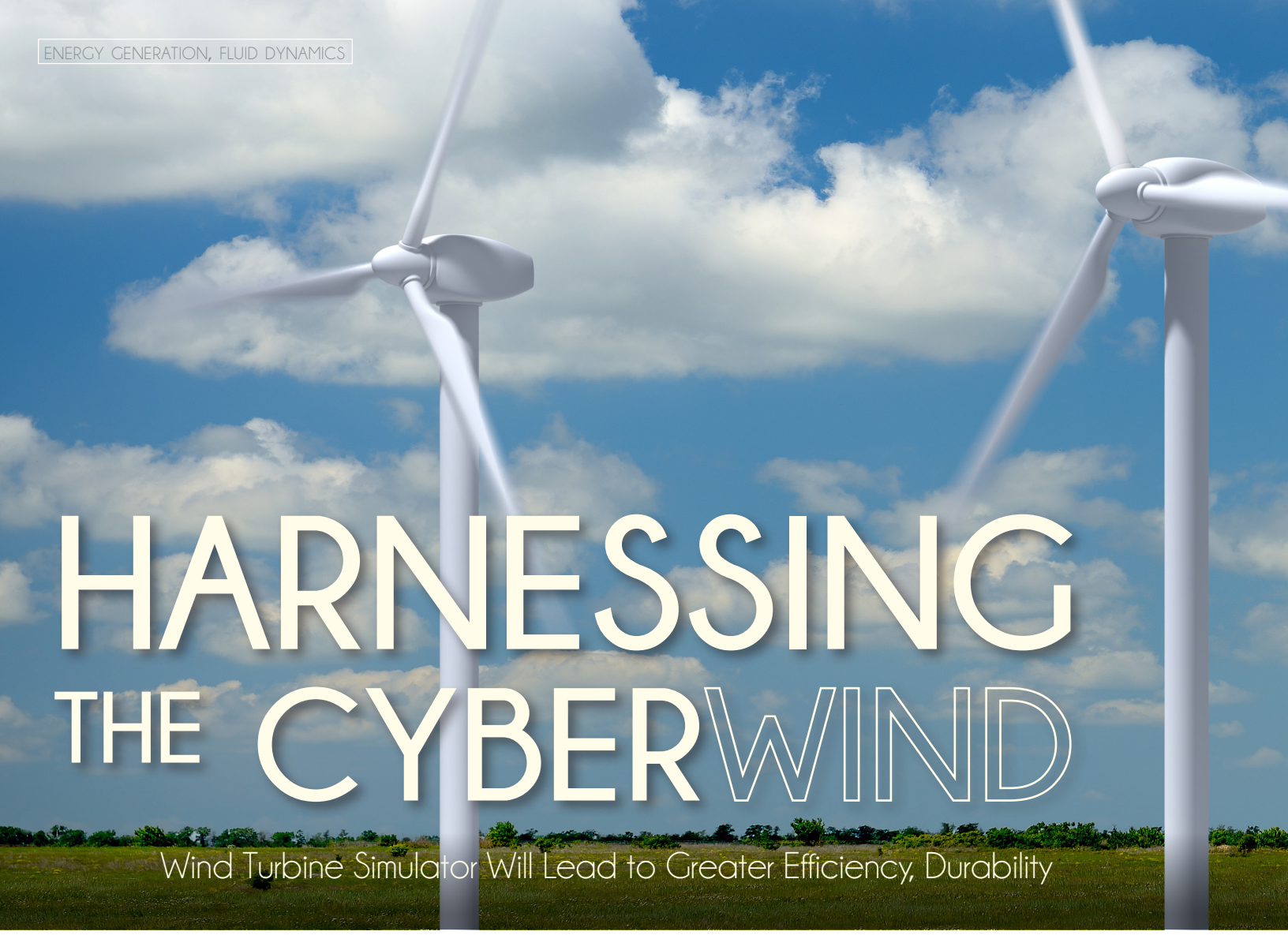
THE **BEST** SCHOOLS

Full bioinformatics courses

- Pittsburgh Public School District SciTech Academy – Dr. Edwina Kinchington
- Our Lady of Sacred Heart High School, Coraopolis, Pa. – Karen Runtich
- Oakland Catholic Girls High School, Pittsburgh, Pa. – Melissa Kavanshansky
- PA Cyber Charter School, Midland, Pa. – Michael Giles

Bioinformatics as part of biology classes

- Ellwood City Area School District, Ellwood City, Pa. – Rachel Evrard
- Central Catholic Boys High School, Pittsburgh, Pa. – Lorraine Horner
- Plum Borough School District, Plum, Pa. – Dr. Shubahada Bhamre
- Fort Cherry School District, McDonald, Pa. – Jeremy Dawson - will be using curriculum in Middle and High school classes
- Frazier High School, Perryopolis, Pa. – Rebecca Day
- Winchester-Thurston High School, Pittsburgh, Pa. –Lara Tukarski (curriculum previously developed by Jared Rashford)



HARNESSING THE CYBERWIND

Wind Turbine Simulator Will Lead to Greater Efficiency, Durability

WHY IT'S IMPORTANT

Wind energy commands outsized promise for power generation. What's not to like about a permanent source of clean energy? But to turn a profit, wind farms need to reach reliability levels better than 90 percent over a 20-year operational life. When you're talking about getting moving parts that approach and exceed the size of a 747 to withstand the buffeting of turbulent winds that can vary by 40 or 50 percent from the average wind speed, that's a tall order.

James Brasseur and Ganesh Vijayakumar of Pennsylvania State University and their colleagues are working on a solution to this problem: The Cyber Wind Facility, an extreme-resolution computer simulation of a wind turbine in turbulent atmospheric winds. This computer model produces detailed results with the quality and usefulness of real-life data.

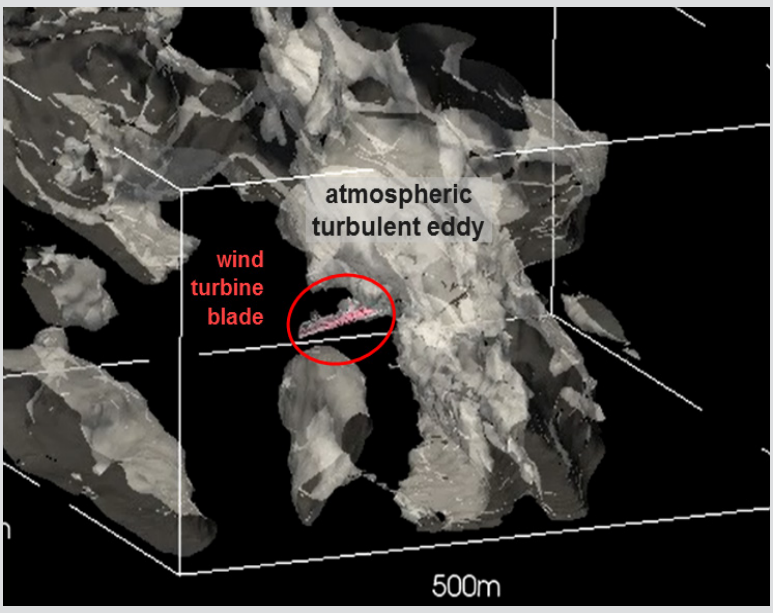
HOW BLACKLIGHT HELPED

To put together a detailed computer model of all the forces that drive a wind turbine, the researchers first had to pull them apart. The model consists of two main simulations. One models the turbulent lower atmosphere, the other simulates wind flow over and around the rotating blades. While the atmosphere simulation must duplicate atmospheric eddies as small as 3 meters, the grid over the surface of the 62-meter blade needs to be precise to the width of a human hair (5 to 10 microns).

The Penn State team used PSC's Blacklight supercomputer and its large shared memory for the critical task of parsing the 160 million computational cells over the thousands of processors required to run the massive simulation. Blacklight's large memory also allowed them to analyze the huge datasets produced in their larger-scale simulations.

"It's one thing to say we can get energy out of the wind, quite another to say that it is profitable to do so … Two issues we want to focus on are power capture—how much energy the turbine can extract from the wind—and reliability—how often components must be repaired or replaced. Since it is very expensive to replace components sitting 80 to 100 meters above the ground, a failure rate of just a few percent among wind turbines in the 20-year lifetime of a wind farm can eliminate profitability."

—James Brasseur, Pennsylvania State University



Output from a Cyber Wind Facility simulation of a single turbine blade cutting through an atmospheric turbulent eddy over a 30-second time period. The simulation showed that surface forces on the blade varied by as much as 35 percent due to the changing wind speed and direction, adding to stress on the blade and potentially decreasing the component's lifespan.

2015 PENNSYLVANIA STATE BUDGET INCLUDES $500,000 FOR PSC

The Commonwealth of Pennsylvania budget for fiscal year 2015 includes $500,000 for the Pittsburgh Supercomputing Center (PSC).

"We are pleased and honored that the state has once again deemed PSC to be worthy of funding in this fiscally challenging year," says Ralph Roskies, scientific director for PSC. "We're grateful to the members of the General Assembly and the Allegheny County delegation."

The state's return on its past investments in PSC has been outstanding. Since PSC's inception, the center has brought over $500 million in outside funds into Pennsylvania, representing a 14:1 return on state funding for PSC.

"An economic study has shown that PSC is responsible for generating 1,600 jobs and over $200 million in annual economic activity," says Cheryl Begandy, PSC's director of education, outreach, and training. "In addition, our place on the leading edge of computing technologies at the largest scale enables us to respond quickly to technological developments, giving the state, its researchers and its small and mid-sized companies a leg up in capitalizing on these advances."

PSC's state funding goes toward science, technology, engineering and math education projects, outreach to serve small- and medium-sized Pennsylvania businesses' computational needs and other workforce development projects and other operational activities.

"Holdem" by Todd Klassy. Licensed under Creative Commons Attribution 2.5 via Wikimedia Commons.

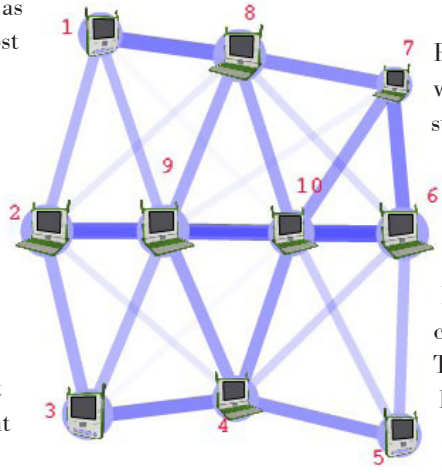
BLACKLIGHT-BUILT COMPUTER PROGRAM WINS 2014 POKER COMPETITION

Tartanian7, a computer poker player developed by Tuomas Sandholm of Carnegie Mellon University and his PhD students Sam Ganzfried and Noam Brown, won first place in the 2014 Annual Computer Poker Competition in both the "Two-player no-limit Instant Runoff" and "Two-player no-limit Total Bankroll" categories. In these two categories the program, which beat all the other contestants by a large margin, demonstrated good performance against the strongest competitors as well as the best aggregate winnings, respectively.

The team used PSC's Blacklight to build Tartanian7 using algorithms rather than human "expert knowledge." Given only the basic rules of poker, the algorithms developed a strategy for playing poker hands, including bluffing and slow play. While chess-playing computers and algorithms such as Deep Blue have captured the public's imagination by beating human chess masters, the incomplete information in poker makes that game far more complex than chess. Blacklight's large memory and computational power allowed the algorithms to make finer distinctions between similar hands while developing its strategy, which proved the winning recipe.

EIGHT PA. COLLEGES AND UNIVERSITIES SELECT PSC AND 3ROX FOR INTERNET, COMPUTATION

A group of eight western Pennsylvania Association of Independent Colleges and Universities of Pennsylvania (AICUP) member-institutions have selected the Three Rivers Optical Exchange (3ROX) as their commodity Internet provider and regional network administrator and PSC for computing and storage resources. This new relationship will provide improved Internet access to these AICUP member institutions as well as access to a host of high performance computing resources through PSC.



"XO classroom network" by OLPC. Licensed under Creative Commons Attribution 2.5 via Wikimedia Commons.

The institutions in the group are Carlow University, Geneva College, Juniata College, La Roche College, Point Park University, Saint Francis University, Washington & Jefferson College and Westminster College. These schools will enjoy substantially increased bandwidth levels with 3ROX/PSC managing Web browser, email and other routine Internet access for the members' faculty, staff and students.

Instead of routing institutional traffic through commodity Internet connections, the new service will allow it to be moved to local links, connecting directly to certain common services without consuming the AICUP members' commodity bandwidth. Members using these services will therefore move more data without incurring the equivalent of overage charges or price premiums charged by other service providers.

Another benefit to AICUP members will be the ability to inexpensively archive large amounts of data in PSC's Data Supercell. The Supercell is a petabyte-scale system that allows users to store data with the security of an archival system and speed of data retrieval of a local hard disk.

$1.2-MILLION SUPPLEMENT TO BROADEN SCIENTIFIC MISSION OF DATA EXACELL

PSC's $7.6-million Data Exacell (DXC) project will add database capabilities thanks to a supplemental grant from the National Science Foundation for $1.2 million.

The DXC is a pilot project to create, deploy, and test software and hardware building blocks designed to develop data-analytic capabilities to advance scientific research. The project focuses on coupling PSC's Big Data analytical resources—Blacklight and Sherlock—with high-performance, multi-petabyte, disk-based storage. The supplemental grant will enhance DXC to provide powerful data representations for database-driven research. It will provide software, hardware, and collaboration necessary to accelerate analytics using state-of-the-art database technologies and transformational, database-driven application architectures.

Pittsburgh Supercomputing Center is a joint effort of Carnegie Mellon University and the University of Pittsburgh. It was established in 1986 and is supported by several federal agencies, the Commonwealth of Pennsylvania and private industry.

PSC gratefully acknowledges significant support from the following:

The National Science Foundation
The National Institutes of Health
The National Energy Technology Laboratory
The National Oceanographic and Atmospheric Administration
The U. S. Department of Defense
The U. S. Department of Energy
The Commonwealth of Pennsylvania
D. E. Shaw Research
Cray Inc.
Silicon Graphics, Inc.
The Buhl Foundation
Bill and Melinda Gates Foundation

DESIGN & PRODUCTION: Shandra Williams
SENIOR WRITER: Ken Chiacchia
MANAGING EDITOR: Vivian Benton
PROJECT COORDINATION: Cheryl Begandy

COVER GRAPHIC: The HERMES model of the Republic of Benin's vaccine supply chain. The PSC/Johns Hopkins team analyzed the West African state's vaccination program, showing how to cut costs while adding new vaccinations. See p. 4.

PRINTING: Knepper Press

Printed on Sappy McCoy Paper, a premium sheet with 10 percent post-consumer waste fiber with vegetable-based inks.

PSC Resources for Data-Driven Science

Blacklight is an SGI Altix® UV1000 supercomputer designed for memory-limited scientific applications in fields as different as biology, chemistry, cosmology, machine learning and economics. Funded by the National Science Foundation (NSF), Blacklight carries this mission out with partitions with as much as 16 terabytes of coherent shared memory.

Sherlock is a YarcData Urika™ (Universal RDF Integration Knowledge Appliance) data appliance with PSC enhancements. It enables large-scale, high-performance graph analytical processing through massive multithreading (128 hardware threads per processor), a shared address space, sophisticated memory access optimizations and support for heterogeneous applications. Sherlock is funded by the NSF.

Anton is a special purpose supercomputer designed to dramatically increase the speed of molecular dynamics simulations, allowing biomedical researchers to understand the motions and interactions of proteins and other biologically important molecules over much longer time periods than previously possible. Designed and built by D. E. Shaw Research (DESRES), the Anton machine hosted at PSC was provided without cost by DESRES for non-commercial use by the national biomedical research community.

The **Data Supercell** (DSC) is a PSC-designed and built system for managing and archiving petabyte-scale data for researchers and industrial users. The DSC provides low-latency, high-capacity, high-reliability, high-bandwidth and low-cost data storage and retrieval.

The **Data Exacell** (DXC) is an NSF-funded pilot project to provide hardware and software building blocks to support data-intensive research projects. DXC is based on the unique, PSC-developed data storage architecture in the *Data Supercell*, combined with the very-large-memory capabilities of *Blacklight* and the graph-analytics capabilities of *Sherlock*, as well as specialized database capabilities. PSC experts are working with multiple research groups both to refine DXC architecture and to extend the NSF's support for new fields of science.

For more information on PSC's resources for data-driven science, see WWW.PSC.EDU